



Étude théorique des saccharides : structures et spectres infrarouges

Loic Barnes

► To cite this version:

Loic Barnes. Étude théorique des saccharides : structures et spectres infrarouges. Autre [cond-mat.other]. Université de Lyon, 2017. Français. NNT : 2017LYSE1152 . tel-01661557

HAL Id: tel-01661557

<https://theses.hal.science/tel-01661557>

Submitted on 12 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N° d'ordre NNT : 2017LYSE1152

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON

opérée au sein de
l'Université Claude Bernard Lyon 1

École Doctorale ED52
Physique & Astrophysique de Lyon

Spécialité de doctorat : Physico-chimie théorique
Discipline : Physique

Soutenue publiquement le 22/09/2017, par :
Loïc Barnes

Étude théorique des saccharides : structures et spectres infrarouges

Devant le jury composé de :

Christophe Morell, Professeur des universités, Université de Lyon

Président

Carine Clavaguéra, Chargée de recherche CNRS, Université Paris Sud

Rapporteure

Aude Simon, Chargée de recherche CNRS, Université de Toulouse

Rapporteure

Anne Zehnacker-Rentien, Directrice de recherches CNRS, Université Paris Sud

Examinatrice

Isabelle Compagnon, Maître de conférences, Université de Lyon

Examinatrice

Abdul-Rahman Allouche, Professeur des universités, Université de Lyon

Directeur de thèse

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université

Président du Conseil Académique

Vice-président du Conseil d'Administration

Vice-président du Conseil Formation et Vie Universitaire

Vice-président de la Commission Recherche

Directrice Générale des Services

M. le Professeur Frédéric FLEURY

M. le Professeur Hamda BEN HADID

M. le Professeur Didier REVEL

M. le Professeur Philippe CHEVALIER

M. Fabrice VALLÉE

Mme Dominique MARCHAND

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard

Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux

Faculté d'Odontologie

Institut des Sciences Pharmaceutiques et Biologiques

Institut des Sciences et Techniques de la Réadaptation

Département de formation et Centre de Recherche en Biologie Humaine

Directeur : M. le Professeur G.RODE

Directeur : Mme la Professeure C. BURILLON

Directeur : M. le Professeur D. BOURGEOIS

Directeur : Mme la Professeure C. VINCIGUERRA

Directeur : M. X. PERROT

Directeur : Mme la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies

Département Biologie

Département Chimie Biochimie

Département GEP

Département Informatique

Département Mathématiques

Département Mécanique

Département Physique

UFR Sciences et Techniques des Activités Physiques et Sportives

Observatoire des Sciences de l'Univers de Lyon

Polytech Lyon

Ecole Supérieure de Chimie Physique Electronique

Institut Universitaire de Technologie de Lyon 1

Ecole Supérieure du Professorat et de l'Education

Institut de Science Financière et d'Assurances

Directeur : M. F. DE MARCHI

Directeur : M. le Professeur F. THEVENARD

Directeur : Mme C. FELIX

Directeur : M. Hassan HAMMOURI

Directeur : M. le Professeur S. AKKOUCHE

Directeur : M. le Professeur G. TOMANOV

Directeur : M. le Professeur H. BEN HADID

Directeur : M. le Professeur J-C PLENET

Directeur : M. Y.VANPOULLE

Directeur : M. B. GUIDERDONI

Directeur : M. le Professeur E.PERRIN

Directeur : M. G. PIGNAULT

Directeur : M. le Professeur C. VITON

Directeur : M. le Professeur A. MOUGNIOTTE

Directeur : M. N. LEBOISNE

Table des matières

I	Introduction	7
II	Méthodes théoriques	21
1	Méthodes de la chimie quantique	23
1.1	Introduction	23
1.2	Hartree-Fock	26
1.2.1	Combinaison Linéaire d'Orbitales Atomiques	32
1.2.2	Les bases atomiques	32
1.2.3	Les équations de Roothan-Hall	34
1.2.4	Résolution par champ auto-cohérent ou SCF	35
1.2.5	Différents schémas de construction de la fonction d'essai	36
1.3	La corrélation électronique	38
1.4	Post Hartree-Fock	39
1.4.1	Interaction de configurations	39
1.4.2	Clusters couplés	42
1.4.3	Théorie de la perturbation de Møller et Plesset	44
1.5	La théorie de la fonctionnelle de la densité	47
1.5.1	Les théorèmes de Hohenberg-Kohn	48
1.5.2	La théorie de Kohn-Sham	50
1.5.3	Approximations de la fonctionnelle de la densité	52
1.6	La DFTB : density functional based tight-binding	55
1.6.1	Résolution de l'hamiltonien $\hat{H}^0$	58
1.6.2	SCC-DFTB : self-consistent charge correction DFTB	59
1.6.3	DFTB3	60
1.7	Les méthodes semi-empiriques NDDO	61
1.7.1	Formalisme	61

1.8	Correction empirique de la dispersion	63
1.9	La mécanique moléculaire	64
1.10	Références	67
2	Dynamique Moléculaire	71
2.1	Introduction	71
2.2	Dynamique moléculaire	72
2.2.1	Thermostat	73
2.2.2	Intégrateur	75
2.3	Recherche de conformations	76
2.4	Références	78
3	Calcul du spectre vibrationnel	79
3.1	Résolution de l'équation de Schrödinger pour les noyaux	80
3.1.1	Calcul des modes et des fréquences de vibration	80
3.1.2	Calcul des probabilités de transition	83
3.2	Anharmonicité	84
3.2.1	Emploi de facteurs correctifs	84
3.2.2	VPT2	85
3.2.3	VSCF	96
3.2.4	Dynamique moléculaire	98
3.3	Références	99
III	Applications	103
4	Anharmonic simulations of the vibrational spectrum of sulfated compounds : application to the glycosaminoglycan fragment glucosamine 6-sulfate[†]	109
5	IRMPD Spectroscopy Sheds New (Infrared) Light on the Sulfate Pattern of Carbohydrates[†]	121
6	Conformational preferences of protonated <i>N</i>-acetylated hexosamines probed by Infrared multiple photon dissociation (IRMPD) spectroscopy and <i>ab initio</i> calculations[†]	141
6.1	Introduction	142
6.1.1	Methods	142
6.2	Results and discussion	143
6.2.1	Identification of the protonation site	143

6.2.2	Conformational analysis of 1- <i>O</i> -methyl <i>N</i> -acetyl hexosamines . . .	145
6.2.3	Analysis of anomeric mixtures	150
6.3	Conclusion	153
6.4	References	153
7	The remarkable structural complexity of simple carbohydrate building blocks : conformational and anomeric heterogeneity of hexosamines resolved by IRMPD spectroscopy and <i>ab initio</i> calculations[†]	157
7.1	Introduction	158
7.2	Methods	158
7.2.1	IRMPD spectroscopy	158
7.2.2	Computational methods	158
7.3	Results and discussion	159
7.3.1	Conformational preference of hexosamines	159
7.3.2	Capture of the dynamic in solution by IRMPD snapshots	163
7.3.3	Elucidation of conformational heterogeneity	165
7.4	Conclusions	165
7.4.1	References	166
8	Étude de la liaison osidique	169
8.1	Introduction	170
8.2	Signature directe de la liaison osidique	170
8.3	Signature indirecte de la liaison osidique	171
8.3.1	Analyse des disaccharides GlcN β (1 $\rightarrow$ 6)GlcNAc et GlcN β (1 $\rightarrow$ 4)GlcNAc	171
8.3.2	Analyse des disaccharides Gal β (1 $\rightarrow$ 6)GlcNAc et Gal β (1 $\rightarrow$ 4)GlcNAc	174
8.4	Conclusion	178
8.5	Références	180
IV	Conclusion et perspectives	181
8.6	Résumé des travaux	183
8.7	Perspectives	184
	Annexe	187
A	Informations supplémentaires pour le chapitre 7	187
B	Liste d'articles	201

Première partie

Introduction

Contexte

Parmi toutes les molécules qui constituent le vivant, il y a en particulier trois classes qui assurent des fonctions essentielles. Ce sont trois classes de biopolymères, c'est-à-dire des macromolécules d'origine organique. Il y a la classe des polymères de nucléotides, d'acides aminés et de monosaccharides. L'acide désoxyribonucléique (ADN) qui est le support du matériel génétique d'une espèce fait partie de la classe des polynucléotides. L'acide ribonucléique (ARN) fait aussi partie de cette classe. Elle intervient dans la production des protéines qui assurent des fonctions dans les cellules ou les tissus. Chaque nucléotide est formé d'une base nucléique ou base azotée (adénine, cytosine, guanine, thymine, uracile, etc) liée par une liaison covalente à un ribose (un sucre cyclique à 5 carbones) qui est lié via son carbone en position 5 à un groupement phosphate. La liaison entre deux nucléotides (ou liaison phosphodiester) est assurée par le groupement phosphate qui se lie à un autre nucléotide sur le carbone 3 (voir fig. 1). Les polypeptides produits via l'ARN

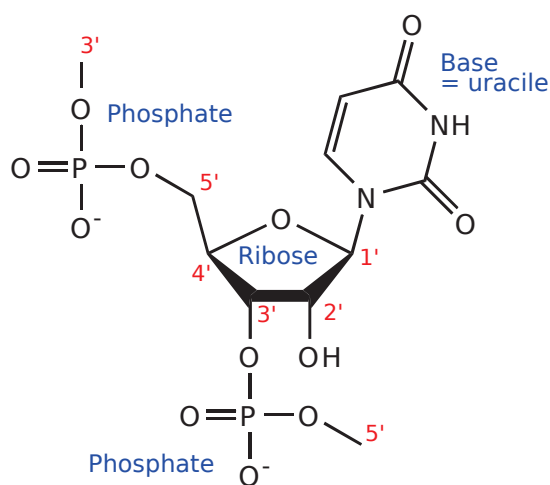


FIGURE 1 – Exemple de nucléotide constitué d'uracile.

sont dépendants de la séquence nucléotidique de l'ARN, c'est-à-dire de la chaîne linéaire de nucléotides.

Les polypeptides font partie de la classe des polymères d'acides aminés tout comme les protéines. Les polypeptides sont des chaînes d'acides aminés liées linéairement et les protéines sont formées par l'association de plusieurs polypeptides. La séquence ou l'enchaînement des acides aminés dans un polypeptide est créé par la séquence nucléotidique de l'ARN qui produit ce polypeptide. La liaison reliant deux acides aminés entre eux est appelée liaison peptidique. Elle intervient entre le carbone α (le carbone le plus près de la fonction de l'acide aminé) de la fonction carboxyle et le carbone α de la fonction amine de l'autre acide aminé (voir fig. 2).

L'autre classe de biopolymères est la classe des polymères de monosaccharides. Ces derniers peuvent servir à former la paroi végétale par exemple dans le bois (cellulose) ou assurer l'apport en énergie aux animaux et aux végétaux (glycogène, amidon). La liaison entre deux monosaccharides est appelée liaison osidique. Elle s'effectue entre le groupement réducteur, c'est-à-dire le groupement hydroxyle du carbone en position 1 (pour les aldoses) ou en position 2 (pour les cétooses) d'un monosaccharide avec tous les autres groupements hydroxyles du second monosaccharide. Le groupement réducteur possède deux configurations lorsque le saccharide est sous forme cyclique, il peut être soit en configuration α soit en configuration β . La particularité de cette classe est que les polymères ne

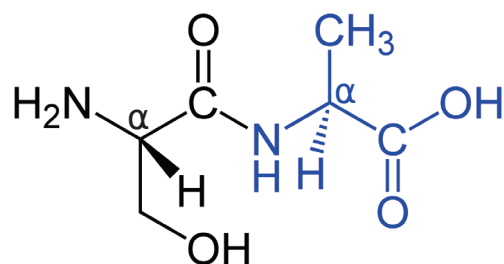


FIGURE 2 – Liaison peptidique entre la sérine et l’alanine

sont pas forcément linéaires et peuvent se ramifier augmentant ainsi la complexité structurelle de ces constituants. En prenant les polynucléotides de l’ADN, on a 4 nucléotides différents. Ainsi pour former un dinucléotide, on a 16 possibilités. Il y a 22 acides aminés ce qui donne 484 possibilités de former un dipeptide. En prenant seulement les 10 monosaccharides les plus abondants chez les mammifères on arrive à un total de 1360 possibilités pour former un disaccharide [1]. De plus les saccharides peuvent subir des modifications, les groupements hydroxyles peuvent être remplacés par d’autres groupements (amines, phosphates, sulfates, etc) ce qui rajoute de la complexité.

La séquence des protéines et des polynucléotides (ADN, ARN) est très importante. Une perturbation dans la séquence de l’ARN peut entraîner la production de nouvelles protéines inadaptées, qui n’arriveront pas à assurer la fonction qu’elles sont sensées avoir conduisant à des carences ou à des maladies. La connaissance de la séquence des protéines et de l’ADN a un intérêt crucial dans le domaine de la santé. De nombreuses méthodes expérimentales [2, 3, 4, 5, 6, 7] pour le séquençage de l’ADN ont vu le jour au siècle dernier et ont débouché sur le séquençage complet du génome de nombreuses espèces (y compris le génome humain achevé en 2001 via le projet HuGO [8]). D’autres méthodes ont été développées [9, 10] pour permettre le séquençage du protéome, de l’ensemble des protéines dans un ensemble biologique (partie d’une cellule, cellule, groupe de cellules). Ces méthodes ont été étendues à l’étude du glycome, l’ensemble des oligosaccharides (un polymère de monosaccharides) présents dans une cellule ou un organisme. Cependant ces méthodes reposent sur la masse, or pour une même masse on peut avoir de nombreux saccharides bien différents (nature du saccharide, séquence, branchement, position des liaisons osidiques, position et nature des modifications). Les méthodes d’étude reposant sur la masse sont limitées pour l’étude du glycome et doivent être couplées avec d’autres approches pour contourner cette difficulté (méthode IRMPD, InfraRed Multiple Photon Dissociation, voir la thèse de Baptiste Schindler [11]). Il existe des méthodes spectroscopiques [12, 13, 14, 15] pour étudier des saccharides mais celles-ci présentent des contraintes très fortes que sont les longues étapes de purification des échantillons et la possibilité que les modifications soient altérées.

Les oligosaccharides sont moins bien caractérisés du fait de la difficulté à trouver une méthode capable de séquencer le glycome. Pourtant cette classe de biopolymères joue un rôle important dans le domaine de la santé tout comme dans les domaines industriels (énergie, matériaux). Développer des outils pour caractériser les oligosaccharides a été déclaré comme nécessaire par l’Europe [16] et les USA [17].

Finalement, une méthode expérimentale prometteuse pour l’analyse des oligosaccharides est la méthode IRMPD basée sur le couplage entre la spectrométrie de masse et la spectroscopie infrarouge. C’est la méthode explorée dans le projet qui a vu le jour au sein de l’Institut Lumière Matière mené par Isabelle Compagnon. Ce projet a pour mission de développer une méthode pour étudier les saccharides. Grâce à la spectrométrie de masse

utilisée dans la méthode IRMPD, les saccharides sont sélectionnés en masse. Enfin les expérimentateurs pourront faire la mesure des spectres infrarouges de saccharides avec une même masse en espérant que ces spectres soient différents. Cependant, la méthode IRMPD seule n'est pas capable d'identifier les anomères ou les conformères au sein de l'échantillon. C'est la comparaison des spectres expérimentaux avec les spectres calculés qui permettra d'identifier des signatures des anomères et des conformères. On pourra savoir quels anomères voire quels conformères contribuent au spectre expérimental. De plus, en mettant en relation les spectres IRMPD et les spectres théoriques, on va pouvoir obtenir des informations sur la structure des saccharides présents dans l'échantillon alors que sans la théorie c'est impossible.

Le projet a débuté en 2013 dans un contexte où les sucres étaient très peu étudiés. Parmi les quelques études présentes avant ce projet, on peut trouver les mesures IRMPD dans la zone FELIX (du nom de l'instrument servant à faire la mesure), c'est-à-dire dans la zone $600\text{-}1800\text{ cm}^{-1}$ effectuées par Oomens [18] et Polfer [19, 20, 21]. Ce sont des mesures effectuées dans une installation particulière de grande taille. Les expériences sont accessibles en faisant des demandes de temps de mesure donc pendant une durée limitée. Simons, Cocinero et Carçabal [22, 23, 24] ont aussi étudié des saccharides en effectuant des mesures en UV-IR sur des sucres neutres et refroidis. Cependant, afin de mener à bien les mesures UV les sucres doivent être modifiés par l'ajout de chromophores. Le projet mené par Isabelle Compagnon est apparu dans ce contexte. Il a pour mission de développer une méthode de routine pour étudier les saccharides. De plus, cette méthode doit permettre d'analyser des saccharides ionisés sans qu'ils ne subissent de modifications notamment par des chromophores. Depuis, d'autres groupes ont commencé à étudier les saccharides. Maître a effectué des mesures IRMPD dans la zone OPO (du nom de l'instrument qui fait les mesures dans la zone $2800\text{-}3800\text{ cm}^{-1}$) en 2015 en utilisant du lithium ($\text{Li}^+, \text{H}_2\text{O}$) [25]. Pagel a étudié très récemment des sucres en utilisant la spectrométrie de masse couplée à la mobilité ionique [26, 27, 28]. Enfin, Rizzo a aussi effectué des mesures sur des sucres en particulier des disaccharides à basse température [29, 30].

Définitions et structures des saccharides

Certains termes comme anomères ou conformères ont été utilisés dans la partie contexte de l'introduction, je vais les définir dans cette partie. J'expliquerai comment sont nommées les différentes structures de sucres [31] et je parlerai de la liaison osidique.

Un sucre simple aussi appelé ose ou monosaccharide est une chaîne dont la formule brute peut s'écrire comme $(\text{CH}_2\text{O})_n$ avec $n \geq 3$. Pour rappel, des molécules qui ont la même formule brute sont appelés des isomères. La chaîne peut être soit ouverte soit fermée. Les monosaccharides peuvent être séparés en deux familles, la famille des aldoses et la famille des cétooses. Par convention, les atomes de carbones sont numérotés afin de nommer les différentes structures que l'on peut rencontrer pour les sucres. Pour la famille des aldoses, le carbone qui possède la fonction aldéhyde est le carbone 1, fig. 3a. En ce qui concerne la famille des cétooses, le carbone qui possède la fonction cétoose est le carbone 2, fig. 3b.

Dans la nature, les sucres sont rarement présents sous forme de chaîne ouverte. La fermeture du cycle peut se faire de deux façons différentes. La première fera intervenir le carbone en position 1 avec le carbone en position 5 pour donner naissance à une structure possédant un cycle à 6 atomes appelée pyranose. Pour la seconde fermeture de cycle possible, ce sont les carbones des positions 1 et 4 qui seront mis en jeu pour donner une

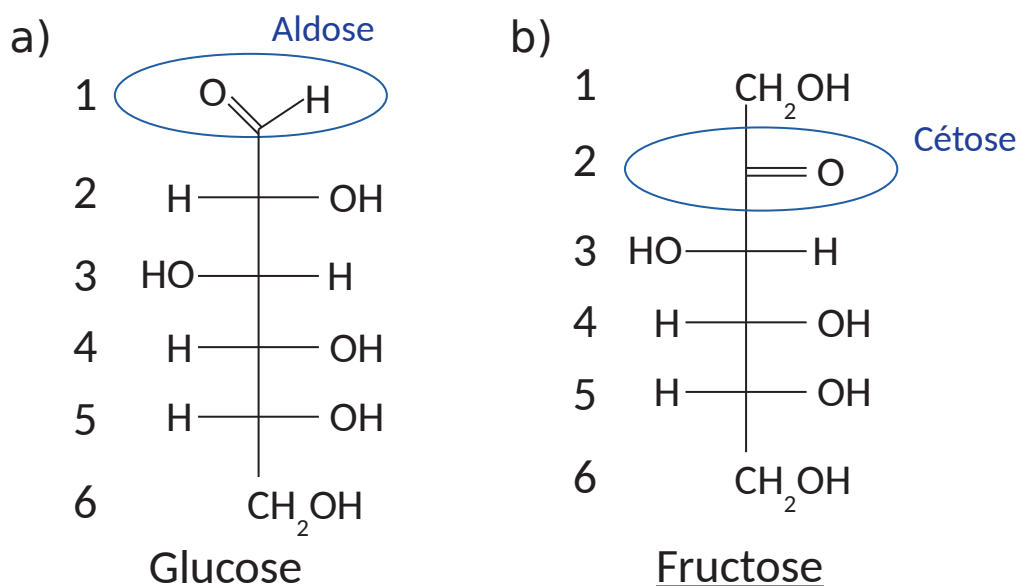


FIGURE 3 – a) Le glucose de la famille des aldoses. b) Le fructose de la famille des cétones.

structure possédant un cycle à 5 atomes appelée furanose. De plus, selon l'orientation du groupement hydroxyle en position 1, deux structures vont pouvoir exister que ce soit pour une forme pyranose ou furanose. Si cet hydroxyle est du même côté du cycle que l'hydroxyle du carbone en position 6 alors on appelle cette forme l'anomère β . Sinon, si les deux groupements ne sont pas du même côté par rapport au cycle, on appelle cette forme l'anomère α . La figure 4 montre l'exemple des anomères α et β pour les formes pyranose et furanose du glucose.

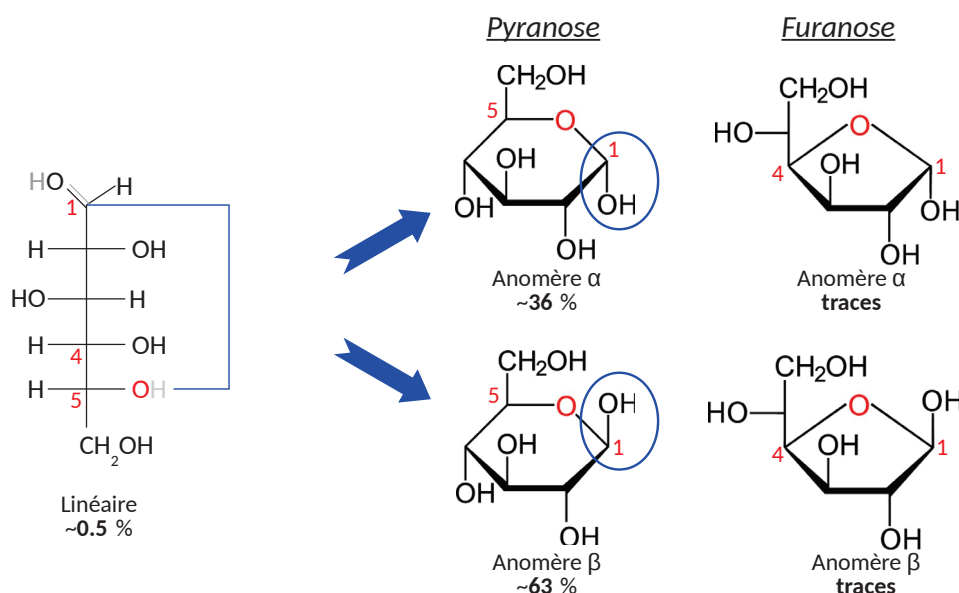


FIGURE 4 – Anomères α et β pour les formes pyranose et furanose du glucose.

Le cycle n'étant pas plan il existe alors différentes configurations spatiales de ce cycle et ce pour chaque anomère. Ces configurations sont appelées des conformères. La figure 5 montre quelques exemples de conformères possibles pour l'anomère β du glucose. Il y a les structures les plus communes, les chaises 4C_1 et 1C_4 . Le 4 en exposant de la structure 4C_1 indique que le carbone en position 4 est au-dessus du plan formé par l'oxygène du

cycle, les carbones en position 2, 3 et 5. Le plan est orienté par le carbone en position 6, dire que le carbone en position 4 est au-dessus du plan signifie qu'il est du même côté que le carbone en position 6 par rapport à ce plan. Le 1 en indice indique que le carbone en position 1 est en-dessous (de l'autre côté du carbone en position 6 par rapport au plan) de ce plan. C'est le contraire pour la structure 1C_4 . La structure ${}^{1-4}B$ est une structure en bateau. Les carbones en position 1 et en position 4 étant au-dessus du plan formés par les autres atomes du cycle explique le 1 – 4 en exposant. La notation des autres structures (1S_5 , 6H_O et 1E) se comprend de la même manière que pour les structures précédentes. Le S indique que la structure est une structure *skew*, le H indique que c'est une demi chaise et le E signifie que c'est une enveloppe. Pour l'enveloppe, on n'a plus qu'un numéro dans le nom cela veut dire que les 5 autres atomes du cycle sont dans le même plan.

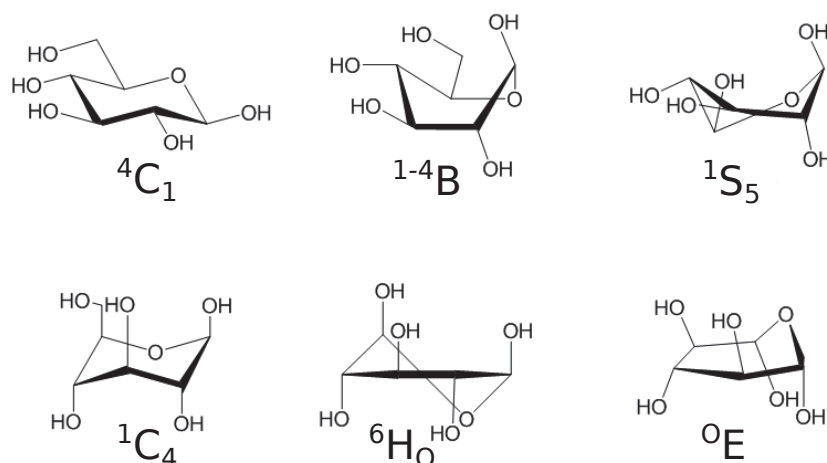


FIGURE 5 – Exemples de conformères pour l'anomère β du glucose

Les monosaccharides peuvent être reliés ensemble par des liaisons osidiques. Des polymères de monosaccharides de petite taille, c'est-à-dire jusqu'à une dizaine de monosaccharides sont appelés des oligosaccharides. Le plus petit oligosaccharide est un disaccharide constitué de deux monosaccharides. Les polymères de plus grand taille sont appelés des polysaccharides. La figure 6 montre toutes les configurations de liaisons osidiques possibles. Sur cette figure les deux monosaccharides sont des glucoses mais cela est valable pour tout autre monosaccharide. La liaison s'effectue entre le groupement hydroxyle en position 1 d'un monosaccharide et l'hydrogène de n'importe quelle autre position possible d'un autre monosaccharide. Une molécule d'eau sera libérée lors de la création de la liaison. La nomenclature de la liaison tient compte de la stéréochimie du groupement en position 1 plus les positions des deux carbones impliqués.

La structure en trois dimensions des oligosaccharides et des polysaccharides sera déterminée par les liaisons osidiques (et les liaisons hydrogènes des groupements hydroxyles ou autres) entre les monosaccharides qui les constituent. Par exemple la cellulose qui compose les parois végétales est une répétition de glucoses liés entre eux par une liaison $1 \rightarrow 4$ en configuration β (voir figure 7). La cellulose forme alors une longue chaîne linéaire. Un autre polysaccharide important pour l'être humain est le glycogène. C'est le réservoir à énergie pour le corps. Il se compose de chaînes de glucoses liés entre eux par une liaison $1 \rightarrow 4$ en configuration α et tous les 8 ou 12 glucoses cette chaîne possède des ramifications. Ce sont d'autres chaînes de glucoses liés entre eux par des liaisons $1 \rightarrow 4$ en configuration α mais la ramification se fait par une liaison $1 \rightarrow 6$ en configuration α (voir figure 8).

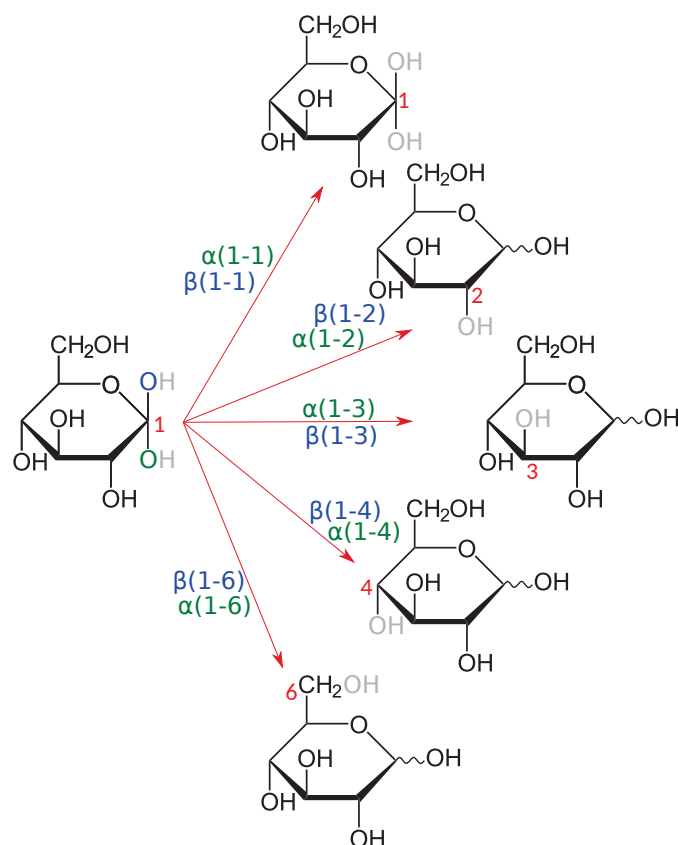


FIGURE 6 – Les différentes configurations possibles de liaisons osidiques entre deux monosaccharides.

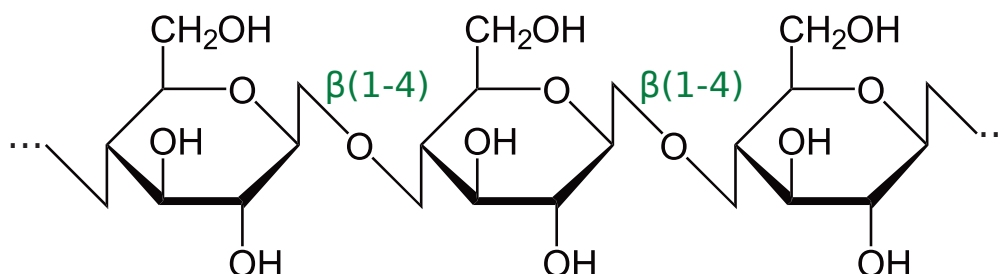


FIGURE 7 – Représentation de la cellulose

L'être humain ne possède pas d'enzymes capables de dégrader la liaison $\beta(1 \rightarrow 4)$ de la cellulose mais possède des enzymes pour dégrader la liaison $\alpha(1 \rightarrow 4)$ du glycogène. La liaison osidique joue donc un rôle important pour les organismes vivants.

Pour connaître un oligosaccharide, il faut connaître les monosaccharides qui les composent et les liaisons osidiques qui relient ces monosaccharides entre eux. Baptiste Schindler a montré dans sa thèse [11] que la nature des monosaccharides et la stéréochimie de la liaison osidique étaient préservées lors de la fragmentation d'un disaccharide. En caractérisant le disaccharide, on va caractériser la liaison osidique. De plus, un oligosaccharide peut être découpé en mono et disaccharides. Ainsi à partir de l'étude des mono et disaccharides on va pouvoir étudier des oligosaccharides. En établissant une base de données de spectres IRMPD et de structures et de spectres calculés correspondants pour des monosaccharides et des disaccharides, on pourra séquencer des oligosaccharides.

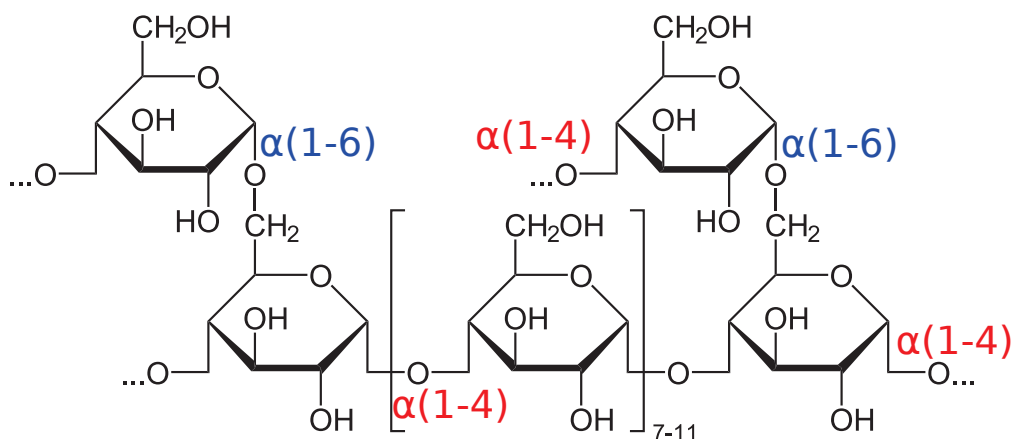


FIGURE 8 – Représentation du glycogène

Organisation du manuscrit

Les résultats sur des monosaccharides et des disaccharides obtenus dans le cadre de ma thèse effectuée dans l'équipe Physico-chimie théorique de l'Institut Lumière Matière seront présentés dans ce manuscrit. Le travail effectué dans cette thèse a été mené en étroite collaboration avec des membres de l'équipe Structure & Dynamique Multi-échelles des Édifices Moléculaires de l'Institut Lumière Matière et de l'équipe de Chimie Organique et Bioorganique de l'Institut de Chimie et Biochimie Moléculaires et Supramoléculaires. Mon travail était consacré exclusivement à l'étude théorique des saccharides alors que Baptiste Schindler s'est chargé des développements expérimentaux. Avant mon arrivée, très peu de résultats théoriques ont été obtenus dans le cadre dans cette collaboration [32]. L'essentiel des travaux théoriques ont donc été obtenus dans le cadre de ma thèse. Le manuscrit est constitué de quatre parties. La première partie présente le contexte général de mon travail. Dans la seconde partie, constituée des chapitres 1 à 3, les méthodes théoriques utilisées seront expliquées et dans la troisième partie, chapitres 4 à 8 nous verrons les résultats obtenus pour des mono et des disaccharides. Et enfin la partie 4 résumera les travaux effectués et proposera des perspectives à ces travaux.

Dans le chapitre 1, en nous plaçant dans le cadre de l'approximation de Born-Oppenheimer nous présenterons plusieurs méthodes pour résoudre l'équation de Schrödinger pour les électrons. La méthode Hartree-Fock sera développée en premier. Nous parlerons ensuite des méthodes post Hartree-Fock telles que la méthode de Moller-Plesset, des interactions de configuration et des clusters couplés. Nous présenterons ensuite la méthode que nous avons essentiellement utilisée pour produire nos résultats, à savoir la méthode de la théorie de la fonctionnelle de la densité (DFT). Les méthodes semi-empiriques (une de ces méthodes a servi pour la recherche de conformation) basées sur la méthode Hartree-Fock seront présentées de même que la théorie "Density-Functional based Tight-Binding" (DFTB) qui est une méthode approchée à partir de la DFT. Finalement, nous présenterons la mécanique moléculaire et nous détaillerons l'expression de l'énergie pour le champ de force MMFF94 (Merck Molecular Force Field 1994, utilisé dans une méthode hybride présentée dans le chapitre 3).

Dans le chapitre 2, nous présenterons brièvement la méthode de la dynamique moléculaire qui repose sur le principe fondamental de la dynamique. Nous discuterons de manière succincte des différentes étapes de la dynamique, des thermostats que l'on peut utiliser et de quelques intégrateurs qui permettent de résoudre les équations de la dynamique

moléculaire. Nous détaillerons ensuite notre méthode de recherche de conformations qui a permis d'obtenir un grand nombre de structures pour chaque saccharide que nous avons étudié. Nous avons utilisé majoritairement la dynamique moléculaire à haute température pour explorer les surfaces d'énergies potentielles et sélectionner des structures

Dans le chapitre 3, toujours dans l'approximation de Born-Oppenheimer nous verrons comment résoudre l'équation de Schrödinger pour la partie nucléaire. C'est cette résolution qui donne le spectre infrarouge de la structure considérée. La double approximation harmonique utilisée pour résoudre l'équation de Schrödinger nucléaire n'est pas suffisante pour reproduire l'expérience. Nous discuterons alors de différentes méthodes pour prendre en compte l'anharmonicité du potentiel. Nous décrirons dans le détail les méthodes que nous avons le plus utilisées comme la méthode des *scaling factor* et la méthode VPT2 et moins la méthode VSCF et la dynamique moléculaire. Nous avons choisi d'étudier des monosaccharides parmi les plus abondants ou les plus importants dans la nature. Cela comprend des monosaccharides avec un groupement amine (GlcN, GalN et ManN), qui est souvent le lieu de nouvelles substitutions, avec un groupement N-acétyl (GlcNAc, GalNAc et ManNAc), très abondants, ou encore avec un groupement sulfate (GlcN-6S et GlcN-3S) qui a un rôle très important dans les fonctions biologiques. La figure 9 montre tous les monosaccharides évoqués. Enfin, nous avons choisi des disaccharides possédant des liaisons osidiques 1 $\rightarrow$ 4 et 1 $\rightarrow$ 6. Tous les mono et disaccharides ont été étudiés dans un état de charge non nul. En effet, la spectroscopie de masse et IRMPD ne peuvent être réalisées que pour des systèmes chargés.

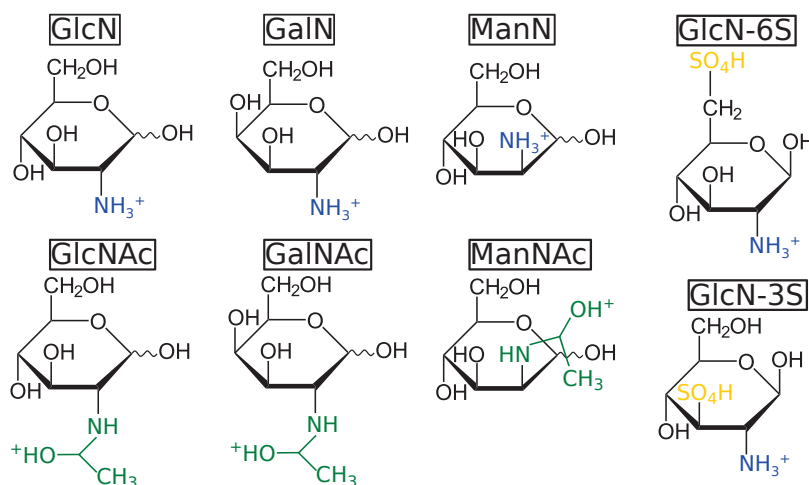


FIGURE 9 – Structures des monosaccharides étudiés dans leur état de charge protoné.

Dans le chapitre 4, nous étudierons le choix de la fonctionnelle et de la base pour reproduire le spectre expérimental de la glucosamine-6 sulfate dans la zone de 700 à 1800 cm^{-1} . La qualité du couple fonctionnelle/base sera évaluée sur les fréquences expérimentales de certains modes de vibration, en particulier les modes d'élongation S-O. Nous discuterons aussi de la prise en compte explicite de l'anharmonicité pour reproduire le motif expérimental en ayant recours à une méthode hybride (introduite dans le chapitre 3).

Dans le chapitre 5, nous nous intéresserons à des monosaccharides substitués par un groupement sulfate dans deux états de charge (protoné et déprotoné). Nous étudierons l'isomérisation de position avec la glucosamine-6 sulfate et la glucosamine-3 sulfate. À l'aide des spectres IRMPD et théoriques nous mettrons en évidence des signatures spécifiques aux isomères dans leur état protoné. Nous montrerons aussi des signatures de l'isomérisation

de position lorsque les monosaccharides sulfatés sont dans leur état de charge déprotoné.

Dans le chapitre 6, nous poursuivrons notre étude sur d'autres monosaccharides non sulfatés mais possédant un groupement N-acétyl en position 2. Ce groupement possédant deux sites de protonation, nous verrons à quelle endroit s'effectue le plus favorablement la protonation. Dans ce chapitre nous aborderons aussi la question du mélange anomérique. Nous montrerons que les signatures des anomères sont différentes et donc que les anomères contribuent différemment au spectre expérimental. Des expériences sur les anomères séparés (après méthylation) confirmeront cela.

Dans le chapitre 7, nous discuterons d'une autre étude sur le mélange dans l'échantillon sur les mêmes monosaccharides du chapitre précédent mais pour lesquels le groupement N-acétyl est remplacé par une fonction amine. Nous verrons que les calculs théoriques indiquent une hétérogénéité de structures dans le mélange. Plusieurs conformations des anomères peuvent être présentes dans l'échantillon. Afin d'avoir une idée des populations de ces conformations nous explorerons une approche pour calculer le pourcentage des conformations. Elle repose sur les spectres IRMPD et les spectres théoriques et sera présentée en détail dans ce chapitre. C'est la première fois qu'une telle approche est utilisée à partir des spectres IRMPD pour obtenir la contribution relative des constituants du mélange.

Dans le chapitre 8, nous discuterons de deux approches pour caractériser la liaison osidique dans les disaccharides. La première approche repose sur l'étude des modes de vibrations C-O-C constituant la signature directe de la liaison osidique. L'autre approche consiste à étudier les modes d'élongation NH et OH soit la signature indirecte. Nous verrons que l'approche directe ne permet pas de caractériser la liaison osidique et que l'approche indirecte est limitée par la résolution expérimentale dans certains cas.

Références

- [1] Daniel B. Werz, René Ranzinger, Stephan Herget, Alexander Adibekian, Claus-Wilhelm von der Lieth, and Peter H. Seeberger. Exploring the structural diversity of mammalian carbohydrates ("glycospace") by statistical databank analysis. *ACS Chem. Biol.*, 2(10) :685–691, 2007.
- [2] F. Sanger. The free amino groups of insulin. *Biochem J*, 39(5) :507–515, 1945.
- [3] F. Sanger. Chemistry of insulin. *Science*, 129(3359) :1340–1344, 1959.
- [4] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids : A structure for deoxyribose nucleic acid. *Nature*, 171(4356) :737–738, 1953.
- [5] F. Sanger, G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison, P. M. Slocombe, and M. Smith. Nucleotide sequence of bacteriophage ϕ X174 DNA. *Nature*, 265(5596) :687–695, 1977.
- [6] F. Sanger, S. Nicklen, and A. R. Coulson. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.*, 74(12) :5463–5467, 1977.

-
- [7] Prescott L. Deininger. Random subcloning of sonicated DNA : Application to shotgun DNA sequence analysis. *Anal. Biochem.*, 129(1) :216–223, 1983.
 - [8] Initial sequencing and analysis of the human genome. *Nature*, 409(6822) :860–921, 2001.
 - [9] Michael. Karas and Franz. Hillenkamp. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.*, 60(20) :2299–2301, 1988.
 - [10] JB Fenn, M Mann, CK Meng, SF Wong, and CM Whitehouse. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926) :64–71, 1989.
 - [11] Baptiste Schindler. *Caractérisation structurale et séquençage de carbohydrates par spectroscopie infrarouge intégrée à la spectrométrie de masse*. PhD thesis, Université Claude Bernard Lyon 1, 2016.
 - [12] L. P. Kuhn. Infrared spectra of carbohydrates. *Anal. Chem.*, 22(2) :276–283, 1950.
 - [13] John J. Cael, Jack L. Koenig, and John Blackwell. Infrared and raman spectroscopy of carbohydrates. *Carbohydr. Res.*, 32(1) :79–91, 1974.
 - [14] Armelle Phalipon, Corina Costachel, Cyrille Grandjean, Audrey Thuizat, Catherine Guerreiro, Myriam Tanguy, Farida Nato, Brigitte Vulliez-Le Normand, Frédéric Bélot, Karen Wright, Véronique Marcel-Peyre, Philippe J. Sansonetti, and Laurence A. Mulard. Characterization of functional oligosaccharide mimics of the shigella flexneri serotype 2a o-antigen : Implications for the development of a chemically defined glycoconjugate vaccine. *J. Immunol.*, 176(3) :1686–1694, 2006.
 - [15] Andrei V. Perepelov, Maria E. Shekht, Bin Liu, Sergei D. Shevelev, Vladimir A. Ledov, Sof’ya N. Senchenkova, Vyacheslav L. L’vov, Alexander S. Shashkov, Lu Feng, Petr G. Aparin, Lei Wang, and Yuriy A. Knirel. Shigella flexneri O-antigens revisited : final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity. *FEMS Immunol. Med. Microbiol.*, 66(2) :201, 2012.
 - [16] Flitsch et al. A roadmap for Glycoscience in Europe - a joint EGSF/IBCarb publication. 2015.
 - [17] National Research Council. Transforming glycoscience : A roadmap for the future. 2012.
 - [18] Darin J. Brown, Sarah E. Stefan, Giel Berden, Jeffrey D. Steill, Jos Oomens, John R. Eyler, and Brad Bendiak. Direct evidence for the ring opening of monosaccharide anions in the gas phase : photodissociation of aldohexoses and aldohexoses derived from disaccharides using variable-wavelength infrared irradiation in the carbonyl stretch region. *Carbohydrate Research*, 346(15) :2469–2481, 2011.
 - [19] Nick C. Polfer, Jose J. Valle, David T. Moore, Jos Oomens, John R. Eyler, and Brad Bendiak. Differentiation of isomers by wavelength-tunable infrared multiple-photon dissociation-mass spectrometry : Application to glucose-containing disaccharides. *Anal. Chem.*, 78(3) :670–679, 2006.
 - [20] Emilio B. Cagmat, Jan Szczepanski, Wright L. Pearson, David H. Powell, John R. Eyler, and Nick C. Polfer. Vibrational signatures of metal-chelated monosaccharide epimers : gas-phase infrared spectroscopy of Rb⁺-tagged glucuronic and iduronic acid. *Phys. Chem. Chem. Phys.*, 12 :3474–3479, 2010.

-
- [21] Cesar S. Contreras, Nicolas C. Polfer, Jos Oomens, Jeffrey D. Steill, Brad Bendiak, and John R. Eyler. On the path to glycan conformer identification : Gas-phase study of the anomers of methyl glycosides of *N*-acetyl-D-glucosamine and *N*-acetyl-D-galactosamine. *International Journal of Mass Spectrometry*, 330–332 :285–294, 2012.
- [22] Emilio J. Cocinero, Pierre Çarçabal, Timothy D. Vaden, Benjamin G. Davis, and John P. Simons. Exploring carbohydrate-peptide interactions in the gas phase : Structure and selectivity in complexes of pyranosides with n-acetylphenylalanine methylamide. *Journal of the American Chemical Society*, 133(12) :4548–4557, 2011. PMID : 21375248.
- [23] E. Cristina Stanca-Kaposta, Pierre Çarçabal, Emilio J. Cocinero, Paola Hurtado, and John P. Simons. Carbohydrate–aromatic interactions : Vibrational spectroscopy and structural assignment of isolated monosaccharide complexes with p-hydroxy toluene and n-acetyl l-tyrosine methylamide. *The Journal of Physical Chemistry B*, 117(27) :8135–8142, 2013. PMID : 23773008.
- [24] Pierre Carcabal, Emilio J. Cocinero, and John P. Simons. Binding energies of microhydrated carbohydrates : measurements and interpretation. *Chem. Sci.*, 4 :1830–1836, 2013.
- [25] Oscar Hernandez, Samantha Isenberg, Vincent Steinmetz, Gary L. Glish, and Philippe Maitre. Probing mobility-selected saccharide isomers : Selective ion-molecule reactions and wavelength-specific IR activation. *J. Phys. Chem. A*, 119(23) :6057–6064, 2015.
- [26] Waldemar Hoffmann, Johanna Hofmann, and Kevin Pagel. Energy-resolved ion mobility-mass spectrometry—a concept to improve the separation of isomeric carbohydrates. *Journal of The American Society for Mass Spectrometry*, 25(3) :471–479, Mar 2014.
- [27] J. Hofmann, H. S. Hahm, P. H. Seeberger, and K. Pagel. Identification of carbohydrate anomers using ion mobility-mass spectrometry. *Nature*, 526(7572) :241–244, Oct 2015. Letter.
- [28] H. Hinneburg, J. Hofmann, W. B. Struwe, A. Thader, F. Altmann, D. Varon Silva, P. H. Seeberger, K. Pagel, and D. Kolarich. Distinguishing n-acetylneuraminic acid linkage isomers on glycopeptides by ion mobility-mass spectrometry. *Chem. Commun.*, 52 :4381–4384, 2016.
- [29] Neelam Khanal, Chiara Masellis, Michael Z. Kamrath, David E. Clemmer, and Thomas R. Rizzo. Glycosaminoglycan analysis by cryogenic messenger-tagging ir spectroscopy combined with ims-ms. *Analytical Chemistry*, 89(14) :7601–7606, 2017. PMID : 28636333.
- [30] Chiara Masellis, Neelam Khanal, Michael Z. Kamrath, David E. Clemmer, and Thomas R. Rizzo. Cryogenic vibrational spectroscopy provides unique fingerprints for glycan identification. *Journal of The American Society for Mass Spectrometry*, 28(10) :2217–2222, Oct 2017.
- [31] J. Lehmann and A.H. Haines. *Carbohydrates : Structure and Biology*. Wiley, 2002.

-
- [32] Baptiste Schindler, Janhavi Joshi, Abdul-Rahman Allouche, Daniel Simon, Stephane Chambert, Vincent Brites, Marie-Pierre Gageot, and Isabelle Compagnon. Distinguishing isobaric phosphated and sulfated carbohydrates by coupling of mass spectrometry with gas phase vibrational spectroscopy. *Phys. Chem. Chem. Phys.*, 16 :22131–22138, 2014.

Deuxième partie

Méthodes théoriques

Chapitre 1

Méthodes de la chimie quantique

1.1 Introduction

En mécanique quantique, l'état d'un système nous est connu grâce à la fonction d'onde de ce système (premier postulat de la mécanique quantique). La connaissance de la fonction d'onde du système permet de calculer l'énergie d'un système et de remonter aux propriétés du système. L'évolution dans le temps et dans l'espace d'un système est régit par l'équation de Schrödinger introduite par Erwin Schrödinger en 1926 (cinquième postulat).

Cette équation s'écrit de la manière suivante :

$$\hat{H}_{mol}\Psi(\vec{r}, t) = (\hat{T}_{mol} + \hat{V}_{mol}(\vec{r}, t))\Psi(\vec{r}, t) = i\hbar\frac{\partial\Psi}{\partial t} \quad (1.1)$$

Dans le cas stationnaire, on a l'équation de Schrödinger indépendante du temps.

$$\hat{H}_{mol}\Psi(\vec{r}) = (\hat{T}_{mol} + \hat{V}_{mol}(\vec{r}))\Psi(\vec{r}) = E(\vec{r})\Psi(\vec{r}) \quad (1.2)$$

L'hamiltonien d'un système moléculaire 1.1 composé de M noyaux et N électrons s'écrit :

$$\begin{aligned} \hat{H}_{mol} = & \underbrace{-\sum_{i=1}^N \frac{1}{2m_e} \nabla_i^2}_{\text{énergie cinétique électronique}} + \underbrace{\sum_{i=1}^N \sum_{j>i}^N \frac{e^2}{4\pi\epsilon_0 |r_1 - r_2|}}_{\text{répulsion électronique}} - \underbrace{\sum_{A=1}^M \sum_{i=1}^N \frac{Z_A e}{4\pi\epsilon_0 |R_A - r_1|}}_{\text{attraction nucléaire}} \\ & \underbrace{-\sum_{A=1}^M \frac{1}{2m_A} \nabla_A^2}_{\text{énergie cinétique nucléaire}} + \underbrace{\sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B e^2}{4\pi\epsilon_0 |R_A - R_B|}}_{\text{répulsion nucléaire}} \\ \hat{H}_{mol} = & T_e + V_{ee} + T_{en} + T_n + V_{nn} \end{aligned} \quad (1.3)$$

avec N le nombre d'électrons du système et M le nombre de noyaux. $\nabla_i = \frac{\partial}{\partial r_i}$ et $\nabla_A = \frac{\partial}{\partial R_A}$ sont les dérivées partielles par rapport aux coordonnées d'espace pour les électrons et les noyaux, respectivement. Les indices i,j indiquent que l'on itère sur les électrons et les indices A et B sur les noyaux. m_e est la masse de l'électron et m_A la masse du noyau A.

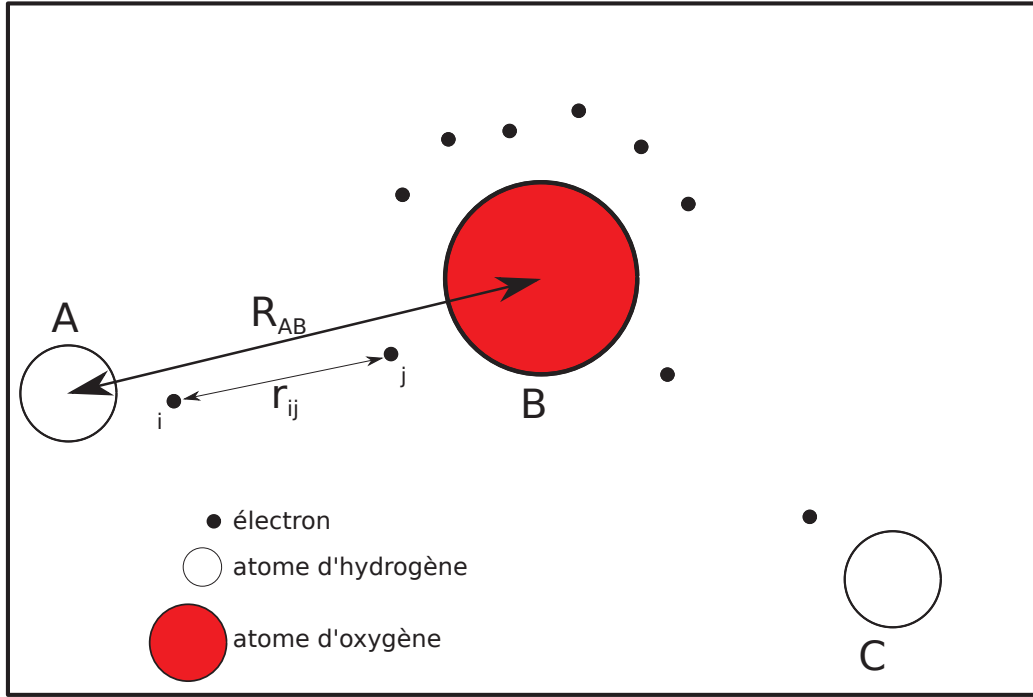


FIGURE 1.1 – Représentation classique d'un système moléculaire pour la molécule d'eau.

L'énergie du système est calculée à partir de la formule suivante :

$$E = \frac{\langle \Psi | \hat{H}_{mol} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

Ce problème ne peut être résolu analytiquement que pour un nombre limité de cas :

- l'atome d'hydrogène, les fonctions d'ondes propres sont les harmoniques sphériques
- la molécule H_2^+

Le problème peut être résolu numériquement pour des systèmes de petite taille seulement. Le coût calculatoire est trop important pour résoudre des systèmes au-delà de quelques atomes. Pour étudier des molécules de plus grosse taille il faut faire des approximations. L'approximation la plus importante sur laquelle repose bon nombre de méthodes de la chimie quantique est l'approximation de Born-Oppenheimer [1]. Du fait de la masse plus importante des noyaux par rapport aux électrons, on considère que les noyaux sont fixes dans le référentiel des électrons. Utiliser cette approximation revient à négliger tout couplage entre les électrons et les noyaux, les électrons s'adapteront immédiatement à n'importe quel changement dans la configuration nucléaire. On peut ainsi séparer le problème électronique du problème nucléaire. Ainsi, dans le problème électronique l'énergie cinétique des noyaux sera prise nulle. Les coordonnées d'espace nucléaires seront des paramètres lors de la résolution du problème électronique. L'énergie cinétique des noyaux sera réintroduite dans le problème nucléaire et l'énergie électronique contribuera au potentiel ressenti par les noyaux. Ainsi la fonction d'onde totale du système, Ψ , est le produit des fonctions d'ondes électronique et nucléaire et l'hamiltonien total peut s'écrire comme la

somme des deux hamiltoniens des sous-problèmes. On obtient les équations suivantes.

$$\begin{aligned}
\Psi &= \Psi_e \cdot \Psi_n = \Psi_e(\{\vec{r}\}; \{\vec{R}\}) \cdot \Psi_n(\{\vec{R}\}) \\
\hat{H}_{mol} &= \hat{H}_e + \hat{H}_n \\
\hat{H}_e \Psi_e &= E_e \Psi_e \\
\hat{H}_n \Psi_n &= E \Psi_n \\
\hat{H}_e &= T_e + V_{ee} + V_{en} \\
\hat{H}_n &= T_n + E_e + V_{nn}
\end{aligned} \tag{1.4}$$

Les problèmes électronique et nucléaire sont résolus séparément. Dans ce chapitre nous décrirons des méthodes de résolution de l'équation de Schrödinger pour la partie électronique, la partie nucléaire sera abordée dans les prochains chapitres.

La fonction d'onde électronique est construite à partir du produit de fonctions d'ondes monoélectronique (c'est l'approximation orbitalaire). Afin de respecter le principe de Pauli (deux électrons ne peuvent pas se trouver dans le même état quantique), la fonction d'onde s'écrit à l'aide d'un déterminant de Slater.

$$\Psi_e = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\vec{r}_1) & \phi_1(\vec{r}_2) & \phi_1(\vec{r}_3) & \cdots & \phi_1(\vec{r}_N) \\ \phi_2(\vec{r}_1) & \phi_2(\vec{r}_2) & \phi_2(\vec{r}_3) & \cdots & \phi_2(\vec{r}_N) \\ \phi_3(\vec{r}_1) & \phi_3(\vec{r}_2) & \phi_3(\vec{r}_3) & \cdots & \phi_3(\vec{r}_N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_N(\vec{r}_1) & \phi_N(\vec{r}_2) & \phi_N(\vec{r}_3) & \cdots & \phi_N(\vec{r}_N) \end{vmatrix}.$$

Nous verrons par la suite que pour certaines méthodes la fonction d'onde électronique est construite à partir de plusieurs déterminants de Slater correspondant à des configurations où certaines électrons sont excités.

Un grand nombre de méthodes ont été développées pour résoudre la structure électronique. L'amélioration de la précision s'accompagne malheureusement d'un temps de calcul plus important. L'une des premières méthodes à avoir été conçue est la méthode de Hartree-Fock. Dans cette méthode, le potentiel inter-électronique qui est le terme qui empêche la résolution analytique du problème est moyenné afin de transformer l'hamiltonien total multiélectronique en un ensemble de systèmes monoélectroniques. La fonction d'onde électronique est construite à partir d'un seul déterminant de Slater. Cependant cette méthode ne prend pas en compte la corrélation électronique qui bien qu'elle ne représente qu'un très faible pourcentage de l'énergie totale n'en reste pas moins fondamentale dans l'obtention de certaines propriétés (par exemple la dissociation de la molécule de H_2). Des méthodes plus robustes ont vu le jour pour combler cette faiblesse. Ce sont les méthodes que l'on appelle méthodes post Hartree-Fock. Pour ces méthodes la fonction d'onde électronique est construite à partir de plusieurs déterminants de Slater. Certaines méthodes reposent sur la théorie de la perturbation ou sur un principe variationnel pour introduire la corrélation électronique. L'inconvénient de ces méthodes est qu'elles présentent un coût en temps de calcul qui peut rapidement devenir prohibitif si l'on veut une grande précision. La théorie de la fonctionnelle de la densité (DFT) a vu le jour afin de prendre en compte différemment la corrélation électronique pour diminuer le temps de calcul sans que la précision s'effondre. La fonction d'onde est prise à nouveau comme un déterminant de Slater mais l'énergie est une fonctionnelle de la densité électronique, une fonction à 3 variables d'espace et une de spin. Cela facilite grandement le problème. Cependant la forme analytique de cette fonctionnelle n'est pas connue alors il faut trouver une expression approchée. Le temps de calcul de la DFT est de l'ordre de celui de la méthode Hartree-Fock,

inférieur à celui des méthodes post Hartree-Fock. Pour des très gros systèmes et une précision convenable elle peut coûter cher en temps de calcul. Il existe des méthodes dites semi-empiriques dérivées de la méthode Hartree-Fock qui négligent certaines intégrales et remplacent les intégrales restantes par des paramètres. Ce sont des méthodes très rapides mais moins précises. La méthode DFTB est une méthode paramétrée dont les paramètres sont ajustés à partir de la DFT. Enfin des méthodes purement empiriques ont vu le jour. Les paramètres sont ajustés sur des résultats *ab initio* ou des mesures expérimentales. Elles sont extrêmement rapides mais moins précises. Elles sont très utilisées pour étudier des grosses protéines par exemple.

Toutes ces méthodes seront discutées dans la suite.

1.2 Hartree-Fock

Nous allons maintenant présenter la méthode de Hartree-Fock. Bien que cette dernière n'a pas été utilisée directement au cours de ma thèse, il est important de la présenter pour comprendre les autres méthodes basées sur celle-ci.

Si on négligeait le terme de couplage entre électrons dans l'hamiltonien évoqué précédemment alors le problème s'en trouverait grandement simplifié. L'hamiltonien ainsi obtenu serait alors constitué de contributions monoélectroniques et des contributions nucléaires que l'on peut écrire comme la somme d'hamiltonien à un électron. La fonction d'onde du système pourrait s'écrire comme le produit des fonctions d'ondes monoélectroniques. C'est la méthode de Hartree.

Dans la méthode Hartree-Fock, on ne néglige pas le terme de couplage inter-électronique mais on écrit tout de même la fonction d'onde du système comme un produit de fonctions d'ondes monoélectroniques, c'est l'approximation orbitale. On cherchera à écrire l'hamiltonien comme la somme d'hamiltoniens monoélectroniques. La fonction d'onde d'un système de fermions doit respecter le principe de Pauli. Elle doit être antisymétrique dans l'échange de deux fermions. La fonction d'onde s'écrit sous la forme d'un déterminant de Slater des fonctions d'ondes monoélectroniques pour respecter ce principe.

Elle peut aussi s'écrire sous une forme condensée.

$$\begin{aligned}\Psi_e &= \frac{1}{\sqrt{N!}} \sum_{p=0}^{N-1} (-1)^p \mathcal{P}_p [\phi_1(\vec{r}_1) \phi_2(\vec{r}_2) \cdots \phi_N(\vec{r}_N)] \\ \Psi_e &= \frac{1}{\sqrt{N!}} [\mathbb{1} - \sum_{ij} P_{ij} + \sum_{ijk} P_{ijk} + \dots] [\phi_1(\vec{r}_1) \phi_2(\vec{r}_2) \cdots \phi_N(\vec{r}_N)] \\ \Psi_e &= \mathcal{A} [\phi_1(\vec{r}_1) \phi_2(\vec{r}_2) \cdots \phi_N(\vec{r}_N)]\end{aligned}$$

$\phi_i(\vec{r}_1)$ est une spin-orbitale, c'est le produit de la fonction spatiale $\varphi_i(\vec{r}_1)$ et la fonction de spin $\sigma_i(s_i)$.

L'opérateur d'antisymétrisation $\mathcal{A}$ est égal à $\frac{1}{\sqrt{N!}} \sum_{p=0}^{N-1} (-1)^p \mathcal{P}_p$. Il s'écrit comme une somme d'opérateurs de permutations des indices des orbitales. Les opérateurs $\mathcal{P}_i$ sont les opérateurs de permutation à $i + 1$ indices. L'opérateur de permutation $\mathcal{P}_1$ est l'opérateur à deux indices. Il permute les indices de deux orbitales. Par exemple, P_{12} appliqué sur $\phi_1(\vec{r}_1) \phi_2(\vec{r}_2)$ conduira au résultat $\phi_2(\vec{r}_1) \phi_1(\vec{r}_2)$. L'opérateur $\mathcal{P}_0$ est simplement l'opérateur identité. Le produit de l'opérateur d'antisymétrisation avec lui-même donne $\mathcal{A}\mathcal{A} = \sqrt{N!}\mathcal{A} = \sum_{p=0}^{N-1} (-1)^p \mathcal{P}_p$. De plus, on peut montrer que $\mathcal{A}$ commute avec $\hat{H}_e$.

L'énergie électronique Hartree-Fock, E_e , s'exprime en fonction de cet opérateur.

$$\begin{aligned}
 E_e &= \langle \Psi_e | \hat{H}_e | \Psi_e \rangle \\
 E_e &= \langle \phi_1(\vec{r}_1) \phi_2(\vec{r}_2) \cdots \phi_N(\vec{r}_N) | \frac{1}{\sqrt{N!}} \sum_{p=0}^{N-1} (-1)^p \mathcal{P}_p \hat{H}_e \frac{1}{\sqrt{N!}} \sum_{p=0}^{N-1} (-1)^p \mathcal{P}_p | \phi_1(\vec{r}_1) \phi_2(\vec{r}_2) \cdots \phi_N(\vec{r}_N) \rangle \\
 E_e &= \langle \phi_1(\vec{r}_1) \phi_2(\vec{r}_2) \cdots \phi_N(\vec{r}_N) | \hat{H}_e \sum_{p=0}^{N-1} (-1)^p \mathcal{P}_p | \phi_1(\vec{r}_1) \phi_2(\vec{r}_2) \cdots \phi_N(\vec{r}_N) \rangle \\
 E_e &= \langle \phi_1(\vec{r}_1) \phi_2(\vec{r}_2) \cdots \phi_N(\vec{r}_N) | \hat{H}_e [\mathbb{1} - \sum_{ij} P_{ij} + \sum_{ijk} P_{ijk} + \dots] | \phi_1(\vec{r}_1) \phi_2(\vec{r}_2) \cdots \phi_N(\vec{r}_N) \rangle
 \end{aligned} \tag{1.5}$$

Pour obtenir la dernière ligne, les propriétés de $\mathcal{A}$ ont été utilisées (commutation avec $\hat{H}_e$ et $\mathcal{A}\mathcal{A} = \sqrt{N!}\mathcal{A}$).

$\hat{H}_e$ est l'hamiltonien du problème électronique évoqué précédemment. Son expression est donnée dans la formule suivante :

$$\hat{H}_e = - \sum_{i=1}^N \frac{1}{2m_e} \nabla_i^2 - \sum_{A=0}^M \sum_{i=1}^N \frac{Z_A e}{4\pi\epsilon_0 |R_A - r_i|} + \sum_{i=0}^N \sum_{j>i}^N \frac{e^2}{4\pi\epsilon_0 |r_i - r_j|} + \sum_{A=0}^M \sum_{B>A}^M \frac{Z_A Z_B e^2}{4\pi\epsilon_0 |R_A - R_B|}$$

Il peut alors s'écrire comme la somme de trois termes : une somme d'opérateurs monoélectroniques plus la somme d'opérateurs biélectroniques et enfin une constante qui correspond aux répulsions coulombiennes entre noyaux.

$$\begin{aligned}
 \hat{H}_e &= \sum_{i=1}^N h_i + \sum_{i=1}^N \sum_{j>i}^N g_{ij} + V_{nn} \\
 h_i &= -\frac{1}{2m_e} \nabla_i^2 - \sum_{A=1}^M \frac{Z_A e}{4\pi\epsilon_0 |R_A - r_i|} \\
 g_{ij} &= \frac{e^2}{4\pi\epsilon_0 |r_i - r_j|} \\
 V_{nn} &= \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B e^2}{4\pi\epsilon_0 |R_A - R_B|}
 \end{aligned} \tag{1.6}$$

On peut ainsi remplacer $\hat{H}_e$ par sa formule dans l'équation de E_e

$$E_e = \langle \phi_1(\vec{r}_1) \cdots \phi_N(\vec{r}_N) | [\sum_{i=1}^N h_i + \sum_{i=0}^N \sum_{j>i}^N g_{ij} + V_{nn}] [\mathbb{1} - \sum_{lm} P_{lm} + \sum_{lmn} P_{lmn} + \dots] | \phi_1(\vec{r}_1) \cdots \phi_N(\vec{r}_N) \rangle \tag{1.7}$$

L'opérateur V_{nn} ne s'applique pas aux orbitales ϕ_i , par conséquent toute permutation des orbitales donnera un produit nul. Seul l'opérateur identité donne un résultat non nul.

En ce qui concerne l'opérateur monoélectronique h_i , toute permutation P_{ij} entraînera des termes du type : $\langle \phi_i(\vec{r}_1) | h_1 | \phi_j(\vec{r}_1) \rangle \langle \phi_j(\vec{r}_2) | \phi_i(\vec{r}_2) \rangle$ qui seront nuls pour obéir à la condition d'orthogonalité des orbitales. Pour toute autre permutation on sera dans la même situation conduisant à des termes qui s'annulent. Encore une fois, seul l'opérateur identité permet d'obtenir des termes non nuls. En suivant le même raisonnement, les termes qui subsisteront pour l'opérateur biélectronique seront ceux des opérateurs identité et P_{ij}

$$\langle \phi_1(\vec{r}_1) \cdots \phi_N(\vec{r}_N) | g_{ij} \cdot 1 | \phi_1(\vec{r}_1) \cdots \phi_N(\vec{r}_N) \rangle = \langle \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) | g_{12} | \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) \rangle = J_{ij} \quad (1.8)$$

$$\langle \phi_1(\vec{r}_1) \cdots \phi_N(\vec{r}_N) | g_{ij} \cdot P_{ij} | \phi_1(\vec{r}_1) \cdots \phi_N(\vec{r}_N) \rangle = \langle \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) | g_{12} | \phi_j(\vec{r}_1) \phi_i(\vec{r}_2) \rangle = K_{ij} \quad (1.9)$$

J_{ij} est appelée intégrale coulombienne parce que la notation de Dirac correspond à l'intégrale suivante :

$$\begin{aligned} J_{ij} &= \int \int \phi_i^*(\vec{r}_1) \phi_j^*(\vec{r}_2) \frac{e^2}{4\pi\epsilon_0 |\vec{r}_1 - \vec{r}_2|} \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) d\vec{r}_1 d\vec{r}_2 \\ &= \int \int |\phi_i(\vec{r}_1)|^2 \frac{e^2}{4\pi\epsilon_0 |\vec{r}_1 - \vec{r}_2|} |\phi_j(\vec{r}_2)|^2 d\vec{r}_1 d\vec{r}_2 \end{aligned} \quad (1.10)$$

C'est la répulsion coulombienne entre deux densités électroniques, $|\phi_i(\vec{r}_1)|^2$ et $|\phi_j(\vec{r}_2)|^2$.

K_{ij} est appelée intégrale d'échange. Il n'a pas d'équivalent classique.

On peut réécrire l'énergie E_e en utilisant les notations J_{ij} et K_{ij} .

$$\begin{aligned} E_e &= \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | h_1 | \phi_i(\vec{r}_1) \rangle + \sum_{i=1}^N \sum_{j>i}^N (J_{ij} - K_{ij}) + V_{nn} \\ &= \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | h_1 | \phi_i(\vec{r}_1) \rangle + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (J_{ij} - K_{ij}) + V_{nn} \end{aligned} \quad (1.11)$$

On vient de poser les équations pour calculer l'énergie électronique E_e . On ne connaît pas la véritable fonction d'onde du système mais on sait que toute fonction d'onde construite selon les orbitales ϕ_i donnera une énergie supérieure ou égale à l'énergie de l'état fondamental. Afin d'avoir la fonction d'onde la plus proche possible de la fonction d'onde réelle du système, il faut trouver le jeu d'orbitales ϕ_i qui minimise l'énergie. On utilise la méthode des multiplicateurs de Lagrange. Notre contrainte sera la conservation de la propriété d'orthonormalité des orbitales ϕ_i .

On doit minimiser le Lagrangien suivant :

$$L = E_e - \sum_{ij} (\lambda_{ij} \langle \phi_i(\vec{r}_1) | \phi_j(\vec{r}_1) \rangle - \delta_{ij}) \quad (1.12)$$

Le principe variationnel stipule que la variation δL doit être nulle. Quelle que soit la variation $|\delta\phi_i\rangle$ des orbitales $|\phi_i\rangle$, l'expression de δL est :

$$\delta L = \delta E_e - \sum_{ij} (\lambda_{ij} \langle \delta\phi_i(\vec{r}_1) | \phi_j(\vec{r}_1) \rangle + \lambda_{ij} \langle \phi_i(\vec{r}_1) | \delta\phi_j(\vec{r}_1) \rangle) \quad (1.13)$$

Il nous reste à calculer la variation de l'énergie électronique.

$$\begin{aligned}
 \delta E_e = & \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | h_1 | \phi_i(\vec{r}_1) \rangle + \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | h_1 | \delta \phi_i(\vec{r}_1) \rangle \\
 & + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N [\langle \delta \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) | g_{12} | \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) \rangle + \langle \phi_i(\vec{r}_1) \delta \phi_j(\vec{r}_2) | g_{12} | \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) \rangle \\
 & + \langle \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) | g_{12} | \delta \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) \rangle + \langle \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) | g_{12} | \phi_i(\vec{r}_1) \delta \phi_j(\vec{r}_2) \rangle \\
 & - (\langle \delta \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) | g_{12} | \phi_j(\vec{r}_1) \phi_i(\vec{r}_2) \rangle + \langle \phi_i(\vec{r}_1) \delta \phi_j(\vec{r}_2) | g_{12} | \phi_j(\vec{r}_1) \phi_i(\vec{r}_2) \rangle \\
 & + \langle \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) | g_{12} | \delta \phi_j(\vec{r}_1) \phi_i(\vec{r}_2) \rangle + \langle \phi_i(\vec{r}_1) \phi_j(\vec{r}_2) | g_{12} | \phi_j(\vec{r}_1) \delta \phi_i(\vec{r}_2) \rangle)] \quad (1.14)
 \end{aligned}$$

Pour simplifier l'expression on peut introduire dans l'équation (1.14) les opérateurs de Coulomb et d'échange.

$$J_i | \phi_j(\vec{r}_2) \rangle = \langle \phi_i(\vec{r}_1) | g_{12} | \phi_i(\vec{r}_1) \rangle | \phi_j(\vec{r}_2) \rangle \quad (1.15)$$

$$K_i | \phi_j(\vec{r}_2) \rangle = \langle \phi_i(\vec{r}_1) | g_{12} | \phi_j(\vec{r}_1) \rangle | \phi_i(\vec{r}_2) \rangle \quad (1.16)$$

A noter que l'application de l'opérateur J_i ne change pas l'orbitale sur laquelle on l'applique ($\phi_j(\vec{r}_2)$) contrairement à l'opérateur K_i qui l'échange avec $\phi_i(\vec{r}_2)$.

$$\begin{aligned}
 \delta E_e = & \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | h_1 | \phi_i(\vec{r}_1) \rangle + \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | h_1 | \delta \phi_i(\vec{r}_1) \rangle \\
 & + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N [\langle \delta \phi_i(\vec{r}_1) | J_j | \phi_i(\vec{r}_1) \rangle + \langle \delta \phi_j(\vec{r}_2) | J_i | \phi_j(\vec{r}_2) \rangle \\
 & + \langle \phi_i(\vec{r}_1) | J_j | \delta \phi_i(\vec{r}_1) \rangle + \langle \phi_j(\vec{r}_2) | J_i | \delta \phi_j(\vec{r}_2) \rangle \\
 & - (\langle \delta \phi_i(\vec{r}_1) | K_j | \phi_i(\vec{r}_1) \rangle + \langle \delta \phi_j(\vec{r}_2) | K_i | \phi_j(\vec{r}_2) \rangle \\
 & + \langle \phi_j(\vec{r}_2) | K_i | \delta \phi_j(\vec{r}_2) \rangle + \langle \phi_i(\vec{r}_1) | K_j | \delta \phi_i(\vec{r}_1) \rangle)] \quad (1.17)
 \end{aligned}$$

$$\begin{aligned}
 \delta E_e = & \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | h_1 | \phi_i(\vec{r}_1) \rangle + \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | h_1 | \delta \phi_i(\vec{r}_1) \rangle \\
 & + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N [\langle \delta \phi_i(\vec{r}_1) | (J_j - K_j) | \phi_i(\vec{r}_1) \rangle + \langle \delta \phi_j(\vec{r}_2) | (J_i - K_i) | \phi_j(\vec{r}_2) \rangle \\
 & + \langle \phi_i(\vec{r}_1) | (J_j - K_j) | \delta \phi_i(\vec{r}_1) \rangle + \langle \phi_j(\vec{r}_2) | (J_i - K_i) | \delta \phi_j(\vec{r}_2) \rangle] \quad (1.18)
 \end{aligned}$$

En sommant sur les indices i et j , on s'aperçoit que des termes sont identiques ce qui fait que le facteur $\frac{1}{2}$ se simplifie. On obtient finalement pour l'expression de la variation de l'énergie δE_e

$$\delta E_e = \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | h_1 + \sum_{j=1}^N (J_j - K_j) | \phi_i(\vec{r}_1) \rangle + \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | h_1 + \sum_{j=1}^N (J_j - K_j) | \delta \phi_i(\vec{r}_1) \rangle \quad (1.19)$$

On pose que $F_i = h_1 + \sum_{j=1}^N (J_j - K_j)$, c'est l'opérateur effectif de Fock monoélectronique.

$$\delta E_e = \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | F_i | \phi_i(\vec{r}_1) \rangle + \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | F_i | \delta \phi_i(\vec{r}_1) \rangle \quad (1.20)$$

Nous pouvons désormais remplacer δE dans l'expression de δL .

$$\delta L = \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | F_i | \phi_i(\vec{r}_1) \rangle + \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | F_i | \delta \phi_i(\vec{r}_1) \rangle - \sum_{ij} (\lambda_{ij} \langle \delta \phi_i(\vec{r}_1) | \phi_j(\vec{r}_1) \rangle + \lambda_{ij} \langle \phi_i(\vec{r}_1) | \delta \phi_j(\vec{r}_1) \rangle) \quad (1.21)$$

δL doit s'annuler quelque soit $|\delta \phi_j\rangle$. $\mathbf{i}|\delta \phi_j\rangle$ est aussi un candidat valable pour que $\delta L = 0$.

Remplaçons $|\delta \phi_j\rangle$ par $\mathbf{i}|\delta \phi_j\rangle$ dans l'équation (1.21) .

$$\begin{aligned} \delta L = & -\mathbf{i} \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | F_i | \phi_i(\vec{r}_1) \rangle + \mathbf{i} \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | F_i | \delta \phi_i(\vec{r}_1) \rangle \\ & - \sum_{ij} (-\mathbf{i} \lambda_{ij} \langle \delta \phi_i(\vec{r}_1) | \phi_j(\vec{r}_1) \rangle + \mathbf{i} \lambda_{ij} \langle \phi_i(\vec{r}_1) | \delta \phi_j(\vec{r}_1) \rangle) \end{aligned} \quad (1.22)$$

Comme les deux expressions, (1.21) et (1.22), sont égales à zéro en les sommant le résultat restera nul.

Calculons (1.21) + $\mathbf{i}$.(1.22)

$$\begin{aligned} 0 = & \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | F_i | \phi_i(\vec{r}_1) \rangle + \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | F_i | \phi_i(\vec{r}_1) \rangle + \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | F_i | \delta \phi_i(\vec{r}_1) \rangle - \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | F_i | \delta \phi_i(\vec{r}_1) \rangle \\ & - \sum_{ij} (\lambda_{ij} \langle \delta \phi_i(\vec{r}_1) | \phi_j(\vec{r}_1) \rangle + \lambda_{ij} \langle \phi_i(\vec{r}_1) | \delta \phi_j(\vec{r}_1) \rangle) - \sum_{ij} (\lambda_{ij} \langle \delta \phi_i(\vec{r}_1) | \phi_j(\vec{r}_1) \rangle - \lambda_{ij} \langle \phi_i(\vec{r}_1) | \delta \phi_j(\vec{r}_1) \rangle) \end{aligned} \quad (1.23)$$

Ainsi,

$$0 = \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | F_i | \phi_i(\vec{r}_1) \rangle - \sum_{ij} \lambda_{ij} \langle \delta \phi_i(\vec{r}_1) | \phi_j(\vec{r}_1) \rangle \quad (1.24)$$

Faisons la même chose maintenant mais en prenant, $\mathbf{i}$.(1.21) + (1.22)

$$\begin{aligned} 0 = & \mathbf{i} \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | F_i | \phi_i(\vec{r}_1) \rangle - \mathbf{i} \sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | F_i | \phi_i(\vec{r}_1) \rangle + \mathbf{i} \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | F_i | \delta \phi_i(\vec{r}_1) \rangle \\ & + \mathbf{i} \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | F_i | \delta \phi_i(\vec{r}_1) \rangle - \sum_{ij} (\mathbf{i} \lambda_{ij} \langle \delta \phi_i(\vec{r}_1) | \phi_j(\vec{r}_1) \rangle + \mathbf{i} \lambda_{ij} \langle \phi_i(\vec{r}_1) | \delta \phi_j(\vec{r}_1) \rangle) \\ & - \sum_{ij} (-\mathbf{i} \lambda_{ij} \langle \delta \phi_i(\vec{r}_1) | \phi_j(\vec{r}_1) \rangle + \mathbf{i} \lambda_{ij} \langle \phi_i(\vec{r}_1) | \delta \phi_j(\vec{r}_1) \rangle) \end{aligned} \quad (1.25)$$

Ainsi,

$$0 = \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | F_i | \delta \phi_i(\vec{r}_1) \rangle - \sum_{ij} \lambda_{ij} \langle \phi_i(\vec{r}_1) | \delta \phi_j(\vec{r}_1) \rangle \quad (1.26)$$

On sait que

$$\begin{aligned} \langle \phi_i(\vec{r}_1) | F_i | \delta \phi_i(\vec{r}_1) \rangle &= \langle \delta \phi_i(\vec{r}_1) | F_i | \phi_i(\vec{r}_1) \rangle^* \\ \langle \phi_i(\vec{r}_1) | \delta \phi_j(\vec{r}_1) \rangle &= \langle \delta \phi_j(\vec{r}_1) | \phi_i(\vec{r}_1) \rangle^* \end{aligned} \quad (1.27)$$

$$\begin{aligned} 0 &= \left(\sum_{i=1}^N \langle \delta \phi_i(\vec{r}_1) | F_i | \phi_i(\vec{r}_1) \rangle - \sum_{ij} \lambda_{ij} \langle \delta \phi_i(\vec{r}_1) | \phi_j(\vec{r}_1) \rangle \right)^* \\ &= \sum_{i=1}^N \langle \phi_i(\vec{r}_1) | F_i | \delta \phi_i(\vec{r}_1) \rangle - \sum_{ij} \lambda_{ji}^* \langle \phi_i(\vec{r}_1) | \delta \phi_j(\vec{r}_1) \rangle \end{aligned} \quad (1.28)$$

Or d'après (1.26),

$$\sum_{i=1}^N \langle \phi_i(\vec{r}_1) | F_i | \delta \phi_i(\vec{r}_1) \rangle = \sum_{ij} \lambda_{ij} \langle \phi_i(\vec{r}_1) | \delta \phi_j(\vec{r}_1) \rangle \quad (1.29)$$

On obtient une équation sur les facteurs de Lagrange.

$$0 = \sum_{ij} (\lambda_{ij} - \lambda_{ji}^*) \langle \phi_i(\vec{r}_1) | \delta \phi_j(\vec{r}_1) \rangle \quad (1.30)$$

Le terme $\lambda_{ji}^* - \lambda_{ij}$ est nul ce qui fait que les éléments λ_{ij} sont réels.

En reprenant (1.24), on a la condition suivante :

$$F_i | \phi_i(\vec{r}_1) \rangle = \sum_j \lambda_{ij} | \phi_j(\vec{r}_1) \rangle \quad (1.31)$$

On peut transformer les orbitales moléculaires (ϕ_i) par une transformation unitaire pour que la matrice des éléments λ_{ij} soit diagonale. Dans ce cas, on dira que les nouvelles orbitales moléculaires obtenues (ϕ'_i) sont canoniques.

$$F_i | \phi'_i(\vec{r}_1) \rangle = \lambda_{ii} | \phi'_i(\vec{r}_1) \rangle = \epsilon_i | \phi'_i(\vec{r}_1) \rangle \quad (1.32)$$

Ces équations permettent d'obtenir les valeurs ϵ_i et les orbitales ϕ'_i . Cependant, l'opérateur de Fock dépend lui-même des orbitales ϕ'_i via les opérateurs de Coulomb et d'échange. Il va donc falloir résoudre ces équations de manière itérative, cette procédure est un *self consistent field* ou SCF.

Pour des systèmes de très petite taille cette résolution est faisable numériquement. En revanche pour des systèmes de plus grosse taille ce n'est pas possible. On exprime alors les orbitales moléculaires en fonction d'orbitales atomiques.

1.2.1 Combinaison Linéaire d'Orbitales Atomiques

La partie spatiale des spin-orbitales est construite sur des combinaisons linéaires d'orbitales atomiques (Linear combinaison of atomic orbitals, LCAO en anglais).

$$\varphi_i = \sum_{\mu} c_{\mu i} \chi_{\mu} \quad (1.33)$$

$c_{\mu i}$ sont les poids des contributions des fonctions de base atomiques χ_{μ} à l'orbitale φ_i . Les fonctions χ_{μ} sont des fonctions centrées sur le noyau correspondant. Ces fonctions peuvent être de plusieurs types : exponentielles de Slater, gaussienne, ondes planes, etc.

1.2.2 Les bases atomiques

Les bases minimales

Un jeu de bases développé par Slater [2], Slater Type Orbital, STO.

$$\chi_{\mu}^{Slater} = STO = N x^m y^n z^o e^{-\zeta r} \quad (1.34)$$

où N est la constante de normalisation, ζ est l'exposant de la fonction de base qui renseigne sur la diffusivité de la fonction, x, y et z sont les coordonnées cartésiennes. Les indices m, n et o sont des entiers tels que $m + n + o = l$ avec l le nombre quantique atomique. Si $l = 0$ alors la STO est une orbitale de type s, si $l = 1$ alors c'est une orbitale de type p. Bien que physiquement intéressantes (proche de la résolution pour l'atome d'hydrogène), d'un point de vue calculatoire, elles ne sont pas du tout optimales. On ne peut pas exprimer le produit de deux STO comme une seule STO.

$$e^{-\alpha|\vec{r}-\vec{R}_A|} \cdot e^{-\beta|\vec{r}-\vec{R}_B|} \neq K \cdot e^{-\gamma|\vec{r}-\vec{R}_C|}$$

Ainsi pour des produits de 4 STO, on doit toutes les calculer. Il peut être difficile d'utiliser ces bases lorsque l'on étudie des systèmes de plus de deux atomes. C'est pourquoi d'autres types d'orbitales sont préférées aux STO. Les bases GTO, Gaussian Type Orbital.

$$\chi_{\mu}^{Gaussian} = GTO = N x^m y^n z^o e^{-\alpha r^2} \quad (1.35)$$

Les indices m, n et o sont des entiers tels que $m + n + o = l$ avec l le nombre quantique atomique. Si $l = 0$ alors la GTO est une orbitale de type s, si $l = 1$ alors c'est une orbitale de type p. Ces orbitales sont attrayantes niveau calcul car le produit de deux GTO peut s'écrire comme une autre GTO multipliée par un facteur. Ainsi pour des produits de 4 GTO, on retombera sur un produit de deux GTO. Cependant, ce type d'orbitales est moins judicieux au niveau physique. En effet, les GTO n'ont pas le même comportement que les STO à une distance nulle et à grande distance. Les STO ont une dérivée non nulle en zéro alors que celle des GTO s'annule. Il est nécessaire de prendre plus de GTO pour atteindre une précision identique avec les STO, il faut au moins trois GTO (voire 6 pour les fonctions du coeur) pour égaler une STO. Pour pallier à ce défaut, on peut envisager de former nos orbitales à l'aide de combinaisons linéaires de GTO. Les GTO qui interviennent dans la combinaison linéaire sont appelées des primitives.

$$\chi_{\mu}^{contractée} = \sum_{w=1}^b d_{wi} \chi_{\mu}^{primitive} \quad (1.36)$$

Utiliser des fonctions contractées permet de réduire le nombre de coefficients $c_{\mu i}$ à calculer. Les exposants des fonctions primitives et les coefficients de contraction d_{wi} ne sont généralement pas calculés pendant la résolution du problème mais ils sont fixés à l'avance. Ils sont choisis par exemple pour reproduire les fonctions de Slater.

Les bases étendues

Les bases décrites au-dessus peuvent se révéler insuffisantes pour une description précise des molécules. Dans la zone interne, les électrons sont proches du noyau et le potentiel nucléaire est quasiment à symétrie sphérique. Les orbitales atomiques sont adaptées. Cependant, il convient de prendre un nombre assez élevé de Gaussiennes du fait que l'énergie est très sensible à la position de l'électron au proche voisinage du noyau. Dans la zone de valence, la densité électronique est délocalisée entre plusieurs atomes. Pour la décrire correctement on utilise la démultiplication de la couche de valence, c'est-à-dire que l'on va remplacer chaque orbitale par plusieurs orbitales de taille différente. On ajoute aussi des orbitales de polarisation, des couches de nombre quantique l supérieur à la couche de valence. Dans la zone diffuse, loin des noyaux et de la symétrie sphérique, on ajoute des orbitales diffuses donc de faible exposant diminuant lentement avec la distance.

Les bases de Pople

John Pople a conçu un ensemble de bases dont la nomenclature est la suivante : n-n'n''n'''...++G(d,p)[3, 4, 5]. n est un entier désignant le nombre de primitives gaussiennes utilisées pour la couche interne, n'n''n''' correspondent au nombre de primitives gaussiennes utilisées dans chaque fonction de valence. Le nombre d'indices n primés indique le nombre de fonctions choisi pour représenter la zone de valence : s'il n'y a que deux indices on aura alors une base à double valence (double zeta), *i.e.* deux fonctions pour représenter une orbitale de valence. La valeur de n primé indique le nombre de primitives gaussiennes pour représenter la fonction. Les signes + ou ++ indiquent si l'on prend un ou deux ensembles d'orbitales diffuses. L'absence de ce symbole indique qu'on n'utilisera pas d'orbitales diffuses. Enfin les lettres entre parenthèses indiquent les nombres quantiques des orbitales que l'on décide d'ajouter aux atomes pour représenter la zone de polarisation. Les lettres à gauche de la virgule sont les fonctions rajoutées sur les atomes de la deuxième période et les lettres à droite de la virgule pour les fonctions sur l'atome d'Hydrogène.

Les bases de Dunning

Ces bases ont été élaboré par Dunning et collaborateurs [6]. Dans ces bases, les exposants des fonctions primitives sont optimisés sur des calculs atomiques où la corrélation électronique a été prise en compte. C'est de là que vient le "cc" dans le nom des bases, pour *correlation consistent*. Des fonctions de polarisation sont incluses dans les bases de Dunning, c'est ce que signifie le "p" dans leur nom. Ensuite l'ordre de la démultiplication de la couche de valence peut aller de 2 à 6 : VDZ, VTZ, VQZ, V5Z et V6Z. On peut rajouter des orbitales diffuses à ces bases. On ajoute alors le préfixe "aug" pour les nommer.

Les bases de Ahlrichs

Ahlrichs et collaborateurs ont développé un premier ensemble de bases [7] pour lesquelles les exposants et les coefficients de contraction ont été optimisés afin d'obtenir un bon compromis entre précision et taille de la base. Ces bases ont pour nom SV, TZV, QZV pour *split-valence*, triple et quadruple. Dans le logiciel avec lequel elles ont été optimisées, elles avaient le préfixe "def" pour *default* en anglais. Ahlrichs et collaborateurs ont ensuite légèrement modifié ces bases [8] et elles sont devenues : def2-SVP, def2-TZVP, def2-QZVP. Des fonctions de polarisation peuvent être rajouter (def2-SVPP, def2-TZVPP, def2-QZVPP) ainsi que des fonctions diffuses (un D est ajouté à la fin de leur nom).

1.2.3 Les équations de Roothan-Hall

C.C.J Roothan et G.G. Hall ont eu indépendamment l'idée d'employer la LCAO pour permettre l'étude des molécules avec la méthode d'Hartree-Fock. Nous allons obtenir des équations qui sont valables pour un système avec un nombre pair d'électrons et pour lequel toutes les orbitales ϕ_i sont occupées par deux électrons (voir section 1.2.5). Revenons à l'équation (1.32) en y injectant la formule LCAO.

$$\sum_{\mu} c_{\mu i} F_i |\chi_{\mu}\rangle = \epsilon_i \sum_{\mu} c_{\mu i} |\chi_{\mu}\rangle \quad (1.37)$$

En multipliant à gauche par $\sum_i \langle \chi_i |$, on obtient les équations de Roothan-Hall.

$$\sum_{\mu, i} c_{\mu i} \langle \chi_i | F_i | \chi_{\mu} \rangle = \epsilon_i \sum_{\mu, i} c_{\mu i} \langle \chi_i | \chi_{\mu} \rangle \quad (1.38)$$

Ces équations peuvent s'écrire sous forme matricielle.

$$\mathbf{FC} = \mathbf{SC}\epsilon \quad (1.39)$$

Avec $\mathbf{F}$, la matrice des éléments de Fock ($F_{ij} = \langle \chi_i | F_i | \chi_{\mu} \rangle$), $\mathbf{C}$ la matrice des éléments $c_{\mu i}$, coefficients de la LCAO, $\mathbf{S}$ la matrice de recouvrement entre les orbitales atomiques ($S_{ij} = \langle \chi_i | \chi_{\mu} \rangle$).

Ce sont ces équations qui sont effectivement résolues dans la méthode Hartree-Fock.

La densité électronique au point $\vec{r}$, $\rho(\vec{r})$ vaut :

$$\rho(\vec{r}) = 2 \sum_i^{\frac{N}{2}} |\phi_i(\vec{r})|^2 \quad (1.40)$$

En développant les orbitales sur une base atomique, on obtient

$$\begin{aligned} \rho(\vec{r}) &= 2 \sum_i^{\frac{N}{2}} \phi_i(\vec{r})^* \phi_i(\vec{r}) = 2 \sum_i^{\frac{N}{2}} \sum_{\mu} c_{\mu i} \chi_{\mu}(\vec{r})^* \sum_{\nu} c_{\nu i} \chi_{\nu}(\vec{r}) \\ \rho(\vec{r}) &= \sum_{\mu\nu} 2 \sum_i^{\frac{N}{2}} c_{\mu i} c_{\nu i} \chi_{\mu}(\vec{r})^* \chi_{\nu}(\vec{r}) \\ \rho(\vec{r}) &= \sum_{\mu\nu} P_{\mu\nu} \chi_{\mu}(\vec{r})^* \chi_{\nu}(\vec{r}) \end{aligned} \quad (1.41)$$

Avec

$$P_{\mu\nu} = 2 \sum_i^{\frac{N}{2}} c_{\mu i} c_{\nu i} \quad (1.42)$$

$P_{\mu\nu}$ est l'élément de la matrice densité.

L'opérateur effectif monoélectronique de Fock s'écrit

$$F(\vec{r}) = h(\vec{r}) + \sum_i (J_i - K_i) \quad (1.43)$$

$$\begin{aligned} F_{\mu\nu} &= \int d\vec{r} \chi_\mu^*(\vec{r}) F(\vec{r}) \chi_\nu(\vec{r}) \\ F_{\mu\nu} &= \int d\vec{r} \chi_\mu^*(\vec{r}) (h(\vec{r}) + \sum_i (J_i - K_i)) \chi_\nu(\vec{r}) \\ F_{\mu\nu} &= \int d\vec{r} \chi_\mu^*(\vec{r}) h(\vec{r}) \chi_\nu(\vec{r}) + \sum_i \int d\vec{r} \chi_\mu^*(\vec{r}) (J_i - K_i) \chi_\nu(\vec{r}) \\ F_{\mu\nu} &= \int d\vec{r} \chi_\mu^*(\vec{r}) h(\vec{r}) \chi_\nu(\vec{r}) + \sum_i \int d\vec{r} d\vec{r}' \chi_\mu^*(\vec{r}) \left(\sum_j \frac{\phi_j(\vec{r}')^* \phi_j(\vec{r}')}{|\vec{r}' - \vec{r}|} \chi_\nu(\vec{r}) - \sum_j \frac{\phi_j(\vec{r}')^* \chi_\nu(\vec{r}')}{|\vec{r}' - \vec{r}|} \phi_j(\vec{r}) \right) \\ F_{\mu\nu} &= H_{\mu\nu}^{core} + \sum_i [(\mu\nu|ii) - (\mu i|\nu i)] \end{aligned} \quad (1.44)$$

Dans la dernière ligne, on a écrit les intégrales coulombiennes selon la notation chimiste.

Si on intègre sur le spin on obtient (voir section 1.2.5)

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_i^{\frac{N}{2}} [2(\mu\nu|ii) - (\mu i|\nu i)] \quad (1.45)$$

On exprime maintenant les fonctions spatiales en fonction des bases atomiques.

$$\begin{aligned} F_{\mu\nu} &= H_{\mu\nu}^{core} + \sum_i^{\frac{N}{2}} \sum_{\lambda, \sigma} C_{\lambda i}^* C_{\sigma i} [2(\mu\nu|\lambda\sigma) - (\mu\lambda|\nu\sigma)] \\ F_{\mu\nu} &= H_{\mu\nu}^{core} + \sum_{\lambda, \sigma} P_{\lambda\sigma} [(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma)] \\ F_{\mu\nu} &= H_{\mu\nu}^{core} + G_{\mu\nu} \end{aligned} \quad (1.46)$$

1.2.4 Résolution par champ auto-cohérent ou SCF

La méthode itérative utilisée pour résoudre les équations de la méthode de Hartree-Fock est résumée dans le pseudo-code 1.2.

-
- 1: Donner les coordonnées nucléaires, la base atomique
 - 2: Calculer les éléments $S_{\mu\nu}$, $H_{\mu\nu}^{core}$ et les intégrales biélectroniques
 - 3: Calculer les matrices $\mathbf{S}^{-1/2}$ et $\mathbf{S}^{1/2}$
 - 4: Obtenir la matrice densité $\mathbf{P}$
 - 5: Calculer la matrice $\mathbf{G}$ pour obtenir la matrice $\mathbf{F}$
 - 6: Diagonaliser la matrice $\mathbf{S}^{-1/2}\mathbf{F}\mathbf{S}^{1/2}$ pour obtenir la matrice des coefficients $\mathbf{C}$ et les valeurs ϵ
 - 7: Calculer la nouvelle matrice $\mathbf{P}$ à partir de la matrice $\mathbf{C}$ obtenue
 - 8: Vérifier si la matrice $\mathbf{P}$ obtenue à l'étape 7 est différente de la précédente selon une certaine précision
 - 9: Si c'est le cas continuer la procédure en revenant à l'étape 5
 - 10: Sinon la convergence est atteinte, utiliser les quantités calculées pour obtenir les propriétés souhaitées
-

FIGURE 1.2 – Pseudo-code : Méthode itérative *self consistent field* (SCF)

1.2.5 Différents schémas de construction de la fonction d'essai

Depuis le début nous n'avons pas ou peu évoqué le spin, omettant le fait que les électrons puissent avoir un spin up ou down. De plus, les systèmes peuvent soit avoir un nombre pair d'électrons qui fait que tous les orbitales moléculaires seront remplies (systèmes à couches fermées) par deux électrons spin up et spin down soit présenter un nombre impair d'électrons (systèmes à couches ouvertes) qui fait que certains électrons peuvent se retrouver seuls sur certaines orbitales moléculaires. Il existe plusieurs schémas pour construire la fonction d'onde du système en prenant en compte ces considérations.

Restricted Hartree-Fock (RHF)

Dans cette situation, on a un nombre pair d'électrons ainsi chaque orbitale moléculaire est complètement remplie et contient deux électrons. Ces deux électrons auront une fonction de spin différente mais la fonction spatiale $\varphi_i(\vec{r})$ sera prise identique (fig. 1.3). On peut résumer cela par :

$$\phi_i(\vec{r}) = \begin{cases} \varphi_i(\vec{r})\alpha(s) \\ \varphi_i(\vec{r})\beta(s) \end{cases} \quad (1.47)$$

Par exemple, pour un système de 4 électrons, le déterminant de Slater s'écrira :

$$\Psi = (4!)^{-\frac{1}{2}} \det |\varphi_1(\vec{r}_1)\alpha(s_1)\varphi_1(\vec{r}_2)\beta(s_2)\varphi_2(\vec{r}_3)\alpha(s_3)\varphi_2(\vec{r}_4)\beta(s_4)| \quad (1.48)$$

C'était dans ce cadre que nous avons dérivé toutes les équations de Hartree-Fock.

Nous allons maintenant dériver les équations de Roothan-Hall pour la partie spatiale de la spin-orbitale.

$$\begin{aligned} F(\vec{r}_1)\phi_i(\vec{r}_1) &= \epsilon_i\phi_i(\vec{r}_1) \\ F(\vec{r}_1)\varphi_i(\vec{r}_1)\alpha(s_1) &= \epsilon_i\varphi_i(\vec{r}_1)\alpha(s_1) \\ F(\vec{r}_1)\varphi_i(\vec{r}_1)\beta(s_1) &= \epsilon_i\varphi_i(\vec{r}_1)\beta(s_1) \end{aligned} \quad (1.49)$$

Dans ce cas-ci (RHF), le résultat sera le même pour un électron α ou β .

Pour obtenir une équation sur la fonction spatiale, il faut intégrer sur la fonction de spin.

$$\int ds_1 \alpha^*(s_1) F(\vec{r}_1) \varphi_i(\vec{r}_1) \alpha(s_1) = f(\vec{r}_1) \varphi_i(\vec{r}_1) = \epsilon_i \varphi_i(\vec{r}_1) \quad (1.50)$$

En remplaçant l'opérateur de Fock par sa valeur puis en considérant que la somme sur les N électrons s'exprime comme deux sommes sur $\frac{N}{2}$ électrons α et β ($\sum^N \rightarrow \sum^{\frac{N}{2}} + \sum^{\frac{N}{2}}$), on a :

$$\begin{aligned} \int ds_1 \alpha^*(s_1) (h(\vec{r}_1) + \sum_j^N (J_j - K_j)) \varphi_i(\vec{r}_1) \alpha(s_1) &= \int ds_1 \alpha^*(s_1) h(\vec{r}_1) \varphi_i(\vec{r}_1) \alpha(s_1) \\ &+ \int ds_1 \alpha^*(s_1) \left(\sum_j^N (J_j - K_j) \right) \varphi_i(\vec{r}_1) \alpha(s_1) \\ &= h(\vec{r}_1) \varphi_i(\vec{r}_1) \\ &+ \int ds_1 \alpha^*(s_1) \sum_j^N \left(\int d\vec{r}_2 \phi_j^*(\vec{r}_2) g_{12} \phi_j(\vec{r}_2) \right) \varphi_i(\vec{r}_1) \alpha(s_1) \\ &- \int ds_1 \alpha^*(s_1) \sum_j^N \left(\int d\vec{r}_2 \phi_j^*(\vec{r}_2) g_{12} \phi_j(\vec{r}_1) \right) \varphi_i(\vec{r}_2) \alpha(s_1) \\ &= h(\vec{r}_1) \varphi_i(\vec{r}_1) \quad (1.51) \\ &+ \int ds_1 ds_2 \alpha^*(s_1) \sum_j^{\frac{N}{2}} \left(\int d\vec{r}_2 \varphi_j^*(\vec{r}_2) \alpha^*(s_2) g_{12} \varphi_j(\vec{r}_2) \alpha(s_2) \right) \varphi_i(\vec{r}_1) \alpha(s_1) \\ &+ \int ds_1 ds_2 \alpha^*(s_1) \sum_j^{\frac{N}{2}} \left(\int d\vec{r}_2 \varphi_j^*(\vec{r}_2) \beta^*(s_2) g_{12} \varphi_j(\vec{r}_2) \beta(s_2) \right) \varphi_i(\vec{r}_1) \alpha(s_1) \\ &- \int ds_1 ds_2 \alpha^*(s_1) \sum_j^{\frac{N}{2}} \left(\int d\vec{r}_2 \varphi_j^*(\vec{r}_2) \alpha^*(s_2) g_{12} \varphi_j(\vec{r}_1) \alpha^*(s_1) \right) \varphi_i(\vec{r}_2) \alpha(s_1) \\ &- \int ds_1 ds_2 \alpha^*(s_1) \sum_j^{\frac{N}{2}} \left(\int d\vec{r}_2 \varphi_j^*(\vec{r}_2) \beta^*(s_2) g_{12} \varphi_j(\vec{r}_1) \beta^*(s_1) \right) \varphi_i(\vec{r}_2) \alpha(s_1) \end{aligned}$$

La dernière ligne vaut zéro parce que les fonctions de spin sont orthogonales. Après intégration sur les spins, les deuxième et troisième termes de la dernière égalité sont les mêmes. On a alors

$$\begin{aligned} f(\vec{r}_1) \varphi_i(\vec{r}_1) &= h(\vec{r}_1) \varphi_i(\vec{r}_1) \\ &+ 2 \sum_j^{\frac{N}{2}} \int d\vec{r}_2 \varphi_j^*(\vec{r}_2) g_{12} \varphi_j(\vec{r}_2) \varphi_i(\vec{r}_1) \\ &- \sum_j^{\frac{N}{2}} \left(\int d\vec{r}_2 \varphi_j^*(\vec{r}_2) g_{12} \varphi_i(\vec{r}_2) \varphi_j(\vec{r}_1) \right) \\ &= \epsilon_i \varphi_i(\vec{r}_1) \end{aligned} \quad (1.52)$$

L'opérateur de Fock s'appliquant sur la partie spatiale des spin-orbitales est :

$$f(\vec{r}_1) = h(\vec{r}_1) + \sum_j^{\frac{N}{2}} (2J_j^x - K_j^x) \quad (1.53)$$

J_j^x et K_j^x sont les opérateurs de Coulomb et d'échange pour les fonctions spatiales.

Unrestricted open-shell Hartree-Fock (UHF)

Dans cette situation, le système comporte un nombre impair d'électrons. La spin-orbitale de chaque électron aura une partie spatiale différente des autres spin-orbitales (fig. 1.3). Cependant, dans cette formulation, la spin-orbitale n'est plus fonction propre de l'opérateur de spin S^2 . Si elle était fonction propre de cet opérateur alors la valeur propre vaudrait $\hbar \left(\frac{1}{2} \left(\frac{1}{2} + 1 \right) \right)$. Lors de la résolution des équations de Pople-Nesbet [9], la généralisation des équations de Roothan-Hall au formalisme UHF, il peut se produire une contamination du spin. C'est-à-dire que la valeur du spin calculée est différente de $\hbar \left(\frac{1}{2} \left(\frac{1}{2} + 1 \right) \right)$.

Les équations de Pople-Nesbet se présentent sous la forme de deux équations, une pour les électrons α et une pour les électrons β .

$$\begin{aligned} \mathbf{F}^\alpha \mathbf{C}^\alpha &= \mathbf{S} \mathbf{C}^\alpha \epsilon^\alpha \\ \mathbf{F}^\beta \mathbf{C}^\beta &= \mathbf{S} \mathbf{C}^\beta \epsilon^\beta \end{aligned} \quad (1.54)$$

Ces équations sont résolues séparément mais l'effet des électrons α sur les électrons β (et inversement) intervient dans la création de l'opérateur de Fock F^β (F^α).

Restricted open-shell Hartree-Fock (ROHF)

Tous les orbitales moléculaires doublement occupées ont toutes la même partie spatiale. En revanche, les orbitales possédant un seul électron n'ont pas la même partie spatiale que les orbitales doublement occupées (voir fig. 1.3). La spin-orbitale est une fonction propre de l'opérateur de spin S^2 .

1.3 La corrélation électronique

L'antisymétrisation de la fonction d'onde via le déterminant de Slater impose que deux électrons qui possèdent la même partie spatiale ne peuvent avoir le même spin. L'approximation orbitalaire et le fait que la fonction d'onde de Hartree-Fock est construite avec un seul déterminant de Slater font que les probabilités de présence de deux électrons sont indépendantes. Les électrons ne sont pas corrélés. Ainsi, rien n'interdit que deux électrons possédant deux spins différents aient la même partie spatiale et soient très proches l'un de l'autre. Dans la méthode de Hartree-Fock, l'interaction entre les électrons n'apparaît que par le biais du champ moyen. C'est une interaction moyennée qui ne prend pas en compte les effets instantanés.

On distingue généralement deux aspects de la corrélation électronique. D'une part, on a la corrélation dynamique ou dite instantanée, elle intervient à courte distance lorsque les deux électrons occupent la même orbitale. Du fait de la répulsion coulombienne, les

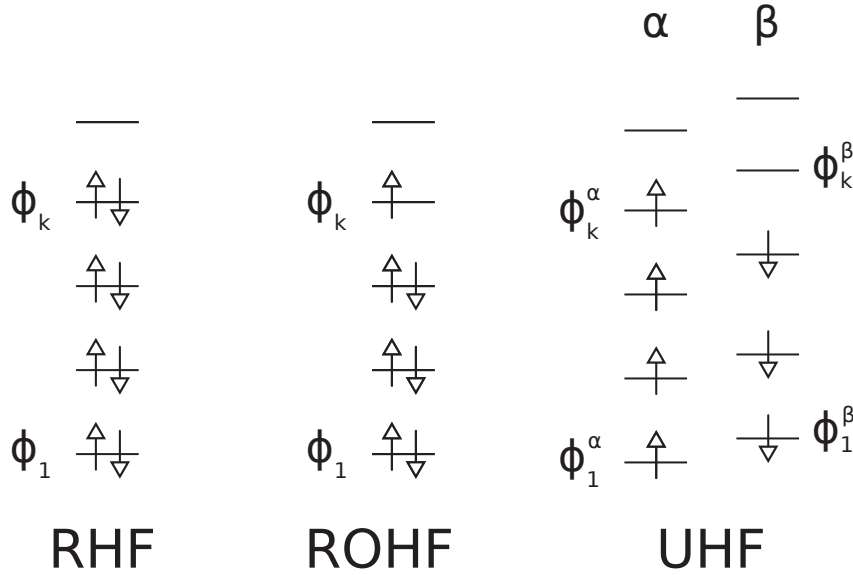


FIGURE 1.3 – Les différents schémas pour la fonction d'onde de Hartree-Fock

électrons doivent se repousser. D'autre part, on a la corrélation statique ou dite permanente, elle intervient à grande distance lorsque les deux électrons occupent des orbitales différentes.

Or dans le cas d'un système fermé et dans le cadre de Hartree-Fock, la corrélation électronique ne peut être prise en compte étant donné que les électrons sont arrangés par paire dans les orbitales moléculaires. On voit que pour retrouver cette corrélation il faut que les électrons soient disposés dans d'autres orbitales. Cela revient donc à exciter ces électrons pour qu'ils se placent dans les orbitales dites inoccupées ou virtuelles. C'est la base des méthodes post Hartree-Fock que nous allons voir dans la partie suivante.

Dans le cas de la molécule de dihydrogène on peut montrer que l'absence de corrélation électronique entraîne une sur-estimation de l'énergie de dissociation. Ainsi il est nécessaire de prendre en compte correctement la corrélation électronique pour traiter convenablement le problème. Autrement, les énergies ou les différentes propriétés calculées ne seront pas du tout fiables et précises.

1.4 Post Hartree-Fock

1.4.1 Interaction de configurations

On a vu que le fait de prendre la fonction d'onde du système sous la forme d'un seul déterminant de Slater ne permettait pas de prendre en compte correctement la corrélation électronique. Pour corriger ce problème on ajoute des déterminants de Slater correspondant à des excitations électroniques dans la fonction d'onde du système.

$$\Psi = C_0 \Phi_0 + \sum_i^{\text{occ}} \sum_a^{\text{vir}} C_i^a \Phi_i^a + \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} C_{ij}^{ab} \Phi_{ij}^{ab} + \sum_{i < j < k}^{\text{occ}} \sum_{a < b < c}^{\text{vir}} C_{ijk}^{abc} \Phi_{ijk}^{abc} + \dots \quad (1.55)$$

Les coefficients $\{C\}$ permettent de déterminer les contributions des différents déterminants à la fonction d'onde totale. Le déterminant Φ_0 correspondant au déterminant non

excité de Hartree-Fock. Les déterminants Φ_i^a correspondent aux mono-excitations, un électron dans l'orbitale occupée ϕ_i se place dans l'orbitale virtuelle ϕ_a . Les déterminants Φ_{ij}^{ab} correspondent aux double-excitations, un électron dans l'orbitale occupée ϕ_i se place dans l'orbitale virtuelle ϕ_a et un autre électron dans l'orbitale occupée ϕ_j se place dans l'orbitale virtuelle ϕ_b . La figure 1.4 illustre cela. Les orbitales de ϕ_1 à ϕ_k sont les orbitales occupées, ϕ_k étant la dernière occupée. Les orbitales ϕ_a à ϕ_{final} sont les orbitales virtuelles, inoccupées, ϕ_{final} étant la dernière inoccupée.

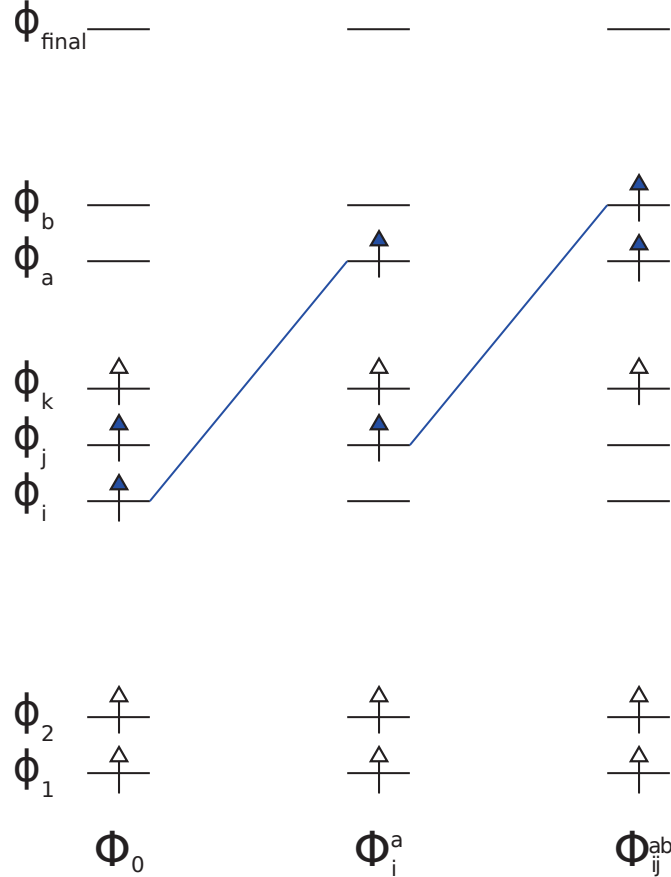


FIGURE 1.4 – Schéma d'occupation électronique des orbitales pour le déterminant de Slater de Hartree-Fock (Φ_0), d'un déterminant mono-excité (Φ_i^a) et d'un déterminant doublement excité (Φ_{ij}^{ab}). Les orbitales indicées de 1 à k sont les orbitales occupées et celles indicées de a à $final$ sont les orbitales virtuelles.

A l'instar de la méthode de Hartree-Fock, la méthode des interactions de configuration est une méthode variationnelle. On va faire varier les coefficients $\{\mathbf{C}\}$ afin de minimiser le Lagrangien en prenant comme contrainte de garder la fonction d'onde du système Ψ normée. Le lagrangien s'écrit donc ainsi :

$$L = \langle \Psi | \hat{H}_e | \Psi \rangle - E(\langle \Psi | \Psi \rangle - 1) \quad (1.56)$$

On peut réécrire la fonction d'onde Ψ comme $|\Psi\rangle = \sum_i C_i |\Phi_i\rangle$. En remplaçant Ψ par son expression,

$$L = \sum_{ij} C_i^* C_j \langle \Phi_i | \hat{H}_e | \Phi_j \rangle - E(\sum_{ij} C_i^* C_j \langle \Phi_i | \Phi_j \rangle - 1) \quad (1.57)$$

En prenant $H_{ij} = \langle \Phi_i | \hat{H}_e | \Phi_j \rangle$ et $S_{ij} = \langle \Phi_i | \Phi_j \rangle$, on a

$$L = \sum_{ij} C_i^* C_j H_{ij} - E \left(\sum_{ij} C_i^* C_j S_{ij} - 1 \right) \quad (1.58)$$

La variation δL du Lagrangien qui doit être nulle selon le principe variationnel vaut donc

$$\delta L = \sum_{ij} \delta C_i^* C_j H_{ij} - E \sum_{ij} \delta C_i^* C_j S_{ij} + \sum_{ij} C_i^* \delta C_j H_{ij} - E \sum_{ij} C_i^* \delta C_j S_{ij} = 0 \quad (1.59)$$

En arrangeant les indices des deux derniers termes, on obtient

$$\delta L = \sum_{ij} \delta C_i^* C_j H_{ij} - E \sum_{ij} \delta C_i^* C_j S_{ij} + \sum_{ij} C_j^* \delta C_i H_{ji} - E \sum_{ij} C_j^* \delta C_i S_{ji} = 0 \quad (1.60)$$

La matrice hamiltonienne est hermitienne par définition et la matrice S, la matrice de recouvrement entre les déterminants de Slater est aussi hermitienne, on peut alors écrire

$$\delta L = \sum_i \delta C_i^* \left[\sum_j H_{ij} C_j - E S_{ij} C_j \right] + c.c. = 0 \quad (1.61)$$

En remplaçant δC_i par $i\delta C_i$ et en calculant (1.59) + $i \cdot$ (1.60), on trouve

$$\sum_j H_{ij} C_j - E S_{ij} C_j = 0 \quad (1.62)$$

Soit en écriture matricielle.

$$\mathbf{HC} = E\mathbf{SC} \quad (1.63)$$

Dans le cas où les déterminants de Slater sont orthonormalisés $\mathbf{S}$ devient la matrice identité.

$$\mathbf{HC} = E\mathbf{C} \quad (1.64)$$

Résoudre le problème revient donc à diagonaliser la matrice $\mathbf{H}$ ($\det|(\mathbf{H} - E \cdot \mathbf{1})\mathbf{C}| = 0$)

Cependant le nombre de déterminants explose très rapidement avec le nombre d'électrons de la molécule. Il est quasiment impossible de prendre en compte tous les déterminants excités à part pour des très petites molécules. Quand on prend en compte tous les déterminants on dit que l'on fait un calcul *Full-CI*. Dans le cas général on tronque la suite de déterminants à un certain ordre, on dit que l'on fait du *truncated-CI*. SDCI voudra donc dire que l'on prend toutes les mono-excitations et les double-excitations, DCI seulement les excitations double, etc. Chaque déterminant de Slater est construit à partir de la base (les orbitales atomiques du LCAO) atomique qui est finie. En *Full-CI* et avec une base atomique infinie on a la fonction d'onde théorique exacte.

1.4.2 Clusters couplés

La fonction d'onde CI vue précédemment peut s'écrire en utilisant un opérateur d'excitation. Soit l'opérateur d'excitations

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_N \quad (1.65)$$

Les opérateurs $\hat{T}_i$ correspondent aux opérateurs à i excitations.

On a donc

$$\Psi_{CI} = C_0 \Phi_0 + \sum_i^{\text{occ}} \sum_a^{\text{vir}} C_i^a \Phi_i^a + \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} C_{ij}^{ab} \Phi_{ij}^{ab} + \cdots = (1 + \hat{T}) \Phi_0 \quad (1.66)$$

Par exemple

$$\hat{T}_1 \Phi_0 = \sum_i^{\text{occ}} \sum_a^{\text{vir}} C_i^a \Phi_i^a = \sum_i^{\text{occ}} \sum_a^{\text{vir}} t_i^a \Phi_i^a \quad (1.67)$$

$$\hat{T}_2 \Phi_0 = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} C_{ij}^{ab} \Phi_{ij}^{ab} = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} t_{ij}^{ab} \Phi_{ij}^{ab} \quad (1.68)$$

Dans la théorie des clusters couplés, on parlera plutôt d'amplitudes t au lieu des coefficients C . De plus, dans cette théorie la fonction d'onde est construite différemment.

$$\Psi_{CC} = e^{\hat{T}} \Phi_0 \quad (1.69)$$

Avec l'opérateur $e^{\hat{T}}$ qui s'écrit en série

$$e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2!} \hat{T}^2 + \frac{1}{3!} \hat{T}^3 + \cdots + \frac{1}{N!} \hat{T}^N = \sum_{k=0}^{\infty} \frac{1}{k!} \hat{T}^k \quad (1.70)$$

$$e^{\hat{T}} = 1 + \hat{T}_1 + (\hat{T}_2 + \frac{1}{2!} \hat{T}_1^2) + (\hat{T}_3 + \hat{T}_2 \hat{T}_1 + \frac{1}{3!} \hat{T}_1^3) + (\hat{T}_4 + \hat{T}_3 \hat{T}_1 + \frac{1}{2!} \hat{T}_1^2 + \frac{1}{2!} \hat{T}_2 \hat{T}_2 + \frac{1}{4!} \hat{T}_1^4) + \cdots \quad (1.71)$$

Sous cette forme apparaissent des produits d'excitations, c'est-à-dire des excitations d'ordre n qui sont le résultat d'excitations d'ordre plus bas. Par exemple, l'excitation triple $\hat{T}_2 \hat{T}_1$ est constituée de l'excitation double d'un électron par l'opérateur $\hat{T}_2$ et l'excitation simple par $\hat{T}_1$. Ces termes ne sont pas présents dans la théorie d'interactions de configurations. On appelle ces excitations des excitations déconnectées. Les autres excitations, c'est-à-dire des excitations d'ordre n engendrées par l'opérateur T_n sont appelées excitations connectées.

L'équation de la théorie des clusters couplés est

$$H_e \Psi_{CC} = H_e e^{\hat{T}} \Phi_0 = E e^{\hat{T}} \Phi_0 \quad (1.72)$$

En notation de Dirac et en multipliant à gauche par $\langle \Phi_0 | e^{-\hat{T}}$, on obtient une équation pour calculer l'énergie de la théorie des clusters couplés :

$$\langle \Phi_0 | e^{-\hat{T}} H_e e^{\hat{T}} | \Phi_0 \rangle = E \langle \Phi_0 | e^{-\hat{T}} e^{\hat{T}} | \Phi_0 \rangle = E \quad (1.73)$$

L'opérateur $e^{-\hat{T}}H_e e^{\hat{T}}$ peut être transformé en utilisant la formule de Baker-Campbell-Hausdorff (BCH)

$$e^{-\hat{T}}H_e e^{\hat{T}} = H_e + [H_e, \hat{T}] + \frac{1}{2}[[H_e, \hat{T}], \hat{T}] + \frac{1}{6}[[[H_e, \hat{T}], \hat{T}], \hat{T}] + \frac{1}{24}[[[[H_e, \hat{T}], \hat{T}], \hat{T}], \hat{T}] \quad (1.74)$$

En remplaçant l'expression de l'opérateur dans l'équation (1.73), on obtient une expression pour l'énergie de la théorie des clusters couplés.

$$\begin{aligned} E &= \langle \Phi_0 | H_e | \Phi_0 \rangle + \langle \Phi_0 | [H_e, \hat{T}] | \Phi_0 \rangle \\ &+ \frac{1}{2} \langle \Phi_0 | [[H_e, \hat{T}], \hat{T}] | \Phi_0 \rangle + \langle \Phi_0 | \frac{1}{6} [[[[H_e, \hat{T}], \hat{T}], \hat{T}], \hat{T}] | \Phi_0 \rangle \\ &+ \langle \Phi_0 | \frac{1}{24} [[[[[H_e, \hat{T}], \hat{T}], \hat{T}], \hat{T}], \hat{T}] | \Phi_0 \rangle \end{aligned} \quad (1.75)$$

Tous les termes en $\langle \Phi_0 | \hat{T} | \Phi_0 \rangle$ sont nuls. En effet, on ne peut pas déexciter l'état fondamental. L'hamiltonien H_e contient un opérateur biélectronique ainsi toute excitation d'ordre supérieur à 2 donnera un terme nul ($\langle \Phi_0 | H_e \hat{T}_i | \Phi_0 \rangle = 0$, $i > 2$). Selon le théorème de Brillouin, les termes $\langle \Phi_0 | H_e \hat{T}_1 | \Phi_0 \rangle$ seront nuls.

Ainsi l'expression de l'énergie se réduit à :

$$E = E_{HF} + \langle \Phi_0 | H_e \hat{T}_2 | \Phi_0 \rangle + \frac{1}{2} \langle \Phi_0 | H_e \hat{T}_1^2 | \Phi_0 \rangle \quad (1.76)$$

L'énergie dépend ainsi uniquement des excitations simples et doubles. Seules les amplitudes de ces excitations contribuent directement à l'énergie. Cependant pour les calculer il faut utiliser les quantités suivantes :

$$\langle \Phi_{ij\dots}^{ab\dots} | e^{-\hat{T}} H_e e^{\hat{T}} | \Phi_0 \rangle = E \langle \Phi_{ij\dots}^{ab\dots} | e^{-\hat{T}} e^{\hat{T}} | \Phi_0 \rangle = E \langle \Phi_{ij\dots}^{ab\dots} | \Phi_0 \rangle = 0 \quad (1.77)$$

$$\begin{aligned} \langle \Phi_{ij\dots}^{ab\dots} | e^{-\hat{T}} H_e e^{\hat{T}} | \Phi_0 \rangle &= \langle \Phi_{ij\dots}^{ab\dots} | H_e | \Phi_0 \rangle + \langle \Phi_{ij\dots}^{ab\dots} | [H_e, \hat{T}] | \Phi_0 \rangle \\ &+ \frac{1}{2} \langle \Phi_{ij\dots}^{ab\dots} | [[H_e, \hat{T}], \hat{T}] | \Phi_0 \rangle + \langle \Phi_{ij\dots}^{ab\dots} | \frac{1}{6} [[[[H_e, \hat{T}], \hat{T}], \hat{T}], \hat{T}] | \Phi_0 \rangle \\ &+ \langle \Phi_{ij\dots}^{ab\dots} | \frac{1}{24} [[[[[H_e, \hat{T}], \hat{T}], \hat{T}], \hat{T}], \hat{T}] | \Phi_0 \rangle \\ &= 0 \end{aligned} \quad (1.78)$$

On obtient un ensemble d'équations non linéaires à résoudre pour trouver toutes les amplitudes. Les amplitudes des excitations simples et doubles dépendront des amplitudes des autres excitations. Au final les autres excitations contribuent indirectement à l'énergie de la théorie des clusters couplés.

De même que pour la méthode CI, on ne peut pas prendre en compte toutes les excitations et on est obligé de couper à un certain ordre. Pour la méthode CCSD par exemple, on considère seulement les excitations simples et doubles tel que l'opérateur $\hat{T} = \hat{T}_1 + \hat{T}_2$. Pour CCSDT, on a $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3$, où les excitations triples ont été ajoutées. Cette dernière est limitée à des petits systèmes. On peut cependant considérer que ces excitations triples ont des amplitudes beaucoup plus faibles que les simples et doubles. Il est alors possible de prendre en compte ces excitations triples par un traitement perturbatif beaucoup moins coûteux. C'est la méthode CCSD(T) qui est actuellement considérée comme une méthode très précise parmi les méthodes de la chimie quantique (avec une base suffisante et pour des états non multi-références).

1.4.3 Théorie de la perturbation de Møller et Plesset

On peut utiliser la théorie des perturbations pour ajouter de la corrélation électronique au problème de Hartree-Fock.

Dans la théorie des perturbations on a un hamiltonien non perturbé, H_0 , dont on connaît les solutions auquel on ajoute une perturbation W . Ici H_0 est pris comme la somme des opérateurs de Fock, $H_0 = \sum_i F_i$. C'est la théorie perturbation de Møller et Plesset [10], MPn, avec n l'ordre de la perturbation. On a vu que l'hamiltonien de Hartree-Fock, sans le terme constant de répulsion nucléaire, est :

$$H = \sum_i^N h_i + \sum_i^N \sum_{i>j}^N g_{ij} \quad (1.79)$$

On cherche à l'écrire sous la forme de l'hamiltonien de référence H_0 plus la perturbation W .

$$H = H_0 + W \quad (1.80)$$

$$H = \sum_i^N F_i + W = \sum_i^N (h_i + \sum_j^N (J_j - K_j)) + W = \sum_i^N h_i + \sum_i^N \sum_{i>j}^N g_{ij} \quad (1.81)$$

Par identification terme à terme, on trouve que la perturbation W est

$$W = \sum_i^N \sum_{i>j}^N g_{ij} - \sum_i^N \sum_j^N (J_j - K_j) \quad (1.82)$$

Dans la théorie des perturbations, l'Hamiltonien $H = H_0 + \lambda W$. La fonction d'onde Ψ s'exprime sous la forme d'une somme de la fonction d'onde propre de l'hamiltonien de référence H_0 plus des corrections, $\Psi = \Phi_0 + \lambda \Psi_1 + \lambda^2 \Psi_2 + \lambda^3 \Psi_3 + \dots$, de même que l'énergie sera $E = E_0 + \lambda E_1 + \lambda^2 E_2 + \lambda^3 E_3 + \dots$. On doit résoudre le problème $H\Psi = E\Psi$. Les fonctions d'ondes Ψ_i s'écrivent en fonction des déterminants de Slater excités solutions de H_0 . Elles forment un ensemble orthonormé.

$$H\Psi = E\Psi$$

$$(H_0 + \lambda W)(\Phi_0 + \lambda \Psi_1 + \lambda^2 \Psi_2 + \dots) = (E_0 + \lambda E_1 + \lambda^2 E_2 + \dots)(\Phi_0 + \lambda \Psi_1 + \lambda^2 \Psi_2 + \dots) \quad (1.83)$$

En développant et en égalisant les termes présentant la même puissance en λ , on obtient que pour

$$\begin{aligned} \lambda^0 : H_0 \Phi_0 &= E_0 \Phi_0 \\ \lambda^1 : H_0 \Psi_1 + W \Phi_0 &= E_0 \Psi_1 + E_1 \Phi_0 \\ \lambda^2 : H_0 \Psi_2 + W \Psi_1 &= E_0 \Psi_2 + E_1 \Psi_1 + E_2 \Phi_0 \end{aligned} \quad (1.84)$$

Φ_0 est le déterminant de Hartree-Fock.

$$\langle \Phi_0 | H_0 | \Phi_0 \rangle = \langle \Phi_0 | \sum_i^N F_i | \Phi_0 \rangle = \sum_i^N \epsilon_i \quad (1.85)$$

A l'ordre zéro, l'énergie obtenue est la somme des énergies des orbitales. Dans cette valeur sont comptées en doubles les interactions entre électrons.

Multiplions l'équation correspondant à l'ordre 1 en λ par Φ_0^* et intégrons.

$$\begin{aligned} \langle \Phi_0 | H_0 | \Psi_1 \rangle + \langle \Phi_0 | W | \Phi_0 \rangle &= \langle \Phi_0 | E_0 | \Psi_1 \rangle + \langle \Phi_0 | E_1 | \Phi_0 \rangle \\ \langle \Psi_1 | H_0 | \Phi_0 \rangle^* + \langle \Phi_0 | W | \Phi_0 \rangle &= E_0 \langle \Phi_0 | \Psi_1 \rangle + E_1 \langle \Phi_0 | \Phi_0 \rangle \\ E_0 \langle \Psi_1 | \Phi_0 \rangle^* + \langle \Phi_0 | W | \Phi_0 \rangle &= E_1 \\ \langle \Phi_0 | W | \Phi_0 \rangle &= E_1 \end{aligned} \quad (1.86)$$

La correction au premier ordre de l'énergie est :

$$\langle \Phi_0 | W | \Phi_0 \rangle = \langle \Phi_0 | \sum_i^N \sum_{i>j}^N g_{ij} - \sum_i^N \sum_j^N (J_j - K_j) | \Phi_0 \rangle \quad (1.87)$$

En manipulant l'opérateur biélectronique g_{ij} et la somme $\sum_{i>j}^N$, on peut montrer (cela a été fait dans la partie sur Hartree-Fock) que

$$\langle \Phi_0 | W | \Phi_0 \rangle = -\frac{1}{2} \sum_i^N \sum_j^N \langle \Phi_0 | (J_j - K_j) | \Phi_0 \rangle \quad (1.88)$$

$$\langle \Phi_0 | H | \Phi_0 \rangle = \langle \Phi_0 | H_0 + W | \Phi_0 \rangle = \sum_i^N \epsilon_i - \frac{1}{2} \sum_i^N \sum_j^N \langle \Phi_0 | (J_j - K_j) | \Phi_0 \rangle \quad (1.89)$$

Au premier ordre, on obtient l'énergie de Hartree-Fock.

Pour obtenir la correction au second ordre de l'énergie on va multiplier à gauche l'expression de l'ordre 2 en λ par Φ_0^* et intégrer. Cela a déjà été évoqué plus haut mais les corrections à la fonction d'onde s'écrivent comme une combinaison linéaire des déterminants de Slater excités (le coefficient pour le déterminant non excité est nul). Ainsi $\Psi_1 = \sum_i c_i \Phi_i$ et $\Psi_2 = \sum_i d_i \Phi_i$, c_0 et d_0 , les coefficients pour le déterminant non excité sont nuls.

$$\begin{aligned} \langle \Phi_0 | H_0 | \Psi_2 \rangle + \langle \Phi_0 | W | \Psi_1 \rangle &= \langle \Phi_0 | E_0 | \Psi_2 \rangle + \langle \Phi_0 | E_1 | \Psi_1 \rangle + \langle \Phi_0 | E_2 | \Phi_0 \rangle \\ \sum_i d_i \langle \Phi_0 | H_0 | \Phi_i \rangle + \sum_i c_i \langle \Phi_0 | W | \Phi_i \rangle &= \sum_i d_i \langle \Phi_0 | E_0 | \Phi_i \rangle + \sum_i c_i \langle \Phi_0 | E_1 | \Phi_i \rangle + \langle \Phi_0 | E_2 | \Phi_0 \rangle \\ \sum_i d_i \langle \Phi_i | H_0 | \Phi_0 \rangle^* + \sum_i c_i \langle \Phi_0 | W | \Phi_i \rangle &= \sum_i d_i E_0 \langle \Phi_0 | \Phi_i \rangle + \sum_i c_i E_1 \langle \Phi_0 | \Phi_i \rangle + E_2 \langle \Phi_0 | \Phi_0 \rangle \\ \sum_i d_i E_0 \langle \Phi_i | \Phi_0 \rangle^* + \sum_i c_i \langle \Phi_0 | W | \Phi_i \rangle &= E_2 \\ \sum_i c_i \langle \Phi_0 | W | \Phi_i \rangle &= E_2 \end{aligned} \quad (1.90)$$

Les coefficients c_i sont indéterminés. Pour trouver leur expression il faut revenir sur l'équation du premier ordre en λ . Au lieu de multiplier à gauche par Φ_0^* on va projeter cette expression sur un déterminant excité. On multiplie donc par Φ_j^* .

$$\begin{aligned}
 \langle \Phi_j | H_0 | \Psi_1 \rangle + \langle \Phi_j | W | \Phi_0 \rangle &= \langle \Phi_j | E_0 | \Psi_1 \rangle + \langle \Phi_j | E_1 | \Phi_0 \rangle \\
 \sum_i c_i \langle \Phi_j | H_0 | \Phi_i \rangle + \langle \Phi_j | W | \Phi_0 \rangle &= \sum_i c_i E_0 \langle \Phi_j | \Phi_i \rangle \\
 \sum_i c_i E_i \langle \Phi_j | \Phi_i \rangle + \langle \Phi_j | W | \Phi_0 \rangle &= \sum_i c_i E_0 \langle \Phi_j | \Phi_i \rangle \\
 c_j E_j + \langle \Phi_j | W | \Phi_0 \rangle &= c_j E_0 \\
 c_j &= \frac{\langle \Phi_j | W | \Phi_0 \rangle}{E_0 - E_j}
 \end{aligned} \tag{1.91}$$

On peut maintenant remplacer l'expression des coefficients c_i dans l'équation (1.90)

$$E_2 = \sum_{i \neq 0} \frac{\langle \Phi_i | W | \Phi_0 \rangle \langle \Phi_0 | W | \Phi_i \rangle}{E_0 - E_i} \tag{1.92}$$

c_0 est égal à zéro par définition donc la somme est sur tous les i différent de zéro. Les Φ_i sont les déterminants d'excitation simple ou plus. Cependant l'opérateur W fait intervenir l'opérateur biélectronique et l'opérateur de Fock. Ainsi seuls les termes correspondant à des excitations simples ou doubles peuvent être non nuls. Evaluons le terme $\langle \Phi_j | W | \Phi_0 \rangle$ pour des excitations simples.

$$\langle \Phi_j | W | \Phi_0 \rangle = \langle \Phi_j | H - H_0 | \Phi_0 \rangle = \langle \Phi_j | H | \Phi_0 \rangle - E_0 \langle \Phi_j | \Phi_0 \rangle \tag{1.93}$$

Le théorème de Brillouin et l'orthogonalité des déterminants fait que ces deux termes sont nuls. Les excitations simples ne contribuent pas au second ordre de l'énergie. Seules les excitations doubles contribuent à cette énergie.

$$E_2 = \frac{1}{4} \sum_{i,j}^{occ} \sum_{a,b}^{vir} \frac{\langle \Phi_{ij}^{ab} | W | \Phi_0 \rangle \langle \Phi_0 | W | \Phi_{ij}^{ab} \rangle}{E_0 - E_{ij}^{ab}} \tag{1.94}$$

$$\langle \Phi_{ij}^{ab} | W | \Phi_0 \rangle = \langle \Phi_{ij}^{ab} | \sum_k^N \sum_{k>l}^N g_{kl} - \sum_k^N \sum_l^N (J_l - K_l) | \Phi_0 \rangle \tag{1.95}$$

Le terme contenant les opérateurs de Coulomb et d'échange s'appliquent à une orbitale, or dans $\langle \Phi_{ij}^{ab} |$ il y a deux orbitales qui diffère par rapport à $|\Phi_0\rangle$ donc ce terme sera nul. Le premier terme c'est l'opérateur coulombien biélectronique. Pour que les termes soient non nuls, il faut appliquer cet opérateur sur les orbitales occupées qui seront excitées et les orbitales virtuelles qui vont être occupées afin que le produit des orbitales restantes soit égal à 1. Il restera donc

$$\begin{aligned}
 \langle \Phi_{ij}^{ab} | W | \Phi_0 \rangle &= \langle \phi_a \phi_b | (1 - P_{12}) g_{12} | \phi_i \phi_j \rangle \\
 \langle \Phi_{ij}^{ab} | W | \Phi_0 \rangle &= \langle \phi_a \phi_b | g_{12} | \phi_i \phi_j \rangle - \langle \phi_a \phi_b | P_{12} \cdot g_{12} | \phi_i \phi_j \rangle \\
 \langle \Phi_{ij}^{ab} | W | \Phi_0 \rangle &= \langle \phi_a \phi_b | \phi_i \phi_j \rangle - \langle \phi_a \phi_b | \phi_j \phi_i \rangle
 \end{aligned} \tag{1.96}$$

De même

$$\langle \Phi_0 | W | \Phi_{ij}^{ab} \rangle = \langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle \quad (1.97)$$

On sait que les déterminants de Slater (de référence et des excitations) sont fonctions propres de l'opérateur de référence $H_0 = \sum_i F_i$, $H_0 \Phi_i = E_i \Phi_i = \sum_j \epsilon_j \Phi_i$ Ainsi

$$\begin{aligned} E_0 &= \sum_i^{occ} \epsilon_i \\ E_{ij}^{ab} &= \sum_{k=0}^{occ} \epsilon_k - \epsilon_i - \epsilon_j + \epsilon_a + \epsilon_b = E_0 - \epsilon_i - \epsilon_j + \epsilon_a + \epsilon_b \\ E_0 - E_{ij}^{ab} &= \epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b \end{aligned} \quad (1.98)$$

On obtient finalement pour l'expression de la correction au second ordre de l'énergie :

$$E_2 = \frac{1}{4} \sum_{i,j}^{occ} \sum_{a,b}^{vir} \frac{(\langle \phi_a \phi_b | \phi_i \phi_j \rangle - \langle \phi_a \phi_b | \phi_j \phi_i \rangle)^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad (1.99)$$

1.5 La théorie de la fonctionnelle de la densité

Les méthodes précédentes, Hartree-Fock et post Hartree-Fock, reposent sur la fonction d'onde du système. C'est une fonction à $4N$ variables composées des 3 variables de position et 1 variable de spin pour chaque électron. Dans la théorie de la densité fonctionnelle, l'énergie devient une fonctionnelle de la densité électronique. Le problème ne dépend plus que de 4 variables (3 d'espace et 1 de spin). De plus le temps de calcul de la DFT est de l'ordre de N_f^3 voire N_f^4 (pour les fonctionnelles hybrides, voir suite) avec N_f le nombre de fonctions de bases et est inférieur à celui des méthodes *ab initio*. Pour Hartree-Fock il est en N_f^4 , pour MP2 il est en N_f^5 , pour CISD et CCSD(T) c'est de l'ordre de N_f^6 . Contrairement à Hartree-Fock qui ne contient pas de corrélation, le potentiel de la DFT contient une contribution explicite de la corrélation électronique.

On peut commencer par définir la matrice densité électronique du système :

$$\gamma_N(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N; \vec{r}_1', \vec{r}_2', \dots, \vec{r}_N') = \Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) \times \Psi^*(\vec{r}_1', \vec{r}_2', \dots, \vec{r}_N') \quad (1.100)$$

Avec $\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)$ la fonction d'onde du système de N électrons. Ainsi que la matrice densité réduite (RDM) :

$$\gamma_p(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_p; \vec{r}_1', \vec{r}_2', \dots, \vec{r}_p') = C_p^N \int \dots \int \gamma_N(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N; \vec{r}_1', \vec{r}_2', \dots, \vec{r}_N') d\vec{r}_{p+1} \dots d\vec{r}_N \quad (1.101)$$

Elle correspond à la densité de N particules prises par groupes de p particules. C_p^N est le coefficient binomial qui représente le nombre de combinaisons possibles de choisir p particules parmi N particules.

On obtient pour la matrice réduite à une particule, aussi appelée 1-RDM :

$$\gamma_1(\vec{r}_1; \vec{r}_1') = N \int \dots \int \gamma_N(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N; \vec{r}_1', \vec{r}_2', \dots, \vec{r}_N') d\vec{r}_2 \dots d\vec{r}_N \quad (1.102)$$

Si on restreint la matrice à sa diagonale, on obtient la densité d'une particule dans l'élément de volume $d\vec{r}_1$.

$$\gamma_1(\vec{r}_1; \vec{r}_1) = N \int \dots \int |\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)|^2 d\vec{r}_2 \dots d\vec{r}_N = \rho_1(\vec{r}_1) \quad (1.103)$$

De même on peut obtenir la matrice réduite à deux particules (2-RDM) et si on la restreint à sa diagonale, on a ainsi accès à la densité de la particule 1 dans l'élément de volume $d\vec{r}_1$ et la densité de la particule 2 dans l'élément de volume $d\vec{r}_2$:

$$\gamma_2(\vec{r}_1, \vec{r}_2; \vec{r}_1, \vec{r}_2) = \rho_2(\vec{r}_1, \vec{r}_2) = C_2^N \int \dots \int |\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)|^2 d\vec{r}_3 \dots d\vec{r}_N \quad (1.104)$$

Hohenberg et Kohn [11] ont montré, en injectant les matrices 1-RDM et 2-RDM dans l'expression de l'énergie, que cette dernière est une fonctionnelle de la densité électronique. La définition générale de l'énergie est :

$$E = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (1.105)$$

avec H l'hamiltonien du système moléculaire déjà défini pour Hartree-Fock. L'expression de cette énergie devient donc :

$$\begin{aligned} E &= \int \left[-\frac{1}{2} \nabla^2 \gamma_1(\vec{r}_1; \vec{r}_1') \right]_{\vec{r}_1=\vec{r}_1'} d\vec{r}_1 - \sum_A \int \left[\frac{Z_A}{|\vec{r}_1 - \vec{R}_A|} \gamma_1(\vec{r}_1; \vec{r}_1') \right]_{\vec{r}_1=\vec{r}_1'} d\vec{r}_1 \\ &\quad + \int \int \left[\frac{\gamma_2(\vec{r}_1, \vec{r}_2; \vec{r}_1', \vec{r}_2')}{|\vec{r}_1 - \vec{r}_2|} \right]_{\substack{\vec{r}_1=\vec{r}_1' \\ \vec{r}_2=\vec{r}_2'}} d\vec{r}_1 d\vec{r}_2 \\ E &= \underbrace{\int -\frac{1}{2} \nabla^2 \rho_1(\vec{r}_1) d\vec{r}_1}_{\text{énergie cinétique}} - \underbrace{\sum_A \int \frac{Z_A}{|\vec{r}_1 - \vec{R}_A|} \rho_1(\vec{r}_1) d\vec{r}_1}_{\text{attraction nucléaire}} + \underbrace{\int \int \frac{\rho_2(\vec{r}_1, \vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2}_{\text{répulsion électron-électron}} \end{aligned} \quad (1.106)$$

Les éléments diagonaux des matrices densité réduites au premier et second ordre déterminent entièrement l'énergie totale du système. Comme énoncé au tout début, le calcul direct de la fonction d'onde du système Ψ à $3N$ variables d'espace (en ne comptant pas le spin) est évité et remplacé simplement par le calcul de ρ_1 et de ρ_2 comptabilisant 6 variables d'espace. En l'état l'équation (1.106) n'est pas utilisable parce qu'on ne sait pas construire la densité électronique à deux électrons (qui contient la corrélation électronique). Il faut procéder d'une autre manière, ce sont les théorèmes de Hohenberg-Kohn et la théorie de Kohn-Sham. Ils ont montré que seule la matrice 1-RDM était nécessaire pour le calcul de l'énergie d'un système.

1.5.1 Les théorèmes de Hohenberg-Kohn

Les deux théorèmes de Hohenberg-Kohn [11] ont permis de concevoir le cadre théorique de la DFT.

Premier théorème

Enoncé : pour tout système de particules en interaction dans un potentiel externe $V_{ext}(\vec{r})$, le potentiel $V_{ext}(\vec{r})$ est uniquement déterminé, à une constante additive près, par la densité $\rho_0(\vec{r})$ à une particule dans son état fondamental.

Démonstration

Supposons que $V_{ext}(\vec{r})$ soit le potentiel externe correspondant à l'hamiltonien H , l'état fondamental non dégénéré est décrit par la fonction d'onde Ψ . Prenons maintenant un autre potentiel externe différent de $V_{ext}(\vec{r})$ par autre chose qu'une constante additive, $V'_{ext}(\vec{r})$. L'hamiltonien correspondant est H' et la fonction d'onde de l'état fondamental est Ψ' . On suppose que les fonctions d'onde Ψ et Ψ' engendrent la même densité électronique ρ_0 . Soit E l'énergie de l'état fondamental pour l'hamiltonien H , alors on a :

$$E = \langle \Psi | H | \Psi \rangle < \langle \Psi' | H | \Psi' \rangle \quad (1.107)$$

Cette inégalité vient du fait que Ψ' n'est pas l'état fondamental pour l'hamiltonien H et on a une égalité stricte parce que l'état fondamental de H n'est pas dégénéré.

$$\begin{aligned} \langle \Psi' | H | \Psi' \rangle &= \langle \Psi' | H' | \Psi' \rangle + \langle \Psi' | H - H' | \Psi' \rangle \\ &= E' + \int (V_{ext}(\vec{r}) - V'_{ext}(\vec{r})) \rho_0(\vec{r}) d\vec{r} \end{aligned} \quad (1.108)$$

Donc

$$E < E' + \int (V_{ext}(\vec{r}) - V'_{ext}(\vec{r})) \rho_0(\vec{r}) d\vec{r} \quad (1.109)$$

Nous avons supposé dans l'énoncé que les fonctions d'onde Ψ et Ψ' engendraient la même densité $\rho_0(\vec{r})$, alors en échangeant les énergies E et E' on aura l'inégalité suivante :

$$E' < E + \int (V'_{ext}(\vec{r}) - V_{ext}(\vec{r})) \rho_0(\vec{r}) d\vec{r} \quad (1.110)$$

On tombe alors sur une contradiction :

$$E' + E < E' + E \quad (1.111)$$

La contradiction entraîne le rejet de l'hypothèse que l'on peut avoir deux potentiels externes qui engendrent une même densité électronique ρ_0 pour un état fondamental non dégénéré. Le potentiel externe est une fonctionnelle unique de la densité $\rho_0(\vec{r})$.

Second théorème

Enoncé : il existe une fonctionnelle universelle $E[\rho]$ exprimant l'énergie en fonction de la densité électronique $\rho(\vec{r})$, valide pour tout potentiel externe $V_{ext}(\vec{r})$. Pour chaque $V_{ext}(\vec{r})$ particulier, l'énergie de l'état fondamental du système est la valeur qui minimise cette fonctionnelle, la densité $\rho(\vec{r})$ qui lui est associée correspond à la densité exacte $\rho_0(\vec{r})$ de l'état fondamental.

Démonstration

Pour toute densité ρ ,

$$\begin{aligned} F_{HK}[\rho] + \int V_{ext}(\vec{r}) \rho(\vec{r}) d\vec{r} &= T[\rho] + E_{ee}[\rho] + \int V_{ext}(\vec{r}) \rho(\vec{r}) d\vec{r} \\ &= \langle \Psi[\rho] | \hat{T} + \hat{W}_{ee} | \Psi[\rho] \rangle + \langle \Psi[\rho] | \hat{V}_{ext} | \Psi[\rho] \rangle \\ &= \langle \Psi[\rho] | \hat{T} + \hat{W}_{ee} + \hat{V}_{ext} | \Psi[\rho] \rangle \\ &= E \geq E_0 \end{aligned} \quad (1.112)$$

$F_{HK}[\rho]$ est la fonctionnelle de Hohenberg-Kohn constituée de l'énergie cinétique des particules et de l'énergie d'interaction entre ces particules. C'est une fonctionnelle universelle qui ne dépend pas du système ni du potentiel externe. L'énergie E du système est supérieure ou égale à l'énergie de l'état fondamental. D'après le premier théorème de Hohenberg et Kohn, si la fonction d'onde est celle de l'état fondamental alors la densité est la densité de l'état fondamental donc l'énergie est l'énergie de l'état fondamental.

$$F_{HK}[\rho_0] + \int V_{ext}(\vec{r})\rho_0(\vec{r})d\vec{r} = \langle \Psi[\rho_0] | \hat{T} + \hat{W}_{ee} + \hat{V}_{ext} | \Psi[\rho_0] \rangle = E_0 \quad (1.113)$$

d'où

$$E_0 = \min_{\rho} \{E[\rho]\} = \min_{\rho} \left\{ F_{HK}[\rho] + \int V_{ext}(\vec{r})\rho(\vec{r})d\vec{r} \right\} \quad (1.114)$$

L'énergie est une fonctionnelle de la densité et pour tout potentiel extérieur, la densité qui minimise cette fonctionnelle est la densité exacte de l'état fondamental.

1.5.2 La théorie de Kohn-Sham

L'expression de l'opérateur $F_{HK} = T + E_{ee}$ qui est une fonctionnelle de la densité électronique n'est pas connue. La résolution n'est pas possible sans la connaissance de son expression. L'astuce de Kohn-Sham [12][13] est de remplacer le système de N électrons interagissant entre eux dans un potentiel extérieur par un système de N électrons non-interagissant entre eux dans un potentiel externe effectif. L'idée est alors de dire que les

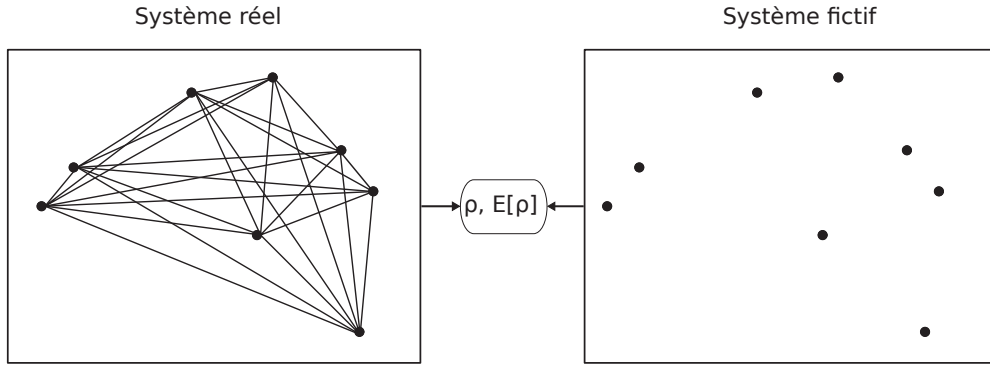


FIGURE 1.5 – Schéma représentatif de la méthode de Kohn-Sham. Le système réel à N particules interagissant entre elles possède la même densité et la même énergie qu'un système fictif à N particules qui n'interagissent pas entre elles.

énergies de ces deux cas de figure sont égales. On est maintenant capable d'écrire l'énergie cinétique de ce système fictif. L'Hamiltonien de ce nouveau système est donc $F_{KS} = t_s + V_s$ où est t_s l'énergie cinétique classique ($-\frac{1}{2}\vec{\nabla}^2$) d'un électron et V_s le potentiel effectif ressenti par cet électron. On considère un système de N électrons dans un potentiel externe V_s ayant ρ pour densité correspondant à la densité de l'état fondamental. La fonction d'onde de l'état fondamental du système est un déterminant de Slater de N spin-orbitales ϕ_i devant chacune satisfaire la nouvelle équation :

$$F_{KS}\phi_i = \left[-\frac{1}{2}\vec{\nabla}^2 + V_s(\vec{r})\right]\phi_i = \epsilon_i\phi_i \quad (1.115)$$

Ces équations sont semblables à celles de Hartree-Fock qui sont obtenues après minimisation du Lagrangien.

Le premier théorème de Hohenberg-Kohn stipule l'existence d'un potentiel extérieur V_s qui génère la densité électronique ρ .

L'énergie cinétique $T_s[\rho]$ et l'énergie totale $E[\rho]$ s'écrivent comme

$$T_s[\rho] = \sum_{i=1}^N -\frac{1}{2} \int \phi_i^* \vec{\nabla}^2 \phi_i d\vec{r} \quad (1.116)$$

$$E[\rho] = \sum_i \epsilon_i = T_s[\rho] + \int \rho(\vec{r}) V_s(\vec{r}) d\vec{r} \quad (1.117)$$

Cette énergie est l'énergie du système fictif, du système d'électrons non interagissant entre eux dans un potentiel effectif. Les opérateurs T et E_{ee} sont universels a contrario du potentiel extérieur V_{ext} qui lui dépend du système, des atomes qui le constituent. L'énergie associée au système réel est :

$$E[\rho] = T[\rho] + E_{ee}[\rho] + \int \rho(\vec{r}) V_{ext}(\vec{r}) d\vec{r} \quad (1.118)$$

L'interaction Coulombienne classique entre électrons est

$$J[\rho] = E_{ee}[\rho] - E_{non\,classique}[\rho] = \frac{1}{2} \int \int \frac{\rho(\vec{r}) \rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' \quad (1.119)$$

où $E_{non\,classique}$ est l'énergie non classique constituée du terme d'échange et du terme de corrélation. On peut faire apparaître dans l'énergie du système réel l'énergie cinétique T_s et le terme d'interaction coulombienne ce qui permet de séparer le tout en deux termes : un terme calculable analytiquement et un terme d'échange-corrélation dont on ne connaît pas l'expression.

$$E[\rho] = \underbrace{\int \rho(\vec{r}) V_{ext}(\vec{r}) d\vec{r} + T_s[\rho] + J[\rho]}_{\text{expr. analytique connue}} + \underbrace{T[\rho] - T_s[\rho] + E_{ee}[\rho] - J[\rho]}_{\text{échange-corrélation}} \quad (1.120)$$

Il y a une contribution de l'énergie cinétique dans le terme d'échange-corrélation.

L'intérêt de faire cette séparation est d'obtenir une équation pour le potentiel effectif V_s lorsqu'on égalisera l'énergie du système réel et celle du système fictif.

On peut écrire que l'énergie d'échange-corrélation est engendrée par un potentiel d'échange-corrélation V_{xc} .

$$E_{xc}[\rho] = T[\rho] - T_s[\rho] + E_{ee}[\rho] - J[\rho] = \int V_{xc}(\vec{r}) \rho(\vec{r}) d\vec{r} \quad (1.121)$$

On obtient ainsi une nouvelle expression pour l'énergie du système réel.

$$E[\rho] = T_s[\rho] + J[\rho] + \int \rho(\vec{r}) V_{ext}(\vec{r}) d\vec{r} + E_{xc}[\rho] \quad (1.122)$$

Le système fictif et le système réel doivent avoir même énergie et même densité électronique, cela fait partie de la théorie de Kohn-Sham. On obtient en égalisant les deux énergies :

$$T_s[\rho] + \int \rho(\vec{r}) V_s(\vec{r}) d\vec{r} = T_s[\rho] + J[\rho] + \int \rho(\vec{r}) V_{ext}(\vec{r}) d\vec{r} + E_{xc}[\rho] \quad (1.123)$$

et en appliquant le principe variationnel par rapport à la densité électronique, on trouve

$$\frac{\delta(T_s[\rho])}{\delta\rho(\vec{r})} + V_s(\vec{r}) = \frac{\delta(T_s[\rho])}{\delta\rho(\vec{r})} + \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' + V_{ext}(\vec{r}) + \frac{\delta(E_{xc}[\rho])}{\delta\rho(\vec{r})} \quad (1.124)$$

$$V_s(\vec{r}) = V_{ext}(\vec{r}) + \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' + V_{xc}(\vec{r}) \quad (1.125)$$

Une fois le potentiel $V_s(\vec{r})$ obtenu, les équations de Kohn-Sham peuvent être résolues. Les solutions obtenues, les différentes orbitales ϕ_i doivent satisfaire les conditions suivantes :

$$\sum_i^N n_i |\phi_i|^2 = \rho(\vec{r}) \quad \text{et} \quad \langle \phi_i | \phi_j \rangle = \delta_{ij} \quad (1.126)$$

A partir des orbitales, on peut calculer la valeur de la densité électronique ρ . Cette valeur de la densité engendre une nouvelle valeur de V_s (qui dépend de ρ). On doit résoudre à nouveau les équations de Kohn-Sham avec cette nouvelle valeur du potentiel V_s et obtenir les orbitales ϕ_i et ainsi de suite. La résolution de ces équations doit se faire de manière itérative à l'aide d'un SCF (déjà vu dans le cadre de Hartree-Fock). On réitère la procédure tant que la convergence n'a pas été atteinte. La valeur finale de ρ nous permet facilement de trouver la valeur de l'énergie en utilisant l'équation (1.122).

La DFT présente un problème important dans le sens où pour calculer V_s il est nécessaire de connaître le potentiel d'échange-corrélation V_{xc} or ce dernier n'est pas connu. La théorie nous dit que ce potentiel existe mais ne nous donne pas son expression. Sinon le résultat serait exact dans le cadre de la théorie. Il a fallu concevoir des fonctionnelles qui approximent ce terme pour résoudre les équations de Kohn-Sham. Le terme E_{xc} peut souvent se mettre sous la forme d'une somme de deux contributions $E_{xc} = E_x + E_c$ avec E_x le terme d'échange et E_c le terme de corrélation.

Il existe plusieurs approximations pour le terme d'échange-corrélation E_{xc} , nous allons présenter ces approximations.

1.5.3 Approximations de la fonctionnelle de la densité

LDA : Approximation de la Densité Locale

Cette approximation consiste à considérer que localement la densité de notre système

est égale à celle d'un gaz uniforme d'électrons .

$$\begin{aligned}
E_x^{LDA}[\rho] &= \int \epsilon_x[\rho(\vec{r})]\rho(\vec{r})d\vec{r} \\
\epsilon_x[\rho(\vec{r})] &= -C_x\rho(\vec{r})^{\frac{1}{3}} \\
E_x^{LDA}[\rho] &= -C_x \int \rho(\vec{r})^{\frac{4}{3}}d\vec{r}
\end{aligned} \tag{1.127}$$

La LSDA [14], l'approximation locale de la densité de spins est la généralisation de la LDA. Ici on prend en compte le spin de l'électron. Les électrons étant de spin α et β le terme d'échange-corrélation de l'énergie devient :

$$\begin{aligned}
E_x[\rho_\alpha, \rho_\beta] &= \int \epsilon_x[\rho_\alpha, \rho_\beta]\rho(\vec{r})d\vec{r} \\
E_x[\rho_\alpha, \rho_\beta] &= -2^{\frac{1}{3}}C_x \int (\rho_\alpha^{\frac{4}{3}} + \rho_\beta^{\frac{4}{3}})d\vec{r}
\end{aligned} \tag{1.128}$$

L'expression de la corrélation pour un gaz d'électrons uniforme a été obtenue analytiquement aux limites de densités faibles et grandes. Pour ce qui est de la zone intermédiaire, il faut ajuster une fonction analytique sur les résultats obtenus par la méthode de Monte Carlo quantique. Vosko, Wilk et Nusair (VWN) [15] et Perdew et Wang (PW) [16] ont produit des ajustements très précis.

GGA : Approximation du gradient généralisé

Dans l'approximation locale de la densité, on considèrerait que la densité électronique variait très faiblement dans l'espace. Dans l'approximation du gradient généralisé, on abandonne cette propriété et on ajoute une dépendance sur le gradient de la densité électronique dans l'énergie d'échange-corrélation.

$$E_{xc}[\rho_\alpha, \rho_\beta] = \int \epsilon_{xc}(\rho_\alpha, \rho_\beta, |\vec{\nabla}\rho_\alpha|, |\vec{\nabla}\rho_\beta|)\rho(\vec{r})d\vec{r} \tag{1.129}$$

L'une des premières fonctionnelles GGA pour l'échange a été développée par A.D.Becke [17] en 1988 pour corriger la fonctionnelle LSDA.

$$\begin{aligned}
\epsilon_x^{B88} &= \epsilon_x^{LDA} + \Delta\epsilon_x^{B88} \\
\Delta\epsilon_x^{B88} &= -\beta\rho^{\frac{1}{3}} \frac{x^2}{1 + 6\beta x \sinh^{-1}x} \\
x &= \frac{|\nabla\rho|}{\rho^{\frac{4}{3}}}
\end{aligned} \tag{1.130}$$

Cette fonctionnelle possède un paramètre, β , qui a été ajusté sur des données d'atomes de gaz rare en phase gazeuse.

Lee, Yang et Parr (LYP) [18] ont développé une fonctionnelle GGA de corrélation. Cette fonctionnelle possède quatre paramètres ajustés sur des données de l'atome d'hélium.

La fonctionnelle d'échange-corrélation BLYP est la combinaison de la fonctionnelle GGA d'échange B88 et de la fonctionnelle GGA de corrélation LYP.

Une autre fonctionnelle GGA importante est la fonctionnelle PBE, développée par Perdew, Burke et Ernzerhof [19]. La partie échange s'écrit comme le produit de la fonctionnelle d'échange LDA et d'un facteur de forme F qui dépend du gradient de la densité (ou gradient réduit x).

$$\begin{aligned}\epsilon_x^{PBE} &= \epsilon_x^{LDA} F(x) \\ F(x) &= 1 + a - \frac{a}{1 + bx^2} \\ x &= \frac{|\nabla\rho|}{\rho^{\frac{4}{3}}}\end{aligned}\tag{1.131}$$

De même pour la partie corrélation, on ajoute un facteur de forme à la fonctionnelle LDA de corrélation.

$$\begin{aligned}\epsilon_c^{PBE} &= \epsilon_c^{LDA} + H(t) \\ H(t) &= cf_3^3 \ln \left[1 + dt^2 \left(\frac{1 + At^2}{1 + At^2 + A^2 t^4} \right) \right] \\ A &= d \left[\exp \left(- \frac{\epsilon_c^{LDA}}{cf_3^3} - 1 \right) \right]^{-1} \\ f_3 &= \frac{1}{2} \left[(1 + \zeta)^{\frac{2}{3}} + (1 - \zeta)^{\frac{2}{3}} \right] \\ \zeta &= \frac{\rho_\alpha - \rho_\beta}{\rho_\alpha + \rho_\beta} \\ t &= \left[2(3\pi^3)^{\frac{1}{3}} f_3 \right]^{-1} x\end{aligned}\tag{1.132}$$

Cette fois-ci les paramètres a , b , c et d ne sont pas ajustés sur des données mais obtenus sur des considérations théoriques.

On peut inclure des ordres supérieurs de la dérivée de la densité électronique dans l'expression de l'énergie d'échange-corrélation. Si on ajoute une dépendance aux dérivées secondes de la densité électronique (son Laplacien) ou une dépendance à la densité d'énergie cinétique des orbitales, on obtiendra une fonctionnelle appartenant aux fonctionnelles méta-GGA.

Méthodes hybrides ou hybrides-GGA

Les fonctionnelles hybrides sont des fonctionnelles LDA ou GGA qui prennent en compte une partie du terme de l'échange exact provenant de la théorie de Hartree-Fock. La fonctionnelle la plus connue parmi les méthodes hybrides est la fonctionnelle B3LYP [17, 18, 20] dont le terme d'échange-corrélation s'écrit :

$$E_{xc}^{B3LYP} = (1 - a_0)E_x^{LDA} + a_0E_x^{HF} + a_x\Delta E_x^{B88} + (1 - a_c)E_c^{LDA} + a_cE_c^{LYP}\tag{1.133}$$

Cette fonctionnelle dépend de 3 paramètres, $a_0 = 0.20$, $a_x = 0.72$ et $a_c = 0.81$ ajustés sur des résultats expérimentaux.

Méthodes LC : méthodes de correction à grande distance

Les méthodes à grande distance ont été introduites pour permettre de corriger les problèmes que l'on rencontre avec des fonctionnelles conventionnelles sur certains métaux. En effet, avec ces dernières, on sous-estime les énergies de transition 4s - 3d de la première

période des métaux de transition. Ces méthodes permettent de mieux traiter les transferts de charge lors d'excitations électroniques. En effet, à grande distance, le potentiel d'échange n'a pas le comportement en $-\frac{1}{r}$ attendu. Cela peut être corrigé en utilisant la séparation d'Ewald [21]

$$\frac{1}{|\vec{r} - \vec{r}'|} = \frac{1 - \text{erf}(\mu|\vec{r} - \vec{r}'|)}{|\vec{r} - \vec{r}'|} + \frac{\text{erf}(\mu|\vec{r} - \vec{r}'|)}{|\vec{r} - \vec{r}'|} \quad (1.134)$$

où μ est un paramètre de coupure qui permet de faire la séparation entre courte distance et grande distance. L'échange calculé par la DFT sera pris en compte à courte distance par le premier terme de l'équation (1.134) et l'échange exact de Hartree-Fock par le second terme.

La fonctionnelle CAM-B3LYP [22] est basée sur la fonctionnelle B3LYP pour laquelle on prend en compte la correction à grande distance. La formule qui corrige le comportement du potentiel d'échange est légèrement modifiée par l'introduction de deux paramètres.

$$\frac{1}{|\vec{r} - \vec{r}'|} = \frac{1 - [\alpha + \beta \cdot \text{erf}(\mu|\vec{r} - \vec{r}'|)]}{|\vec{r} - \vec{r}'|} + \frac{\alpha + \beta \cdot \text{erf}(\mu|\vec{r} - \vec{r}'|)}{|\vec{r} - \vec{r}'|} \quad (1.135)$$

Cette écriture permet de prendre en compte l'échange DFT et l'échange exact dans les deux zones (courte et grande distance) pondérés par les coefficients α et β . Pour CAM-B3LYP, $\mu = 0.33$, $\alpha = 0.19$ et $\beta = 0.46$.

Méthodes double hybrides On a vu que pour les méthodes hybrides une partie du terme d'échange provient du terme d'échange exact d'Hartree-Fock. Les méthodes double hybrides consiste à calculer une partie de la corrélation via MP2. C'est le terme E_2 de MP2 (1.99) qui est calculé à l'aide des orbitales de Kohn-Sham et une fraction de ce terme est ensuite ajouté au terme d'échange-corrélation E_{xc} de la fonctionnelle. Cette incorporation permet d'améliorer le traitement de la corrélation statique. La formule générale pour les fonctionnelles double hybride est :

$$E_{xc}^{DH} = (1 - a_x)E_x + a_x E_x^{HF} + (1 - a_c)E_c + a_c E_c^{PT2} \quad (1.136)$$

L'une des plus utilisées et l'une des meilleures fonctionnelles double hybrides est la fonctionnelle B2PLYP. Le terme d'échange-corrélation est décomposé selon :

$$E_{xc}^{DH} = (1 - 0.53)E_x^{B88} + 0.53E_x^{HF} + (1 - 0.27)E_c^{LYP} + 0.27E_c^{PT2} \quad (1.137)$$

1.6 La DFTB : density functional based tight-binding

L'emploi du formalisme de "tight-binding" (TB) pour le calcul des structures électroniques a débuté avec les travaux de J. Slater et G. Koster [23]. Les états propres de l'Hamiltonien sont décrits avec une base d'orbitales atomiques et l'Hamiltonien est remplacé par un Hamiltonien paramétré dont les éléments de matrice dépendent seulement des distances internucléaires (les intégrales à plus de deux centres doivent être négligées) et des symétries des orbitales. La DFTB est une méthode approchée et paramétrée à partir de la DFT. Nous avons utilisé la DFTB afin d'évaluer les énergies et les fréquences harmoniques obtenus avec la DFT. Cependant, les résultats n'étaient pas satisfaisants. Nous avons alors optimisé

certain paramètre de la DFTB afin d'améliorer les énergies et les fréquences harmoniques de la DFTB. Bien que les résultats soient meilleurs après cette étape d'optimisation, cette amélioration n'est pas suffisante pour que l'on puisse utiliser la DFTB pour comparer avec l'expérience. Nous avons aussi utilisé la DFTB dans le cadre des méthodes hybrides pour des calculs anharmoniques (présentées dans le chapitre 3).

L'équation (1.122) qui donne l'énergie de la DFT peut se mettre sous la forme suivante :

$$E[\rho] = \sum_i^N n_i \langle \phi_i | -\frac{1}{2} \nabla^2 + V_{ext}(\vec{r}) + \frac{1}{2} \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' | \phi_i \rangle + E_{xc}[\rho] + E_{nn} \quad (1.138)$$

Cela équivaut à :

$$E[\rho] = \sum_i^N n_i \langle \phi_i | -\frac{1}{2} \nabla^2 | \phi_i \rangle + \int V_{ext}(\vec{r}) \rho(\vec{r}) d\vec{r} + \frac{1}{2} \iint \frac{\rho(\vec{r}') \rho(\vec{r})}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + E_{xc}[\rho] + E_{nn} \quad (1.139)$$

Où E_{nn} est le terme de répulsion nucléaire que l'on a réintroduit dans l'équation.

On écrit la densité électronique au point $\vec{r}$ comme la somme de la densité de référence $\rho_0(\vec{r})$ et d'une faible fluctuation $\delta\rho(\vec{r})$: $\rho(\vec{r}) = \rho_0(\vec{r}) + \delta\rho(\vec{r})$

On injecte maintenant cette relation dans l'équation précédente et on obtient :

$$\begin{aligned} E[\rho_0 + \delta\rho] &= \sum_i^N n_i \langle \phi_i | -\frac{1}{2} \nabla^2 + V_{ext}(\vec{r}) + \frac{1}{2} \int \frac{\rho'_0}{|\vec{r} - \vec{r}'|} d\vec{r}' | \phi_i \rangle \\ &\quad + \frac{1}{2} \sum_i^N n_i \langle \phi_i | \int \frac{\delta\rho'}{|\vec{r} - \vec{r}'|} d\vec{r}' | \phi_i \rangle + E_{xc}[\rho_0 + \delta\rho] + E_{nn} \end{aligned} \quad (1.140)$$

ρ' est la valeur de ρ en $\vec{r}'$

$$\begin{aligned} \sum_i^N n_i \langle \phi_i | \frac{1}{2} \int \frac{\rho'_0}{|\vec{r} - \vec{r}'|} d\vec{r}' | \phi_i \rangle &= \sum_i^N n_i \langle \phi_i | \int \frac{\rho'_0}{|\vec{r} - \vec{r}'|} d\vec{r}' | \phi_i \rangle - \frac{1}{2} \sum_i^N n_i \langle \phi_i | \int \frac{\rho'_0}{|\vec{r} - \vec{r}'|} d\vec{r}' | \phi_i \rangle \\ \sum_i^N n_i \langle \phi_i | \frac{1}{2} \int \frac{\rho'_0}{|\vec{r} - \vec{r}'|} d\vec{r}' | \phi_i \rangle &= \sum_i^N n_i \langle \phi_i | \int \frac{\rho'_0}{|\vec{r} - \vec{r}'|} d\vec{r}' | \phi_i \rangle - \frac{1}{2} \iint \frac{\rho'_0 \rho}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' \end{aligned} \quad (1.141)$$

$$\begin{aligned} E[\rho_0 + \delta\rho] &= \sum_i^N n_i \langle \phi_i | -\frac{1}{2} \nabla^2 + V_{ext}(\vec{r}) + \int \frac{\rho'_0}{|\vec{r} - \vec{r}'|} d\vec{r}' + V_{xc}[\rho_0] | \phi_i \rangle \\ &\quad - \frac{1}{2} \iint \frac{\rho'_0 (\rho_0 + \delta\rho)}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' - \int V_{xc}[\rho_0] (\rho_0 + \delta\rho) d\vec{r} \\ &\quad + \frac{1}{2} \iint \frac{\delta\rho' (\rho_0 + \delta\rho)}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + E_{xc}[\rho_0 + \delta\rho] + E_{nn} \end{aligned} \quad (1.142)$$

On a ajouté le potentiel d'échange-corrélation dans le premier et le deuxième terme mais les deux occurrences n'ont pas le même signe pour qu'elles puissent s'annuler.

La prochaine étape consiste à développer le terme d'échange-corrélation en série de Taylor :

$$E_{xc}[\rho_0 + \delta\rho] = E_{xc}[\rho_0] + \int \frac{\delta E_{xc}}{\delta\rho}|_{\rho_0} \delta\rho d\vec{r} + \frac{1}{2} \iint \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'}|_{\rho_0} \delta\rho\delta\rho' d\vec{r}d\vec{r}' + \frac{1}{6} \iiint \frac{\delta^3 E_{xc}}{\delta\rho\delta\rho'\delta\rho''}|_{\rho_0} \delta\rho\delta\rho'\delta\rho'' d\vec{r}d\vec{r}'d\vec{r}'' + \dots \quad (1.143)$$

En utilisant le fait que $\frac{\delta E_{xc}}{\delta\rho}|_{\rho_0} = V_{xc}[\rho_0]$ et le développement de Taylor, on écrit l'équation sous la forme suivante :

$$\begin{aligned} E &= E_{bnd} + E_{rep}[\rho_0] + E_{2nd}[\rho_0, \delta\rho] \\ E &= \sum_i^N n_i \langle \phi_i | -\frac{1}{2} \nabla^2 + V_{ext}(\vec{r}) + \int \frac{\rho_0'}{|\vec{r} - \vec{r}'|} d\vec{r}' + V_{xc}[\rho_0] | \phi_i \rangle \\ &\quad - \frac{1}{2} \iint \frac{\rho_0' \rho_0}{|\vec{r} - \vec{r}'|} d\vec{r}d\vec{r}' + E_{xc}[\rho_0] + E_{nn} - \int V_{xc}[\rho_0] \rho_0 d\vec{r} \\ &\quad + \frac{1}{2} \iint \left(\frac{\delta\rho' \delta\rho}{|\vec{r} - \vec{r}'|} + \frac{\delta^2 E_{xc}}{\delta\rho' \delta\rho}|_{\rho_0} \delta\rho\delta\rho' \right) d\vec{r}d\vec{r}' \end{aligned} \quad (1.144)$$

$$\begin{aligned} E_{bnd} &= \sum_i^N n_i \langle \phi_i | -\frac{1}{2} \nabla^2 + V_{ext}(\vec{r}) + \int \frac{\rho_0'}{|\vec{r} - \vec{r}'|} d\vec{r}' + V_{xc}[\rho_0] | \phi_i \rangle \\ E_{rep}[\rho_0] &= -\frac{1}{2} \iint \frac{\rho_0' \rho_0}{|\vec{r} - \vec{r}'|} d\vec{r}d\vec{r}' + E_{xc}[\rho_0] + E_{nn} - \int V_{xc}[\rho_0] \rho_0 d\vec{r} \\ E_{2nd}[\rho_0, \delta\rho] &= \frac{1}{2} \iint \left(\frac{\delta\rho' \delta\rho}{|\vec{r} - \vec{r}'|} + \frac{\delta^2 E_{xc}}{\delta\rho' \delta\rho}|_{\rho_0} \delta\rho\delta\rho' \right) d\vec{r}d\vec{r}' \end{aligned} \quad (1.145)$$

Le terme E_{bnd} ne dépend que de la densité électronique de référence et est la somme des énergies de toutes les orbitales occupées. Le second terme, $E_{rep}[\rho_0]$, correspond à la partie répulsive de l'Hamiltonien. Enfin le dernier terme, $E_{2nd}[\rho_0, \delta\rho]$, contient les corrections dues à la variation de la densité électronique.

La densité électronique de référence peut se mettre sous la forme d'une somme de pseudo-densités atomiques : $\rho_0(\vec{r}) = \sum_{\alpha=1}^M \rho_0^\alpha(\vec{r}_\alpha)$, $\vec{r}_\alpha = \vec{r} - \vec{R}_\alpha$ où M est le nombre d'atomes de la molécule. Ainsi la partie répulsion de l'énergie s'écrit :

$$E_{rep} = \sum_{\alpha}^M E_{rep}[\rho_0^\alpha] + \frac{1}{2} \sum_{\beta}^M \sum_{\alpha \neq \beta}^M (E_{rep}[\rho_0^\alpha + \rho_0^\beta] - E_{rep}[\rho_0^\alpha] - E_{rep}[\rho_0^\beta]) + \text{termes à 3 centres et plus} \quad (1.146)$$

Les termes à trois centres et plus peuvent être négligés. Cela est justifié par l'écrantage coulombien puisque ρ_0^α est la densité électronique d'un atome neutre, les termes d'interaction electron-electron de plus de deux centres sont annulés par les interactions noyaux-noyaux. Les interactions à deux centres sont de courte portée.

A grandes distances interatomiques, l'énergie de répulsion tend vers une valeur limite qui est :

$$\lim_{R_{\alpha\beta} \rightarrow +\infty} E_{rep}[\rho_0] = \sum_{\alpha}^M E_{rep}[\rho_0^\alpha] \quad (1.147)$$

Pour qu'elle dépende uniquement des contributions à deux centres, on suppose que :

$$E_{rep}[\rho_0] \approx \frac{1}{2} \sum_{\alpha, \beta}^M U[\rho_0^\alpha, \rho_0^\beta] \quad (1.148)$$

L'énergie E_{bnd} est obtenue en diagonalisant la matrice de l'Hamiltonien $\hat{H}^0 = -\frac{1}{2}\nabla^2 + V_{ext}(\vec{r}) + \int \frac{\rho_0'}{|\vec{r}-\vec{r}'|} d\vec{r}' + V_{xc}[\rho_0]$ de sorte que $E_{bnd} = \sum_i^N n_i \epsilon_i$.

L'énergie de répulsion E_{rep} n'est pas calculée à partir de valeurs connues de ρ_0^α mais ajustée entre la différence de l'énergie DFT et l'énergie E_{bnd} . Elle est ajustée à l'aide d'une série de splines.

1.6.1 Résolution de l'hamiltonien $\hat{H}^0$

Si on néglige les termes de fluctuations des charges atomiques alors $E_{DFTB} = \sum_i^N n_i \langle \phi_i | \hat{H}^0 | \phi_i \rangle + E_{rep}$.

Les orbitales de Kohn-Sham sont une combinaison linéaire d'orbitales atomiques centrées sur l'atome : $\phi_i(\vec{r}) = \sum_\mu C_{i\mu} \chi_\mu(\vec{r} - \vec{R}_\alpha)$.

On obtient alors le problème suivant à résoudre :

$$\sum_\mu^N C_{i\mu} (H_{\mu\nu}^0 - \epsilon_i S_{\mu\nu}) = 0 \quad (1.149)$$

avec

$$\begin{aligned} H_{\mu\nu}^0 &= \langle \chi_\mu | \hat{H}^0 | \chi_\nu \rangle \\ S_{\mu\nu} &= \langle \chi_\mu | \chi_\nu \rangle \end{aligned} \quad (1.150)$$

Comme E_{rep} est ajusté à des résultats de la DFT, il ne reste plus que le premier terme à calculer.

$$\sum_i^N n_i \langle \phi_i | \hat{H}^0 | \phi_i \rangle = \sum_i^N n_i \sum_{\mu, \nu}^N C_{i\mu} C_{i\nu} \langle \chi_\mu | \hat{T} + V_{KS}[\rho_0] | \chi_\nu \rangle = tr(P \cdot \hat{H}^0) \quad (1.151)$$

$$\text{où } P_{\mu\nu} = \sum_i^N n_i C_{i\mu} C_{i\nu} \text{ est la matrice densité}$$

A partir de là, on peut utiliser deux approches différentes mais dont les résultats sont similaires. La première (utilisée à partir des années 80) c'est d'écrire le potentiel de Kohn-Sham V_{KS} comme la somme des potentiels atomiques : $V_{KS} = \sum_\alpha^M V_\alpha(\vec{r}_\alpha)$. On obtient alors

$$H_{\mu\nu}^0 = \langle \chi_\mu | -\frac{1}{2}\nabla^2 + V_\alpha + (1 - \delta_{\alpha\beta})V_\beta | \chi_\nu \rangle, \mu \in \{\alpha\}, \nu \in \{\beta\} \quad (1.152)$$

La deuxième approche [24] consiste à considérer que le potentiel ne peut pas s'écrire de cette manière à cause du terme d'échange-corrélation qui n'est pas linéaire et d'utiliser la superposition des densités.

$$H_{\mu\nu}^0 = \langle \chi_\mu | -\frac{1}{2}\nabla^2 + V_{KS}[\rho_0^\alpha + \rho_0^\beta] | \chi_\nu \rangle, \mu \in \{\alpha\}, \nu \in \{\beta\} \quad (1.153)$$

Les orbitales χ_μ et les pseudo-densités atomiques de référence sont obtenues en résolvant l'équation de Schrödinger

$$[\hat{T} + V_{KS}[\rho_0^\alpha] + (\frac{r}{r_0})^2]\chi_\mu(\vec{r}) = \epsilon_\mu \chi_\mu(\vec{r}) \quad (1.154)$$

Au final, l'élément de matrice $H_{\mu\nu}^0$ est

$$H_{\mu\nu}^0 = \begin{cases} \epsilon_\mu^{\text{atome libre}} & \mu = \nu \\ \langle \chi_\mu | -\frac{1}{2}\nabla^2 + V_{KS}[\rho_0^\alpha + \rho_0^\beta] | \chi_\nu \rangle & \mu \in \{\alpha\}, \nu \in \{\beta\}, \alpha \neq \beta \\ 0 & \text{sinon} \end{cases} \quad (1.155)$$

La DFTB sans prendre en compte la fluctuation de la densité électronique est plutôt adapté à des systèmes très fortement ionisés ou avec des liaisons covalentes homonucléaires.

1.6.2 SCC-DFTB : self-consistent charge correction DFTB

La SCC-DFTB [25] est utilisée pour étudier des systèmes présentant des liaisons sensibles aux charges des atomes de la liaison, des systèmes tels que des molécules hétéronucléaires. Les contributions du deuxième ordre E_{2nd} sont prises en compte pour corriger la densité électronique.

La variation de la densité électronique est décomposée en une somme de contributions proches de contributions atomiques dans une approche de tight-binding : $\delta\rho = \sum_\alpha^M \delta\rho_\alpha$.

Les contributions $\delta\rho_\alpha$ décroissent très rapidement en s'éloignant du centre atomique correspondant. En utilisant l'approximation du monopôle, on peut réécrire cette contribution :

$$\delta\rho_\alpha \approx \Delta q_\alpha F_{00}^\alpha Y_{00} \quad (1.156)$$

Ici, Δq_α est la charge de Mulliken [26], donc la différence entre la population atomique de Mulliken q_α et le nombre d'électrons de valence de l'atome neutre q_α^0 . F_{00}^α est la dépendance radiale de la fluctuation de densité de l'atome α et Y_{00} la fonction angulaire sphérique (on s'est placée dans cette approximation).

En prenant en compte ces termes de fluctuations, on prend en compte le phénomène de transfert de charges.

En utilisant l'approximation du monopôle, l'énergie due aux fluctuations de la densité électronique peut se réécrire :

$$E_{2nd} \approx \frac{1}{2} \sum_{\alpha,\beta}^M \Delta q_\alpha \Delta q_\beta \underbrace{\iint \left(\frac{1}{|\vec{r} - \vec{r}'|} + \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'} |_{\rho_0} \right) F_{00}^\alpha F_{00}^\beta Y_{00}^2 d\vec{r} d\vec{r}'}_{\gamma_{\alpha\beta}} \quad (1.157)$$

L'élément $\gamma_{\alpha\beta}$ est relié à la dureté chimique η_α par la relation : $\gamma_{\alpha\alpha} = 2\eta_\alpha = U_\alpha$. Les éléments diagonaux sont en fait paramétrés, ce sont les paramètres de Hubbard U_α .

Finalement l'énergie s'écrit

$$E_{SCC-DFTB} = \sum_i^N n_i \langle \phi_i | \hat{H}^0 | \phi_i \rangle + \frac{1}{2} \sum_{\alpha,\beta}^M \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta + E_{rep} \quad (1.158)$$

Les fluctuations de charge sont calculées à l'aide de l'analyse de population de Mulliken :

$$q_\alpha = \frac{1}{2} \sum_i^N n_i \sum_{\mu \in \alpha}^N \sum_{\nu}^N (C_{i\mu} C_{i\nu} S_{\mu\nu} + C_{i\nu} C_{i\mu} S_{\nu\mu}) \quad (1.159)$$

Ainsi l'élément de matrice $H_{\mu\nu}$ devient

$$H_{\mu\nu} = \langle \phi_i | \hat{H}^0 | \phi_i \rangle + \frac{1}{2} S_{\mu\nu} \sum_{\zeta}^M (\gamma_{\alpha\zeta} + \gamma_{\beta\zeta}) \Delta q_{\zeta} \quad (1.160)$$

On doit résoudre une équation similaire à l'équation (1.149).

Les charges atomiques dépendent des coefficients C_{ij} . En utilisant les C_{ij} , on construit la matrice $\hat{H}$, la diagonalisation de cette matrice nous donne des nouvelles valeurs des C_{ij} , donc des nouvelles orbitales ϕ_i . Ces nouveaux coefficients conduisent à de nouvelles charges atomiques. Il faut résoudre le problème en utilisant une procédure de SCF.

1.6.3 DFTB3

La méthode DFTB3 [27] correspond à l'ajout des dérivées troisièmes de l'énergie d'échange corrélation dans l'énergie SCC-DFTB. Ce nouveau terme s'écrit :

$$\begin{aligned} E^{3\text{ème}} &= \frac{1}{6} \iiint \frac{\delta^3 E_{xc}}{\delta \rho \delta \rho' \delta \rho''} \Big|_{\rho_0} \delta \rho \delta \rho' \delta \rho'' d\vec{r} d\vec{r}' d\vec{r}'' \\ E^{3\text{ème}} &= \frac{1}{6} \iiint \frac{\delta}{\delta \rho} \left[\frac{\delta^2 E_{xc}}{\delta \rho \delta \rho'} \right]_{\rho_0} \delta \rho \delta \rho' \delta \rho'' d\vec{r} d\vec{r}' d\vec{r}'' \end{aligned} \quad (1.161)$$

Or d'après 1.157,

$$\gamma_{\alpha\beta} = \iint \left(\frac{1}{|\vec{r} - \vec{r}'|} + \frac{\delta^2 E_{xc}}{\delta \rho \delta \rho'} \Big|_{\rho_0} \right) F_{00}^\alpha F_{00}^\beta Y_{00}^2 d\vec{r} d\vec{r}' \quad (1.162)$$

$$\begin{aligned} E^{3\text{ème}} &= \frac{1}{6} \sum_{\alpha\beta\nu} \Delta q_\alpha \Delta q_\beta \Delta q_\nu \int \frac{\delta}{\delta \rho} \left[\frac{\delta^2 E_{xc}}{\delta \rho \delta \rho'} \right]_{\rho_0} F_{00}^\alpha F_{00}^\beta F_{00}^\nu Y_{00}^3 d\vec{r} d\vec{r}' d\vec{r}'' \\ E^{3\text{ème}} &= \frac{1}{6} \sum_{\alpha\beta\nu} \Delta q_\alpha \Delta q_\beta \Delta q_\nu \int \frac{\delta}{\delta \rho} (\gamma_{\alpha\beta} - \iint \frac{1}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}') F_{00}^\nu Y_{00} d\vec{r}'' \\ E^{3\text{ème}} &= \frac{1}{6} \sum_{\alpha\beta\nu} \Delta q_\alpha \Delta q_\beta \Delta q_\nu \int \frac{\delta \gamma_{\alpha\beta}}{\delta \rho} F_{00}^\nu Y_{00} d\vec{r}'' \end{aligned} \quad (1.163)$$

La dérivée de la fonction d'Hubbard par rapport à la densité ρ est approximée par une dérivée par rapport à la charge q_ν . En intégrant $F_{00}^\nu Y_{00}$ sur tout l'espace, on obtient l'unité. On a alors :

$$E^{3\text{ème}} = \frac{1}{6} \sum_{\alpha\beta\nu} \Delta q_\alpha \Delta q_\beta \Delta q_\nu \frac{d\gamma_{\alpha\beta}}{dq_\nu} \quad (1.164)$$

De plus, les termes avec $\alpha \neq \beta \neq \nu$, c'est-à-dire les termes à trois centres sont négligés.

$$E^{3\text{ème}} = \frac{1}{6} \sum_{\alpha} \Delta q_\alpha^3 \frac{d\gamma_{\alpha\alpha}}{dq_\alpha} + \frac{1}{6} \sum_{\alpha \neq \beta} \Delta q_\alpha \Delta q_\beta \left[\Delta q_\alpha \frac{d\gamma_{\alpha\beta}}{dq_\alpha} + \Delta q_\beta \frac{d\gamma_{\alpha\beta}}{dq_\beta} \right] \quad (1.165)$$

La résolution se fera par une méthode auto cohérente.

1.7 Les méthodes semi-empiriques NDDO

Toute une gamme de méthodes semi-empiriques a été développée pour contourner les temps de calcul beaucoup trop importants des méthodes *ab initio*. Le grand nombre d'intégrales à calculer rend ses méthodes très coûteuses en terme de temps de calcul. L'idée des méthodes semi-empiriques est de négliger certaines intégrales et de paramétrer les restantes sur des calculs *ab initio* ou des données expérimentales pour avoir des résultats précis. A l'instar de la DFTB, nous avons évalué plusieurs potentiels NDDO sur les énergies et les fréquences harmoniques de la DFT. De même, les résultats n'étaient pas satisfaisants. L'optimisation des paramètres NDDO n'a pas permis d'améliorer suffisamment la qualité des résultats. Dans le cadre des méthodes hybrides pour le calcul du spectre anharmonique, nous avons utilisé des potentiels NDDO. Les résultats de ces méthodes hybrides seront discutés dans le chapitre 3. Nous avons testé la méthode SCP-NDDO [28] et nous avons dû implémenter dans CP2K la possibilité d'optimiser les paramètres de cette méthode. Elle a été développée pour étudier des agrégats de molécules d'eau qui présentent de nombreuses liaisons hydrogène. Pour les plus petits agrégats, les résultats obtenus par les auteurs de cette méthode sont encourageants. Ainsi, cette méthode permet de décrire correctement ces liaisons hydrogène. Or, nos objets d'étude qui sont les sucres contiennent eux aussi beaucoup de liaisons hydrogène. C'est pour cette raison que nous avons décidé de la tester. Cependant, les paramètres existent uniquement pour les atomes d'hydrogène et d'oxygène. Nous n'avons pas réussi à obtenir des jeux de paramètres qui donnaient de bons résultats. De plus, d'après nos nombreux tests, nous avons remarqué que la méthode est instable. En effet, les résultats dépendaient de l'algorithme d'optimisation de la géométrie.

1.7.1 Formalisme

Les méthodes de type NDDO (Neglected of Diatomic Differential Overlap) sont basées sur le formalisme de Hartree-Fock. Les orbitales électroniques sont développées sur une base minimale d'orbitales atomiques telles que

$$\psi_i = \sum_{\mu} c_{\mu i} \phi_{\mu} \quad (1.166)$$

$$0 = \sum_{\nu} [F_{\mu\nu} - S_{\mu\nu} \epsilon_i] c_{\nu i} = \sum_{\nu} [F_{\mu\nu} - \delta_{\mu\nu} \epsilon_i] c_{\nu i} \quad (1.167)$$

$$S_{\mu\nu} = \delta_{\mu\nu} \quad (1.168)$$

Les éléments de matrice, pour des atomes identiques, de l'opérateur Fock sont

$$\begin{aligned} F_{\mu^A \nu^A} = & H_{\mu^A \nu^A}^{core} + \sum_{\lambda^A} \sum_{\sigma^A} P_{\lambda^A \sigma^A} [(\mu^A \nu^A, \lambda^A \sigma^A) - \frac{1}{2}(\mu^A \lambda^A, \nu^A \sigma^A)] \\ & + \sum_B \sum_{\lambda^B} \sum_{\sigma^B} P_{\lambda^B \sigma^B} (\mu^B \nu^B, \lambda^B \sigma^B) \end{aligned} \quad (1.169)$$

Les éléments de matrice, pour des atomes différents, de l'opérateur de Hartree-Fock sont

$$F_{\mu^A \nu^B} = H_{\mu^A \nu^B}^{core} - \frac{1}{2} \sum_{\lambda^A} \sum_{\sigma^B} P_{\lambda^A \sigma^B} (\mu^A \nu^A, \lambda^B \sigma^B) \quad (1.170)$$

$P_{\mu\nu} = \sum_i n_i c_{i\mu} c_{i\nu}$ est la matrice densité.

Dans ces équations ne figurent pas les intégrales à plus de trois centres (trois atomes) parce qu'elle ont été négligées.

Les indices $\mu, \nu, \lambda, \sigma$ désignent des orbitales et A, B des atomes.

Les termes $(\mu\nu, \lambda\sigma)$ sont des intégrales à deux électrons pouvant être à un ou deux centres.

On a déjà vu dans la méthode Hartree-Fock que l'élément de matrice de Fock peut s'écrire comme :

$$F_{\mu\nu} = H_{\mu\nu}^{core} + G_{\mu\nu} \quad (1.171)$$

Où $H_{\mu\nu}^{core}$ correspond à l'Hamiltonien du coeur (électrons des couches internes) et $G_{\mu\nu}$ à l'Hamiltonien des électrons de valence qui comprend les termes coulombien et d'échange.

$$H_{\mu^A\nu^A}^{core} = U_{\mu^A\nu^A} + \sum_B V_{\mu^A\nu^A}^B \quad (1.172)$$

Le terme $H_{\mu^A\nu^A}^{core}$ comprend des intégrales monoélectroniques mettant en jeu les électrons de l'atome A.

$$\sum_B V_{\mu^A\nu^A}^B = \sum_B \langle \phi_\mu^A | \frac{Z_B}{r_{AB}} | \phi_\nu^A \rangle \approx \sum_B -Z_B (\mu^A\nu^A, s^B s^B) \quad (1.173)$$

On considère que l'interaction des électrons de l'atome A avec les atomes B équivaut à l'interaction des électrons de l'atome A avec des électrons du coeur pris dans la couche s des atomes B.

L'énergie totale de la molécule est la somme de l'énergie électronique et de l'énergie de répulsion atomique de sorte que

$$E_{tot} = E_{elec} + \sum_{A \neq B} E_{coeur}^{AB} \quad (1.174)$$

L'énergie électronique est calculée lors de la résolution par le champ auto-cohérent mais l'énergie de répulsion atomique est évaluée par une relation empirique

$$E_{coeur}^{AB} = Z_A Z_B (s^A s^A, s^B s^B) [1 + f(R_{AB})] \text{ avec } f(R_{AB}) = \exp(-\alpha_A R_{AB}) + \exp(-\alpha_B R_{AB}) \quad (1.175)$$

Si les atomes considérés sont O et H ou N et H alors l'expression est légèrement modifiée.

$$f(R_{XH}) = R_{XH} \exp(-\alpha_X R_{XH}) + \exp(-\alpha_H R_{XH}) \text{ avec } X = O \text{ ou } N \quad (1.176)$$

Plusieurs méthodes reposant sur ce formalisme ont été développées. Les différences principales entre elles peuvent venir de la forme de la répulsion interatomique, de la façon dont ont été optimisés les paramètres de la méthode, des données de référence. Chaque méthode tente de corriger les erreurs des précédentes par exemple sur la prédiction de l'enthalpie de formation ΔH_f , la géométrie de certains systèmes, la modélisation de solides.

AM1

La méthode AM1 [29] (Austin Model 1) a été implémentée pour corriger l'erreur trop importante faite sur la partie répulsion entre les atomes dans la méthode MNDO[30].

Pour corriger la surestimation de la répulsion atomique, des termes gaussiens ont été ajoutés dans la fonction de répulsion de répulsion atomique, la fonction $f(R_{AB})$ de l'équation (1.175).

$$f(R_{AB}) = Z_A Z_B (s^A s^A, s^B s^B) [1 + F(A) + F(B)]$$

$$\text{avec } F(A) = \exp(-\alpha_A r_{AB}) + \sum_i K_{Ai} \exp[L_{Ai}(r_{AB} - M_{Ai})^2] \quad (1.177)$$

i pouvant aller de 2 à 4 selon les atomes considérés.

PM3

La méthode PM3 [31] (Parametric Model number 3) intègre seulement deux gaussiennes pour corriger la fonction de répulsion atomique. De plus, les paramètres de cette méthode ont été obtenus d'une manière différente à celle de AM1. Par exemple, AM1 utilise des données spectroscopiques quand PM3 utilise des paramètres optimisés pour reproduire des résultats expérimentaux.

PM6

La méthode PM6 [32] (Parametric Model number 6) a été mise au point après qu'on se soit aperçu que certaines données expérimentales sur lesquelles reposaient les paramétrisations des méthodes antérieures étaient erronées. Des données de référence ont été obtenues par des calculs ab initio. PM6 présente plusieurs fonctions de répulsion atomique selon la paire atomique considérée, C-C, N-O et H-O, Si-O. Par exemple :

$$f(R_{AB}) = Z_A Z_B (s^A s^A, s^B s^B) [1 + x_{AB} \exp(-\alpha_A (r_{AB} + 0.0003 r_{AB}^6)) + 9.28 \exp(-5.98 r_{AB})] \quad (1.178)$$

pour le C - C.

$$f(R_{AB}) = Z_A Z_B (s^A s^A, s^B s^B) [1 + x_{AB} \exp(-\alpha_A (r_{AB}^2))] \quad (1.179)$$

pour le N - H et O - H.

PM7

La méthode PM7 [33] (Parametric Model number 7) est apparue pour corriger PM6 qui ne parvenait pas à décrire correctement les solides. En effet les erreurs s'ajoutaient et devenaient donc trop importantes voire infinies. L'approximation sur les intégrales biélectroniques a été changée pour que les intégrales puissent converger vers la valeur exacte lorsque la distance augmente. Des termes de dispersion et de correction pour la liaison d'hydrogène ont été implémentés dans cette méthode. Les paramètres de la méthode ont été optimisés pour reproduire des quantités physiques autres que celles de PM6.

1.8 Correction empirique de la dispersion

La DFT, la DFTB et les méthodes semi-empiriques ne traitent pas ou très mal les interactions de dispersion, c'est-à-dire les interactions électrostatiques (en R^{-6} , avec R la

distance entre les deux charges. Il existe une méthode robuste pour ajouter ces interactions mais son temps de calcul est important. Des méthodes empiriques ont été conçues pour ajouter de la dispersion et permettre aussi sa prise en compte pour des gros systèmes. Grimme est un concepteur de ce genre de méthodes empiriques. L'une des méthodes les plus utilisées est la méthode D3 [34].

L'idée consiste à ajouter à l'énergie obtenue par un calcul de chimie quantique (*ab initio*, semi-empiriques, empiriques) un terme de dispersion qui est :

$$E_{disp} = -\frac{1}{2} \sum_{n=6,8} s_n \sum_{\substack{atomes \\ A \neq B}} \frac{C_n^{AB}}{R_{AB}^n} f_n^{damp}(R_{AB}) \quad (1.180)$$

$$f_n^{damp}(R_{AB}) = \frac{1}{1 + 6 \left(\frac{R_{AB}}{s_{r,n} R_{0,AB}} \right)^{-\alpha_n}}$$

$f_n^{damp}(R_{AB})$ est une fonction permettant d'amortir la dispersion à courtes distances. Les coefficients sont calculés selon certaines formules analytiques et sont dépendant des systèmes via un nombre de coordination. Les paramètres α_6 , α_8 et $s_{r,8}$ sont fixés et valent 14, 16 et 1, respectivement. Les paramètres s_6 , s_8 et $s_{r,6}$ sont des paramètres ajustables.

1.9 La mécanique moléculaire

Il existe des modèles totalement empiriques pour étudier les molécules. On appelle ces modèles des champs de force. Les paramètres dans ces modèles sont optimisés pour reproduire les résultats de calculs *ab initio* ou bien directement pour reproduire des données expérimentales. Ces modèles se basent sur des types d'atomes, par exemple un carbone sp^3 n'aura pas les mêmes paramètres qu'un carbone sp^2 .

Dans ces modèles les électrons ne sont pris en compte que par le biais des charges atomiques. La charge atomique d'un atome sera son nombre atomique plus la portion de charge électronique de l'atome. La charge atomique d'un atome dans une molécule n'est pas forcément zéro car l'électronégativité n'est pas la même pour tous les atomes.

Au début de notre projet, il existait des champs de force uniquement pour des molécules neutres. Nous avons alors optimisé les paramètres d'un champ de force de type AMBER [35] pour nos systèmes qui sont des ions. Afin d'aller au-delà de l'approximation des charges atomiques fixes, nous avons employé le modèle ACKS2 [36] (qui repose sur le formalisme de la DFT) pour calculer les charges atomiques en fonction de la configuration spatiale atomique. Cependant, aucun jeu de paramètres ne permettait d'obtenir des géométries, des énergies et des fréquences précises simultanément.

L'énergie calculée par la mécanique moléculaire peut s'écrire sous la forme générale

$$E_{MM} = E_{liaison} + E_{angle} + E_{dihédre} + E_{électrostatique} + E_{Van\ der\ Waals} \quad (1.181)$$

$$\begin{aligned}
E_{\text{liaison}} &= \frac{1}{2} \sum_l k_l \cdot (d - d_{0,l})^2 \\
E_{\text{angle}} &= \frac{1}{2} \sum_a k_a \cdot (\theta - \theta_{0,a})^2 \\
E_{\text{dihédre}} &= \sum V_n (1 \pm \cos(n(\Phi \pm \Phi_0))) \\
E_{\text{électrostatique}} &= \sum_{AB} \frac{q_A q_B}{4\pi\epsilon_0 r_{AB}^2} \\
E_{\text{Van der Waals}} &= \sum_{AB} \left[\left(\frac{c_{AB}}{r_{AB}} \right)^{12} - \left(\frac{d_{AB}}{r_{AB}} \right)^6 \right]
\end{aligned} \tag{1.182}$$

Les termes k_l , k_a , V_n , c_{AB} et d_{AB} dépendent du type des atomes.

E_{liaison} :

C'est l'énergie associée à l'élongation de la liaison covalente. Tout écart par rapport à une longueur de référence donnera une énergie non nulle. Généralement, la liaison est considérée comme un oscillateur harmonique. Dans plusieurs champs de force, le caractère anharmonique d'une liaison diatomique est pris en compte en ajoutant des termes d'ordre supérieur de la distance séparant les deux atomes. Le potentiel de Morse, $E_{\text{Morse}} = D(1 - e^{-\alpha(r-r_0)})$, plus proche du potentiel réel peut aussi être utilisé. Dans les champs de force AMOEBA [37] et MMFF94 [38] par exemple, des termes jusqu'à l'ordre 4 sont inclus.

E_{angle} :

Cette énergie fait intervenir trois atomes liés deux à deux entre eux. Elle correspond à la déformation de l'angle entre ces deux liaisons. Tout écart par rapport à un angle de référence donnera une énergie non nulle. De même que pour l'énergie de liaison vue précédemment, l'expression de base de cette énergie est prise comme un oscillateur harmonique. Des termes d'ordres supérieurs peuvent être ajoutés pour tenir compte du caractère anharmonique de la déformation.

On peut aussi ajouter des termes de couplage entre l'élongation des liaisons et la déformation angulaire de ces liaisons. En effet, lors d'une déformation angulaire la longueur des liaisons peut elle aussi être modifiée et réciproquement. Par exemple au premier ordre l'énergie sera :

$$E_{\text{liaison-angle}} = k_{d\theta}[(d - d_0) + (d' - d'_0)](\theta - \theta_0) \tag{1.183}$$

$E_{\text{dihédre}}$:

Cette énergie est associée à la torsion de liaisons chimiques. L'énergie est calculée à partir de l'angle diédral correspondant aux quatre atomes, 3 liaisons et 2 angles impliqués. Elle s'écrit comme la somme de cosinus. Cette écriture permet d'obtenir 1 ou plusieurs minimums comme on peut le voir sur l'exemple de la figure 1.6.

$E_{\text{électrostatique}}$:

C'est l'énergie électrostatique entre deux atomes qui sont séparés de plus de 3 liaisons. Elle est calculée à partir de la charge atomique et non pas du nombre atomique des atomes. Cette énergie est prise en compte différemment si les deux atomes sont séparés

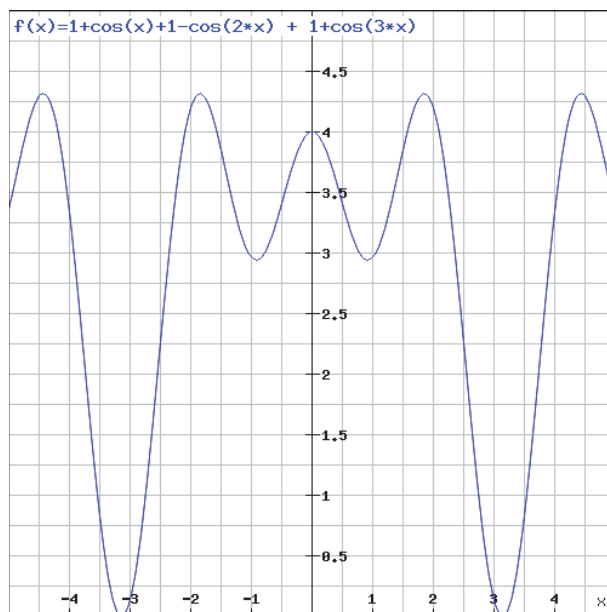


FIGURE 1.6 – Illustration d'un potentiel de torsion

de 3 liaisons ou plus : si les deux atomes sont séparés par exactement 3 liaisons alors on prend en compte seulement une fraction du terme électrostatique. On considère en effet que l'autre partie est incluse dans les termes évoqués précédemment. Si les deux atomes sont séparés de plus de 3 liaisons alors on prend en compte entièrement le terme électrostatique. Certains champs de force ont des coefficients différents dans le cas où les atomes sont séparés de 3, 4, 5 liaisons.

$E_{\text{Van der Waals}}$:

A l'instar de l'énergie électrostatique, cette énergie est calculée pour des atomes séparés de plus de 3 liaisons. Une fraction de l'énergie sera prise en compte selon le nombre de liaisons séparant les deux atomes. Elle est constituée d'une partie répulsive qui simule la difficulté qu'ont deux nuages électroniques à se rapprocher (terme en r^{-12}) et d'une partie attractive qui correspond à la dispersion (terme en r^{-6}). On dit que ce potentiel est un potentiel 6-12 faisant référence aux exposants des termes attractif et répulsif. Il existe d'autres potentiels pour modéliser le terme de Van der Waals : un potentiel 7-14, un potentiel exp(6) (le terme répulsif est une exponentielle et le terme attractif en r^{-6}), etc.

Il existe d'autres types de contributions comme la contribution dite hors du plan (*out-of-plane* en anglais). On l'ajoute lorsque l'on veut pénaliser un atome qui sort du plan formé par trois autres atomes. L'énergie correspondante est $E_{oop} = k_{oop}\chi^2$. χ est l'angle entre une liaison d'un atome du plan avec l'atome hors du plan et le plan lui-même comme montré sur la figure 1.6.

On va s'intéresser au champ de force nommé MMFF94 ou Merck Molecular Force Field 1994 parce qu'on montrera dans un prochain chapitre que l'on peut calculer des spectres anharmoniques avec ce potentiel.

L'énergie MMFF94 est :

$$E_{MMFF94} = E_{\text{liaison}} + E_{\text{angle}} + E_{\text{liaison-angle}} + E_{\text{dihédre}} + E_{\text{hors du plan}} + E_{\text{électrostatique}} + E_{\text{Van der Waals}} \quad (1.184)$$

L'expression des différents termes est donnée en-dessous.

$$\begin{aligned}
E_{\text{liaison}} &= \sum_{ij} [143,9325 \frac{K_{l,ij}}{2} (r_{ij} - r_{0,ij})^2 (1 + cs(r_{ij} - r_{0,ij}) + \frac{7}{12} cs^2 (r_{ij} - r_{0,ij})^2)] \\
E_{\text{angle}} &= \sum_{ijk} [0,043844 \frac{K_{a,ijk}}{2} (\theta_{ijk} - \theta_{0,ijk})^2 (1 + cb(\theta_{ijk} - \theta_{0,ijk}))] \\
E_{\text{liaison-angle}} &= \sum_{ijk} 2,51210 [K_{la,ijk} (r_{ij} - r_{0,ij}) + K_{la,kji} (r_{kj} - r_{0,kj})] (\theta_{ijk} - \theta_{0,ijk}) \\
E_{\text{dihédre}} &= \sum_{ijkl} 0,5 [V_1 (1 + \cos(\Phi)) + V_2 (1 - \cos(2\Phi)) + V_3 (1 + \cos(3\Phi))] \text{ avec } V_u = V_u(i, j, k, l) \\
E_{\text{hors du plan}} &= \sum_{ijk;l} 0,043844 \frac{K_{oop,ijk;l}}{2} \chi_{ijk;l}^2 \\
E_{\text{électrostatique}} &= \sum_{ij} 332,0716 \frac{q_i q_j}{D(r_{ij} + \delta)^n} \text{ avec } q_i = q_{0,i} + \sum_k \omega_{ki} \\
E_{\text{Van der Waals}} &= \sum_{ij} \epsilon_{ij} \left(\frac{1,07 r_{ij}^*}{r_{ij} + 0,07 r_{ij}^*} \right)^7 \left(\frac{1,12 r_{ij}^{*7}}{r_{ij}^7 + 0,12 r_{ij}^{*7}} - 2 \right)
\end{aligned} \tag{1.185}$$

$K_{l,ij}$ est la constante de force pour l'énergie de liaison. Elle s'exprime en md/Å. cs est la constante d'élongation cubique qui vaut -2Å^{-1} . $K_{a,ijk}$ est la constante de force pour l'énergie de déformation. Elle s'exprime en md Å/rad². cb est la constante de déformation cubique qui vaut $-0,4 \text{ rad}^{-1}$. $K_{la,ijk}$ et $K_{la,kji}$ sont des constantes de force pour le couplage entre la déformation et l'élongation qui s'expriment en md/rad. $K_{oop,ijk;l}$ est la constante de force pour l'énergie hors du plan (en md Å/rad²). V_1 , V_2 et V_3 sont des constantes. ϵ_{ij} et r_{ij}^* dépendent des polarisabilités selon les formules données dans la référence [38].

Dans le champ de force MMFF94, les énergies de Van der Waals et électrostatique sont calculées lorsque les atomes n'appartiennent pas au même fragment ou sont séparés par 3 liaisons ou plus. Les interactions mettant en jeu des atomes séparés par 3 liaisons sont prises en compte entièrement.

1.10 Références

- [1] Max Born and Robert Oppenheimer. On the Quantum Theory of Molecules (translation by S. M. Blinder). *Ann. Phys. (Berlin)*, 84, 1927.
- [2] J. C. Slater. Atomic Shielding Constants. *Phys. Rev.*, 36 :57–64, 1930.
- [3] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.*, 56(5) :2257–2261, 1972.
- [4] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople. Self consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.*, 72(1) :650–654, 1980.

-
- [5] Michael J. Frisch, John A. Pople, and J. Stephen Binkley. Self Consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.*, 80(7) :3265–3269, 1984.
- [6] Thom H. Dunning Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90(2) :1007–1023, 1989.
- [7] Ansgar Schäfer, Hans Horn, and Reinhart Ahlrichs. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.*, 97(4) :2571–2577, 1992.
- [8] Florian Weigend and Reinhart Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn : Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7 :3297–3305, 2005.
- [9] J. A. Pople and R. K. Nesbet. Self-Consistent Orbitals for Radicals. *J. Chem. Phys.*, 22(3) :571–572, 1954.
- [10] Chr. Møller and M. S. Plesset. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.*, 46 :618–622, 1934.
- [11] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.*, 136 :B864–B871, 1964.
- [12] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, 140 :A1133–A1138, 1965.
- [13] R.G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. International Series of Monographs on Chemistry. 1989.
- [14] G. L. Oliver and J. P. Perdew. Spin-density gradient expansion for the kinetic energy. *Phys. Rev. A*, 20 :397–403, 1979.
- [15] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations : a critical analysis. *Can. J. Phys.*, 58(8) :1200–1211, 1980.
- [16] John P. Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45 :13244–13249, 1992.
- [17] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38 :3098–3100, 1988.
- [18] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37 :785–789, 1988.
- [19] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77 :3865–3868, 1996.
- [20] AD Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98 :5648–5652, 1993.
- [21] Takeshi Yanai Hisayoshi Iikura, Takao Tsuneda and Kimihiko Hirao. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *J. Chem. Phys.*, 115(8) :3540 – 3544, 2001.

-
- [22] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange-correlation functional using the coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.*, 393(1-3) :51 – 57, 2004.
- [23] J. C. Slater and G. F. Koster. Simplified LCAO method for the periodic potential problem. *Phys. Rev.*, 94 :1498–1524, 1954.
- [24] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58 :7260–7268, p 1998.
- [25] M. Elstner. The scc-dftb method and its application to biological systems. *Theoretical Chemistry Accounts*, 116(1) :316–325, Aug 2006.
- [26] R. S. Mulliken. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.*, 23(10) :1833–1840, 1955.
- [27] Michael Gaus, Qiang Cui, and Marcus Elstner. DFTB3 : Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.*, 7(4) :931–948, 2011.
- [28] Daniel T. Chang, Gregory K. Schenter, and Bruce C. Garrett. Self-consistent polarization neglect of diatomic differential overlap : Application to water clusters. *The Journal of Chemical Physics*, 128(16) :164111, 2008.
- [29] Michael J. S. Dewar, Eve G. Zoebisch, Eamonn F. Healy, and James J. P. Stewart. Development and use of quantum mechanical molecular models. 76. AM1 : a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, 107(13) :3902–3909, 1985.
- [30] Michael J. S. Dewar and Walter Thiel. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.*, 99(15) :4899–4907, 1977.
- [31] James J. P. Stewart. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.*, 10(2) :209–220, 1989.
- [32] James J. P. Stewart. Optimization of parameters for semiempirical methods V : Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.*, 13(12) :1173–1213, 2007.
- [33] James J. P. Stewart. Optimization of parameters for semiempirical methods VI : more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.*, 19(1) :1–32, 2013.
- [34] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132(15) :154104, 2010.
- [35] Scott J. Weiner, Peter A. Kollman, Dzung T. Nguyen, and David A. Case. An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry*, 7(2) :230–252, 1986.

-
- [36] T. Verstraelen, P. W. Ayers, V. Van Speybroeck, and M. Waroquier. Acks2 : Atom-condensed kohn-sham dft approximated to second order. *The Journal of Chemical Physics*, 138(7) :074108, 2013.
- [37] Yue Shi, Zhen Xia, Jiajing Zhang, Robert Best, Chuanjie Wu, Jay W. Ponder, and Pengyu Ren. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.*, 9(9) :4046–4063, 2013.
- [38] Thomas A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.*, 17(5-6) :490–519, 1996.

Dynamique Moléculaire

2.1 Introduction

Dans le chapitre précédent nous avons discuté de plusieurs méthodes pour résoudre l'équation de Schrödinger pour la structure électronique. Pour connaître l'énergie totale du système, il faut résoudre l'équation de Schrödinger pour les noyaux. La résolution nous donne accès au spectre vibrationnel (et/ou rotationnel et/ou translationnel, selon les termes présents dans l'hamiltonien des noyaux) du système. Les spectres vibrationnels obtenus expérimentalement sont généralement mesurés sur des systèmes correspondant au minimum global de la surface d'énergie potentielle. Pour pouvoir comparer les spectres théoriques et expérimentaux il faut que les structures théoriques soient optimisées. On part d'une structure, c'est-à-dire d'une configuration spatiale des noyaux et on déplace ces derniers jusqu'à ce que les forces exercées sur les noyaux et que la variation d'énergie (ΔE_e) soient plus faible qu'un certain seuil. Cette étape d'optimisation permet d'obtenir un extremum de la surface d'énergie potentielle. La résolution de l'équation de Schrödinger pour les noyaux s'effectue alors sur cet extremum. S'il y a au moins une fréquence imaginaire (la dérivée seconde de l'énergie est négative donc la fréquence qui est la racine carrée est donc imaginaire) alors on ne se trouve pas sur un minimum. En revanche, si toutes les fréquences sont réelles alors la structure correspond bien à la structure d'un minimum (local, pas forcément global). La résolution de l'équation de Schrödinger pour les noyaux sera discutée dans le chapitre suivant. Pour certaines molécules des arguments physico-chimiques sont suffisants pour connaître la structure correspondant au minimum global mais en général il est très difficile de le connaître voire impossible. La stratégie pour obtenir le minimum global consiste à obtenir plusieurs structures, optimiser leurs géométries et calculer l'énergie (E_e puis l'énergie totale E après résolution du problème nucléaire). Si toute la surface d'énergie potentielle a été explorée alors la géométrie ayant l'énergie totale la plus basse est le minimum global. La dynamique moléculaire est une méthode qui permet de calculer de nombreuses propriétés pour des systèmes notamment le calcul du spectre vibrationnel infrarouge que nous décrirons brièvement dans le chapitre suivant. Nous avons essentiellement utilisé la dynamique moléculaire pour sonder la surface d'énergie potentielle pour obtenir un grand nombre de minima.

Dans un premier temps nous allons décrire le principe de la dynamique moléculaire puis nous verrons comment l'utiliser pour obtenir des minima de la surface.

2.2 Dynamique moléculaire

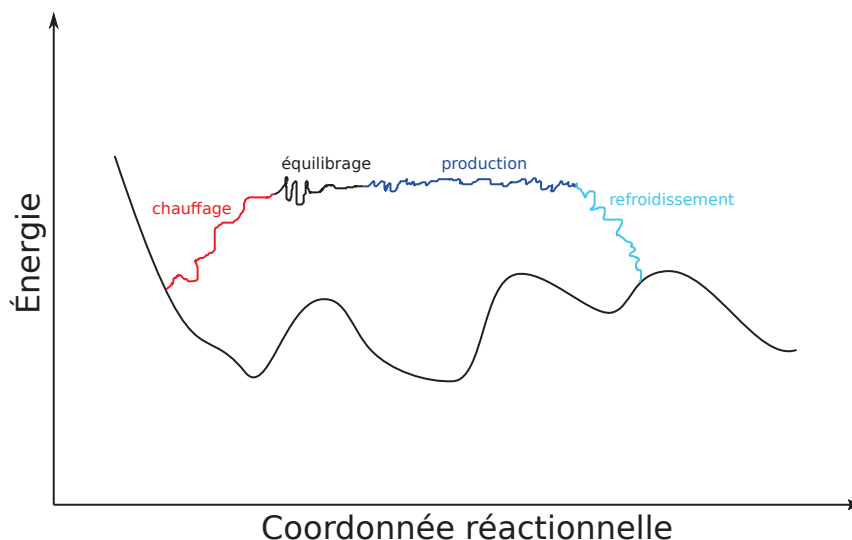


FIGURE 2.1 – Schéma de principe de la dynamique moléculaire

La dynamique moléculaire permet d'étudier l'évolution dans le temps d'un système. Elle repose sur le principe fondamental de la dynamique, $\vec{F} = \sum \vec{F}_i = m\vec{a}$. Les forces exercées sur chaque noyau sont calculées à partir du potentiel utilisé (*ab initio*, semi-empirique ou empirique) selon la formule $\vec{F}_{\text{noyau}_A} = -\nabla_{\vec{q}_A} V(\{\vec{q}\})$. Les équations du mouvement sont :

$$\begin{aligned} \dot{q}_i &= \frac{p_i}{m_i} \\ \dot{p}_i &= -\frac{\partial}{\partial q_i} V(\{\vec{q}\}) \end{aligned} \quad (2.1)$$

La résolution analytique de la dynamique moléculaire n'est possible que pour des systèmes à 2 corps. Pour des systèmes de plus grande taille il faut résoudre numériquement les équations du mouvement. En utilisant les positions et vitesses à un instant donné des noyaux et en calculant les forces qu'ils subissent, on peut à l'aide d'un intégrateur calculer les vitesses et les positions à l'instant d'après. La dynamique moléculaire s'effectue dans différents ensembles thermodynamiques, par exemple dans l'ensemble microcanonique (ou NVE pour N le nombre de particules, V le volume et E l'énergie, ces trois variables étant conservées) ou dans l'ensemble canonique, en présence d'un thermostat (NVT, T est la température).

En pratique la dynamique moléculaire s'effectue en plusieurs étapes. La première étape consiste à chauffer notre système jusqu'à une certaine température. La phase de chauffage est une phase progressive qui permet de donner une certaine vitesse aux noyaux correspondant à la température souhaitée. On peut s'affranchir de cette dernière en faisant appel à une distribution des vitesses correspondant à une certaine température. On peut par exemple utiliser la distribution de Boltzmann pour attribuer des vitesses aux noyaux.

$$p(v_{ix}) = \left(\frac{m_i}{2\pi kT}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{m_i v_{ix}^2}{kT}\right) \quad (2.2)$$

C'est la probabilité pour l'atome i d'avoir la vitesse selon l'axe x , v_{ix} , à la température T . La seconde étape est une étape d'équilibrage. L'étape suivante est l'étape que l'on

appelle l'étape de production. C'est au cours de cette phase que l'on explore la surface d'énergie potentielle et que l'on peut sélectionner des géométries. On peut effectuer une ultime étape de refroidissement si nécessaire.

2.2.1 Thermostat

Afin de contrôler la température du système on utilise un thermostat pendant la dynamique moléculaire. Nous allons brièvement présenter quelques thermostats.

Thermostat d'Andersen

Andersen [1] a montré que l'on pouvait générer d'autres ensembles que des ensembles microcanoniques lors de calculs de dynamique moléculaire. Générer des ensembles canoniques est important pour se rapprocher des conditions expérimentales. A température constante, l'énergie du système fluctue. Pour prendre en compte ce phénomène on ajoute des forces aléatoires agissant sur les noyaux et modifiant leurs vitesses. On ajoute un terme de collision stochastique dans la formule de la dérivée de la quantité de mouvement.

$$\dot{p}_i = F_i + R(t) \quad (2.3)$$

On introduit un paramètre ν qui est le taux moyen de collision pour une particule. C'est la probabilité qu'un noyau subisse une collision entre l'instant t et l'instant $t + \Delta t$, avec Δt un petit intervalle de temps. Pour savoir à quels moments un noyau subit une collision on emploie des nombres aléatoires. Lorsqu'il y a collision la vitesse du noyau est modifiée en suivant une distribution de Boltzmann à la température T .

Il s'avère que ce thermostat peut introduire des discontinuités dans la trajectoire la rendant non physique.

Thermostat de Nosé-Hoover

Dans ce thermostat [2, 3, 4], le système est couplé au bain de chaleur par l'introduction de nouvelles variables (s , p_s et Q). L'hamiltonien modifié est alors :

$$H_{\text{Nosé}} = V(q) + \sum_i \frac{p_i^2}{2m_i} + (g+1)kT \ln s + \frac{p_s^2}{2Q} \quad (2.4)$$

Le paramètre Q agit comme une masse fictive pour le bain. s est le degré de liberté pour le bain et p_s le moment correspondant. $g+1$ est le nombre de degrés de liberté du système. On introduit une nouvelle variable η telle que :

$$\eta = \ln s \quad (2.5)$$

Les équations du mouvement pour le thermostat de Nosé-Hoover sont :

$$\begin{aligned}
 &\text{pour le système :} \\
 &\dot{q}_i = \frac{p_i}{m_i} \\
 &\dot{p}_i = -\frac{\partial}{\partial q_i} V(\{\vec{q}\}) - p_i \frac{p_\eta}{Q} \\
 &\text{et pour le bain :} \\
 &\dot{\eta} = \frac{p_\eta}{Q} \\
 &\dot{p}_\eta = 2 \sum_i \frac{p_i^2}{2m_i} - gkT
 \end{aligned} \tag{2.6}$$

Finalement les vitesses sont multipliées par le facteur suivant :

$$\hat{s} = \exp - \left[\frac{\Delta t}{2} \frac{p_\eta + \frac{\Delta t}{4} (\sum_i p_i^2 / m_i - gkT)}{Q} \right] \tag{2.7}$$

Δt est le pas de temps de la discrétisation de la dynamique moléculaire.

Thermostat de Berendsen

Dans ce thermostat [5] on considère que le bain est couplé au système par un terme de friction.

$$\dot{p}_i = F_i - m_i \gamma_i v_i \tag{2.8}$$

γ_i est le facteur de friction.

A partir de la variation temporelle de l'énergie cinétique E_{cin} :

$$\frac{dE_{cin}}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\sum_i \frac{1}{2} m_i v_i^2(t + \Delta t) - \sum_i \frac{1}{2} m_i v_i^2(t)}{\Delta t} \tag{2.9}$$

On peut montrer que les vitesses sont multipliées par un facteur λ qui vaut :

$$\lambda = \sqrt{1 + 2\gamma\Delta t \left(\frac{T_0}{T} - 1 \right)} \tag{2.10}$$

Le coefficient de friction est pris identique pour tous les degrés de liberté : $\gamma_i = \gamma$.

Thermostat de Bussi

Le thermostat de Bussi [6] est une approche pour modifier les vitesses avec la méthode de réajustement des vitesses. Dans laquelle les vitesses sont multipliées par le coefficient

$$\lambda = \sqrt{\frac{\bar{E}_{cin}}{E_{cin}}} \tag{2.11}$$

De cette manière on force l'énergie cinétique, E_{cin} , à être égale à l'énergie cinétique moyenne pour la température que l'on cherche à atteindre, $\overline{E}_{cin}$. Pour le thermostat de Bussi, on modifie l'expression de ce coefficient afin d'obtenir des ensembles canoniques par un processus stochastique.

Finalement, la nouvelle expression pour le carré du coefficient multiplicateur est :

$$\lambda^2 = e^{-\Delta t/\tau} + \frac{\overline{E}_{cin}}{N_f E_{cin}} (1 - e^{-\Delta t/\tau}) (R_1^2 + \sum_{i=2}^{N_f} R_i^2) + 2e^{-\Delta t/2\tau} \sqrt{\frac{\overline{E}_{cin}}{N_f E_{cin}} (1 - e^{-\Delta t/\tau})} R_1 \quad (2.12)$$

Où τ est un temps caractéristique, R_i est un nombre tiré à partir d'une distribution gaussienne et N_f est le nombre de degrés de liberté.

2.2.2 Intégrateur

Nous avons déjà évoqué le fait que la dynamique moléculaire repose sur le principe fondamental de la dynamique.

$$\sum_i \vec{F}_i = \sum_i m_i \vec{a}_i = \sum_i m_i \ddot{\vec{x}}_i \quad (2.13)$$

La résolution analytique de la dynamique moléculaire n'est pas possible pour des systèmes comportant plus de quelques atomes, la résolution numérique s'impose alors.

Les méthodes numériques utilisées pour intégrer les équations du mouvement sont appelées des intégrateurs. Il en existe plusieurs qui ont chacun leurs avantages et leurs inconvénients. Nous allons présenter quelques-uns de ces intégrateurs.

Intégrateur de Verlet-vitesse

L'un des intégrateurs les plus utilisés est l'intégrateur de Verlet-vitesse [7]. Pour cet intégrateur la position à l'instant $t + \Delta t$ est obtenue à partir de

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \quad (2.14)$$

et la vitesse au même instant :

$$v(t + \Delta t) = v(t) + \frac{1}{2} [a(t) + a(t + \Delta t)] \Delta t + \mathcal{O}(\Delta t^2) \quad (2.15)$$

La formule de Verlet-vitesse pour la vitesse à l'instant $t + \Delta t$ dépend de l'accélération en ce même instant. Cette accélération est calculée à partir du potentiel en prenant comme positions pour les noyaux les positions calculées avec l'algorithme de Verlet-vitesse, c'est-à-dire $r(t + \Delta t)$.

$$a(t + \Delta t) = -\frac{1}{m} \nabla V(r(t + \Delta t)) \quad (2.16)$$

Cet intégrateur est numériquement stable et possède la propriété physique d'être réversible dans le temps.

Intégrateur de Beeman

Un autre intégrateur beaucoup utilisé est l'intégrateur de Beeman [8, 9]. Pour cet intégrateur la position à l'instant $t + \Delta t$ est obtenue à partir de

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{6} [4a(t) - a(t - \Delta t)] \Delta t^2 + \mathcal{O}(\Delta t^4) \quad (2.17)$$

et la vitesse au même instant :

$$v(t + \Delta t) = v(t) + \frac{1}{6} [5a(t) + 2a(t + \Delta t) - a(t - \Delta t)] \Delta t + \mathcal{O}(\Delta t^3) \quad (2.18)$$

Dans cet intégrateur les positions et les vitesses à l'instant $t + \Delta t$ sont calculés à partir de l'instant $t - \Delta t$. Ainsi par rapport à l'intégrateur de Verlet vitesse l'intégrateur de Beeman requiert plus de mémoire puisqu'il faut stocker en plus les positions à l'instant $t - \Delta t$.

2.3 Recherche de conformations

Comme la dynamique moléculaire permet d'explorer la surface d'énergie potentielle, on peut s'en servir pour obtenir des structures. Dans le cadre de notre étude sur les saccharides, nous avons systématiquement fait appel à la dynamique moléculaire pour obtenir un grand nombre de structures. Nous allons maintenant présenter notre démarche. La figure 2.2 schématise la procédure que nous avons suivie. La première étape consiste à faire de la dynamique moléculaire à l'aide du potentiel semi-empirique PM7. L'emploi de méthodes plus poussées est impossible à cause du coût calculatoire prohibitif de telles méthodes. De plus, PM7 a été conçu avec l'objectif d'améliorer le traitement des liaisons hydrogène qui sont très présentes dans nos systèmes d'intérêt, à savoir les saccharides. La méthode PM7 semble meilleure [10] que certains champs de force de mécanique moléculaire pour reproduire les géométries et leurs énergies relatives par rapport à la DFT. Nous effectuons une étape de chauffage pour atteindre une température de 3000 K. On choisit une température très haute pour que le système ne reste pas dans une zone de minima locaux. Cette température permet de franchir toutes les barrières entre minima locaux et d'augmenter les chances pour trouver le minimum global. Une fois le système équilibré, on passe à l'étape de production. Au cours de cette étape on va sélectionner régulièrement des géométries. L'étape de refroidissement n'est pas intéressante pour nous puisque nous voulons seulement obtenir des structures. L'exploration de la surface d'énergie potentielle au cours de cette trajectoire peut ne pas être suffisante. Dans ce cas, il est possible de lancer d'autres trajectoires à partir de structures sélectionnées dans la première trajectoire. Une fois la ou les trajectoires terminées, on optimise toutes les géométries avec le potentiel PM7. La sélection des structures peut être telle qu'après optimisation des géométries on se retrouve avec des géométries similaires, il faut alors éliminer les doublons. Cette élimination repose sur deux critères : un critère énergétique et un critère géométrique. Ici, nous avons choisi un critère énergétique de $0,5 \text{ kcal/mol}$. Si deux géométries ont une différence d'énergie au-dessus de ce critère alors elles ne sont pas identiques. Pour le critère géométrique, on calcule toutes les distances interatomiques (atomes liés par une liaison covalente ou non) et si une distance est supérieure à $0,2 \text{ \AA}$ alors les deux géométries ne sont pas identiques. On pourrait s'arrêter ici et résoudre le problème nucléaire pour les géométries restantes mais la précision de la méthode PM7 n'est pas suffisante pour le calcul des spectres vibrationnels. En revanche la DFT est plus précise. On va alors réoptimiser les géométries

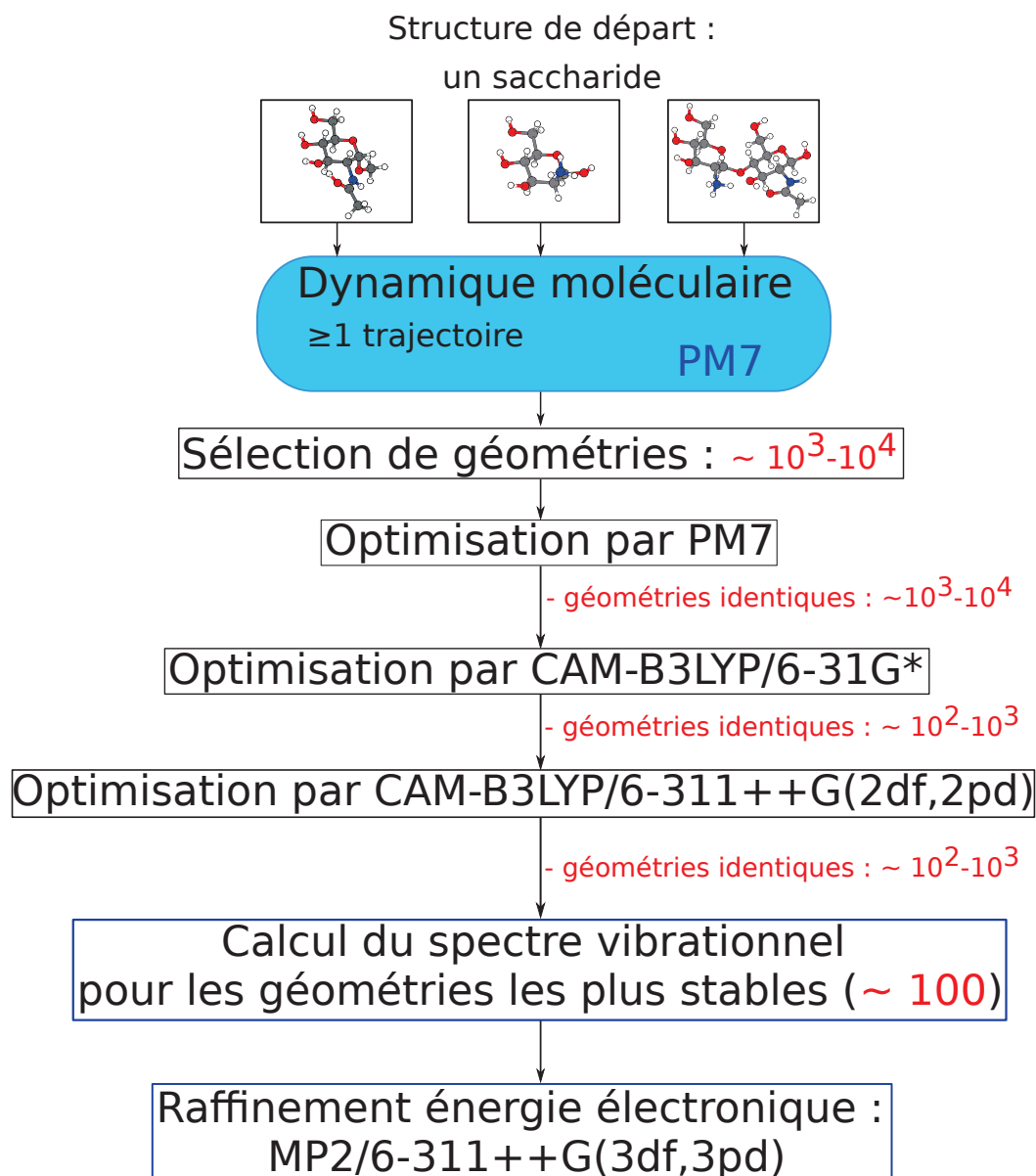


FIGURE 2.2 – Schéma résumant les différentes étapes de notre méthode de recherche de conformations.

obtenues avec une fonctionnelle de la DFT. Nous avons choisi la fonctionnelle CAM-B3LYP. On effectue deux étapes d'optimisation avec cette fonctionnelle. Une première étape qui consiste à optimiser les géométries avec une petite base 6-31G* pour éliminer les géométries identiques que PM7 n'a pu déceler. Et la dernière et ultime étape où les géométries seront optimisées avec la base 6-311++G(2df,2pd). Dans cette dernière étape on vérifie à nouveau s'il y a des géométries similaires. Pour améliorer les énergies on procède à un calcul de l'énergie électronique avec la méthode MP2 et la base 6-311++G(3df,3pd) sur les structures optimisées par CAM-B3LYP/6-311++G(2df,2pd) (les structures ne sont pas réoptimisées en MP2). Le choix de la fonctionnelle CAM-B3LYP est discuté dans la partie III et le chapitre 4.

2.4 Références

- [1] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72(4) :2384–2393, 1980.
- [2] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1) :511–519, 1984.
- [3] Shuichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52(2) :255–268, 1984.
- [4] William G. Hoover. Canonical dynamics : Equilibrium phase-space distributions. *Phys. Rev. A*, 31 :1695–1697, 1985.
- [5] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8) :3684–3690, 1984.
- [6] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1) :014101, 2007.
- [7] Loup Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, 159 :98–103, 1967.
- [8] P. Schofield. Computer simulation studies of the liquid state. *Comput. Phys. Commun.*, 5(1) :17–23, 1973.
- [9] D. Beeman. Some multistep methods for use in molecular dynamics calculations. *J. Comput. Phys.*, 20(2) :130 – 139, 1976.
- [10] Geoffrey R. Hutchison Ilana Y. Kanai, John A. Keith. A Sobering Assessment of Classical Force Field Methods for Low Energy Conformer Predictions. <https://arxiv.org/abs/1705.04308>, 2017.

Calcul du spectre vibrationnel

Dans les chapitres précédents nous avons montré comment résoudre le problème électronique et comment obtenir des minima grâce à la dynamique moléculaire. Pour pouvoir calculer l'énergie totale du système considéré l'équation de Schrödinger nucléaire doit être résolue. La résolution de cette équation permet de calculer le spectre vibrationnel du minima considéré qui pourra être comparé à des spectres expérimentaux. L'énergie potentielle des noyaux n'étant pas connue analytiquement elle doit donc être approximée. L'énergie potentielle est développée en série de Taylor. Nous verrons qu'en première approximation nous pouvons arrêter le développement de la série à l'ordre 2, c'est l'approximation harmonique du potentiel. Cependant cette approximation n'offre pas de résultats comparables aux mesures expérimentales. Pour comparer les spectres infrarouges théoriques aux spectres mesurés expérimentalement, il est impératif de prendre en compte l'anharmonicité du potentiel, c'est-à-dire inclure des termes d'ordre supérieur dans le développement en série de Taylor. Il existe plusieurs méthodes pour tenir compte de cette anharmonicité. Dans ce chapitre nous discuterons de la résolution de l'équation de Schrödinger pour les noyaux dans le cadre de l'approximation harmonique. Puis nous discuterons des différentes méthodes existantes pour inclure l'anharmonicité du potentiel. Nous présenterons tout d'abord la méthode des *scaling factors*. C'est une méthode rapide (temps de calcul nul) et simple d'un point de vue conceptuel mais qui malgré sa simplicité reste une méthode très employée offrant des résultats très satisfaisants. Nous présenterons ensuite deux méthodes qui prennent en compte explicitement les termes anharmoniques dans le potentiel. La méthode GVPT2 traite de manière perturbative le couplage faible entre les modes de vibration et de manière variationnelle les couplages forts. L'autre méthode est la méthode VSCF dans laquelle le couplage entre modes de vibration se fait par un champ moyen. Finalement, nous verrons que la dynamique moléculaire peut être utilisée pour produire des spectres vibrationnels. Pour cette dernière méthode, l'anharmonicité est contenue dans le potentiel qui sert à calculer les forces exercées sur les noyaux. Nous allons développer plus en détail la méthode des *scaling factor* et la méthode GVPT2 car ce sont les méthodes que nous avons le plus utilisées au cours de la thèse. La méthode VSCF et la dynamique moléculaire seront vues de manière plus succinctes.

3.1 Résolution de l'équation de Schrödinger pour les noyaux

3.1.1 Calcul des modes et des fréquences de vibration

Dans le cadre de l'approximation de Born-Oppenheimer, on a considéré les électrons comme allant beaucoup plus vite que les noyaux. Ainsi du point de vue des électrons les noyaux étaient comme fixes. Du point de vue des noyaux, le mouvement des électrons est un mouvement moyenné qui ne dépend que des coordonnées nucléaires. Pour résoudre la partie nucléaire, on réintroduit l'énergie cinétique des noyaux.

L'hamiltonien nucléaire est :

$$\begin{aligned} H_n &= T_n + E_e + V_{nn} \\ H_n &= T_n + V_n(\{\vec{R}\}) \\ V_n(\{\vec{R}\}) &= E_e + V_{nn} \end{aligned} \quad (3.1)$$

Lorsque l'on se place à l'équilibre, c'est-à-dire lorsque les forces exercées sur les atomes sont nulles, on peut écrire le potentiel V_n comme :

$$V_n(\{\vec{R} - \vec{R}_{eq}\}) = V_n(\{\vec{x}\}) = V_n(0) + \sum_i \left(\frac{\partial V_n}{\partial x_i} \right) x_i + \frac{1}{2} \sum_{ij} \left(\frac{\partial^2 V_n}{\partial x_i \partial x_j} \right) x_i x_j + \dots \quad (3.2)$$

Le potentiel $V_n(0)$ est pris égal à zéro et les dérivées premières sont nulles parce qu'on se place à l'équilibre (les forces F sont reliées au gradient du potentiel). Dans un premier temps on peut faire l'approximation harmonique (ou approximation harmonique mécanique) qui consiste à arrêter l'expansion du potentiel à l'ordre 2. Le potentiel V_n s'écrit alors :

$$V_n(\{\vec{x}\}) = \frac{1}{2} \sum_{ij} \left(\frac{\partial^2 V_n}{\partial x_i \partial x_j} \right) x_i x_j = \frac{1}{2} \sum_{ij} k_{ij} x_i x_j \quad (3.3)$$

Les dérivées secondes $\frac{\partial^2 V_n}{\partial x_i \partial x_j}$ où k_{ij} sont appelées constantes de force.

L'énergie cinétique T_n s'écrit :

$$\begin{aligned} T_n &= \frac{1}{2} \sum_i m_i v_i^2 = \frac{1}{2} \sum_i m_i \left(\frac{dx_i}{dt} \right)^2 \\ T_n &= \sum_i \frac{1}{2m_i} p_i^2 \\ T_n &= -\frac{1}{2} \sum_i m_i \frac{\partial^2}{\partial x_i^2} \text{ selon le principe de correspondance} \end{aligned} \quad (3.4)$$

Ainsi l'hamiltonien nucléaire est :

$$H_n = T_n + V_n = -\frac{1}{2} \sum_i m_i \frac{\partial^2}{\partial x_i^2} + \frac{1}{2} \sum_{ij} k_{ij} x_i x_j \quad (3.5)$$

Dans le terme cinétique la masse m_i de l'atome i apparaît mais n'apparaît pas dans le terme du potentiel. On peut introduire un changement de coordonnées pour éliminer cette masse. On passe en coordonnées de masse réduite :

$$q_i = m_i^{1/2} x_i \quad (3.6)$$

A noter que les coordonnées x_i et donc q_i correspondent à des déplacements nucléaires.

L'hamiltonien nucléaire s'écrit alors en fonction des nouvelles coordonnées :

$$H_n = -\frac{1}{2} \sum_i \frac{\partial^2}{\partial q_i^2} + \frac{1}{2} \sum_{ij} K_{ij} q_i q_j \quad (3.7)$$

$$K_{ij} = \frac{k_{ij}}{(m_i m_j)^{1/2}}$$

On voit aussi que le terme du potentiel contient des termes croisés $q_i q_j$ pour $i \neq j$. On peut construire de nouvelles coordonnées Q_i à partir de combinaisons linéaires des coordonnées de masse réduite q_i pour avoir un potentiel ne contenant pas de termes croisés de ces nouvelles coordonnées.

$$Q_i = \sum_{k=0}^{3M} l_{ik} q_k$$

$$q_i = \sum_{k=0}^{3M} l_{ik}^t Q_k \quad (3.8)$$

$$\mathbf{Q} = \mathbf{l} \mathbf{q}$$

$$\mathbf{q} = \mathbf{l}^t \mathbf{Q}$$

La matrice $\mathbf{l}$ est une matrice réelle orthogonale. $\mathbf{Q}$ est le vecteur des nouvelles coordonnées Q_i , $\mathbf{Q} = (Q_1, Q_2, \dots, Q_{3M})$ et $\mathbf{q}$ est le vecteur des coordonnées q_i , $\mathbf{q} = (q_1, q_2, \dots, q_{3M})$. M est le nombre d'atomes dans la molécule.

Ces nouvelles coordonnées sont appelées les coordonnées normales. Les vibrations correspondant aux déplacements selon ces coordonnées sont les modes normaux.

$$H_n = -\frac{1}{2} \frac{\partial^2}{\partial \mathbf{q}^t \partial \mathbf{q}} + \frac{1}{2} \mathbf{q}^t \mathbf{K} \mathbf{q}$$

$$H_n = -\frac{1}{2} \frac{\partial^2}{\partial (\mathbf{l}^t \mathbf{Q})^t \partial (\mathbf{l}^t \mathbf{Q})} + \frac{1}{2} (\mathbf{l}^t \mathbf{Q})^t \mathbf{K} \mathbf{l}^t \mathbf{Q}$$

$$H_n = -\frac{1}{2} \frac{\partial^2}{\partial (\mathbf{Q}^t \mathbf{l}) \partial (\mathbf{l}^t \mathbf{Q})} + \frac{1}{2} \mathbf{Q}^t \mathbf{l} \mathbf{K} \mathbf{l}^t \mathbf{Q} \quad (3.9)$$

$$H_n = -\frac{1}{2} \frac{\partial^2}{\partial \mathbf{Q}^t \partial \mathbf{Q}} + \frac{1}{2} \mathbf{Q}^t \mathbf{l} \mathbf{K} \mathbf{l}^t \mathbf{Q}$$

$$H_n = -\frac{1}{2} \frac{\partial^2}{\partial \mathbf{Q}^t \partial \mathbf{Q}} + \frac{1}{2} \mathbf{Q}^t \mathbf{\Lambda} \mathbf{Q}$$

Il faut diagonaliser la matrice $\mathbf{K}$ pour obtenir les coordonnées normales Q_i et les fréquences de vibrations $\omega_i = \sqrt{\kappa_i}$.

Parmi les $3M$ coordonnées il y a 3 coordonnées qui correspondent à la translation de la molécule et il y a 3 autres coordonnées qui décrivent la rotation (seulement 2 si la molécule est linéaire). Ces coordonnées auront une valeur de $\kappa_i = 0$. Dans la suite on ne considère plus ces coordonnées, il restera donc $3M - 6$ coordonnées normales ($3M - 5$ si la molécule est linéaire).

Finalement l'hamiltonien nucléaire s'exprime comme la somme d'hamiltonien mettant en jeu un seul mode normal.

$$H_n = \sum_i H_i = -\frac{1}{2} \sum_i \frac{\partial^2}{\partial Q_i^2} + \frac{1}{2} \sum_i \kappa_i Q_i^2 \quad (3.10)$$

Comme l'hamiltonien est une somme d'hamiltonien à un mode, la fonction d'onde s'écrit comme le produit de fonction d'onde à un mode.

$$\Psi = \prod_i \Psi_{\nu_i}(Q_i) \quad (3.11)$$

L'équation de Schrödinger est :

$$\begin{aligned} H_n \Psi &= E \Psi \\ &= \sum_i H_i \prod_i \Psi_{\nu_i}(Q_i) = \sum_i H_i \Psi_{\nu_i}(Q_i) = \sum_i E_{\nu_i} \Psi_{\nu_i}(Q_i) \end{aligned} \quad (3.12)$$

Chaque équation

$$H_i \Psi_{\nu_i}(Q_i) = E_{\nu_i} \Psi_{\nu_i}(Q_i) \quad (3.13)$$

est l'équation d'un oscillateur harmonique dont l'énergie est

$$E_{\nu_i} = (\nu_i + \frac{1}{2}) \hbar \omega_i \text{ avec } \omega_i = \kappa_i^{\frac{1}{2}} \text{ et } \nu_i = 0, 1, 2 \dots \quad (3.14)$$

Les fonctions d'ondes $\Psi_{\nu_i}(Q_i)$ s'expriment en fonction des polynômes d'Hermite, P^h .

$$\Psi_{\nu_i}(Q_i) = N_{\nu_i} P_{\nu_i}^h(y_i) e^{-y_i^2/2} \text{ avec } y_i = (\frac{\omega_i}{\hbar})^{\frac{1}{2}} Q_i \quad (3.15)$$

L'énergie totale E vaut alors :

$$E = \sum_i E_{\nu_i} = \sum_i (\nu_i + \frac{1}{2}) \hbar \omega_i \quad (3.16)$$

Un mode de vibration de la molécule sera défini par l'ensemble des entiers ν_i : $|\nu_1 \nu_2 \dots \nu_{3M-6}\rangle$ (si la molécule est non linéaire), où M est le nombre d'atomes. L'état fondamental est l'état pour lequel tous les ν_i sont nuls : $|0_1 0_2 \dots 0_{3M-6}\rangle$. Il a pour énergie :

$$E_0 = \sum_i \frac{1}{2} \hbar \omega_i \quad (3.17)$$

Cette énergie est appelée l'énergie au point zéro.

Si la molécule reçoit par exemple un photon dont l'énergie est identique ou très proche de l'un des $\hbar \omega_i$ alors la molécule peut absorber cette énergie qui sera redistribuée sur le

mode correspondant faisant vibrer les atomes concernés. Si l'énergie d'un photon est juste égale à $\hbar\omega_i$ et que la molécule est dans son état fondamental alors la molécule subira une transition vibrationnelle passant de l'état fondamental à l'état $|0_1 0_2 \dots 1_i \dots 0_{3M-6}\rangle$. Ce type de transitions est appelé transitions fondamentales.

On peut calculer la probabilité à laquelle peut se produire cette transition en regardant comment varie le dipôle électrique de la molécule. En effet, si le dipôle électrique de la molécule varie lors d'un déplacement atomique alors il peut se coupler avec le champ électromagnétique.

3.1.2 Calcul des probabilités de transition

La probabilité de transition de l'état $|\nu_i\rangle$ vers l'état $|\nu'_i\rangle$ est :

$$P_{|\nu_i\rangle \rightarrow |\nu'_i\rangle} = \langle \nu'_i | \vec{\mu} | \nu_i \rangle \quad (3.18)$$

où $\vec{\mu}$ est le dipôle électrique de la molécule.

L'intensité I de cette transition est calculée à partir de cette probabilité de transition :

$$I \propto |\langle \nu'_i | \vec{\mu} | \nu_i \rangle|^2 \quad (3.19)$$

On peut exprimer le dipôle électrique en fonction des coordonnées normales selon une série de Taylor autour de l'équilibre :

$$\vec{\mu} = \vec{\mu}_0 + \sum_j \left(\frac{\partial \vec{\mu}}{\partial Q_j} \right) Q_j + \dots \quad (3.20)$$

$$\begin{aligned} \langle \nu'_i | \vec{\mu} | \nu_i \rangle &= \langle \nu'_i | \vec{\mu}_0 | \nu_i \rangle + \langle \nu'_i | \sum_j \left(\frac{\partial \vec{\mu}}{\partial Q_j} \right) Q_j | \nu_i \rangle + \dots \\ &= \vec{\mu}_0 \langle \nu'_i | \nu_i \rangle + \left(\frac{\partial \vec{\mu}}{\partial Q_i} \right) \langle \nu'_i | Q_i | \nu_i \rangle + \dots \\ &= 0 + \left(\frac{\partial \vec{\mu}}{\partial Q_i} \right) \langle \nu'_i | Q_i | \nu_i \rangle + \dots \\ &\approx \left(\frac{\partial \vec{\mu}}{\partial Q_i} \right) \langle \nu'_i | Q_i | \nu_i \rangle \end{aligned} \quad (3.21)$$

Dans le cadre de l'approximation harmonique électrique, on arrête le développement à l'ordre 1.

Les états vibrationnels étant orthogonaux, le premier terme qui fait intervenir le dipole de la molécule ($\vec{\mu}_0$) est nul. Lorsqu'on arrête l'expansion du dipôle électrique au premier ordre, les probabilités de transition dépendent des dérivées premières du dipôle électrique par rapport aux coordonnées normales. Pour un oscillateur harmonique les règles de sélection pour les transitions sont $\Delta\nu_i = \nu'_i - \nu_i = \pm 1$

Les transitions électriques dipolaires ne sont possibles que si le dipole varie avec le déplacement Q_i . Les transitions pour lesquelles le dipôle électrique ne varie pas sont dites inactives dans l'infrarouge.

3.2 Anharmonicité

Lorsque l'on compare les fréquences calculées (les différents ω_i) à des valeurs expérimentales, l'écart est important. Une partie de l'erreur vient directement de la méthode employée qui n'est pas exacte (Hartree-Fock, DFT, CCSD(T), etc) mais surtout vient de l'approximation harmonique utilisée pour le potentiel nucléaire. En réalité, cette approximation est grossière, le potentiel est anharmonique. Il faut rajouter des termes d'ordres supérieurs dans le développement du potentiel et du dipôle pour corriger les fréquences et les intensités. Cependant le problème devient plus compliqué et les calculs sont beaucoup plus longs. De plus, dans le cadre de la double approximation harmonique (pour le potentiel et le dipôle), seules les transitions fondamentales sont obtenues et ont une intensité non nulle. Les autres modes harmoniques (seconde harmonique, les modes overtones) et les combinaisons de bandes (deux modes excités simultanément) ne sont pas calculés. Expérimentalement ces derniers peuvent être accessibles et pour pouvoir comprendre les spectres mesurés il faut prendre en compte l'anharmonicité. Différentes méthodes existent pour prendre en compte l'anharmonicité du potentiel et du dipôle.

3.2.1 Emploi de facteurs correctifs

La méthode la moins coûteuse en temps de calcul consiste à d'utiliser des facteurs empiriques pour corriger les fréquences harmoniques. La méthode a d'abord été testée pour des calculs Hartree-Fock par Pople [1] puis étendue à d'autres méthodes [2, 3]. Après un calcul des fréquences harmoniques, il suffit simplement de les multiplier par un ou des facteurs empiriques. Ces facteurs empiriques sont optimisés pour reproduire les fréquences expérimentales mesurées pour certaines molécules. Etant donné que toutes les méthodes donnent rarement les mêmes fréquences on aura alors un ou plusieurs facteurs par méthode de calcul. La même méthode mais avec deux bases différentes peut aussi présenter des facteurs différents. Il existe de nombreux facteurs déjà optimisés et prêt à être utilisés pour certaines méthodes. Cependant comme les données expérimentales sont assez restreintes, l'optimisation des facteurs est effectuée sur un ensemble de molécules limité. Ils ne seront pas forcément transférables pour d'autres molécules.

Ce facteur corrige l'erreur que la méthode fait sur les valeurs harmoniques des fréquences et sur l'anharmonicité. De plus les effets anharmoniques sont moyennés dans ce cas. Ainsi, pour deux modes de vibration différents l'anharmonicité ne sera pas forcément la même. Si on multiplie les fréquences de deux modes qui ont une anharmonicité différente par le même facteur alors on fera forcément une erreur sur la fréquence anharmonique d'au moins un des deux modes. On peut très bien employer deux facteurs différents mais utiliser plusieurs facteurs différents rend le modèle moins facile d'emploi car le nombre de modes augmente très rapidement avec le nombre d'atomes. Les modes n'étant pas forcément les mêmes entre les molécules, si l'on veut prendre correctement l'anharmonicité de chaque mode il faut très vite utiliser un grand nombre de facteurs. De plus, il faut pouvoir identifier les modes correctement pour pouvoir appliquer le bon facteur correctif. Généralement, on utilise un facteur pour corriger toutes les fréquences voire deux facteurs correspondant à deux zones distinctes du spectre.

En effet, la figure 3.1 montre le rapport entre les fréquences expérimentales harmoniques et les fréquences expérimentales fondamentales d'une trentaine de petites molécules. On voit qu'il y a deux zones distinctes, la zone I entre 500 et 2500 cm^{-1} et la zone II entre 2500 et 4500 cm^{-1} . Dans chacune de ces deux zones les données peuvent être ajustées

par une courbe $y = \text{constante}$ dont la valeur de la constante n'est pas la même pour les deux zones. Pour la zone I la constante (donc le rapport entre fréquences harmoniques et fréquences anharmoniques) vaut 0.980 et pour la zone II il vaut 0.955. La dispersion n'est pas dans la même dans les deux zones. Dans la zone I, la dispersion est plus grande et cela s'explique par le fait que les types de modes de vibration ne sont pas tous les mêmes. Dans cette zone, on peut avoir des modes d'élongation comme des modes de déformation. Dans la zone II la dispersion est plus petite parce que les modes de vibration sont tous des modes d'élongation. Enfin il y a deux modes (un mode de la molécule CH_3OH et un mode de la molécule NH_3) qui sont très écartés des courbes de tendance, cela s'explique par le fait que ces modes sont très anharmoniques. Pour conclure, on voit que le fait d'employer des facteurs multiplicatifs pour corriger les fréquences harmoniques théoriques est justifié.

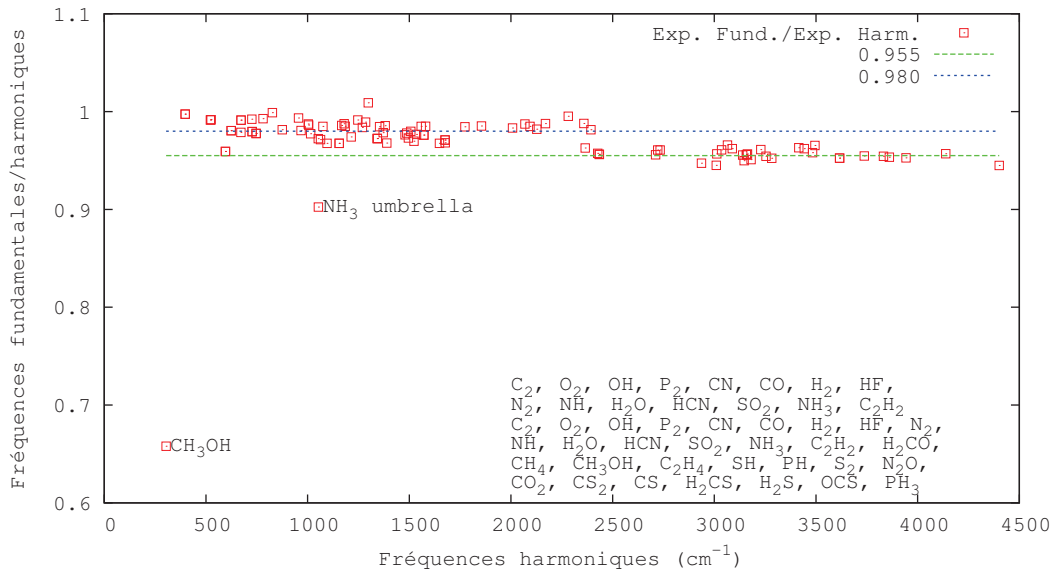


FIGURE 3.1 – Tracé du rapport entre les fréquences fondamentales expérimentales et les fréquences harmoniques expérimentales en fonction des fréquences expérimentales harmoniques

3.2.2 VPT2

Une approche pour calculer l'anharmonicité du potentiel est la théorie de la perturbation de Van Vleck [4, 5]. Cette méthode a été popularisée par Vincenzo Barone qui l'a implémentée dans le logiciel Gaussian [6]. L'avantage de cette méthode est qu'elle traite de manière perturbative les interactions faibles (états non dégénérés) et de manière variationnelle les interactions fortes (état quasi-dégénérés). Les états vibrationnels sont classés en deux catégories, les états de classe 1 sont les modes interagissant très fortement entre eux et qui peuvent conduire à des résonances et les états de classe 2 sont les modes qui interagissent faiblement entre eux. La méthode consiste à transformer l'hamiltonien vibrationnel pour obtenir un nouvel hamiltonien diagonal par blocs.

$$H = \begin{pmatrix} H_2 & H_{1-2} \\ H_{2-1} & H_1 \end{pmatrix} \rightarrow \tilde{H} = \begin{pmatrix} \tilde{H}_2 & 0 \\ 0 & \tilde{H}_1 \end{pmatrix} \quad (3.22)$$

Où H_2 est l'hamiltonien pour les états appartenant à la classe 2, H_1 l'hamiltonien des états appartenant à la classe 1 et H_{1-2} l'hamiltonien des interactions entre les états de

ces deux classes. La transformation T opérée sur l'hamiltonien H est telle que :

$$\tilde{H} = THT^{-1} \quad (3.23)$$

Pour le développement en série du potentiel V_n (voir 3.1.1), on va au-delà de l'approximation harmonique mécanique. L'hamiltonien vibrationnel s'exprime généralement en fonction des coordonnées sans dimension $\{p\}$ et $\{y\}$ et s'écrit :

$$\begin{aligned} H &= H^0 + H' + H'' \\ H &= \frac{hc}{2} \sum_k \theta_k (p_k^2 + y_k^2) + \frac{hc}{3!} \sum_{klm} \phi_{klm} y_k y_l y_m + \frac{hc}{4!} \sum_{klmn} \phi_{klmn} y_k y_l y_m y_n \\ H^0 &= \frac{hc}{2} \sum_k \theta_k (p_k^2 + y_k^2) \\ H' &= \frac{hc}{3!} \sum_{klm} \phi_{klm} y_k y_l y_m \\ H'' &= \frac{hc}{4!} \sum_{klmn} \phi_{klmn} y_k y_l y_m y_n \end{aligned} \quad (3.24)$$

H^0 est l'hamiltonien harmonique considéré comme l'hamiltonien non perturbé. H' est l'hamiltonien de la perturbation à l'ordre 1 et H'' à l'ordre 2. θ_k est la pulsation angulaire telle que $\kappa_k^{1/2} = \omega_k = 2\pi c\theta_k$ (voir partie calcul des vibrations, κ_k est la constante de force en coordonnées réduites et ω_k la fréquence de vibration). Enfin ϕ_{klm} et ϕ_{klmn} sont les constantes de force, c'est-à-dire les dérivées troisième et quatrième du potentiel par rapport aux coordonnées y . Originellement, la méthode a été dérivée en tenant compte des effets rotationnels sur les niveaux d'énergie vibrationnels. Ici, pour simplifier, on ne prend pas en compte ces termes.

La surface d'énergie potentielle est représentée par un champ de force quartique (QFF) puisque le potentiel est développé jusqu'à l'ordre 4 dans la série de Taylor. En pratique, pour l'hamiltonien H'' , tous les termes ne sont pas retenus. En effet, les termes ϕ_{klmn} à 4 indices différents ne sont pas calculés, ce qui revient à négliger les interactions entre 4 modes normaux différents. Ces termes ne sont pas calculés pour réduire le temps de calcul des dérivées sans trop détériorer la précision du calcul. Le champ de force est alors appelé champ de force semi-quartique. Au maximum 3 modes différents peuvent interagir, on appelle un tel champ de force, un champ de force 3MR-QFF pour *3 modes-representation quartic force field*.

Dans le cadre la théorie de la perturbation, on écrit l'hamiltonien

$$H = H^0 + \lambda H' + \lambda^2 H'' + \dots \quad (3.25)$$

avec λ le paramètre d'ordre. Après transformation l'hamiltonien est

$$\tilde{H} = \tilde{H}^0 + \lambda \tilde{H}' + \lambda^2 \tilde{H}'' + \dots \quad (3.26)$$

L'opérateur de la transformation s'écrit comme :

$$T = e^{i\lambda S} = 1 + i\lambda S - \frac{\lambda^2}{2} S^2 + \dots \quad (3.27)$$

Avec S une matrice hermitienne.

En remplaçant (3.27) et (3.25) dans (3.26) et en identifiant les puissances en λ , on obtient que :

$$\begin{aligned}\lambda^0 : \tilde{H}^0 &= H^0 \\ \lambda^1 : \tilde{H}' &= H' + i(SH^0 - H^0S) \\ \lambda^2 : \tilde{H}'' &= H'' + SH^0S - \frac{1}{2}(H^0S^2 + S^2H^0) + i(SH' - H'S)\end{aligned}\tag{3.28}$$

La transformation de Van Vleck doit vérifier les deux propriétés suivantes : Les interactions au sein d'une même classe ne doivent pas être altérées. La matrice S doit donc vérifier les conditions suivantes :

$$\begin{aligned}S_{ab} &= 0, \text{ a et b } \in \text{ classe 1} \\ S_{\gamma\sigma} &= 0, \text{ } \gamma \text{ et } \sigma \in \text{ classe 2}\end{aligned}\tag{3.29}$$

La transformation doit conduire à $\tilde{H}_{1-2} = 0$. C'est-à-dire que l'hamiltonien transformé est diagonal par blocs. La matrice S doit donc vérifier que

$$\begin{aligned}\tilde{H}'_{a\gamma} &= 0, \text{ a } \in \text{ classe 1 et } \gamma \in \text{ classe 2} \\ \tilde{H}''_{a\gamma} &= 0\end{aligned}\tag{3.30}$$

En gardant à l'esprit ces conditions on peut dériver toutes les expressions pour les termes diagonaux $\tilde{H}_{ii}$ et non diagonaux $\tilde{H}_{ij}$ avec i l'indice d'un état pouvant appartenir soit la classe 1 ou la classe 2. On ne va pas détailler le calcul de ces termes mais toutes les expressions se trouvent dans la littérature [5].

L'énergie d'un état vibrationnel peut se mettre sous la forme d'une expression de Dunham :

$$\tilde{H}_{nn} = E_n = \chi_0 + \sum_i^N \omega_i(n_i + \frac{1}{2}) + \sum_i^N \sum_j^N \chi_{ij}(n_i + \frac{1}{2})(n_j + \frac{1}{2})\tag{3.31}$$

Les éléments χ_{ij} dans le terme de l'énergie (3.31) sont les contributions anharmoniques. Lorsque l'on développe les termes $\tilde{H}_{nn}$, on va obtenir des termes qui sont multipliés par les nombres quantiques $(n_i + \frac{1}{2})(n_j + \frac{1}{2})$. Ces termes sont regroupés dans la matrice χ . Leur expression est la suivante, pour les termes diagonaux

$$16\chi_{ii} = \phi_{iiii} - \sum_j \phi_{ij}^2 \left[\frac{2}{\omega_j} + \frac{1}{2(2\omega_i + \omega_j)} - \frac{1}{2(2\omega_i - \omega_j)} \right]\tag{3.32}$$

et pour les termes non diagonaux :

$$\begin{aligned}4\chi_{ij} &= \phi_{iijj} - \sum_k \frac{\phi_{ijk}^2}{2} \left[\frac{1}{\omega_i + \omega_j + \omega_k} + \frac{1}{\omega_i + \omega_j - \omega_k} + \frac{1}{\omega_i - \omega_j + \omega_k} + \frac{1}{\omega_i - \omega_j - \omega_k} \right] \\ &\quad - \sum_k \phi_{iik} \phi_{jjk} \frac{1}{\omega_k}\end{aligned}\tag{3.33}$$

χ_0 est la contribution au point zéro. χ_0 regroupe les termes des éléments $\tilde{H}_{nn}$ qui ne sont pas multipliés par les nombres quantiques. L'expression de χ_0 est :

$$64\chi_0 = \sum_i \phi_{iiii} - \frac{7}{9} \sum_i \phi_{iii}^2 + \frac{3}{2} \sum_{j \neq i} \phi_{ij}^2 \left[\frac{1}{2\omega_i - \omega_j} - \frac{1}{2\omega_i + \omega_j} \right] - \frac{1}{3} \sum_{i \neq j \neq k} \phi_{ijk}^2 \left[\frac{1}{\omega_i + \omega_j + \omega_k} + \frac{1}{\omega_i + \omega_j - \omega_k} + \frac{1}{\omega_i - \omega_j + \omega_k} + \frac{1}{\omega_i - \omega_j - \omega_k} \right] \quad (3.34)$$

A partir de l'expression de l'énergie (3.31) on peut remonter aux fréquences des transitions fondamentales, aux premiers overtones et aux combinaisons de bandes.

$$\text{Transition fondamentale : } \nu_{1_i} = \omega_i + 2\chi_{ii} + \sum_{j \neq i} \frac{\chi_{ij}}{2}$$

$$\text{Transition overtone : } \nu_{2_i} = 2\omega_i + 6\chi_{ii} + \sum_{j \neq i} \chi_{ij}$$

$$\text{Transition de combinaison : } \nu_{1_i 1_j} = \omega_i + \omega_j + 2\chi_{ii} + 2\chi_{jj} + 2\chi_{ij} + \frac{1}{2} \sum_{k \neq i, j} (\chi_{ik} + \chi_{jk}) \quad (3.35)$$

Certains des termes figurant dans les expressions de χ_{ii} , χ_{ij} et χ_0 ont des dénominateurs qui peuvent diverger si des états ont des énergies très proches ($\pm\omega_i \pm \omega_j \pm \omega_k \approx 0$). Et lorsque c'est le cas on parle de résonance. Il existe deux types de résonances selon le nombre de quanta qui interagissent ensemble.

Les résonances de Fermi

Ces résonances correspondent à un changement total de 3 quanta (n_i). Soit la fréquence du mode est environ égale à deux fois la fréquence d'un autre mode soit la fréquence d'un mode est quasiment égale à la somme des fréquences de deux autres modes. Cela revient à dire qu'un quantum d'un mode est éliminé pour en créer deux autres dans un même mode (résonance de type I) ou pour en créer deux autres dans deux modes différents (résonance de type II). On appelle ces résonances, des résonances 1-2.

Les différents dénominateurs posant problème sont :

$$\frac{1}{2(2\omega_i - \omega_j)} \quad \frac{1}{\omega_i - \omega_j - \omega_k} \quad (3.36)$$

Ces résonances proviennent de la perturbation au premier ordre (H').

Les résonances de Darling et Dennison

Ces résonances correspondent à un changement total de 4 quanta (n_i). Soit la somme des fréquences de deux modes est quasiment égale à la somme des fréquences de deux autres

modes. Cela revient à dire qu'un quantum est éliminé dans deux modes pour en créer deux autres dans deux modes différents. On appelle ces résonances, des résonances 2-2. Ces résonances proviennent de la perturbation au second ordre (H''). Elles sont donc plus faibles que les résonances de Fermi.

Traitement des résonances

DVPT2

En laissant les termes résonants dans l'expression des constantes anharmoniques χ_{ij} on obtiendrait des résultats totalement aberrants. La manière la plus simple de les traiter est de tout simplement les enlever de l'expression. On enlève ces termes de l'expression après les avoir identifiés via des critères empiriques (voir la suite). Cette méthode est appelée DVPT2 pour *Deperturbed VPT2* au sens où les termes divergents issus de la théorie de la perturbation sont enlevés.

GVPT2

Si l'on s'arrête ici on néglige une partie de l'interaction entre certains modes (ceux qui interviennent dans la résonance donc des modes de classe 1 qui interagissent fortement entre eux) et le résultat peut être faussé. Pour obtenir une bonne précision il faut pouvoir réintroduire cette interaction d'une autre manière. Cela se fait par un traitement variationnel. Pour cela on effectue la diagonalisation de la matrice de l'hamiltonien effectif (on a enlevé des termes) des états mis en jeu dans la résonance de Fermi. En effet, pour des états a et b de classe 1 le terme $\tilde{H}'_{ab}$ ne s'annule pas, il faut donc diagonaliser le bloc. Cette méthode est discutée dans l'annexe A de l'article de Martin [7]. La méthode qui consiste à éliminer les termes résonants et à les réintroduire variationnellement est appelée GVPT2 pour *Generalized VPT2*. Dans le même article de Martin, des conditions ont été dérivées pour l'identification des résonances de Fermi, c'est le test de Martin. Ces conditions sont les suivantes :

$$\begin{aligned}
 &\text{Pour le type I : } |\omega_i - 2\omega_j| \leq \Delta\omega \\
 &\Delta_{ij} = \frac{\phi_{ijj}^4}{256(2\omega_j - \omega_i)^3} \geq A^I \\
 &\text{Pour le type II : } |\omega_i - (\omega_j + \omega_k)| \leq \Delta\omega \\
 &\Delta_{ijk} = \frac{\phi_{ijk}^4}{64(\omega_j + \omega_k - \omega_i)^3} \geq A^{II}
 \end{aligned} \tag{3.37}$$

Avec A^I et A^{II} des paramètres empiriques qui valent (par défaut dans Gaussian) tous les deux 1 cm^{-1} et $\Delta\omega = 200 \text{ cm}^{-1}$.

Ces paramètres influencent la qualité des résultats puisqu'ils permettent l'identification des résonances de Fermi. C'est là l'un des points faibles de la méthode GVPT2. Un autre point faible de la méthode GVPT2 est le temps nécessaire à la diagonalisation de la matrice lors du traitement variationnel.

DCPT2

Une méthode a été développée pour contourner ces limitations. C'est la méthode DCPT2 [8, 9] pour *Degeneracy-corrected VPT2*. Dans cette méthode, les termes potentiellement résonants, c'est-à-dire tous les termes contenant un pôle sont réécrits selon une certaine formule qui lève la divergence.

Pour comprendre la méthode on part d'un hamiltonien simplifié sous la forme :

$$H = \begin{pmatrix} A - \epsilon & k \\ k & A + \epsilon \end{pmatrix} \quad (3.38)$$

Les valeurs propres sont :

$$E_{\pm} = A \pm \epsilon \sqrt{1 + \frac{k^2}{\epsilon^2}} \quad (3.39)$$

Maintenant on écrit la racine carrée en une série de Taylor à l'ordre 2.

$$E_{\pm} = A \pm \epsilon \pm \frac{k^2}{2\epsilon} + \dots \quad (3.40)$$

En égalisant les équations (3.39) et (3.40), on obtient :

$$\begin{aligned} \pm \frac{k^2}{2\epsilon} &= \pm(\sqrt{\epsilon^2 + k^2} - \epsilon) \\ \frac{Sk^2}{2\epsilon} &= S(\sqrt{\epsilon^2 + k^2} - \epsilon) \end{aligned} \quad (3.41)$$

k^2 et ϵ sont pris positifs par définition. S est le signe qui peut être soit $+1$ soit -1 .

Regardons maintenant un exemple d'application sur le terme potentiellement résonant dans la constante d'anharmonicité χ_{ii} :

$$\begin{aligned} \frac{\phi_{ij}^2}{2(2\omega_j - \omega_i)} &= \frac{Sk^2}{2\epsilon} \\ k^2 &= \phi_{ij}^2 \\ \epsilon &= |2\omega_j - \omega_i| \\ S &= \text{sign}(2\omega_j - \omega_i) \end{aligned} \quad (3.42)$$

En remplaçant le terme $\frac{Sk^2}{2\epsilon}$ par l'expression de (3.41), le terme potentiellement résonant se réécrit :

$$\frac{\phi_{ij}^2}{2(2\omega_j - \omega_i)} = \text{sign}(2\omega_j - \omega_i) \left[\sqrt{(2\omega_j - \omega_i)^2 + \phi_{ij}^2} - |2\omega_j - \omega_i| \right] \quad (3.43)$$

Ainsi quand ϵ soit $|2\omega_j - \omega_i|$ tend vers zéro il n'y a plus de divergence.

Tous les autres termes potentiellement résonants sont réécrits de la même manière. Il n'y a alors plus aucun terme divergent. On n'a plus à diagonaliser la matrice réduite formée par les états en résonance comme pour la méthode GVPT2. Avec la réécriture des termes potentiellement résonants tout se passe comme si les matrices avaient déjà été diagonalisées. L'avantage de cette méthode, c'est de ne plus avoir à compter sur des facteurs empiriques pour identifier les résonances puisque tous les termes potentiellement

résonants sont réécrits sous une forme qui ne contient plus de résonance. Cependant la réécriture des termes s'effectue même si on est loin de la résonance ce qui peut détériorer le résultat. Or loin de la résonance la méthode VPT2 est parfaitement valide (pas de dégénérescence donc pas de termes divergents). C'est avec cette méthode qu'il faudrait calculer les termes lorsque l'on est loin de la résonance.

HDCPT2

Une nouvelle méthode [9] a vu le jour pour prendre en compte ces considérations. C'est la méthode HDCPT2 pour *Hybrid DCPT2*. Dans cette méthode le terme potentiellement résonant est calculé de la manière suivante :

$$f_{HDCPT2} = \Lambda f_{VPT2} + (1 - \Lambda) f_{DCPT2} \quad (3.44)$$

Où Λ est une fonction de transition :

$$\Lambda = \frac{1}{2}(\tanh(\alpha [\sqrt{k^2 \epsilon^2} - \beta]) + 1) \quad (3.45)$$

α et β sont de nouveaux paramètres mais qui n'ont que peu d'influence sur les résultats [9] ($\alpha = 1$ et $\beta = 5.10^5$). Lorsque l'on est loin de la résonance, on a $\epsilon \rightarrow \infty$ alors $\Lambda \rightarrow 1$. On prend totalement ou en très grande partie la contribution du terme VPT2. Lorsqu'il y a résonance, on a $\epsilon \rightarrow 0$ alors $\Lambda \rightarrow 0$. On prend alors majoritairement la contribution du terme de DCPT2.

Calcul des intensités

Dans le cadre de l'approximation harmonique, l'expression du dipôle électrique est limitée au premier ordre.

$$\vec{\mu} = \vec{\mu}_0 + \sum_i \left(\frac{\partial \vec{\mu}}{\partial Q_i} \right) Q_i \quad (3.46)$$

Pour prendre en compte l'anharmonicité on ajoute les deuxième et troisième ordres. Ainsi le dipôle électrique s'écrit :

$$\begin{aligned} \vec{\mu} &= \vec{\mu}_0 + \sum_i \left(\frac{\partial \vec{\mu}}{\partial Q_i} \right) Q_i + \frac{1}{2} \sum_{ij} \left(\frac{\partial^2 \vec{\mu}}{\partial Q_i \partial Q_j} \right) Q_i Q_j + \frac{1}{6} \sum_{ijk} \left(\frac{\partial^3 \vec{\mu}}{\partial Q_i \partial Q_j \partial Q_k} \right) Q_i Q_j Q_k \\ \vec{\mu} &= \vec{\mu}_0 + \sum_i P_i Q_i + \frac{1}{2} \sum_{ij} P_{ij} Q_i Q_j + \frac{1}{6} \sum_{ijk} P_{ijk} Q_i Q_j Q_k \end{aligned} \quad (3.47)$$

Le moment de transition est calculé à partir de $\langle \nu'_i | \vec{\mu} | \nu_i \rangle$. En utilisant les fonctions d'ondes vibrationnelles obtenues avec la théorie de la perturbation de Van Vleck, on obtient des expressions qui ne seront pas détaillées ici (voir [10] pour plus de détails). L'expression de cette intensité dépend des dérivées premières, deuxième et troisième du dipôle et des constantes de force ϕ_{ijk} . Comme pour les énergies, certains termes possèdent des dénominateurs qui peuvent diverger. Mais contrairement aux énergies, actuellement, les résonances pour les intensités ne sont pas traitées. Le schéma DVPT2 est utilisé pour les résonances intervenant dans le calcul des intensités. C'est-à-dire que les termes résonants sont simplement enlevés de l'expression.

Calcul des constantes de force et des dérivées du dipôle

Il y a deux façons de calculer les dérivées du potentiel et par la même occasion les dérivées du dipôle.

Calcul des dérivées d'ordre 3 et 4 du potentiel à partir du hessian analytique

Les dérivées d'ordre 3 et 4 peuvent être obtenues numériquement à partir des dérivées secondes analytiques comme c'est le cas avec le logiciel Gaussian. Elles s'expriment avec les coordonnées normales [11] (les modes normaux Q_i) comme :

$$\phi_{ijk} = \frac{1}{3} \left[\sqrt{\frac{\hbar}{2\pi c\omega_i}} \frac{\phi_{jk}(+\delta Q_i) + \phi_{jk}(-\delta Q_i)}{2\delta Q_i} + \sqrt{\frac{\hbar}{2\pi c\omega_j}} \frac{\phi_{ik}(+\delta Q_j) + \phi_{ik}(-\delta Q_j)}{2\delta Q_j} + \sqrt{\frac{\hbar}{2\pi c\omega_k}} \frac{\phi_{ij}(+\delta Q_k) + \phi_{ij}(-\delta Q_k)}{2\delta Q_k} \right] \quad (3.48)$$

$$\phi_{iijk} = \frac{\hbar}{2\pi c\omega_i} \frac{\phi_{jk}(+\delta Q_i) + \phi_{jk}(-\delta Q_i) - 2\phi_{jk}(0)}{\delta Q_i^2} \quad (3.49)$$

$$\phi_{iiij} = \frac{1}{2} \left[\frac{\hbar}{2\pi c\omega_i} \frac{\phi_{jj}(+\delta Q_i) + \phi_{jj}(-\delta Q_i) - 2\phi_{jj}(0)}{\delta Q_i^2} + \frac{\hbar}{2\pi c\omega_j} \frac{\phi_{ii}(+\delta Q_j) + \phi_{ii}(-\delta Q_j) - 2\phi_{ii}(0)}{\delta Q_j^2} \right] \quad (3.50)$$

En ce qui concerne les dérivées secondes et troisièmes, elles sont calculées numériquement à partir des dérivées premières analytiques du dipôle.

$$\vec{P}_{ji} = \frac{1}{2} \left[\sqrt{\frac{\hbar}{2\pi c\omega_i}} \frac{\vec{P}_j(+\delta Q_i) + \vec{P}_j(-\delta Q_i)}{2\delta Q_i} + \sqrt{\frac{\hbar}{2\pi c\omega_j}} \frac{\vec{P}_i(+\delta Q_j) + \vec{P}_i(-\delta Q_j)}{2\delta Q_j} \right] \quad (3.51)$$

$$\vec{P}_{jji} = \left[\frac{\hbar}{2\pi c\omega_j} \frac{\vec{P}_i(+\delta Q_j) + \vec{P}_i(-\delta Q_j) - 2\vec{P}_i(0)}{\delta Q_j^2} \right] \quad (3.52)$$

δQ_i est un petit déplacement selon le mode normal Q_i . Par défaut dans le logiciel Gaussian, le pas δQ_i est le même pour tous les modes et vaut 0,01 Å. Pour calculer toutes les constantes de forces il faut calculer un hessian par déplacement δQ_i . En tout on a $3M - 6$ modes si la molécule n'est pas linéaire ($3M - 5$ sinon), où M est le nombre d'atomes. De plus on se déplace de $\pm \delta Q_i$, ainsi pour chaque mode on calcule deux hessians selon le signe du déplacement. Ce qui nous fait $6M - 12$ hessians à calculer. Il faut compter le hessian à calculer correspondant à la configuration à l'équilibre. Au total il faut calculer $6M - 11$ hessians ($6M - 9$ si la molécule est linéaire) pour obtenir toutes les dérivées nécessaires au calcul des fréquences et des intensités anharmoniques. (Il faut calculer $2f + 1$ hessians avec f le nombre de modes). Une fois que l'on a calculé les fréquences harmoniques, il reste à calculer $2f$ hessians. Ces calculs sont indépendants les uns des autres ce qui fait que cette partie peut être en théorie parallélisable (si l'implémentation de la méthode dans le logiciel de calcul le permet). Dans le meilleur des cas (nombre de processeurs et de noeuds suffisant, c'est-à-dire $2f$ calculs simultanés) le calcul des $2f$ hessians peut être effectué en un temps égal à celui nécessaire pour calculer un hessian.

Calcul numérique des dérivées du potentiel et du dipôle à partir de l'énergie et du dipôle

Les dérivées d'ordre 3 et 4 du potentiel et les dérivées d'ordre 2 et 3 du dipôle peuvent être obtenues entièrement numériquement à partir de l'énergie et du dipôle. Les dérivées sont calculées à partir de la formule de dérivation à 7 points présentée par Yagi [12]. Pour alléger les expressions suivantes, on considère que δ_i est le petit déplacement δQ_i . Le déplacement δQ_i peut être calculé à partir des coordonnées réduites y_i . On a

$$y_i = \left(\frac{\omega_i}{\hbar}\right)^{\frac{1}{2}} Q_i \quad (3.53)$$

En regardant un petit déplacement

$$\delta y_i = \left(\frac{\omega_i}{\hbar}\right)^{\frac{1}{2}} \delta Q_i \quad (3.54)$$

En choisissant δy_i identique pour tous les modes (δy) on obtient un déplacement δQ_i qui va dépendre de la fréquence du mode en question.

$$\delta Q_i = \left(\frac{\omega_i}{\hbar}\right)^{-\frac{1}{2}} \delta y_i = \left(\frac{\omega_i}{\hbar}\right)^{-\frac{1}{2}} \delta y \quad (3.55)$$

On peut aussi choisir de prendre un δQ_i fixe pour tous les modes.

Les expressions pour les dérivées troisièmes sont :

$$\begin{aligned} \phi_{iii} &= \frac{1}{8\delta_i^3} [-V(+3\delta_i) + 8V(+2\delta_i) - 13V(+\delta_i) + 13V(-\delta_i) - 8V(-2\delta_i) + V(-3\delta_i)] \\ \phi_{iij} &= \frac{1}{2\delta_i^2\delta_j} \{V(+\delta_i, +\delta_j) + V(-\delta_i, +\delta_j) - 2V(+\delta_j) \\ &\quad - [V(+\delta_i, -\delta_j) + V(-\delta_i, -\delta_j) - 2V(-\delta_i)]\} \\ \phi_{ijk} &= \frac{1}{8\delta_i\delta_j\delta_k} \{V(+\delta_i, +\delta_j, +\delta_k) - V(-\delta_i, +\delta_j, +\delta_k) - V(+\delta_i, -\delta_j, +\delta_k) \\ &\quad - V(+\delta_i, +\delta_j, -\delta_k) + V(+\delta_i, -\delta_j, -\delta_k) + V(-\delta_i, -\delta_j, +\delta_k) \\ &\quad + V(-\delta_i, +\delta_j, -\delta_k) - V(-\delta_i, -\delta_j, -\delta_k)\} \end{aligned} \quad (3.56)$$

Et pour les dérivées quatrièmes :

$$\begin{aligned} \phi_{iiii} &= \frac{1}{6\delta_i^4} [-V(+3\delta_i) + 12V(+2\delta_i) - 39V(+\delta_i) + 56V_0 - 39V(-\delta_i) + 12V(-2\delta_i) - V(-3\delta_i)] \\ \phi_{iiij} &= \frac{1}{16\delta_i^3\delta_j} \{[V(+3\delta_i, +\delta_j) - 3V(+\delta_i, +\delta_j) + 3V(-\delta_i, +\delta_j) - V(-3\delta_i, +\delta_j)] \\ &\quad - [V(+3\delta_i, -\delta_j) - 3V(+\delta_i, -\delta_j) + 3V(-\delta_i, -\delta_j) - V(-3\delta_i, -\delta_j)]\} \\ \phi_{iiijj} &= \frac{1}{\delta_i^2\delta_j^2} \{[V(+\delta_i, +\delta_j) + V(-\delta_i, +\delta_j) + V(+\delta_i, -\delta_j) + V(-\delta_i, -\delta_j)] \\ &\quad - 2[V(\delta_i) + V(-\delta_i) + V(\delta_j) + V(-\delta_j)] + 4V_0\} \\ \phi_{iiijk} &= \frac{1}{4\delta_i^2\delta_j\delta_k} \{[V(+\delta_i, +\delta_j, +\delta_k) - V(+\delta_i, -\delta_j, +\delta_k) - V(+\delta_i, +\delta_j, -\delta_k) \\ &\quad + V(+\delta_i, -\delta_j, -\delta_k) + V(-\delta_i, +\delta_j, +\delta_k) - V(-\delta_i, -\delta_j, +\delta_k) \\ &\quad - V(-\delta_i, +\delta_j, -\delta_k) + V(-\delta_i, -\delta_j, -\delta_k) - 2[V(+\delta_j, +\delta_k) \\ &\quad - V(-\delta_j, +\delta_k) - V(+\delta_j, -\delta_k) + V(-\delta_j, -\delta_k)]\} \end{aligned} \quad (3.57)$$

$$\begin{aligned}
\vec{P}_i &= \frac{1}{60\delta_i} [\vec{\mu}(+3\delta_i) - 9\vec{\mu}(+2\delta_i) + 45\vec{\mu}(+\delta_i) - 45\vec{\mu}(-\delta_i) + 9\vec{\mu}(-2\delta_i) - \vec{\mu}(-3\delta_i)] \\
\vec{P}_{ii} &= \frac{1}{180\delta_i^2} [2\vec{\mu}(+3\delta_i) - 27\vec{\mu}(+2\delta_i) + 270\vec{\mu}(+\delta_i) - 490\vec{\mu}_0 + 270\vec{\mu}(-\delta_i) - 27\vec{\mu}(-2\delta_i) + 2\vec{\mu}(-3\delta_i)] \\
\vec{P}_{ij} &= \frac{1}{4\delta_i\delta_j} [\vec{\mu}(+\delta_i, +\delta_j) - \vec{\mu}(+\delta_i, -\delta_j) - \vec{\mu}(-\delta_i, +\delta_j) + \vec{\mu}(-\delta_i, -\delta_j)]
\end{aligned} \tag{3.58}$$

Pour calculer toutes les dérivées nécessaires il faut calculer N_{QFF} énergies (le dipôle est calculé en même temps). Le nombre d'énergies à calculer suit la formule suivante :

$$N_{QFF} = 1 + 6f + 6f(f-1) + \frac{4}{3}f(f-1)(f-2) \tag{3.59}$$

avec f le nombre de modes. Le nombre d'énergies à calculer est en f^3 mais selon le nombre de processeurs et/ou de noeuds disponibles, la parallélisation du calcul des dérivées peut être beaucoup plus efficace que pour la méthode précédente.

VPT2 hybride

Que ce soit avec l'une ou l'autre méthode, le calcul des dérivées peut prendre beaucoup de temps. Pour des molécules de grande taille le calcul n'est pas réalisable. Des méthodes hybrides ont alors été testées pour lesquelles les dérivées sont calculées à l'aide d'un niveau de théorie plus bas que celui utilisé pour obtenir les valeurs harmoniques afin de réduire le temps de calcul. Les résultats de ces méthodes dépendent fortement du haut niveau (HN) utilisé pour calculer les fréquences harmoniques. Il faut que la précision du haut niveau soit bonne si l'on veut des bons résultats.

Il existe des méthodes hybrides suivant la première méthode pour calculer les dérivées. La géométrie est optimisée, les fréquences sont calculées et les modes normaux sont obtenus avec un haut niveau de théorie. Avec le bas niveau (BN) on réoptimise la géométrie, on calcule les fréquences et on obtient les nouveaux modes normaux. On associe ensuite les modes normaux du bas niveau à ceux du haut niveau [13]. On calcule ensuite les dérivées cubiques et quartiques du potentiel avec le bas niveau. Pour obtenir les fréquences anharmoniques il faut utiliser les fréquences calculées du haut niveau auxquelles on ajoute les dérivées du bas niveau :

$$\begin{aligned}
\text{Transition fondamentale : } \nu_{1_i} &= \omega_i^{HN} + 2\chi_{ii}^{BN} + \sum_{j \neq i} \frac{1}{2}\chi_{ij}^{BN} \\
\text{Transition overtone : } \nu_{2_i} &= 2\omega_i^{HN} + 6\chi_{ii}^{BN} + \sum_{j \neq i} \chi_{ij}^{BN} \\
\text{Transition de combinaison : } \nu_{1_i 1_j} &= \omega_i^{HN} + \omega_j^{HN} + 2\chi_{ii}^{BN} + 2\chi_{jj}^{BN} + 2\chi_{ij}^{BN} + \frac{1}{2} \sum_{k \neq i,j} (\chi_{ik}^{BN} + \chi_{jk}^{BN})
\end{aligned} \tag{3.60}$$

Cependant, il faut que la géométrie et les modes obtenus avec le bas niveau soient proches de ceux trouvés avec le haut niveau. Néanmoins avec un bon choix pour le bas niveau cette

-
- 1: Choix de la méthode comme haut niveau
 - 2: Optimisation de la géométrie avec le haut niveau
 - 3: Calcul des fréquences et des intensités harmoniques avec le haut niveau
 - 4: Choix de la méthode bas niveau
 - 5: Choix des modes à considérer dans le calcul
 - 6: Génération de toutes les géométries utiles au calcul des dérivées cubiques et quartiques
 - 7: **Pour** chaque fichier généré **Faire** ▷ Cette étape peut s'effectuer de manière parallèle
 - 8: Calculer l'énergie et le dipôle
 - 9: En utilisant les énergies calculées, obtenir les dérivées cubiques et quartiques
 - 10: En utilisant les dipôles calculés, obtenir les dérivées secondes et cubiques du dipôle
 - 11: En utilisant ces dérivées, calculer les fréquences et les intensités anharmoniques en utilisant VPT2, GVPT2, DCPT2 ou HDCPT2.
-

FIGURE 3.2 – Principe de fonctionnement de la méthode hybride

méthode produit de bons résultats [13, 14, 15, 16, 17]. Ce type de méthodes hybrides est implémenté dans Gaussian. Cependant son implémentation force le fait que la géométrie est réoptimisée par le bas niveau.

Il existe une autre manière de calculer les fréquences anharmoniques avec des méthodes hybrides plus générale. Elles reposent sur la différentiation numérique à 7 points introduite par Yagi.

Le principe de fonctionnement de la méthode est résumé dans la figure 3.2 . Les géométries

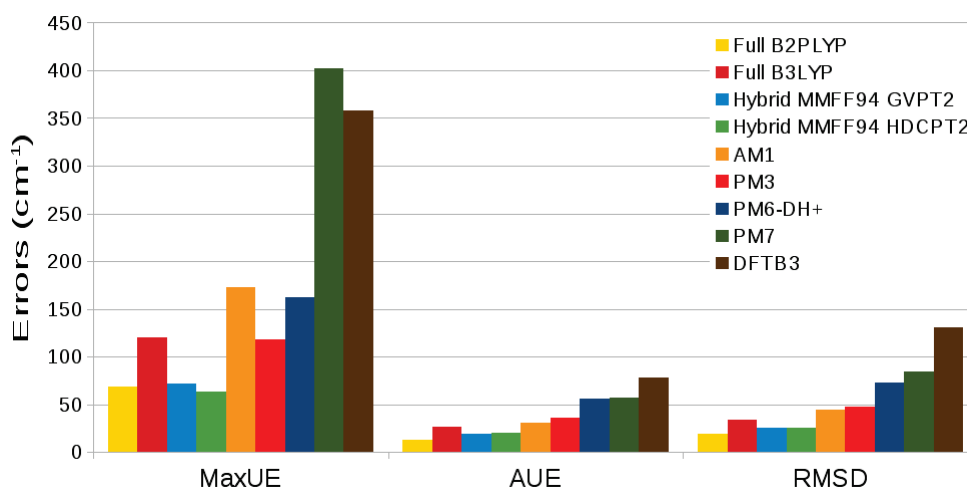


FIGURE 3.3 – Histogramme des erreurs sur les fréquences anharmoniques calculées par différentes méthodes hybrides par rapport aux fréquences expérimentales pour 15 molécules de taille inférieure à 6 atomes

de l'étape 6 sont obtenues à partir de la géométrie à l'équilibre du haut niveau et des déplacements selon les modes normaux du haut niveau.

Nous avons testé plusieurs méthodes hybrides en prenant B2PLYP/def2-TZVPP comme haut niveau et plusieurs méthode semi-empiriques ou empiriques comme bas niveau. Les résultats obtenus avec cette méthode sont discutés dans notre article [18]. Les résultats sont résumés dans la figure 3.3. Le champ de force MMFF94 est la méthode de bas niveau qui donne les meilleurs résultats.

3.2.3 VSCF

Une autre méthode pour prendre en compte l'anharmonicité est la méthode VSCF [19] ou *vibrational self consistent field* en anglais. Les vibrations sont traités à l'aide d'un SCF. Le principe de cette méthode est le même que pour la méthode Hartree-Fock. Les modes normaux vont évoluer dans le potentiel moyen généré par les autres modes normaux.

Comme l'approximation harmonique, la fonction d'onde nucléaire d'essai est prise comme le produit des fonctions d'onde à un mode.

$$\Psi = \prod_i \Psi_i(Q_i) \quad (3.61)$$

Le principe variationnel mène au résultat suivant :

$$\left[-\frac{1}{2} \sum_i \frac{\partial^2}{\partial Q_i^2} + \bar{V}_i(Q_i) \right] \Psi_i(Q_i) = \epsilon_i \Psi_i(Q_i) \quad (3.62)$$

$\bar{V}_i(Q_i)$ est le potentiel effectif moyen pour le mode Q_i . Il vaut :

$$\bar{V}_i(Q_i) = \left\langle \prod_{j \neq i}^N \Psi_j(Q_j) \left| V(Q_1, \dots, Q_N) \right| \prod_{j \neq i}^N \Psi_j(Q_j) \right\rangle \quad (3.63)$$

Ainsi l'énergie pour un état n est :

$$E^{(n)} = \sum_i^N \epsilon_i^{(n)} + (N-1) \left\langle \prod_{j \neq i}^N \Psi_j^{(n)}(Q_j) \left| V(Q_1, \dots, Q_N) \right| \prod_{j \neq i}^N \Psi_j^{(n)}(Q_j) \right\rangle \quad (3.64)$$

Le potentiel effectif $\bar{V}_i(Q_i)$ fait intervenir des intégrales à plusieurs dimensions via la forme du potentiel $V(Q_1, \dots, Q_N)$. Ce terme est alors difficile à calculer.

Le potentiel $V(Q_1, \dots, Q_N)$ est généralement pris sous la forme d'un polynôme à l'ordre 4. On appelle ce type de potentiel un champ de force quartique. Pour alléger les calculs sans perdre en précision, on ne calcule pas tous les termes quartiques. Seuls les termes $Q_i^2 Q_j^2$ sont effectivement calculés.

Une autre approche est de prendre le potentiel $V(Q_1, \dots, Q_N)$ comme un potentiel par paires (*pair wise potential*).

$$\begin{aligned} V(Q_1, \dots, Q_N) &= \sum_i^N V_i^{\text{diag}}(Q_i) + \sum_j \sum_{i>j} W_{ij}^{\text{coup}}(Q_i, Q_j) \\ V_i^{\text{diag}}(Q_i) &= V(0, \dots, Q_i, \dots, 0) \\ W_{ij}^{\text{coup}}(Q_i, Q_j) &= V(0, \dots, Q_i, \dots, Q_j, \dots, 0) - V_i^{\text{diag}}(Q_i) - V_j^{\text{diag}}(Q_j) \end{aligned} \quad (3.65)$$

Les termes $V_i^{\text{diag}}(Q_i)$ et $W_{ij}^{\text{coup}}(Q_i, Q_j)$ sont directement calculés par la méthode de calcul choisie (DFT, HF, etc).

La méthode VSCF est basée sur la méthode de champ moyen et par conséquent hérite de ses faiblesses : le couplage entre les modes de vibrations Q_i est moyenné. Il existe des méthodes post-VSCF pour améliorer le couplage entre les modes normaux.

VSCF-PT2

Pour corriger le couplage entre les modes normaux dans la méthode VSCF on peut faire appel à la théorie de la perturbation. On considère que l'hamiltonien VSCF est très proche de l'hamiltonien exact et que donc la différence entre les deux étant très faible sera traitée comme une perturbation.

$$\begin{aligned}
H &= \sum_i -\frac{1}{2} \frac{\partial^2}{\partial Q_i^2} + V(Q_1, \dots, Q_N) = H^{VSCF} - \sum_i \bar{V}_i(Q_i) + V(Q_1, \dots, Q_N) \\
H &= H^{VSCF} + \Delta V(Q_1, \dots, Q_N) \\
\text{avec } H^{VSCF} &= \sum_i -\frac{1}{2} \frac{\partial^2}{\partial Q_i^2} + \sum_i \bar{V}_i(Q_i)
\end{aligned} \tag{3.66}$$

Où la différence de potentiel $\Delta V(Q_1, \dots, Q_N)$ est la perturbation et vaut :

$$\Delta V(Q_1, \dots, Q_N) = V(Q_1, \dots, Q_N) - \sum_i \bar{V}_i(Q_i) \tag{3.67}$$

Le déroulement de la théorie de la perturbation nous donne que l'énergie de VSCF-PT2 est :

$$E_{PT2}^{(n)} = E_{VSCF}^{(n)} + \sum_{m \neq n} \frac{\left| \left\langle \prod_j^N \Psi_j^{(n)}(Q_j) \middle| \Delta V \middle| \prod_j^N \Psi_j^{(m)}(Q_j) \right\rangle \right|^2}{E_0^{(n)} - E_0^{(m)}} \tag{3.68}$$

L'énergie $E_0^{(n)}$ vaut

$$E_0^{(n)} = \sum_j^N \epsilon_j^{(m),n} \tag{3.69}$$

où $\epsilon_j^{(m),n}$ est calculée par $\left[-\frac{1}{2} \frac{\partial^2}{\partial Q_j^2} + \bar{V}_j^{(m)}(Q_j) \right] \Psi_j^{(n)}(Q_j) = \epsilon_j^{(m),n} \Psi_j^{(n)}(Q_j)$

Pour $E_0^{(m)}$ c'est la même chose en changeant les indices m et n .

La méthode VSCF-PT2 est aussi appelée CC-VSCF pour *correlation-consistent* VSCF. L'introduction du terme perturbatif $E_{PT2}^{(n)}$ ajoute de la corrélation entre les modes qui n'était présente que par le champ moyen dans la méthode VSCF. La formule CC-VSCF est valide pour des états non dégénérés. Pour des états dégénérés ou quasi-dégénérés, le couplage est très fort entre ces états et il faut utiliser d'autres pour les décrire correctement.

VCI

La méthode développée pour traiter les fortes corrélations entre les modes vibrationnels est la méthode VCI pour *Vibrational Configuration Interaction*. C'est l'équivalent de la méthode CI (interactions de configuration) pour la structure électronique pour les modes de vibrations.

La fonction d'onde VCI est une combinaison linéaire des fonctions d'ondes des états vibrationnels.

$$\Psi^{\text{VCI}} = \sum_L C_L \Psi^{(L)} \quad (3.70)$$

$\Psi^{(L)}$ est la fonction d'onde pour le mode vibrationnel L . Le problème se résout variationnellement, on obtient alors une relation similaire à celle trouvée pour la méthode CI pour le problème électronique :

$$\mathbf{H}\mathbf{C} = \mathbf{E}\mathbf{S}\mathbf{C} \quad (3.71)$$

Où $\mathbf{H}$ est la matrice des éléments $\langle \Psi^{(L)} | H | \Psi^{(L')} \rangle$ avec H l'hamiltonien constitué de la somme des hamiltoniens à un mode, $\mathbf{S}$ est la matrice de recouvrement entre les fonctions d'ondes des états vibrationnels. $\mathbf{C}$ et $\mathbf{E}$ sont les vecteurs des coefficients de la combinaison linéaire et des énergies des états, respectivement.

De même que pour la méthode CI, on ne peut pas prendre en compte toutes les excitations possibles sinon le coût du calcul devient prohibitif. Il faut choisir les types d'excitations vibrationnelles que l'on souhaite intégrer dans le calcul.

VSCF hybride

Une méthode VSCF dite hybride a été développée afin de réduire le temps de calcul des vibrations sans trop réduire la précision. Le potentiel $V(Q_1, \dots, Q_N)$ est évalué par une méthode dite de bas niveau telle que PM3 [20, 21], Hartree-Fock [22], etc. Cependant le potentiel bas niveau (BN) n'est pas pris tel quel mais il est ajusté (BNA) sur des calculs de plus haut niveau (HN) tels que DFT, post Hartree-Fock.

$$V^{\text{BNA}}(Q_1, \dots, Q_N) = V^{\text{BN}}(\lambda_1 Q_1, \dots, \lambda_N Q_N) \quad (3.72)$$

$V^{\text{BNA}}(Q_1, \dots, Q_N)$ est le potentiel ajusté calculé à partir du potentiel bas niveau. $V^{\text{BN}}(Q_1, \dots, Q_N)$ est le potentiel bas niveau. Les modes normaux obtenus par le bas niveau sont corrigés par un facteur multiplicatif calculés à partir des fréquences du haut niveau tel que :

$$\lambda_i = \frac{\omega_i^{\text{HN}}}{\omega_i^{\text{BN}}} \quad (3.73)$$

L'association des facteurs λ aux modes normaux repose sur le recouvrement entre les modes normaux du bas et du haut niveau.

3.2.4 Dynamique moléculaire

En utilisant la dynamique moléculaire il est possible de produire un spectre dans lequel les effets anharmoniques ont été pris en compte [23]. Pour cela il faut récupérer la valeur du dipôle à chaque instant au cours de la dynamique. Le spectre est alors calculé à partir de la fonction d'autocorrélation du dipôle électrique $\vec{\mu}(t)$ de la molécule :

$$I(\omega) = \frac{2\pi\omega^2}{3cVkT} \int_{-\infty}^{\infty} dt \langle \vec{\mu}(t) \cdot \vec{\mu}(0) \rangle e^{i\omega t} \quad (3.74)$$

ω est la fréquence, V le volume, T la température et c la vitesse de la lumière dans le vide. À noter que l'on peut utiliser un filtre $f(t)$ (par exemple une fonction gaussienne) pour

éliminer le bruit numérique provenant du fait que la transformée de Fourier s'effectue sur un domaine fini [23]. Le terme $\langle \vec{\mu}(t) \cdot \vec{\mu}(0) \rangle \rightarrow \langle \vec{\mu}(t) \cdot \vec{\mu}(0) \rangle f(t)$ Dans cette méthode il n'y a pas de double approximation harmonique (mécanique et électrique) ni de traitement perturbatif ou variationnel pour calculer les effets anharmoniques. Le potentiel est celui de la méthode de chimie quantique choisie ainsi l'anharmonicité du potentiel est entièrement présente. Ce potentiel est évalué lorsque les atomes se déplacent ainsi si le potentiel est anharmonique alors les effets sont directement pris en compte. Toutes les transitions (fondamentales, overtones, combinaisons de bandes) sont alors obtenues par la dynamique moléculaire. Comme on se déplace le long du potentiel lors de la dynamique, il est possible d'obtenir d'autres conformations. Ainsi plusieurs conformations peuvent contribuer au spectre vibrationnel grâce à la dynamique moléculaire ce qui n'est pas possible avec les méthodes statiques.

De plus, la dynamique permet d'inclure les effets de la température sur le spectre vibrationnel. Pour les méthodes statiques (double approximation harmoniques, VSCF, VPT2) le spectre vibrationnel est toujours obtenu pour une température de 0 K.

Actuellement, la méthode de chimie quantique la plus utilisée est la fonctionnelle GGA BLYP. Les méthodes hybrides-GGA comme la fonctionnelle B3LYP ne sont pas encore utilisables pour le moment. Elles sont trop coûteuses d'un point de vue calculatoire. Ainsi la précision de la dynamique moléculaire pour calculer des spectres infrarouges est limitée par la qualité de BLYP.

3.3 Références

- [1] J. A. Pople, H. B. Schlegel, R. Krishnan, D. J. Defrees, J. S. Binkley, M. J. Frisch, R. A. Whiteside, R. F. Hout, and W. J. Hehre. Molecular orbital studies of vibrational frequencies. *Int. J. Quantum Chem.*, 20(S15) :269–278, 1981.
- [2] Anthony P. Scott and Leo Radom. Harmonic Vibrational Frequencies : An Evaluation of Hartree-Fock, Møller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.*, 100(41) :16502–16513, 1996.
- [3] Pankaj Sinha, Scott E. Boesch, Changming Gu, Ralph A. Wheeler, and Angela K. Wilson. Harmonic Vibrational Frequencies : Scaling Factors for HF, B3LYP, and MP2 Methods in Combination with Correlation Consistent Basis Sets. *J. Phys. Chem. A*, 108(42) :9213–9217, 2004.
- [4] Vincenzo Barone. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.*, 122 :014108, 2005.
- [5] Andreana M. Rosnik and William F. Polik. VPT2+K spectroscopic constants and matrix elements of the transformed vibrational Hamiltonian of a polyatomic molecule with resonances using Van Vleck perturbation theory. *Mol. Phys.*, 112(2) :261–300, 2014.
- [6] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Cari-

-
- cato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision D.01, 2009. Gaussian Inc. Wallingford CT 2009.
- [7] Jan M. L. Martin, Timothy J. Lee, Peter R. Taylor, and Jean-Pierre François. The anharmonic force field of ethylene, C_2H_4 , by means of accurate ab initio calculations. *J. Chem. Phys.*, 103 :2589–2602, 1995.
 - [8] Kathleen M. Kuhler, Donald G. Truhlar, and Alan D. Isaacson. General method for removing resonance singularities in quantum mechanical perturbation theory. *J. Chem. Phys.*, 104(12) :4664–4671, 1996.
 - [9] Julien Bloino, Malgorzata Biczysko, and Vincenzo Barone. General Perturbative Approach for Spectroscopy, Thermodynamics, and Kinetics : Methodological Background and Benchmark Studies. *J. Chem. Theory Comput.*, 8(3) :1015–1036, 2012.
 - [10] Vincenzo Barone, Malgorzata Biczysko, and Julien Bloino. Fully anharmonic IR and Raman spectra of medium-size molecular systems : accuracy and interpretation. *Phys. Chem. Chem. Phys.*, 16 :1759–1787, 2014.
 - [11] Julien Bloino. A VPT2 Route to Near-Infrared Spectroscopy : The Role of Mechanical and Electrical Anharmonicity. *J. Phys. Chem. A*, 119(21) :5269–5287, 2015.
 - [12] Kiyoshi Yagi, Kimihiko Hirao, Tetsuya Taketsugu, Michael W. Schmidt, and Mark S. Gordon. Ab initio vibrational state calculations with a quartic force field : Applications to H_2CO , C_2H_4 , CH_3OH , CH_3CCH , and C_6H_6 . *J. Chem. Phys.*, 121 :1383–1389, 2004.
 - [13] Malgorzata Biczysko, Pawel Panek, Giovanni Scalmani, Julien Bloino, and Vincenzo Barone. Harmonic and Anharmonic Vibrational Frequency Calculations with the Double-Hybrid B2PLYP Method : Analytic Second Derivatives and Benchmark Studies. *J. Chem. Theory Comput.*, 6(7) :2115–2125, 2010.
 - [14] Cristina Puzzarini, Malgorzata Biczysko, and Vincenzo Barone. Accurate Harmonic/Anharmonic Vibrational Frequencies for Open-Shell Systems : Performances of the B3LYP/N07D Model for Semirigid Free Radicals Benchmarked by CCSD(T) Computations. *J. Chem. Theory Comput.*, 6(3) :828–838, 2010. PMID : 26613310.
 - [15] Vincenzo Barone, Malgorzata Biczysko, Julien Bloino, and Cristina Puzzarini. Accurate structure, thermodynamic and spectroscopic parameters from CC and CC/DFT schemes : the challenge of the conformational equilibrium in glycine. *Phys. Chem. Chem. Phys.*, 15 :10094–10111, 2013.
 - [16] Vincenzo Barone, Malgorzata Biczysko, Julien Bloino, Paola Cimino, Emanuele Penedocchio, and Cristina Puzzarini. CC/DFT Route toward Accurate Structures and

Spectroscopic Features for Observed and Elusive Conformers of Flexible Molecules : Pyruvic Acid as a Case Study. *J. Chem. Theory Comput.*, 11(9) :4342–4363, 2015.

- [17] Loïc Barnes, Baptiste Schindler, Abdul-Rahman Allouche, Daniel Simon, Stephane Chambert, Jos Oomens, and Isabelle Compagnon. Anharmonic simulations of the vibrational spectrum of sulfated compounds : application to the glycosaminoglycan fragment glucosamine 6-sulfate. *Phys. Chem. Chem. Phys.*, 17, 2015.
- [18] Loïc Barnes, Baptiste Schindler, Isabelle Compagnon, and Abdul-Rahman Allouche. Fast and accurate hybrid QM//MM approach for computing anharmonic corrections to vibrational frequencies. *J. Mol. Model.*, 22(11) :285, 2016.
- [19] Joel M. Bowman. Self-consistent field energies and wavefunctions for coupled oscillators. *J. Chem. Phys.*, 68 :608–610, 1978.
- [20] Adeyemi A. Adesokan and R. B. Gerber. Anharmonic vibrational spectroscopy calculations for proton-bound amino acid dimers. *J. Phys. Chem. A*, 113(10) :1905–1912, 2009.
- [21] Brina Brauer, Galina M. Chaban, and R. Benny Gerber. Spectroscopically-tested, improved, semi-empirical potentials for biological molecules : Calculations for glycine, alanine and proline. *Phys. Chem. Chem. Phys.*, 6 :2543–2556, 2004.
- [22] Tapta Kanchan Roy, Rahul Sharma, and R. Benny Gerber. First-principles anharmonic quantum calculations for peptide spectroscopy : VSCF calculations and comparison with experiments. *Phys. Chem. Chem. Phys.*, 18 :1607–1614, 2016.
- [23] Marie-Pierre Gageot. Theoretical spectroscopy of floppy peptides at room temperature. A DFTMD perspective : gas and aqueous phase. *Phys. Chem. Chem. Phys.*, 12 :3336–3359, 2010.

Troisième partie

Applications

Introduction

Notre travail a pour objectif de fournir une méthode pour permettre le séquençage des oligosaccharides, qui sont des biopolymères jouant un rôle clé dans le domaine de la santé comme dans certains domaines industriels tel que l'énergie ou l'alimentation. Les méthodes utilisées pour séquencer les protéines telle que la spectrométrie de masse ne sont pas utilisables dans l'étude des oligosaccharides. En revanche, l'alliance entre la méthode expérimentale de spectroscopie d'action (IRMPD : InfraRed MultiPhoton Dissociation) et les calculs théoriques est une méthode prometteuse pour le séquençage des oligosaccharides.

Dans cette partie, nous allons présenter les résultats obtenus sur les monosaccharides et les disaccharides. Les monosaccharides ont été choisis pour leur abondance dans la nature ou leur rôle important, tel que des monosaccharides avec un groupement sulfate qui intervient dans de nombreuses fonctions biologiques. Nous avons choisis quelques disaccharides présentant des liaisons osidiques $1 \rightarrow 4$ et $1 \rightarrow 6$. Cette partie sera organisée autour des articles publiés ou en cours de rédaction pour les monosaccharides et de résultats non publiés sur les disaccharides. Les articles publiés ont été rédigés de manière collective avec les autres acteurs du projet. J'ai rédigé les articles correspondant au chapitre 6 (publié) et au chapitre 7 (non publié). Tous les résultats théoriques (recherche de conformation, spectres théoriques, benchmark) présentés dans les chapitres suivants sont le fruit de mon travail.

Les premiers spectres vibrationnels étaient obtenus dans notre laboratoire avec la fonctionnelle B3LYP et la base 6-311+G(d). Cette méthode a permis d'élucider la structure de la glucosamine-6 sulfate et de la glucosamine-6 phosphate [1] en étudiant la zone entre 3000 et 3700 cm^{-1} , appelée zone OPO. Cependant, en analysant la zone comprise entre 500 et 1700 cm^{-1} , appelée zone FELIX, les spectres des structures de la glucosamine-6 sulfate qui avaient été validées pour la zone OPO ne reproduisaient pas le spectre expérimental dans cette zone. Nous avons alors évalué plusieurs fonctionnelles et différentes bases afin de reproduire des modes de vibration (élongation C-H, élongation S-O, élongation O-H, etc) dont les fréquences expérimentales sont connues. Le meilleur couple fonctionnelle et base pour reproduire les fréquences expérimentales est CAM-B3LYP/6-311++G(2df,2pd). La fonctionnelle B3LYP avec une plus grande base n'est pas suffisante pour reproduire le spectre de la zone FELIX. Nous avons par la suite toujours conservé cette méthode pour obtenir les spectres infrarouges des structures de basse énergie. Cette étude a permis d'obtenir de nouveaux scaling factor pour les différentes fonctionnelles et de montrer l'importance de la prise en compte explicite de l'anharmonicité pour reproduire le spectre expérimental dans la zone FELIX. Ces résultats seront présentés dans le chapitre 4.

En appliquant notre méthode (allier le calcul des spectres théoriques à l'IRMPD) nous avons dans le chapitre 5, poursuivi notre étude sur des monosaccharides substitués par un groupement sulfate. En particulier nous avons étudié l'isomérisation de position du groupement sulfate. Des signatures distinctes ont été mises en évidence pour la glucosamine-6 sulfate et pour la glucosamine-3 sulfate toutes les deux protonées. Or ces saccharides sont présents dans les tissus biologiques dans un état déprotoné. L'étude de ces saccharides déprotonés a alors été menée et nous avons aussi mis en évidence des signatures qui dépendent de la position du groupement sulfate. Même si du point de vue théorique les calculs statiques reproduisent mal l'élargissement des bandes expérimentales néanmoins ils permettent de distinguer les deux isomères.

Nous nous sommes intéressés à d'autres monosaccharides qui sont cette fois substitués par un groupement N-acétyl. Nous les avons étudiés dans leur état protoné or ce groupement N-acétyl présente deux sites de protonation potentiels. Il a alors fallu étudier sur quel site s'effectuait la protonation (voir chapitre 6). Nous avons montré que la protonation s'effectue sur l'oxygène du groupement carboxylique et que le proton peut être fortement partagé avec un groupement hydroxyle selon les conformations (certaines conformations ne forment pas de liaisons hydrogènes entre le proton et les autres groupements). Le plus souvent les spectres expérimentaux sont obtenus pour un mélange d'anomères d'un même saccharide. Les résultats théoriques ont montré que les contributions des anomères au spectre expérimental étaient différentes. Pour vérifier cela nous avons d'abord étudié les saccharides en ne considérant qu'un seul anomère. On peut séparer les anomères en les modifiant chimiquement, c'est-à-dire en effectuant une méthylation. Cela consiste à remplacer l'hydrogène du groupement hydroxyle anomérique par un groupement méthyl. De cette manière on peut isoler les anomères et à l'aide de l'IRMPD faire un spectre sur un seul anomère (qui est donc modifié chimiquement). Cette approche nous a permis de confirmer que les signatures anomériques sont bien différentes.

Nous avons effectué un autre travail sur le mélange d'anomères. Nous avons étudié 3 monosaccharides présentant un groupement amine protoné (voir chapitre 7). L'étude des structures et des signatures vibrationnelles a montré une complexe hétérogénéité des structures pour reproduire les bandes expérimentales. C'est-à-dire qu'en plus des anomères, plusieurs conformations d'un même anomère peuvent être présentes dans le mélange. L'information sur les structures des constituants du mélange est importante mais connaître le pourcentage de ces derniers dans le mélange l'est tout autant. Nous avons exploré une approche pour obtenir des informations sur les pourcentages des constituants du mélange en utilisant les spectres IRMPD et les spectres théoriques. Les pourcentages obtenus ont été comparés à des données de RMN et les résultats sont très encourageants. C'est la première fois que des données obtenues par IRMPD sont utilisées pour obtenir le pourcentage des constituants présent dans l'échantillon. Nous avons montré que les mesures IRMPD permettent de remonter à la configuration du mélange en solution.

Dans le chapitre 8, nous présentons les résultats obtenus sur la caractérisation de la liaison osidique de certains disaccharides. Il y a deux approches possibles pour étudier la liaison osidique, soit on se base sur la signature directe de la liaison qui correspond aux modes de vibration C-O-C soit on se base sur la signature indirecte, autrement dit les modes d'élongation NH et OH. Nous montrerons sur un exemple que la première approche ne permet pas de caractériser la liaison osidique. Enfin, nous montrerons que la seconde approche permet d'obtenir des informations sur la liaison osidique dans les cas où la résolution expérimentale est suffisante.

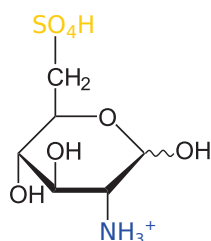
Bibliographie

- [1] Baptiste Schindler, Janhavi Joshi, Abdul-Rahman Allouche, Daniel Simon, Stephane Chambert, Vincent Brites, Marie-Pierre Gageot, and Isabelle Compagnon. Distinguishing isobaric phosphated and sulfated carbohydrates by coupling of mass spectrometry with gas phase vibrational spectroscopy. *Phys. Chem. Chem. Phys.*, 16 :22131–22138, 2014.

Anharmonic simulations of the vibrational spectrum of sulfated compounds : application to the glycosaminoglycan fragment glucosamine 6-sulfate[†]

Highlights

- Finding the best couple functional basis for predicting sulfate frequencies of a set of benchmark species
- Exploration of the performance of GVPT2 calculation
- Hybrid strategy to reduce computational cost of GVPT2 calculation



glucosamine-6 sulfate :
GlcN-6S

[†] Published, DOI : 10.1039/C5CP02079D



Cite this: *Phys. Chem. Chem. Phys.*,
2015, 17, 25705

Received 9th April 2015,
Accepted 22nd June 2015

DOI: 10.1039/c5cp02079d

www.rsc.org/pccp

Anharmonic simulations of the vibrational spectrum of sulfated compounds: application to the glycosaminoglycan fragment glucosamine 6-sulfate†

Loïc Barnes,^{abc} Baptiste Schindler,^{abc} Abdul-Rahman Allouche,^{*abc} Daniel Simon,^{abc}
Stéphane Chambert,^{abd} Jos Oomens^{ef} and Isabelle Compagnon^{abcg}

Mid-infrared spectroscopy coupled with mass spectrometry is an appealing tool for the sequencing and structural elucidation of functional modifications in biopolymers, as it offers direct spectroscopic identification of the functionality where the traditional mass spectrometric approach is insufficient. Whereas the gas phase vibrational spectroscopy of peptides (and to a lesser extent saccharides) has been widely investigated, sulfation has attracted much less attention, despite its prevalence in natural polymers. The simulation of the vibrational spectra of such functionalized compounds is however notoriously challenging, which impairs the interpretation of spectroscopic data in terms of structure. Driven by a striking case of such a failure for a sulfated glycosaminoglycan fragment, we elaborate on an original hybrid GVPT2 anharmonic approach. This strategy offers a significantly improved accuracy in the description of the sulfate modes, without the recourse to empirical scaling factors, and with a greatly reduced computational cost which is otherwise prohibitive for molecules of this size. Alternatively, we propose a selection of reasonably accurate harmonic methods with adequate scaling factors optimized on a set of benchmark compounds.

Introduction

Sulfation and phosphorylation are among the most common post-translational modifications of biomolecules such as proteins and they control their biological function.^{1–4} Yet, their characterization remains a challenge in proteomics due to their lability, versatility and isobaricity. Indeed, fragile functional modifications tend to be lost upon traditional tandem mass spectrometry analysis, which is virtually blind to their structure. Functionalization is also a burning question in glycosciences, which have recently risen as a strategic priority.⁵ Elucidation of the structural features of saccharides is highly challenging due to their unique complexity among biomolecules, and the development of a suite of analytical tools analogous to these available for proteomics is timely. Namely, an accurate description of an oligosaccharide includes: the monosaccharide content, which is complicated by the frequent occurrence of isobaric monomers, the identification of the nature and position of functional modifications, and the branched structures. While mass spectrometry (MS) has proven its use as an accurate sequencing tool for other biopolymers (proteins, DNA), it is not straightforwardly transferred to the saccharides family due to their equivocal MS and MS/MS signatures.^{6–9} To date, the full development of glycomics is impaired by this ambiguity, and orthogonal structural tools are greatly sought after. Among them, the coupling of MS with ion mobility or laser spectroscopy appears to develop as very promising hyphenated methods, each offering direct structural information, with complementary resolutions. While the overall shape can be derived from ion mobility cross section measurements, laser spectroscopy combined with *ab initio* calculations can provide details of the local chemical arrangement. Recent examples of electronic spectroscopy,¹⁰ and vibrational spectroscopy¹¹ coupled with mass spectrometry have demonstrated the relevance of gas phase spectroscopy for the structural characterization of isolated, mass selected carbohydrates ions.

In this context, we are developing a suite of experimental and theoretical tools for the structural characterization of

^a Université de Lyon, F-69622, Lyon, France. E-mail: isabelle.compagnon@univ-lyon1.fr

^b Université Lyon 1, Villeurbanne, France

^c Institut Lumière Matière, UMR5306 Université Lyon 1-CNRS,
Université de Lyon 69622 Villeurbanne Cedex, France

^d Laboratoire de Chimie Organique et Bioorganique, INSA Lyon, CNRS, UMR5246,
ICBMS, Bât. J. Verne, 20 Avenue A. Einstein, 69621 Villeurbanne Cedex, France

^e Radboud University, Institute for Molecules and Materials, FELIX Laboratory,
Toernooiveld 7, 6525ED Nijmegen, The Netherlands

^f Van't Hoff Institute for Molecular Sciences, University of Amsterdam,
Science Park 904, 1098XH Amsterdam, The Netherlands

^g Institut Universitaire de France IUF, 103 Blvd St Michel, 75005 Paris, France

† Electronic supplementary information (ESI) available: Experimental methods; cpu time for dimethyl sulfate; experimental frequencies and theoretical variation; geometry of the benchmark species; vibrational spectra of the benchmark species; anharmonic PBE0 spectrum of dimethyl sulfate; application of the hybrid method to glucosamine 6-phosphate. See DOI: 10.1039/c5cp02079d

glycosaminoglycans (GAGs), a major class of sulfated carbohydrate polymers expressed in extracellular matrices and on cell surfaces. GAGs are involved in diverse processes essential to biomedical research (e.g. regulation of coagulation,¹² development¹³ or cancer¹⁴). They are characterized by a large variety of structures originating from the different carbohydrate constituting monomers, the different possible linkages between them and also their modification patterns, including sulfation, which is essential to their biological activities. The understanding of those patterns is thus of considerable biological interest.¹⁵ We have recently used IR-MPD (InfraRed-Multiple Photon Dissociation) spectroscopy to distinguish between isobaric sulfated and phosphorylated monosaccharides isolated in an ion trap.¹⁶ In this previous work, we probed the OH and NH stretching vibrations accessible in the mid-IR region with a table-top tunable IR laser covering the frequency range between 2900 and 3700 cm^{-1} . Despite their identical MS and MS/MS signatures we were able to unambiguously – although somewhat indirectly – identify the nature of the functional modification (sulfate *vs.* phosphate), and the conformation.

In order to obtain a direct signature of the sulfate pattern in the fingerprint region, and more generally to heavy atoms-containing functional modifications, it is desirable to extend the IR-MPD approach toward lower frequencies.^{17,18} From an experimental point of view, this is feasible using Free Electron Lasers. On the other hand, the interpretation of such experimental data raises acute theoretical questions. Firstly, the widely used harmonic approximation of *ab initio* electronic potentials typically yields overestimated simulated vibrational frequencies. While it is generally accepted to apply an empirical scaling factor to reduce the calculated frequencies and match experiments, the use of several empirical scaling factors to account for different families of vibrations presenting distinct anharmonic behavior is strongly debated. As we intend to investigate a wide IR spectroscopic range – including as various vibrational modes as stretchings, bendings and torsions involving light and heavy atoms – this issue cannot be ignored. Furthermore, we face a dramatic failure of traditional harmonic simulations in the case of our model glycosaminoglycan component 6-O-sulfated glucosamine^{19,20} in the fingerprint region, as described below.

Glucosamine 6-sulfate: a case of failure of harmonic simulations

It is noteworthy that the DFT B3LYP/6-311+G(d) harmonic frequencies, which previously supported an unambiguous structural assignment of glucosamine 6-sulfate in the high frequency range, utterly fail to account for its IR-MPD spectrum measured in the fingerprint region (Experimental detail in ESI[†]), as seen in Fig. 1.

A most remarkable discrepancy is the absence of photo-fragmentation between 1200 and 1420 cm^{-1} , while an intense SO_2 asymmetric stretching mode is predicted at 1337 cm^{-1} . Another inconsistency is the intense feature measured at 1180 cm^{-1} , which is not accounted for in the simulations. Finally, the region below 850 cm^{-1} , with only two weak bands measured at 560 cm^{-1} and 780 cm^{-1} is poorly reproduced.

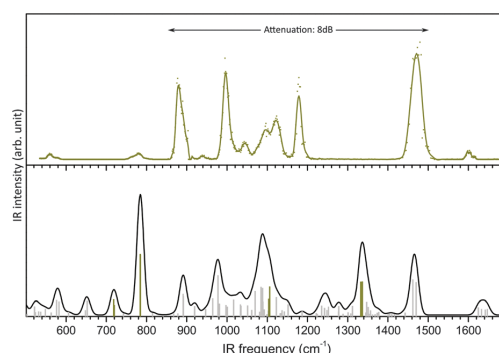


Fig. 1 Mid-IR spectra of glucosamine 6-sulfate. Top panel: measured IR-MPD spectrum. Lower panel: B3LYP/6-311+G(d) harmonic frequencies scaled by 0.965 (sulfate modes are highlighted in dark yellow) and convoluted spectrum using a fwhm of 20 cm^{-1} (line).

This constitutes a striking case of impossible structural assignment, which is not due to an insufficient conformational exploration (the conformation was indeed established with confidence in ref. 16), but instead because of a thoroughly misleading simulated vibration pattern. Similar issues were encountered in the case of phosphorylated compounds, which remain to date a challenge for theoreticians.²¹ This suggests that the theoretical treatment of sulfation and phosphorylation is a general problem which critically impairs the structural characterization of a broad range of functionalized biomolecules. In this context, gas phase vibrational spectroscopy of peptides phosphorylation (and to a lesser extent saccharides phosphorylation) is widely investigated. On the other hand, harmonic and anharmonic simulations of the vibrational spectra of sulfate clusters related to aerosol seeding have attracted a great deal of attention.^{22,23} However, analogous studies addressing sulfated biomolecules are scarce despite their considerable biological relevance. Hence we believe that a strategy for the reliable simulation of the vibrational frequencies of sulfated biomolecules is of general interest.

Outline

In order to tackle the inconsistencies raised by sulfated vibrational frequencies, we report an extensive investigation of the accuracy of popular DFT functionals and basis sets for the prediction of the mid-IR harmonic frequencies of a series of benchmark species and we reckon suitable scaling factors. Then, the merits *vs.* computational costs of GVPT2 anharmonic corrections is assessed. The great advantage of the GVPT2 is that the anharmonic effect is taken into account without any empirical parameter. However, as the computational cost of the GVPT2 approach becomes prohibitive for larger species, we devised an original hybrid strategy, consisting of a large basis set for the calculation of the harmonic frequencies, combined with a smaller basis set for the calculations of the cubic and quartic terms of the potential. This method offers a significant reduction of the computational cost, bringing anharmonic correction to an accessible level for biomolecules without loss





of precision and without empirical parameters. The performance of our hybrid method is emphasized in the first section for the case of glucosamine 6-sulfate, for which we show that the issue of the miscalculated vibrational pattern can be resolved. In the second section the elaboration of the hybrid method and a “how-to” are described in detail for facile implementation by others. Should one favor time-saving harmonic calculations, we also propose suitable combinations of functional and basis sets for sulfated compounds, together with appropriate scaling factors.

Anharmonic simulations of glucosamine 6-sulfate

Influence of the functional and basis set on the harmonic frequencies

As shown in Fig. 1 and 2a, scaled harmonic B3LYP/6-311+G(d) simulations fail to account for the spectrum measured in the fingerprint region (see ESI† for Experimental details). This is particularly critical for the sulfate modes (dark yellow bars in Fig. 1) which do not match any experimental feature.

First, in order to illustrate the influence of the choice of functional and basis set on the accuracy of the sulfate pattern, a series of harmonic calculations was performed for the conformer present in the experiment (a 4C_1 chair, as previously established¹⁶), and scaled with our best scaling factors (reported in section Theoretical methods and How-to). The scaled harmonic spectra

obtained with the widely used B3LYP functional and the most accurate CAM-B3LYP functional (see justification in section Theoretical methods and How-to); and a modest and a large basis sets (6-311++G** and 6-311++G(2df,2pd), respectively) are shown in Fig. 2b. Note that only the α -anomer is shown for clarity.†

With B3LYP functional, increasing the size of the basis set results in a blueshift of the sulfate modes (SO(C) stretch, SO(H) stretch and SO₂ asymmetric stretch) from 760, 840 and 1420 cm⁻¹ to 770, 860 and 1430 cm⁻¹, respectively. Using the smaller basis set 6-311+G(d) and a standard scaling factor of 0.965, these modes were previously predicted at 720, 785 and 1337 cm⁻¹, respectively and were thus beyond recognition (Fig. 2a). With the blueshift trend, it becomes plausible to associate the two red sulfate modes with the experimental features at 780 and 880 cm⁻¹ (Fig. 2d). An other recognizable region is the NH₃ scissor doublet at 1600 cm⁻¹ in the IR-MPD spectrum, which is poorly reproduced. Overall, it is striking that B3LYP systematically underestimates the sulfate frequencies. This anomalous behavior is somewhat attenuated by a counter-intuitive scaling factor greater than 1.0, which in turn deteriorates other ordinary overestimated frequencies (such as the NH₃ pattern in this example). This oddity is not present in CAM-B3LYP simulations, which are consequently scaled with more traditional scaling factors smaller than 1.0. The blueshift of sulfate modes is consistently reinforced. With the larger basis set SO(C) and SO(H) stretches reach near the experimental positions. The SO₂ asymmetric stretch and the NH₃ umbrella mode almost merge at

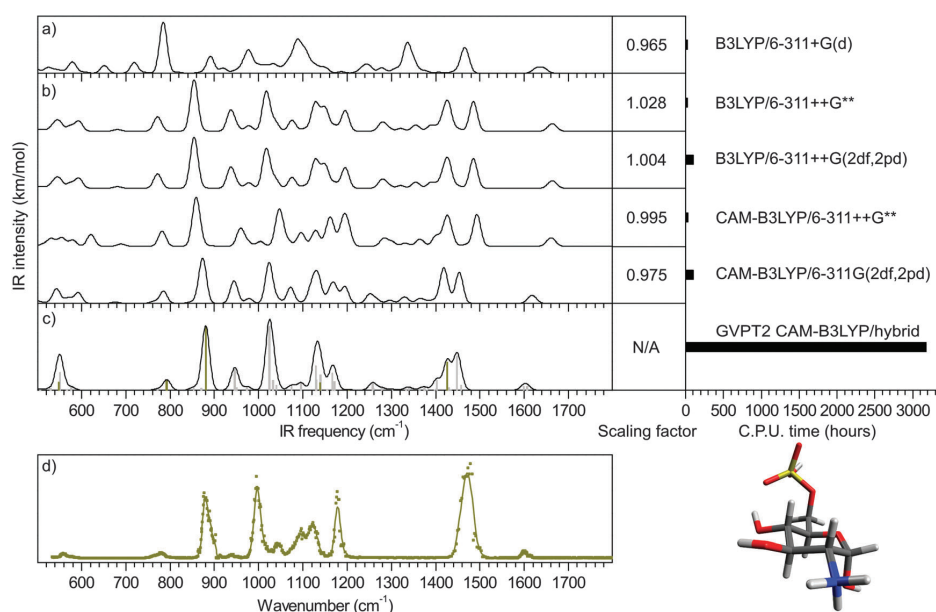


Fig. 2 Mid-IR spectra of α -glucosamine 6-sulfate. (a) Convolved harmonic spectrum calculated with B3LYP/6-311+G(d). Fwhm = 20 cm⁻¹. Standard scaling factor: 0.965. (b) Convolved harmonic spectra calculated with B3LYP/6-311++G**, B3LYP/6-311++G(2df,2pd); CAM-B3LYP/6-311++G** and CAM-B3LYP/6-311++G(2df,2pd). Fwhm = 20 cm⁻¹. The computational times and our best scaling factors are given in the right panel. (c) GVPT2 CAM-B3LYP/hybrid anharmonic frequencies (sulfate modes in dark yellow) and convolved spectrum with fwhm = 20 cm⁻¹ (d) IR-MPD spectrum.

1420 and 1450 cm^{-1} . Between 900 and 1030 cm^{-1} , the two dominant carbohydrate modes predicted at 940 and 1020 cm^{-1} reasonably match the experimental pattern with features at 940 and 1000 cm^{-1} (note that the band measured at 940 cm^{-1} is very weak. This is due to the attenuation applied to record the most intense bands of the spectrum without saturation). In contrast, the 1030–1200 cm^{-1} range, which consists of partly unresolved bands, poorly accounts for the experimental pattern. The low frequency range ($<700 \text{ cm}^{-1}$) is equally unconvincing. The two NH scissor modes are predicted around 1610 cm^{-1} , in good agreement with the experimental value. Finally, it appears that both the change of functional and the use of increased basis set improve the simulation of the sulfate pattern. Not unexpectedly, the improvement of the carbohydrate pattern is more debatable: as a matter of fact, B3LYP is very popular for its excellent performances for organic compounds. In fine, the CAM-B3LYP/6-311++G(2df,2pd) method provides the best simulation of the sulfate and NH scissor modes. The rest of the spectrum is equally satisfying between 900 and 1030 cm^{-1} , and equally disappointing below 700 cm^{-1} and in the 1030–1200 cm^{-1} range regardless of the method.

Anharmonic corrections

Hybrid GVPT2 anharmonic simulation was performed in a second stage (Fig. 2c). It consists in the correction of the CAM-B3LYP/6-311++G(2df,2pd) harmonic spectrum at the GVPT2 CAM-B3LYP/6-311++G** level. With excellent prediction of the SO(C) and SO(H) stretching frequencies and significant alterations of the overall CAM-B3LYP/6-311++G(2df,2pd) harmonic pattern, the hybrid spectrum provides the most refined representation of the experimental data. Firstly, the series of weak bands predicted between 500 and 700 cm^{-1} narrows down to a single feature centered around 550 cm^{-1} , which perfectly matches the single feature measured in this region. As already mentioned, the SO(C) and SO(H) stretching frequencies, predicted at 790 and 880 cm^{-1} , are in excellent agreement with the corresponding experimental bands. The next bands observed at 940 and 1000 cm^{-1} , which do not involve sulfate, are also well accounted for (942 and 1022 cm^{-1}), although not better than with the harmonic approximation. The 1030–1200 cm^{-1} pattern is improved with the detachment of a band at 1170 cm^{-1} from the rest of the unresolved pattern, in good agreement with the experimental feature at 1180 cm^{-1} . The rest of this range is dominated by a feature at 1130 cm^{-1} in both experimental and hybrid spectra but the shapes of the left part of this band mismatch. The hybrid spectrum shows a weak activity between 1200 and 1400 cm^{-1} , which is not observed in the experimental spectrum. As for the low intensity of the band at 940 cm^{-1} , this is reasonably justified by the experimental conditions. The NH_3 umbrella mode and the SO_2 asymmetric stretch have now fully merged into a broad band (fwhm = 40 cm^{-1}) centered around 1435 cm^{-1} , which is associated to the intense band measured at 1470 cm^{-1} with a fwhm of 45 cm^{-1} . The shape of this broad feature is only correctly reproduced by the hybrid method, although its position is underestimated by 35 cm^{-1} which falls slightly off the experimental resolution. Finally, the doublet of NH_3 scissor mode is perfectly reproduced.

To summarize, the best harmonic approximation was obtained with the CAM-B3LYP/6-311++G(2df,2pd) method and the optimized scaling factor 0.975. Using the hybrid anharmonic approach, the IR spectrum is further refined without the recourse to any empirical scaling factor and reaches a close-to-perfect agreement with the experimental data, thus resolving the case of glucosamine 6-sulfate. The computational cost of the hybrid anharmonic corrections was 35 times greater than the harmonic frequencies: 3160 hCPU vs. 89 hCPU. In contrast the full anharmonic simulation at the CAM-B3LYP/6-311++G(2df,2pd) level was impractically long. We could however estimate its computational cost to 18 024 hCPU (extrapolated from the calculation of a reduced number of frequencies). This represent a gain of six, which is not only a quantitative gain, but brings otherwise prohibitive anharmonic corrections to an accessible level.

Theoretical methods and how-to

The hybrid anharmonic approach was elaborated and validated on a set of benchmark systems of modest size (Chart 1). First, we assessed the accuracy of a series of functionals and basis sets for the prediction of sulfate vibrational frequencies. In order to improve the performances of the harmonic approximation, we reckoned pairs of scaling factor minimising the error on the ensemble of mid-IR modes. A cutoff was applied at 2800 cm^{-1} to best account for the high frequency modes (CH, NH and OH stretches) and the fingerprint modes, including the sulfate vibrations. Then, we assessed the precision of the functionals and basis sets for the GVPT2 corrections. Finally we devised a time-saving hybrid GVPT2 approach, thus making anharmonic corrections accessible for larger molecules. The method is fully illustrated on one of the benchmark species (*i.e.* dimethyl sulfate (4), the results obtained for the other benchmark species are shown in ESI†) for facile implementation by the interested reader. In order to verify the versatility of the hybrid method, it was also tested against phosphate frequencies – which are notoriously difficult to simulate – in the case of glucosamine 6-phosphate.

Benchmark systems

The set of reference systems was selected to feature a sulfur atom in a variety of environments relevant to the context of our

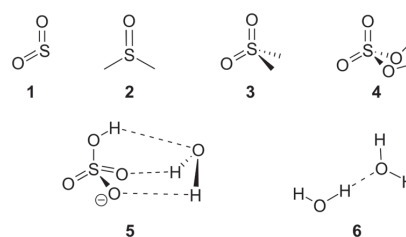


Chart 1 Benchmark species: sulfur dioxide (1) dimethyl sulfoxide (2) dimethyl sulfone (3) dimethyl sulfate (4) hydrated bisulfate anion (5) and water dimer (6).





work on sulfated biomolecules, *i.e.* free and H-bonded S=O groups, SOC and SOH groups. It includes sulfur dioxide (1), dimethyl sulfoxide (2), dimethyl sulfone (3) and dimethyl sulfate (4), as well as the hydrated bisulfate anion (5). The water dimer (6) was also included to examine the anharmonicity of H-bonded OH groups. The reference experimental IR spectra and geometries (shown in ESI†) were obtained from literature.^{24–27}

Harmonic simulations

The geometries were optimized with Gaussian09²⁸ using seven functionals (BLYP,^{29,30} B3LYP,^{30–32} CAM-B3LYP,³³ M06-2X,^{34,35} ω B97X-D,³⁶ PBE0,³⁷ LC-PBE^{38,39} with another range separation parameter, $\omega = 0.25$ instead of 0.47, as per a previous study,⁴⁰ referred to as LC-PBE), one post-Hartree Fock method (MP2⁴¹) combined with eleven basis sets, including five triple-zeta Pople's basis (6-311+G(d), 6-311++G**, 6-311++G(df,pd), 6-311++G(2df,2pd) and 6-311++G(3df,3pd));^{42,43} four Dunning's basis sets (cc-pVDZ, cc-pVTZ, cc-pVQZ, aug-cc-pVTZ);^{44–46} and SNSD.^{47–49} Then, the harmonic frequencies were calculated at the same levels. The analysis of normal modes was carried out with Gabedit.⁵⁰ To assess the reliability of the computed S=O, SOC and SOH frequencies against the set of benchmark spectra, the signed errors ($\text{freq}_{\text{DFT}} - \text{freq}_{\text{Exp}}$) summed over all the concerned modes are reported in Fig. 3 (the list of experimental and computed frequencies is given in ESI†). It is expected that an accurate representation of the electronic potential yields harmonic frequencies greater than the experimental values. Thus, the negative errors obtained with BLYP, B3LYP with all basis sets and LC-PBE with most basis sets are anomalous and suggest a deficient representation of the electronic potential of sulfated compounds. More precisely, as the optimized geometries are not inaccurate, this indicates that the minimum of the electronic potential is realistic, while its curvature is underestimated. In contrast, ω B97X-D and M06-2X yield positive, physically acceptable errors for all basis sets with exception of the smallest one (SNSD). The behavior of CAM-B3LYP, PBE0 and MP2 is more basis set dependent: the inclusion of d and f orbitals is essential for Pople's basis sets to yield acceptable results and similarly, only the largest Dunning's basis sets are acceptable. Note that in this picture, an error close to zero might not be an indicator of quality. Instead, the methods yielding the best representations of the potential are likely to show errors "somewhere" above zero. Without a more quantitative criterion, we do not attempt to compare the accuracy of the methods in the positive domain at this stage.

Optimization of the scaling factors

The explicit simulation of the anharmonic terms of the electronic potential is not a widely used procedure to date. Instead, it is generally accepted to use an empirical correcting scaling factor which reduces the values of the harmonic frequencies. To further quantify the performances of the selected functionals and basis sets for sulfated compounds, we have reckoned pairs of scaling factors minimizing the RMS (root-mean squared) deviation over all frequencies of the benchmark set (the list of experimental and computed frequencies is given in ESI†) in the high frequency and low frequency ranges. The cutoff was applied at 2800 cm^{-1} to best account for the CH, NH

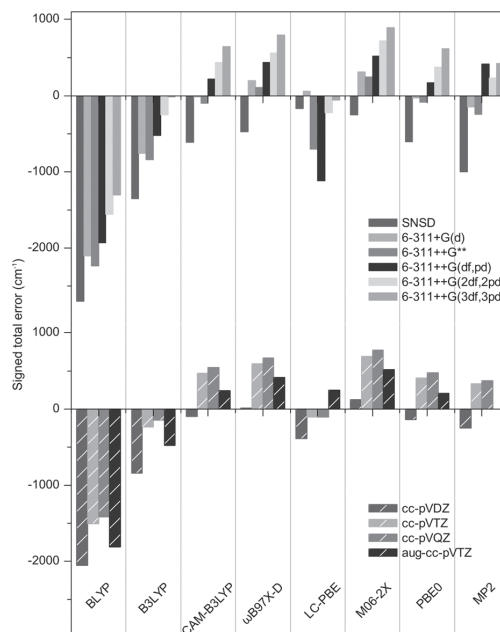


Fig. 3 Signed total error on the S=O, SOC and SOH frequencies of the benchmark species.

and OH stretches region and the fingerprint region. The minimal RMS deviations obtained with our optimized scaling factors are shown in Fig. 4, and the scaling factors are given in Table 1.

Conforming to the previous section, the methods consistently underestimating the S=O, SOC and SOH frequencies require scaling factors greater than 1.0. Not only is this an arbitrary way to correct for the incorrect representation of the sulfate modes, it also leads to a poorer match for the other modes (as illustrated in Fig. 2b for NH scissor modes). The RMS's of these methods are not further discussed and are plotted with empty bars in Fig. 4 for clarity.

For the methods consistently yielding errors in the positive domain, the minimized RMS deviations are hardly functional dependent. For Pople's basis sets, the precision steadily increases with the size of the basis set and the best results are obtained with the 6-311++G(2df,2pd) and the 6-311++G(3df,3pd) basis sets (RMS = 26 to 35 cm^{-1}). The Dunning's basis sets show similar performances with RMS deviations ranging from 25 to 30 cm^{-1} . In conclusion, the precision of the harmonic frequencies corrected with our optimized scaling factors ranges from 25 to 35 cm^{-1} , which is slightly below the experimental resolution (typically 20 cm^{-1} in the fingerprint region and 10 cm^{-1} in the high frequency range).

Anharmonic corrections

The cubic and quartic terms of the potential were calculated with Gaussian09 using the GVTP2 method.⁵¹ The RMS deviations thus obtained are shown in Fig. 5. In contrast to scaled harmonic simulations, it is remarkable that the result is highly functional dependent. In particular, the precision obtained with

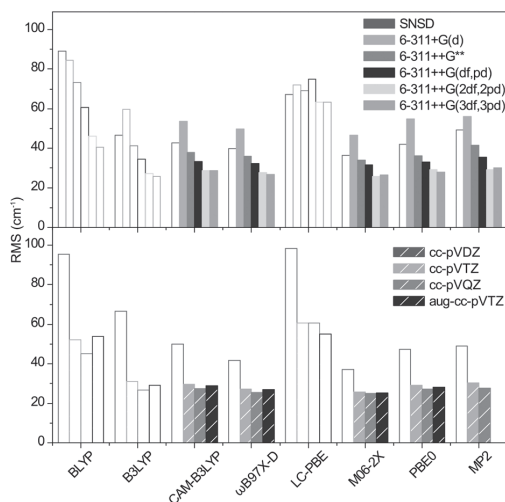


Fig. 4 RMS deviations of the harmonic frequencies of the benchmark set scaled with our optimized scaling factors. Empty bars are used to fade out the methods consistently underestimating the sulfate frequencies.

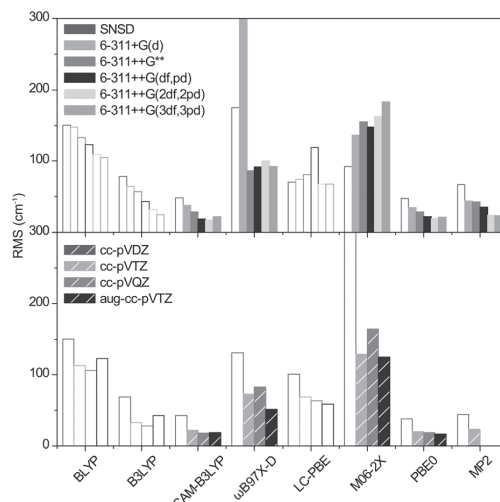


Fig. 5 RMS deviations of the anharmonic frequencies of the benchmark set. The y scale is cropped at 300 cm⁻¹ for clarity (for ωB97X-D/6-311+G(d) RMS = 400). Empty bars are used to fade out the methods consistently underestimating the sulfate frequencies.

ωB97X-D and M06-2X (RMS = 50 to 400 cm⁻¹) deteriorates dramatically. This is also true, to a lesser extent for MP2. With CAM-B3LYP and PBE0, the trend is opposite: the RMS is reduced to 18 and 19 cm⁻¹, respectively, with 6-311++G(2df,2pd). For Dunning's basis sets, it ranges from 22 to 17 cm⁻¹ as the size of the basis set increases. With a systematic improvement of 10 cm⁻¹, the precision now falls very near the experimental resolution.

The computational cost of 6-311++G(2df,2pd) calculation is comparable for both CAM-B3LYP and PBE0 functionals (292 and 203 hCPU for dimethyl sulfate (4), for instance). For Dunning's basis sets, the cost is similar for cc-pVTZ but increases by a factor of 10 for cc-pVQZ and aug-cc-pVTZ.

Considering the gain in precision vs. the computational cost associated with the anharmonic corrections, we select a combination of 6-311++G(2df,2pd) with either CAM-B3LYP or PBE0 for the best representation of vibrational frequencies in sulfated compounds.

The hybrid method

Furthermore, in the prospect of extending the application of the GVPT2 approach to more complex molecules, we attempted to

introduce a hybrid method, aiming at reducing the computational cost of the anharmonic corrections with minimal loss of precision.

The hybrid method consists of computing the harmonic frequencies with the best combination of functional and basis set (*i.e.* CAM-B3LYP or PBE0 with 6-311++G(2df,2pd)), followed by calculation of the time consuming cubic and quartic terms of the potential with a smaller basis set. The procedure is explained in ESI.† In short, the normal modes obtained in the harmonic simulations at the 6-311++G(2df,2pd) level and at the lower level are matched. Then, hybrid anharmonic frequency is obtained by adding the low level anharmonic correction to the corresponding high level harmonic frequency.

With the highest gain in computational time and a preserved precision (RMS = 19 cm⁻¹), the 6-311++G** basis set gave the best results.

Illustration: dimethyl sulfate

Using the CAM-B3LYP functional, the performance of the hybrid approach in terms of precision and speed is illustrated

Table 1 Optimized scaling factors (low frequency/high frequency)

	BLYP	B3LYP	CAM-B3LYP	ωB97X-D	LC-PBE	M06-2X	PBE0	MP2
SNSD	1.049/0.996	1.049/0.962	1.018/0.953	1.011/0.945	0.997/0.978	1.003/0.945	1.018/0.95	1.036/0.95
6-311+G(d)	1.049/1.004	1.015/0.968	0.984/0.958	0.974/0.951	0.981/0.984	0.97/0.949	0.983/0.956	0.988/0.949
6-311++G**	1.049/0.994	1.028/0.961	0.995/0.952	0.986/0.946	1.029/0.976	0.982/0.944	0.994/0.949	1/0.942
6-311++G(df,pd)	1.049/0.994	1.014/0.962	0.982/0.953	0.973/0.946	1.049/0.993	0.971/0.943	0.984/0.95	0.971/0.94
6-311++G(2df,2pd)	1.049/0.994	1.004/0.962	0.975/0.953	0.969/0.946	1.009/0.978	0.963/0.946	0.977/0.95	0.983/0.945
6-311++G(3df,3pd)	1.049/0.996	0.994/0.963	0.966/0.955	0.96/0.947	1.002/0.979	0.957/0.946	0.967/0.952	0.976/0.947
cc-pVDZ	1.049/1.011	1.021/0.972	0.992/0.962	0.987/0.952	1.002/0.986	0.984/0.953	0.994/0.958	0.997/0.947
cc-pVTZ	1.049/0.999	1.002/0.965	0.972/0.956	0.966/0.948	1.002/0.98	0.964/0.948	0.975/0.953	0.978/0.948
cc-pVQZ	1.049/0.998	0.999/0.965	0.97/0.956	0.964/0.947	1.004/0.981	0.962/0.948	0.973/0.953	0.978/0.949
aug-cc-pVTZ	1.049/0.999	1.015/0.966	0.983/0.957	0.975/0.949	0.984/0.982	0.973/0.949	0.984/0.955	

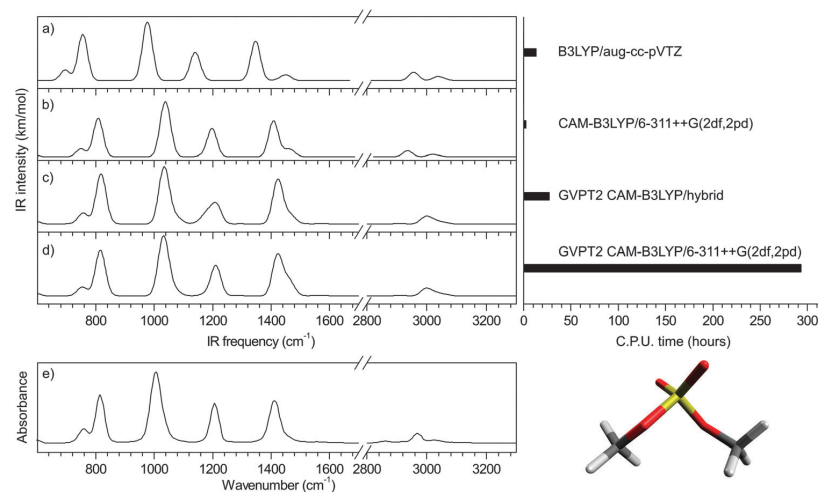


Fig. 6 Illustration: dimethyl sulfate (**4**) (a) scaled B3LYP harmonic spectrum with NIST scaling factor, (b) scaled CAM-B3LYP harmonic spectrum with optimized scaling factor, (c) hybrid anharmonic spectrum, (d) anharmonic spectrum and (e) experimental spectrum from NIST.

in Fig. 6 for dimethyl sulfate (**4**) (all other benchmark species are shown in ESI†). Both the position and relative intensities of the vibrational modes calculated with the CAM-B3LYP/6-311++G(2df,2pd) and CAM-B3LYP/hybrid (Fig. 6c and d) are in close agreement with the experimental spectrum (Fig. 6e), the latter offering a substantial gain of a factor of 11 in computational time without significant loss of precision. For comparison, two scaled harmonic spectra are presented: (i) a popular combination of B3LYP with an augmented Dunning's basis set^{17,27} with a standard scaling factor of 0.968²⁴ which yields unacceptably underestimated sulfate frequencies (Fig. 6a), as previously discussed; and (ii) CAM-B3LYP/6-311++G(2df,2pd) (Fig. 6b) with our optimized scaling factor of 0.975, which offers the most accurate harmonic approximation of the experimental spectrum. The gain of the anharmonic calculations is twofold: an increase in precision on individual frequencies, and a consequential alteration of the harmonic pattern, in better agreement with the experimental spectrum. The results obtained with the PBE0 functional are shown in ESI†. Additionally, the versatility of the hybrid method is illustrated in ESI† in the case of glucosamine 6-phosphate.

Conclusion

A striking case of failure of DFT harmonic simulation of sulfate vibrational frequencies was observed in the case of our model glycosaminoglycan fragment glucosamine 6-sulfate, critically impairing its structural characterization and somewhat echoing similar situations reported for phosphorylated compounds. Driven by this observation, we have explored the accuracy of 79 methods for predicting the sulfate frequencies of a set of benchmark species. We report the anomalous behavior of the BLYP, B3LYP and LC-PBE functionals, as well as of small basis

sets, yielding atypical underestimated frequencies. With an adequate choice of method and scaling factor, we show that a precision of $\sim 30 \text{ cm}^{-1}$ can be obtained in the harmonic approximation, approaching the experimental resolution. In a second step, we explored the performance of GVPT2 anharmonic corrections. The best results were obtained with CAM-B3LYP/6-311++G(2df,2pd), offering a precision of 18 cm^{-1} within the spectroscopic resolution in the fingerprint region. Finally, as the computational cost of explicit anharmonic corrections becomes prohibitive for species of increasing size, we devised an alternative "hybrid" strategy which brings down the computational cost to an accessible level for functionalized biomolecules, without significant loss of precision. Using this original approach, we could solve the case of glucosamine 6-sulfate. The robustness of the hybrid method was further illustrated on a phosphorylated species. We expect that this novel theoretical tool will unlock the potential of mid-IR spectroscopy coupled to mass spectrometry for the sequencing and structural characterization of functionalized biopolymers.

Acknowledgements

This work was supported by the LABEX iMUST (ANR-10-LABX-0064) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). Experiments in the mid-IR region were performed at the free electron laser facility FELIX in Netherlands with the skillful assistance of the FELIX staff, in particular Dr Britta Redlich and Lex van der Meer. Travel costs to the Netherlands were supported by the Dutch-French Van Gogh program. In this work, we were granted access to the HPC resources of the FLMSN, "Fédération Lyonnaise de Modélisation et Sciences Numériques", partner of EQUIPEX EQUIP@MESO

and the HPC Resources from GENCI-CINES (Grant 2013–2014 [087025]).

Notes and references

‡ The two pyranoside anomers (α and β) co-exist in the measured sample. NMR experiments were performed in a deuterated water/methanol (1:1) mixture to measure the anomeric ratio (method and spectra in ESI†). The sugar was found predominantly in the alpha configuration immediately after the preparation of the solution sample, it reached an equilibrium after few hours with a ratio of 70% of the α -anomer for 30% of the β -anomer. We have verified that introduction of a fraction of β -anomer does not significantly affect the shape of the simulated spectra.

- 1 A. Alonso, J. Sasin, N. Bottini, I. Friedberg, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon and T. Mustelin, *Cell*, 2004, **117**, 699–711.
- 2 T. Pawson and J. D. Scott, *Science*, 1997, **278**, 2075–2080.
- 3 W. B. Huttner, *Nature*, 1982, **299**, 273–276.
- 4 A. Tempez, M. Ugarov, T. Egan, J. A. Schultz, A. Novikov, S. Della-Negra, Y. Lebeyec, M. Pautrat, M. Caroff, V. S. Smentkowski, H.-Y. J. Wang, S. N. Jackson and A. S. Woods, *J. Proteome Res.*, 2005, **4**, 540–545.
- 5 *Transforming Glycoscience: A Roadmap for the Future*, The National Academies Press, Washington, DC, 2012.
- 6 M. J. Kailemia, L. R. Ruhaak, C. B. Lebrilla and I. J. Amster, *Anal. Chem.*, 2014, **86**, 196–212.
- 7 D. J. Harvey, *Mass Spectrom. Rev.*, 1999, **18**, 349–450.
- 8 J. Zaia, *OMICS: J. Integr. Biol.*, 2010, **14**, 401–418.
- 9 J. Zaia, *Mass Spectrom. Rev.*, 2004, **23**, 161–227.
- 10 A. Racaud, R. Antoine, L. Joly, N. Mesplet, P. Dugourd and J. Lemoine, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 1645–1651.
- 11 E. B. Cagmat, J. Szczepanski, W. L. Pearson, D. H. Powell, J. R. Eyler and N. C. Polfer, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3474–3479.
- 12 B. Casu, M. Guerrini and G. Torri, *Curr. Pharm. Des.*, 2004, **10**, 939–950.
- 13 G. K. Dhoot, M. K. Gustafsson, X. Ai, W. Sun, D. M. Standiford and C. P. Emerson Jr, *Science*, 2001, **293**, 1663–1666.
- 14 R. Sasisekharan, Z. Shriver, G. Venkataraman and U. Narayanasami, *Nat. Rev. Cancer*, 2002, **2**, 521–528.
- 15 S. M. Muthana, C. T. Campbell and J. C. Gildersleeve, *ACS Chem. Biol.*, 2012, **7**, 31–43.
- 16 B. Schindler, J. Joshi, A.-R. Allouche, D. Simon, S. Chambert, V. Brites, M.-P. Gaigeot and I. Compagnon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 22131–22138.
- 17 A. L. Patrick, C. N. Stedwell, B. Schindler, I. Compagnon, G. Berden, J. Oomens and N. C. Polfer, *Int. J. Mass Spectrom.*, 2015, **379**, 26–32.
- 18 R. Paciotti, C. Coletti, N. Re, D. Scuderi, B. Chiavarino, S. Fornarini and M. E. Crestoni, *Phys. Chem. Chem. Phys.*, 2015, DOI: 10.1039/C5CP01409C.
- 19 L. Wang, J. R. Brown, A. Varki and J. D. Esko, *J. Clin. Invest.*, 2002, **110**, 127–136.
- 20 X. Ai, *J. Cell Biol.*, 2003, **162**, 341–351.
- 21 A. Sharma, G. Ohanessian and C. Clavaguera, *J. Mol. Model.*, 2014, **20**, 1–9; F. Turecek, C. L. Moss, I. Pikalov, R. Pepin, K. Gulyuz, N. Polfer, M. F. Bush, J. Brown, J. Williams and K. Richardson, *Int. J. Mass Spectrom.*, 2013, **354–355**, 249–256.
- 22 C. J. Johnson and M. A. Johnson, *J. Phys. Chem. A*, 2013, **117**, 13265–13274.
- 23 Y. Miller, G. M. Chaban, J. Zhou, K. R. Asmis, D. M. Neumark and R. Benny Gerber, *J. Chem. Phys.*, 2007, **127**, 094305.
- 24 NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, 2005.
- 25 T. I. Yacovitch, T. Wende, L. Jiang, N. Heine, G. Meijer, D. M. Neumark and K. R. Asmis, *J. Phys. Chem. Lett.*, 2011, **2**, 2135–2140.
- 26 R. Kalescky, W. Zou, E. Kraka and D. Cremer, *Chem. Phys. Lett.*, 2012, **554**, 243–247.
- 27 A. Borba, A. Gómez-Zavaglia, P. N. N. L. Simões and R. Fausto, *Spectrochim. Acta, Part A*, 2005, **61**, 1461–1470.
- 28 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, Gaussian, Inc, Wallingford, CT, USA, 2009.
- 29 A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100.
- 30 C. T. Lee, W. T. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
- 31 S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200–1211.
- 32 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
- 33 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.
- 34 Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **125**, 194101.
- 35 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
- 36 J. D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.
- 37 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.
- 38 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 39 H. Iikura, T. Tsuneda, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2001, **115**, 3540–3544.





PCCP

[View Article Online](#)

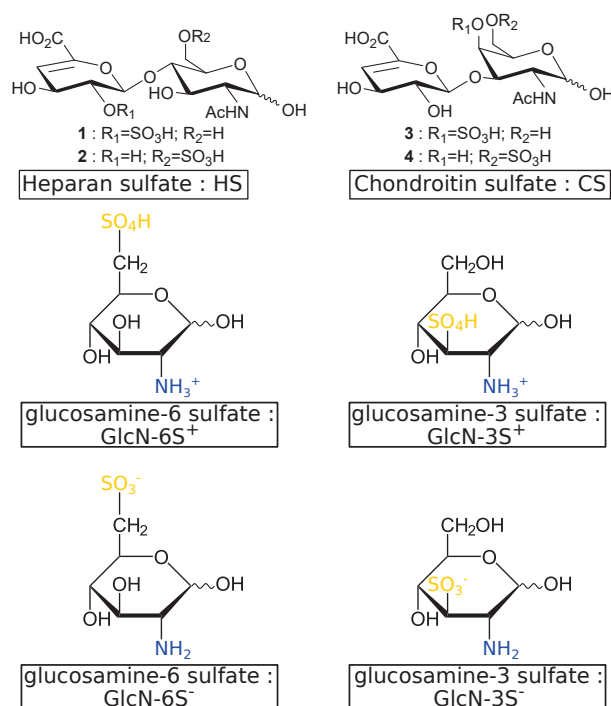
Communication

- 40 L. Barnes, S. Abdul-Al and A. R. Allouche, *J. Phys. Chem. A*, 2014, **118**, 11033–11046.
- 41 C. Moller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 0618–0622.
- 42 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
- 43 R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.*, 1980, **72**, 650–654.
- 44 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.
- 45 R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.
- 46 D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, 1994, **100**, 2975–2988.
- 47 V. Barone and P. Cimino, *Chem. Phys. Lett.*, 2008, **454**, 139–143.
- 48 V. Barone, P. Cimino and E. Stendardo, *J. Chem. Theory Comput.*, 2008, **4**, 751–764.
- 49 V. Barone and P. Cimino, *J. Chem. Theory Comput.*, 2009, **5**, 192–199.
- 50 A.-R. Allouche, *J. Comput. Chem.*, 2011, **32**, 174–182.
- 51 V. Barone, *J. Chem. Phys.*, 2005, **122**, 014108.

IRMPD Spectroscopy Sheds New (Infrared) Light on the Sulfate Pattern of Carbohydrates[†]

Highlights

- Differentiation of positional isomers of sulfated carbohydrates.
- Assignment of the structure for glucosamine 3-sulfate and glucosamine 6-sulfate.
- DFT simulation demonstrates that the features of the IRMPD spectrum clearly depend on the sulfate position.
- Resolution of isomeric disaccharides derived from the two GAG polysaccharides heparan and chondroitin sulfate.



[†] Published, DOI : 10.1021/acs.jpca.6b11642

IRMPD Spectroscopy Sheds New (InfraRed) Light on the Sulfate Pattern of Carbohydrates

B. Schindler,[†] L. Barnes,[†] C. J. Gray,[‡] S. Chambert,[§] S. L. Flitsch,[‡] J. Oomens,[¶] R. Daniel,^{*#} A.R. Allouche[†] and I. Compagnon^{*†a}

[†] Univ Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-69622, VILLEURBANNE, France.

[‡] School of Chemistry & Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

[§] Univ Lyon, INSA-Lyon, CNRS, Université Lyon 1, CPE Lyon, ICBMS, UMR 5246, Bâtiment Jules Verne, 20 avenue Albert Einstein, F-69621 Villeurbanne, France

[¶] Radboud University, Institute for Molecules and Materials, FELIX Laboratory, Toernooiveld 7, 6525ED Nijmegen, The Netherlands

[◇] Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098XH Amsterdam, The Netherlands

[#] LAMBE UMR 8587 CNRS, Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, Université d'Evry Val d'Essonne, 91025 Evry, France

^a Institut Universitaire de France IUF, 103 Blvd St Michel, 75005 Paris, France

ABSTRACT: IR spectroscopy of gas phase ions is proposed to resolve positional isomers of sulfated carbohydrates. Mass spectrometric fingerprints and gas phase vibrational spectra in the near and mid IR regions were obtained for sulfated monosaccharides, yielding unambiguous signatures of sulfated isomers. We report the first systematic exploration of the biologically relevant but notoriously challenging deprotonated state in the near IR region. Remarkably, anions displayed very atypical vibrational profiles, which challenge the well-established DFT (Density Functionnal Theory) modeling. The proposed approach was used to elucidate the sulfate patterns in glycosaminoglycans – a ubiquitous class of mammalian carbohydrates – which is regarded as a major challenge in carbohydrate structural analysis. Isomeric glycosaminoglycan disaccharide from heparin and chondroitin sources were resolved, highlighting the potential of InfraRed Multiple Photon Dissociation spectroscopy as a novel structural tool for carbohydrates.

INTRODUCTION

The analysis of the glycome, covering the entire repertoire of carbohydrate structures produced by cells, attracts growing attention.^{1,2} A major challenge in the characterization of glycosidic chains is the versatility of structures as a consequence of dynamic modifications as well as of their non-gene driven nature. In this field, mass spectrometry (MS) has become a primary technique for characterization, especially when only small quantities or complex samples are available. However, application of MS in glycomics is also notorious for ambiguities raised by the large numbers of isomers, as the carbohydrate building blocks can differ only by the stereochemistry of a carbon atom. Other key isomeric features making structural elucidation by MS difficult include (i) the re-

giochemistry of the glycosidic bond linking two carbohydrate moieties; (ii) the anomer stereochemistry (α or β); (iii) possible epimerization; (iv) the structure of the monomeric ring (v) the identification of the position of commonly observed modifications such as sulfation.

The elucidation of sulfate patterns is regarded as a major challenge in carbohydrate analysis, and in particular in the case of glycosaminoglycans (GAGs). This major class of bioactive sulfated carbohydrates illustrates the biological significance of sulfation, for example in the case of the fine tuning of the protein binding properties of heparin (Hp)/heparan sulfate (HS).³ On the one hand, the 3-O-sulfation site was found to be crucial for the blood anticoagulant activity of hepa-

rin⁴ and on the other hand, the more common 6-O-sulfation pattern on HS was found to be responsible for controlling cell proliferation.^{5,6}

In this context, continuous efforts are made to develop MS strategies for obtaining structural information on carbohydrate structures, usually involving chromatographic or electrophoretic separation associated or not with chemical derivation, gas phase reactivity with metals, ion mobility spectrometry, or fragmentation modes alternative to collision induced dissociation (CID), such as electron detachment dissociation (EDD) and Infra-Red Multiple Photon Dissociation (IRMPD).⁷⁻¹⁴ On the other hand, spectroscopic approaches provide direct and refined structural information but are generally technically incompatible with mass spectrometry. IRMPD spectroscopy combines the advantages of MS and spectroscopy, as the mass and the IR fingerprint of mass selected ions are recorded simultaneously. Since the seminal paper of Ohanessian *et al.*,¹⁵ IRMPD spectroscopy has become a key physical chemistry approach for the conformational analysis of biomolecules. However it has never been established as an hyphenated MS analytical tool despite its evident potential for the disambiguation of isomers. Following the pioneer work of John Simons on the gas phase IR spectroscopy of sugars¹⁶ and our earlier report of the first application of IRMPD for the disambiguation of isobaric sulfated and phosphorylated saccharides,¹⁷ we propose the first exploration of positional isomers of sulfated carbohydrates.

In the first section of this article, we report the structure of protonated glucosamine 3-S and glucosamine 6-S, supported by theory, then we show that these species display diagnostic IRMPD signatures in their native state, i.e. deprotonated. In the second section, we aim at exploiting the analytical potential of a frugal IRMPD approach to discriminate sulfated isomers of two pairs of disaccharides of biological interest, namely heparan sulfate and chondroitin sulfate.

RESULTS AND DISCUSSION

Spectroscopic diagnostic of positional isomers of sulfated monosaccharides

As a proof of principle of the resolving power of IRMPD spectroscopy, we present the diagnostic signatures of glucosamine 3-S and 6-S in their protonated charge state; followed by a generic IRMPD analysis, that is a full conformational investigation supported by quantum chemistry.

The IRMPD signatures of the pair of isomers glucosamine 6-sulfate and 3-sulfate measured at the free electron laser facility FELIX (800-1700 cm⁻¹) and using a tabletop IR system (2700-3700 cm⁻¹) are reported in Figure 1. They are clearly distinct in both the long and short wavelength spectral ranges and allow for facile differentiation bet-

ween the two isomers. To further exploit these IRMPD patterns in terms of molecular conformation and to assign spectral features to specific molecular vibration, a thorough conformational search followed by DFT simulation of the frequencies at the CAMB3LYP/6-311++G(2df,2pd) level was carried out. Around 200 conformers were obtained for each anomer of glucosamine 3-sulfate, including 19 conformers of the α anomer and 17 conformers of β anomer below 25 kJ/mol. These are further classified in three types of stable ring conformations: 4C_1 chair, 1S_3 skew and 1C_4 chair. These conformations are characterized by a H-bond between one of the protons of the amine group and one oxygen atom of the sulfate moiety. For both anomers, the lowest energy structure is a 4C_1 chair stabilized by a H-bond between OH(4) and OH(6). Other 4C_1 chairs are observed at higher energies with small variations in OH interactions. The 1S_3 conformation stabilized by a H-bond between OH(1) and OH(6) is only present for the β anomer, and the third ring conformation, the 1C_4 chair, does not show any OH interactions for the α anomer; the β anomer in this conformation lies at relative energies higher than 25 kJ/mol. The vibrational spectra for these conformations are then computed in the harmonic approximation and compared to the experimental spectrum in Figure S1. The best match is obtained for a combination of the two anomeric configurations of the lowest energy conformation: a 4C_1 chair. The conformational analysis of glucosamine 6-sulfate was previously reported.^{18,19} It was found that it adopts the same 4C_1 chair ring conformation. The differences between the two IRMPD spectra thus originate solely from the different positions of the sulfate group and the induced differences in intramolecular hydrogen bonding.

Once the stable conformation is established, the spectra can be interpreted in terms of molecular vibration: the carbohydrate OH stretches between 3500 and 3680 cm⁻¹, a distinctive NH₃⁺ pattern between 3100 and 3400 cm⁻¹ and four diagnostic modes of the neutral sulfate moiety: the (S)OH stretch, the asymmetric SO₂ stretch, the symmetric SO₂ stretch and the SO(H) stretch, which range from 800 to 3600 cm⁻¹. Their experimental frequencies are reported in Figure 1 and diagnostic bands are highlighted in contrasting color. In the FEL region, the pattern of sulfate vibrational bands appears as a distinctive diagnostic of the positional isomers, with frequency shifts ranging from -47 to +21 between protonated glucosamine 3-sulfate and 6-sulfate. In the high energy range, indirectly, the pattern of carbohydrate OH stretching bands is also a straightforward signature of the sulfate position: the functionalization of O3 and O6 results in the disappearance of the OH(3) stretch at 3600 cm⁻¹ and the OH(6) stretch at 3666 cm⁻¹, respectively (a detail-

led mode analysis is shown in Figures S2 and S3 of the Supporting Information).

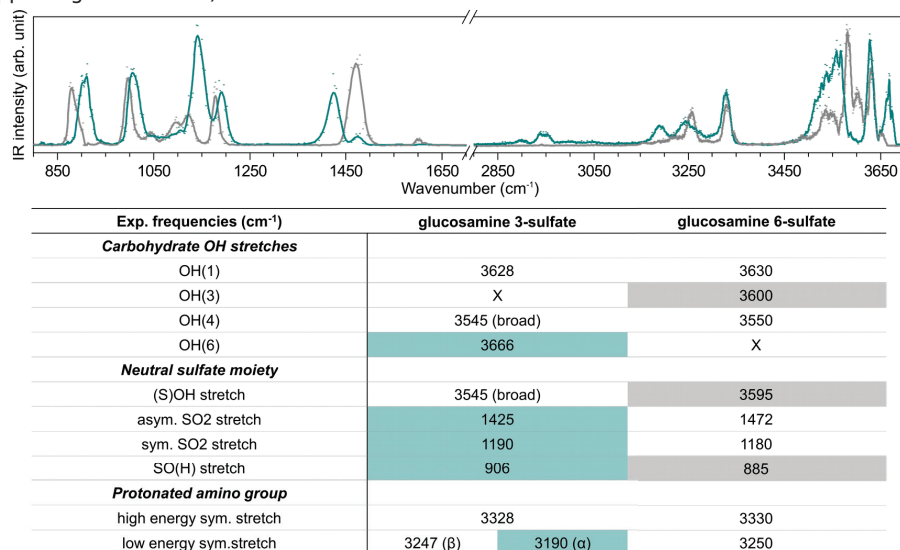


Figure 1: IRMPD spectra of protonated glucosamine 6-sulfate (grey: [GlcN6S + H]⁺) and 3-sulfate (cyan: [GlcN3S + H]⁺) in the 800-1700 cm⁻¹ and 2800-3700 cm⁻¹ spectral ranges and table of experimental IR frequencies for protonated glucosamine 3-sulfate and 6-sulfate.

Sulfated biomolecules however, are generally deprotonated in their native state. Although IRMPD spectroscopy of deprotonated ions can be done using the high fluence available at free electron laser facilities,¹⁹⁻²² it is notoriously difficult to achieve with a tabletop IR system.²³⁻²⁵ This is regarded as a major limitation of the IRMPD technique and constitute a severe obstacle to its rise as a routine analytical tool. The performance of our IRMPD setup were enhanced by tighthening the focus of the laser beam (15 cm focus). This allowed to overcome this limitation and to perform systematic analysis of deprotonated ions, not only using a FEL, but also using a compact setup.

Surprisingly, the FEL spectral range does not appear particularly informative in this study. Indeed, the spectrum of deprotonated glucosamine 6-sulfate in the 800-1700 cm⁻¹ spectral range shown in Figure 2 is remarkably similar to this of other deprotonated sulfated compounds such as fucose 3-sulfate (shown in Fig. 2) or even sulfosyringine (from ref ²⁵) They essentially display two doublets of bands, one centered around 1025 cm⁻¹ and the other around 1250 cm⁻¹. Hence, the infrared signature in the 800-1700 cm⁻¹ spectral range does not offer a clear signature of sulfated compounds. A theoretical exploration of the IRMPD signature of anionic sulfated compounds in the FEL range is shown in Fig. S4. It reveals that these features correspond to the a pattern of sulfate

modes, which is poorly isomer or conformer dependent.

In contrast, the spectra of anionic glucosamine 6-sulfate and 3-sulfate recorded in the 2700-3700 cm⁻¹ spectral range, displayed in figure 2 (right panel), show a number of distinctive features. In the OH stretching range, resolved features are observed at 3576 and 3495 cm⁻¹, for glucosamine 6-sulfate and 3-sulfate, respectively. A broad active region corresponding to unresolved H-bonded OH stretch bands is observed with maximal absorption at 3200 and 3305 cm⁻¹, respectively, likely convoluted with NH₂ stretch modes expected in the same region. The CH stretch patterns are both centered around 2900 cm⁻¹. Thus, in absence of diagnostic in the low frequency range, the high frequency range offers an excellent alternative for the identification of the sulfate position. A conformational search followed by DFT simulation of the frequencies at the CAMB3LYP/6-311++G(2df,2pd) level was carried out. The calculated spectrum obtained for the lowest energy conformer of GlcN6S, a ⁴C₁ chair, is shown in S5 (GlcN3S in S6). The main groups of features, namely the free and/or weakly H-bonded OH groups (OH1 and OH3) at high energy, the NH modes around 3400 cm⁻¹, and the strongly H-bonded OH4 group around 3200 cm⁻¹ reasonably match the main features observed in the experimental spectrum. It is noteworthy, however, that based on our experience of IRMPD spectroscopy of posi-

tively charged species, these spectra of anionic species are very atypical. Indeed only four OH modes and two NH modes are responsible for the broad feature between 3000 and 3600 cm^{-1} and it is very surprising that they are not better resolved. Further studies of anionic species in this spectral range will be essential to build new chemical intuition and better interpret these atypical patterns.

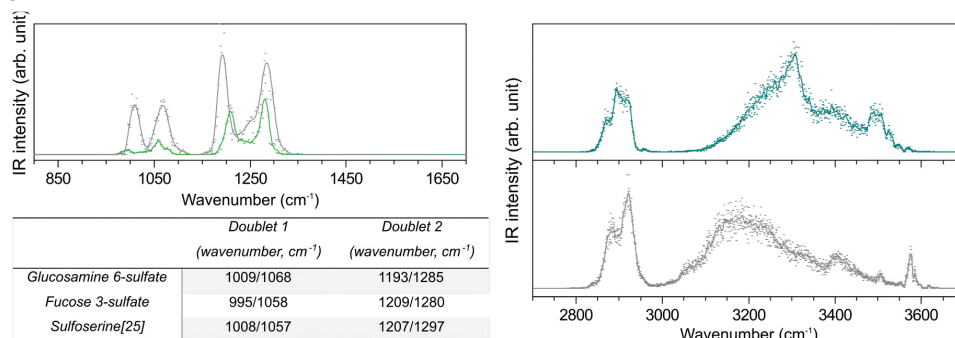


Figure 2. Left: IRMPD spectra of deprotonated glucosamine 6-sulfate (grey: [GlcN6S - H] $^{-}$) and fucose 3-sulfate (green: [Fuc3S - H] $^{-}$) in the 800-1700 cm^{-1} spectral range. Table of IR frequencies observed in the 800-1700 cm^{-1} spectral range. Right: IRMPD spectra of deprotonated glucosamine 6-sulfate (grey: [GlcN6S - H] $^{-}$) and glucosamine 3-sulfate (cyan: [GlcN3S - H] $^{-}$) in the 2700-3700 cm^{-1} spectral range.

Application to the analysis of sulfate patterns in glycosaminoglycan fragments

The high sensitivity of IRMPD spectroscopy for distinguishing positional isomers of sulfated monosaccharides appears promising for studying more complex sulfated carbohydrates like GAGs. They are expressed at the surface of mammalian cells and in the extracellular matrix (e.g. heparan sulfate (HS) and chondroitin sulfate (CS)) and are involved in a wide range of biological processes such as signaling,²⁶ development,²⁷ inflammation,²⁸ cancer,^{29,30} and coagulation.³¹ One of the key questions in the understanding of the biological activity of GAGs is the identification of the sulfate position. They appear naturally as deprotonated biomolecules, therefore, the capacity of our technic to record IRMPD spectra of deprotonated species is particularly relevant. Four GAG-derived isomeric Δ -unsaturated monosulfated disaccharide are compared in this study and shown in the scheme in fig. 3: two derived from heparin $\Delta\text{UA}(2\text{S})$ -[1 $\rightarrow$ 4]-GlcNAc (**1**, Hp III-A) and ΔUA -[1 $\rightarrow$ 4]-GlcNAc(6S) (**2**, Hp II-A) and two derived from chondroitin sulfate ΔUA -[1 $\rightarrow$ 3]-GalNAc-4S (**3**, CS-A) and ΔUA -[1 $\rightarrow$ 3]-GalNAc-6S (**4**, CS-C).

The two heparin monosulfated disaccharides Hp II-A and Hp III-A are characterized by sulfation on the N-acetyl glucosamine or on the unsaturated uronic acid, respectively. Their IRMPD spectra in the negative ion mode between 550-1850 cm^{-1} and 2700-3700 cm^{-1} are shown in Figure 3.

In the low frequency range, both spectra display quite distinct patterns in contrast to the monosaccharides discussed above. The IRMPD spectrum of Hp II-A displays three of the deprotonated sulfate specific modes observed also in the spectra of the monosaccharides: at 1280 (free S=O stretch), 1180 (H-bonded S=O stretch) and 1030 cm^{-1} (SO₃ sym). The intense C-O stretch observed around 1060 cm^{-1} in monosaccharides is not observed here. For the IRMPD spectrum of Hp III-A, the pattern is different from the sulfated monosaccharide reference. The band at 1033 cm^{-1} is similar to that in the spectrum of Hp II-A, but the bands at 1323 cm^{-1} and 1230 cm^{-1} are shifted to higher frequencies by 50 cm^{-1} .

For both isomers, the absence of the characteristic neutral sulfate modes around 900 and 1450 cm^{-1} , indicates deprotonation of the sulfate group, in agreement with chemical intuition. Additionally carboxylate function would show a diagnostic mode near 1630 cm^{-1} ³² which is not observed here. This confirms the deprotonation of the sulfate group.

In the high frequency range, IRMPD spectra of Hp II-A and Hp III-A display the typical features of negatively charged sulfated saccharides as established above for sulfated monosaccharides, that is: a CH-stretch absorption pattern around 2900 cm^{-1} , and a broad and unresolved region at higher energy. They also display two partially resolved bands at 3510 and 3586 cm^{-1} . The overall shape is distinctive of the isomer: Hp III-A shows a domi-

nant feature between 3300 and 3500 cm^{-1} and a single feature at 2940 cm^{-1} . In contrast, Hp II-A shows a broad active region between 3100 and 3500 cm^{-1} and two C-H features at 2916 and 2944 cm^{-1} . IR activity at 2915 and 3200 cm^{-1} is thus specific of Hp II-A.

The IRMPD signatures in the low and the high frequency range appear diagnostic of the sulfate position in these oligosaccharides.

The case of unsaturated disaccharides derived from chondroitin sulfate is a more difficult challenge as they possess two very close sites of sulfation on the N-acetyl galactosamine that is O6 and O4. These two disaccharides derived from CS possess notoriously ambiguous MS/MS signatures with variable ratios of Z and Y ions depending on the experimental conditions,³³⁻³⁵ it is hence challenging to identify them by MS/MS.

Samples of CS-A and CS-C are studied and their IRMPD spectra in the high frequency range are shown in Figure 3. They display a similar pattern over the whole range, with CH features around 2930 cm^{-1} and a broad unresolved active region between 3100 and 3550 cm^{-1} .

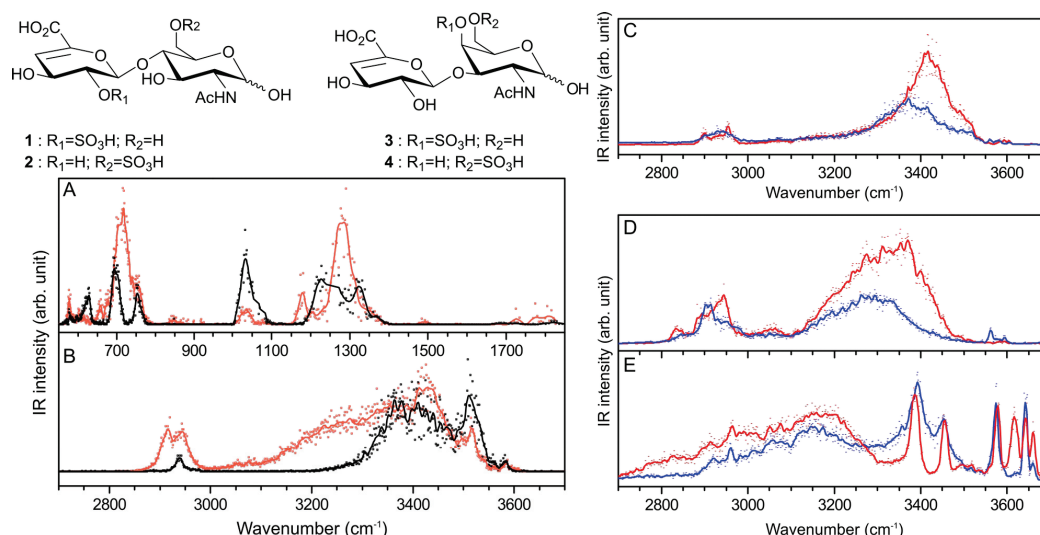


Figure 3. Scheme: Δ -unsaturated disaccharides derived from HS (left) and CS (right). A&B: IRMPD spectra of deprotonated Hp II-A (red: $[\text{Hp II-A} - \text{H}]^-$) and Hp III-A (black: $[\text{Hp III-A} - \text{H}]^-$) in the 550-1850 cm^{-1} and the 2700-3700 cm^{-1} spectral ranges. Right panels: IRMPD spectra of CS-A (red) and CS-C (blue) in different charge states. C: singly deprotonated $[\text{CS-A} - \text{H}]^-$ and $[\text{CS-C} - \text{H}]^-$. D: doubly deprotonated $[\text{CS-A} - 2\text{H}]^{2-}$ and $[\text{CS-C} - 2\text{H}]^{2-}$. E: NH_4^+ cationic complexes $[\text{CS-A} + \text{NH}_4]^+$ and $[\text{CS-C} + \text{NH}_4]^+$

Nonetheless, the IRMPD spectrum of CS-A shows two bands in the CH region at 2954 and 2902 cm^{-1} and a broad active region with a maximum absorption at 3405 cm^{-1} (HWHM= 218 cm^{-1}), whereas the IRMPD signature of the CS-C displays

a unique band in the CH region around 2930 cm^{-1} and a maximum absorption at 3370 cm^{-1} (HWHM= 140 cm^{-1}) in the broad unresolved active region.

Additionally, the IRMPD spectra of unsaturated disaccharides derived from Hp are distinct from

their isomers derived from CS. Hp show more structure with a partially resolved band at 3510 cm^{-1} which is not present for CSs. This exploration of GAG fragments is thus promising for GAG structural analysis. The mid-IR spectral range accessible with FELs offers a wealth of diagnostic features for Hp II-A and Hp III-A, but requires the FEL infrastructure which makes it impracticable for routine analysis. Instead, the high-energy range accessible with a table top laser system offers a relevant alternative for routine analysis: Using a compact IRMPD setup, we have been able to discriminate two Hp and two CS isomeric fragments.

The IRMPD signatures of CS fragments discussed above were recorded in their natural charge state. For analytical perspectives, it might be interesting to explore the spectroscopic resolving power of other charge states and counter ions accessible in MS. As an example doubly deprotonated and NH_4^+ adduct are shown in fig 3. Both spectra of doubly deprotonated species display similar features with a CH stretching region between 2800 and 3000 cm^{-1} and a broad active region in the $3100\text{--}3500\text{ cm}^{-1}$ spectral range. Nonetheless the IRMPD spectrum of doubly deprotonated CS-C shows a specific absorption band at 3562 cm^{-1} . In the case of NH_4^+ adduct, the two IRMPD spectra show more structure: a large unresolved region between 2800 and 3320 cm^{-1} and two resolved bands observed at 3388 and 3455 cm^{-1} . Remarkably the OH stretching range becomes highly diagnostic: the IRMPD spectrum of CS-A display four intense bands at 3576 , 3618 , 3642 and 3660 cm^{-1} whereas in the IRMPD spectrum of CS-C, only two intense bands are observed at 3574 , 3641 cm^{-1} .

CONCLUSION

We have used a combination of gas phase IRMPD spectroscopy with mass spectrometry to address the differentiation of positional isomers of sulfated carbohydrates on monosaccharide standards and GAG disaccharides. The structure of protonated glucosamine 3-sulfate and glucosamine 6-sulfate could both be assigned to a ${}^4\text{C}_1$ chair conformation using DFT calculations. DFT simulation demonstrates that the features of the IRMPD spectrum clearly depend on the sulfate position. Differentiation of the biologically more relevant deprotonated compounds is more challenging as compared to protonated compounds. Indeed, we showed that the IRMPD signature of sulfated sugars in the FEL region is ambiguous and glucosamine 3-sulfate and glucosamine 6-sulfate could only be resolved in the OPO region. Moreover the DFT simulation of the IR spectrum of deprotonated carbohydrates appears as a new challenge for theoretical modelling.

Our exploration of anionic GAG fragments validates the potential of the spectroscopic approach to resolve isomeric disaccharides derived from the two GAG polysaccharides heparin and chondroitin sulfate. We have demonstrated that the resolving power of the spectroscopic signature can be enhanced by tuning the charge state of the analyte. The added value of the spectroscopic approach will be particularly useful when other MS-based analytical methods fail to resolve isomeric structures.

METHODS

Samples

Samples of glucosamine 6-sulfate and 3-sulfate were obtained from Sigma Aldrich. Fucose 3-sulfate was from Grampian Enzymes (UK), heparin disaccharides Hp II-A and Hp III-A were purchased from Dextra Laboratories (UK) and chondroitin sulfate disaccharides CS-A and CS-C were obtained from Iduron (UK).

MS and IRMPD experiments

Solutions were prepared for mass spectrometry at a concentration of 250 μM in water/methanol solution (1:1). 0.1% of acetic acid was added to optimize ion signal in direct infusion.

IRMPD spectra of mass-selected ions were obtained using a modified ion trap mass spectrometer with an electrospray source from Thermo Finnigan (LCQ classic). Modifications allow the injection of a focused laser beam into the mass analyzer in order to irradiate the ion cloud, as described in our previous paper.¹⁷ An IRMPD spectroscopic scheme is used (InfraRed Multiple Photon Dissociation). IRMPD spectra were measured in two different regions, the 3 μm region (2700–3700 cm^{-1}) with a YAG-pumped tunable IR optical parametric oscillator/amplifier (OPO/A) (LaserVision) and the fingerprint region (550–1850 cm^{-1}) with the beamline of the Free Electron Laser for Infrared eXperiments (FELIX).³⁶ Mass-selected ions were fragmented upon resonant IR excitation and the photofragmentation yield was plotted as a function of the wavenumber to obtain IRMPD spectra. Each experiment was performed at a repetition rate of 10 Hz and averaged 3 times with typical irradiation times of 400 ms and 700 ms for FELIX and OPO experiments, respectively.

Theoretical methods

A conformational search was achieved using PM7³⁷ in the same way as in our previous paper,¹⁷ and optimization of the geometries and harmonic frequencies were computed using Gaussian09.³⁸ Density Functional Theory at the CAMB3LYP³⁹ level of theory with the 6-311++G(2df,2pd)^{40,41} basis set was chosen to take in account the difficulties to simulate vibrational spectra of sulfated species.¹⁸

The mode analysis was performed using Gabe-dit⁴² and an empirical scaling factor of 0.947 was used to correct the frequency of stretching modes in the 2700–3700 cm^{-1} spectral range.

ASSOCIATED CONTENT

Methods. NMR experiments. Theoretical IR spectra. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

* isabelle.compagnon@univ-lyon1.fr

ACKNOWLEDGMENT

This work was supported by the LABEX iMUST (ANR-10-LABX-0064) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). Experiments in the mid-IR region were performed at the free electron laser facility FELIX in Netherlands with the skillful assistance of the FELIX staff, in particular Dr Britta Redlich and Lex

van der Meer. Travel costs to the Netherlands were supported by the Dutch-French Van Gogh program. In this work, we were granted access to the HPC resources of the FLMSN, “Fédération Lyonnaise de Modélisation et Sciences Numériques”, partner of EQUIPEX EQUIP@MESO and the HPC Resources from GENCI-CINES (Grant 2013-2014[087025]).

REFERENCES

- (1) Hart, G. W.; Copeland, R. J. Glycomics Hits the Big Time. *Cell* **2010**, *143*, 672–676.
- (2) Cummings, R. D.; Pierce, J. M. The Challenge and Promise of Glycomics. *Chem. Biol.* **2014**, *21*, 1–15.
- (3) Kreuger, J. Interactions between Heparan Sulfate and Proteins: The Concept of Specificity. *J. Cell Biol.* **2006**, *174*, 323–327.
- (4) Lindahl, U.; Bäckström, G.; Thunberg, L.; Leder, I. G. Evidence for a 3-O-Sulfated D-Glucosamine Residue in the Antithrombin-Binding Sequence of Heparin. *Proc. Natl. Acad. Sci.* **1980**, *77*, 6551–6555.
- (5) Sugaya, N.; Habuchi, H.; Nagai, N.; Ashikari-Hada, S.; Kimata, K. 6-O-Sulfation of Heparan Sulfate Differentially Regulates Various Fibroblast Growth Factor-Dependent Signalings in Culture. *J. Biol. Chem.* **2008**, *283*, 10366–10376.
- (6) Vivès, R. R.; Seffouh, A.; Lortat-Jacob, H. Post-Synthetic Regulation of HS Structure: The Yin and Yang of the Sulf in Cancer. *Front. Oncol.* **2014**, *3*, 331.
- (7) Staples, G. O.; Zaia, J. Analysis of Glycosaminoglycans Using Mass Spectrometry. *Curr. PROTEOMICS* **2011**, *8*, 325–336.
- (8) Wührer, M. Glycomics Using Mass Spectrometry. *Glycoconj. J.* **2013**, *30*, 11–22.
- (9) Ortiz, D.; Enjalbert, Q.; MacAleese, L.; Dugourd, P.; Salpin, J.-Y. Effects of Calcium Complexation on Heparin-like Disaccharides. A Combined Theoretical, Tandem Mass Spectrometry and Ultraviolet Experiment: Calcium Complexation of Heparin-like Disaccharides. *Rapid Commun. Mass Spectrom.* **2015**, *29*, 1135–1144.
- (10) Williams, J. P.; Grabenauer, M.; Holland, R. J.; Carpenter, C. J.; Wormald, M. R.; Giles, K.; Harvey, D. J.; Bateman, R. H.; Scrivens, J. H.; Bow-ers, M. T. Characterization of Simple Isomeric Oligosaccharides and the Rapid Separation of Glycan Mixtures by Ion Mobility Mass Spectrometry. *Int. J. Mass Spectrom.* **2010**, *298*, 119–127.
- (11) Harvey, D. J.; Scarff, C. A.; Crispin, M.; Scanlan, C. N.; Bonomelli, C.; Scrivens, J. H. MALDI-MS/MS with Traveling Wave Ion Mobility for the Structural Analysis of N-Linked Glycans. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1955–1966.
- (12) Seo, Y.; Andaya, A.; Leary, J. A. Preparation, Separation, and Conformational Analysis of Differentially Sulfated Heparin Octasaccharide Isomers Using Ion Mobility Mass Spectrometry. *Anal. Chem.* **2012**, *84*, 2416–2423.
- (13) Leach, F. E.; Arungundram, S.; Al-Mafraji, K.; Venot, A.; Boons, G.-J.; Amster, I. J. Electron Detachment Dissociation of Synthetic Heparan Sulfate Glycosaminoglycan Tetrasaccharides Varying in Degree of Sulfation and Hexuronic Acid Stereochemistry. *Int. J. Mass Spectrom.* **2012**, *330–332*, 152–159.

- (14) Kailemia, M. J.; Park, M.; Kaplan, D. A.; Venot, A.; Boons, G.-J.; Li, L.; Linhardt, R. J.; Amster, I. J. High-Field Asymmetric-Waveform Ion Mobility Spectrometry and Electron Detachment Dissociation of Isobaric Mixtures of Glycosaminoglycans. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 258–268.
- (15) Kapota, C.; Lemaire, J.; Maitre, P.; Ohanessian, G. Vibrational Signature of Charge Solvation vs Salt Bridge Isomers of Sodiated Amino Acids in the Gas Phase. *J. Am. Chem. Soc.* **2004**, *126*, 1836–1842.
- (16) Talbot F.; Simons J. P. Sugars in the Gas Phase: the Spectroscopy and Structure of Jet-Cooled Phenyl β -D-Glucopyranoside. *Phys. Chem. Chem. Phys.*, **2002**, *4*, 3562–3565.
- (17) Schindler, B.; Joshi, J.; Allouche, A.-R.; Simon, D.; Chambert, S.; Brites, V.; Gaigeot, M.-P.; Compagnon, I. Distinguishing Isobaric Phosphated and Sulfated Carbohydrates by Coupling of Mass Spectrometry with Gas Phase Vibrational Spectroscopy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 22131–22138.
- (18) Barnes, L.; Schindler, B.; Allouche, A.-R.; Simon, D.; Chambert, S.; Oomens, J.; Compagnon, I. Anharmonic Simulations of the Vibrational Spectrum of Sulfated Compounds: Application to the Glycosaminoglycan Fragment Glucosamine 6-Sulfate. *Phys Chem Chem Phys* **2015**, *17*, 25705–25713.
- (19) Wight, C. A.; Beauchamp, J. L. Infrared Spectra of Gas-Phase Ions and Their Use in Elucidating Reaction Mechanisms. Identification of C7H7-Structural Isomers by Multiphoton Electron Detachment Using a Low-Power Infrared Laser. *J. Am. Chem. Soc.* **1981**, *103*, 6499–6501.
- (20) Steill, J. D.; Oomens, J. Gas-Phase Deprotonation of *p*-Hydroxybenzoic Acid Investigated by IR Spectroscopy: Solution-Phase Structure is Retained upon ESI. *J. Am. Chem. Soc.* **2009**, *131*, 13570–13571.
- (21) Oomens, J.; Steill, J. D.; Redlich, B. Gas-Phase IR Spectroscopy of Deprotonated Amino Acids. *J. Am. Chem. Soc.* **2009**, *131*, 4310–4319.
- (22) Crestoni, M. E.; Chiavarino, B.; Steinmetz, V.; Fornarini, S. Communication: Vibrational Study of a Benzyl Carbanion: Deprotonated 2,4-Dinitrotoluene. *J. Chem. Phys.* **2012**, *137*, 181101.
- (23) Almasian, M.; Grzetic, J.; van Maurik, J.; Steill, J. D.; Berden, G.; Ingemann, S.; Buma, W. J.; Oomens, J. Non-Equilibrium Isomer Distribution of the Gas-Phase Photoactive Yellow Protein Chromophore. *J. Phys. Chem. Lett.* **2012**, *3*, 2259–2263.
- (24) Schinle, F.; Jacob, C. R.; Wolk, A. B.; Greisch, J.-F.; Vonderach, M.; Weis, P.; Hampe, O.; Johnson, M. A.; Kappes, M. M. Ion Mobility Spectrometry, Infrared Dissociation Spectroscopy, and Ab Initio Computations toward Structural Characterization of the Deprotonated Leucine-Enkephalin Peptide Anion in the Gas Phase. *J. Phys. Chem. A* **2014**, *118*, 8453–8463.
- (25) Paciotti, R.; Coletti, C.; Re, N.; Scuderi, D.; Chiavarino, B.; Fornarini, S.; Crestoni, M. E. Serine O-Sulfation Probed by IRMPD Spectroscopy. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25891–25904.
- (26) Couchman, J. R. Transmembrane Signaling Proteoglycans. *Annu. Rev. Cell Dev. Biol.* **2010**, *26*, 89–114.
- (27) Haltiwanger, R. S.; Lowe, J. B. Role of Glycosylation in Development. *Annu. Rev. Biochem.* **2004**, *73*, 491–537.
- (28) Taylor, K. R.; Gallo, R. L. Glycosaminoglycans and Their Proteoglycans: Host-Associated Molecular Patterns for Initiation and Modulation of Inflammation. *FASEB J.* **2006**, *20*, 9–22.
- (29) Afratis, N.; Gialeli, C.; Nikitovic, D.; Tseggenidis, T.; Karousou, E.; Theocharis, A. D.; Pavão, M. S.; Tzanakakis, G. N.; Karamanos, N. K. Glycosaminoglycans: Key Players in Cancer Cell Biology and Treatment. *FEBS J.* **2012**, *279*, 1177–1197.
- (30) Yip, G. W.; Smollich, M.; Götte, M. Therapeutic Value of Glycosaminoglycans in Cancer. *Mol. Cancer Ther.* **2006**, *5*, 2139–2148.
- (31) Casu, B.; Guerrini, M.; Torri, G. Structural and Conformational Aspects of the Anticoagulant and Antithrombotic Activity of Heparin and Dermatan Sulfate. *Curr. Pharm. Des.* **2004**, *10*, 939–950.
- (32) Oomens, J.; Steill, J. D. Free Carboxylate Stretching Modes. *J. Phys. Chem. A* **2008**, *112*, 3281–3283.
- (33) Desaire, H.; Leary, J. A. Detection and Quantification of the Sulfated Disaccharides in Chondroitin Sulfate by Electrospray Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 916–920.
- (34) Zaia, J.; Costello, C. E. Compositional Analysis of Glycosaminoglycans by Electrospray Mass Spectrometry. *Anal. Chem.* **2001**, *73*, 233–239.
- (35) Flangea, C.; Serb, A. F.; Schiopu, C.; Tudor, S.; Sisu, E.; Seidler, D. G.; Zamfir, A. D. Discrimination of GalNAc (4S/6S) Sulfation Sites in Chondroitin Sulfate Disaccharides by Chip-Based Nanoelectrospray Multistage Mass Spectrometry. *Cent. Eur. J. Chem.* **2009**, *7*, 752–759.
- (36) Oepts, D.; van der Meer, A. F. G.; van Amersfoort, P. W. The Free-Electron-Laser User Facility FELIX. *Infrared Phys. Technol.* **1995**, *36*, 297–308.
- (37) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters. *J. Mol. Model.* **2013**, *19*, 1–32.
- (38) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, USA, 2009.
- (39) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange Correlation Functional Using the Coulomb Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (40) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (41) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self Consistent Molecular Orbital Methods .20. Basis Set for Correlated Wave Functions. *J. Chem. Phys.* **1980**, *72*, 650–654.

-
- (42) Allouche, A.-R. Gabedit-A Graphical User Interface for Computational Chemistry Softwares. *J. Comput. Chem.* **2011**, 32, 174-182.

METHODS AND SUPPLEMENTARY INFORMATION: IRMPD Spectroscopy Sheds New (InfraRed) Light on the Sulfate Pattern of Carbohydrates

B. Schindler,[†] L. Barnes,[†] C. J. Gray,[‡] S. Chambert,[§] S. L. Flitsch,[‡] J. Oomens,^{⊥||} R. Daniel,^{*#} A.R. Allouche[†] and
I. Compagnon^{*†a}

[†] Univ Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-69622, VILLEURBANNE, France.

[‡] School of Chemistry & Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

[§] Univ Lyon, INSA-Lyon, CNRS, Université Lyon 1, CPE Lyon, ICBMS, UMR 5246, Bâtiment Jules Verne, 20 avenue Albert Einstein, F-69621 Villeurbanne, France

[⊥] Radboud University, Institute for Molecules and Materials, FELIX Laboratory, Toernooiveld 7, 6525ED Nijmegen, The Netherlands

^{||} Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098XH Amsterdam, The Netherlands

[#] LAMBE UMR 8587 CNRS, Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, Université d'Evry Val d'Essonne, 91025 Evry, France

^a Institut Universitaire de France IUF, 103 Blvd St Michel, 75005 Paris, France

METHODS

NMR experiments

NMR spectra were recorded on Bruker AVANCE III 300 spectrometers at 300.13 MHz for ^1H NMR. Chemical shifts (δ) are measured in parts per million (ppm) and were calibrated relatively to the residual signal of methanol ($\delta = 3.31$ ppm). Glucosamine 3-Sulfate was dissolved at a 5 mg.mL^{-1} concentration in a $\text{CD}_3\text{OD}/\text{D}_2\text{O}$ mixture (1:1). After the mutarotation equilibrium was attained, the NMR 1D experiment was performed at 20°C . H(2) protons were ideally located on each side of the residual peak of methanol for measuring the α/β ratio. The α - and β -anomer were identified as doublets of doublets at $\delta=3.45$ ($J=10.3$ and 3.6 Hz) and $\delta=3.16$ ($J=10.4$ and 8.4 Hz), respectively (see Figure S16).[1]

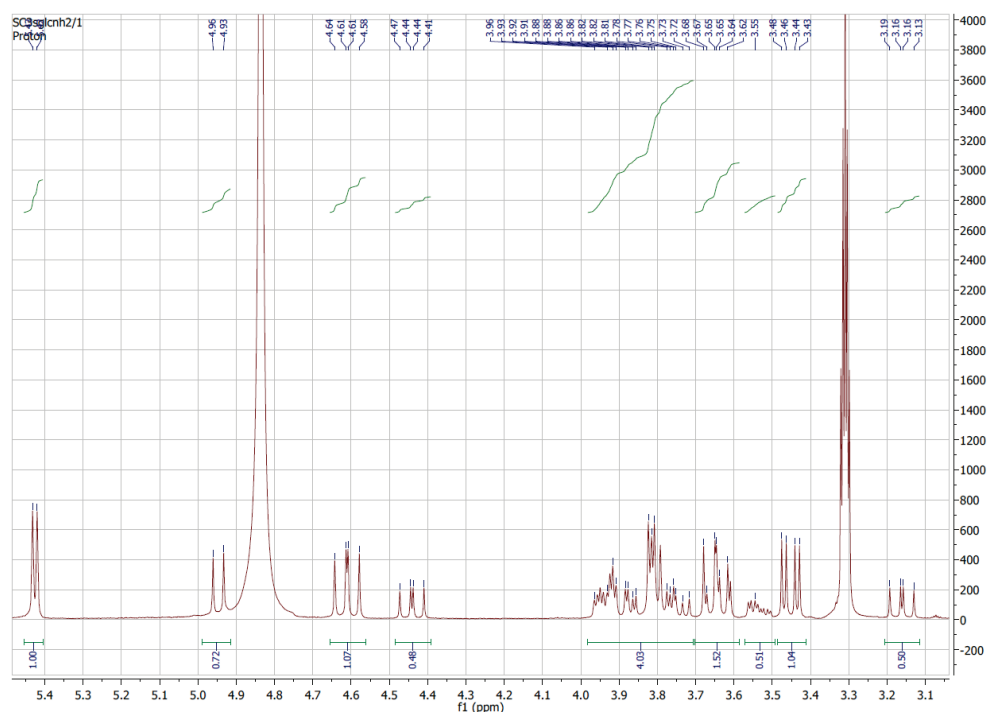


Figure M1: NMR spectrum of glucosamine 3-sulfate

SUPPLEMENTARY FIGURES AND DATA

Structural analysis of protonated glucosamine 3-sulfate

The experimental gas phase IRMPD spectrum in positive ion mode of protonated glucosamine 3-sulfate in the 3 μm region is recorded and shown in Figure S11. It displays intense spectroscopic features attributed to the neutral sulfate (S)OH stretch at 3562 cm^{-1} , three carbohydrate OH(i) (with $i=1, 4, 6$) stretches between 3500 and 3680 cm^{-1} and a distinctive NH_3^+ pattern between 3100 and 3400 cm^{-1} .

To extract molecular structure information from IRMPD spectra, comparison with theory is generally needed. In short, a conformational search was performed for both anomers α and β of D-glucosamine 3-sulfate. Around 200 conformers were obtained for each anomer, resulting in three ring conformations below 25 kJ/mol: a ${}^4\text{C}_1$ chair, a ${}^1\text{S}_3$ skew and a ${}^1\text{C}_4$ chair. These conformations are characterized by a H-bond between one of the protons of the amine group and one oxygen atom of the sulfate moiety. For both anomers, the lowest energy structure is a ${}^4\text{C}_1$ chair stabilized by a H-bond between OH(4) and OH(6). Other ${}^4\text{C}_1$ chairs are observed at higher energies with small variations in OH interactions. The ${}^1\text{S}_3$ conformation stabilized by a H-bond between OH(1) and OH(6) is only present for the β anomer, and the third ring conformation, the ${}^1\text{C}_4$ chair, does not show any OH interactions for the α anomer; the β anomer in this conformation lies at relative energies higher than 25 kJ/mol. The vibrational spectra for these conformations are then computed in the harmonic approximation and shown in Figure S1.

The harmonic spectra of the lowest energy conformation of α and β anomers are compared with the experimental spectrum in Figure S1. The match is excellent in the high energy range and the combination of the calculated frequencies account for the experimental features between 3100 and 3400 cm^{-1} . This shows that α and β anomers coexist in the gas phase in a ${}^4\text{C}_1$ chair conformation. For comparison, their ratio in solution is $\alpha/\beta = 67/33$ (NMR method and details in supporting information). An analysis of the normal modes shows two free OH stretches, OH(6) and OH(1), corresponding to the experimental bands at 3665 and 3628 cm^{-1} , respectively. H-bonded OH(4) and sulfate OH stretches around 3545 cm^{-1} results in a broad, unresolved feature. The broadening observed for of H-bonded modes in IRMPD was discussed in a previous work [2]. Finally, the experimental bands observed at 3190 , 3247 and 3328 cm^{-1} correspond to the NH stretch modes of the two structures identified.

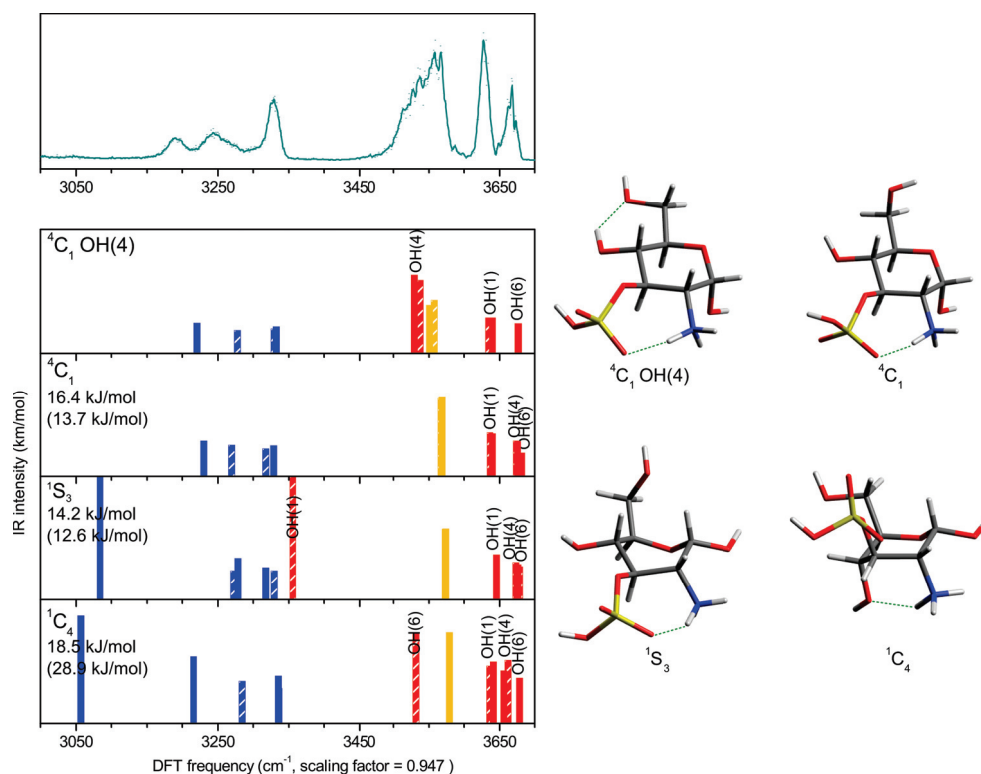


Figure S1: Top panel: IRMPD spectrum of protonated glucosamine 3-sulfate : $[\text{GlcN3S} + \text{H}]^+$ between 3000 and 3700 cm^{-1} . Middle panels: theoretical harmonic IR spectra of the four lowest energy structures of protonated glucosamine 3-sulfate, obtained with CAMB3LYP/6-311++G(2df,2pd) and scaled with a scaling factor of 0.947. In each panel the α anomer is represented with solid bars and the β anomer is represented with dashed bars and energy in brackets. A color code is used to highlight the mode analysis: OH stretch from group of sulfate function is indicated in orange. NH's are indicated in blue, carbohydrate OH's in red with individual labels indicating their position on the pyranose ring. Four conformations were selected, a ${}^4\text{C}_1$ chair with an interaction between OH(4) and OH(6), a ${}^4\text{C}_1$ chair without interaction, a ${}^1\text{S}_3$ skew and a ${}^1\text{C}_4$ chair conformation. The energies of geometries were computed with DFT with MP2/6-311++G(3df,3pd). Lower panels: Boltzmann 12.5 kJ/mol

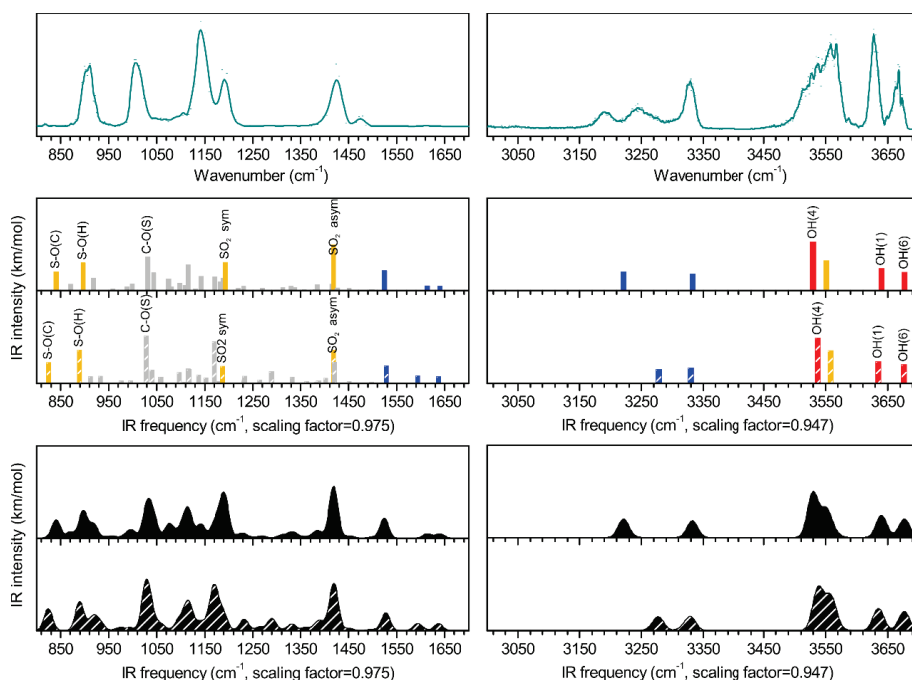


Figure S2: Top panels: IRMPD spectra of protonated glucosamine 3-sulfate $[\text{GlcN3S} + \text{H}]^+$ in the FEL and table-top ranges. Middle panel : theoretical harmonic IR spectra of the lowest energy structure of protonated glucosamine 3-sulfate (${}^4\text{C}_1(\text{OH}_4)$), as shown in Fig. S1) obtained with CAMB3LYP/6-311++G(2df,2pd). A scaling factor of 0.947 was used for the 3000-3700 cm^{-1} spectral range and a scaling factor of 0.975 for the 800-1700 cm^{-1} spectral range. The α anomer is represented with solid bars and the β anomer is represented with dashed bars. A color code is used to highlight the mode analysis: OH stretch from group of sulfate function is indicated in orange. NH's are indicated in blue, carbohydrate OH's in red with individual labels indicating their position on the pyranose ring. Lower panel: Convolved spectra (10 cm^{-1} width) of the contributions of all stables structures below 12.5 kJ/mol using the Boltzmann contributions at 300K (solid: α anomer; dashed: β anomer).

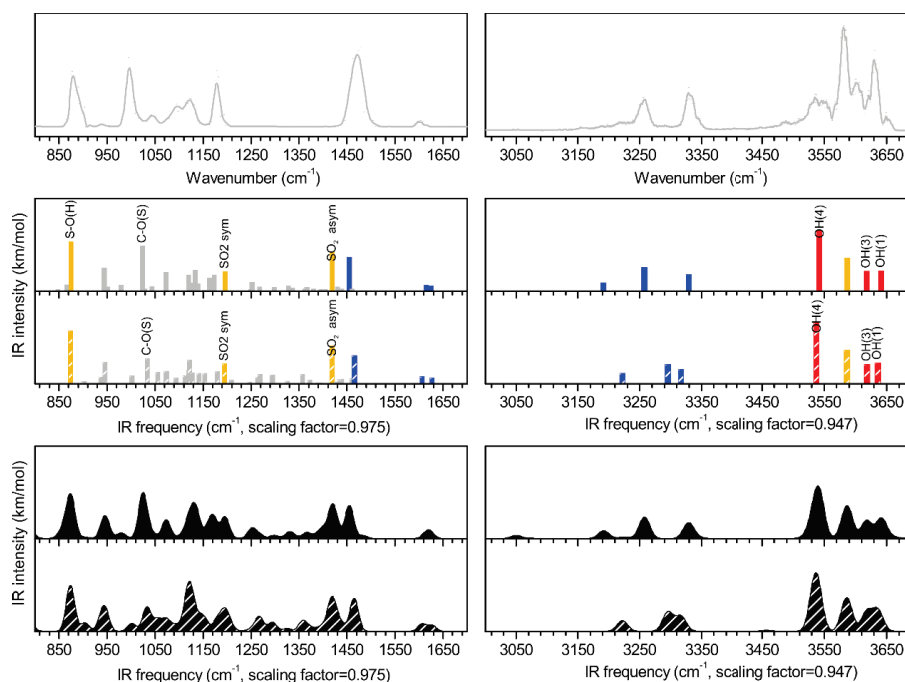


Figure S3: Top panels: IRMPD spectra of protonated glucosamine 3-sulfate $[\text{GlcN3S} + \text{H}]^+$ in the FEL and table-top ranges. Middle panel : theoretical harmonic IR spectra of the lowest energy structure of protonated glucosamine 3-sulfate ($^4\text{C}_1(\text{OH}_4)$, as described in reference [A]) obtained with CBMY/ L6P2-/ 11++G(Adf,Apd). B scaling factor of 0.947 was used for the 700-1000 cm^{-1} spectral range and a scaling factor of 0.975 for the 800-1700 cm^{-1} spectral range. The α anomer is represented with solid bars and the β anomer is represented with dashed bars. B color code is used to highlight the mode analysis: OH stretch from group of sulfate function is indicated in orange. NH's are indicated in blue, carbohydrate OH's in red with individual labels indicating their position on the pyranose ring. Lower panel: Convolved spectra (10 cm^{-1} width) of the contributions of all stables structures below 1A.5 kJ2mol using the Boltzmann contributions at 00K (solid: α anomer; dashed: β anomer).

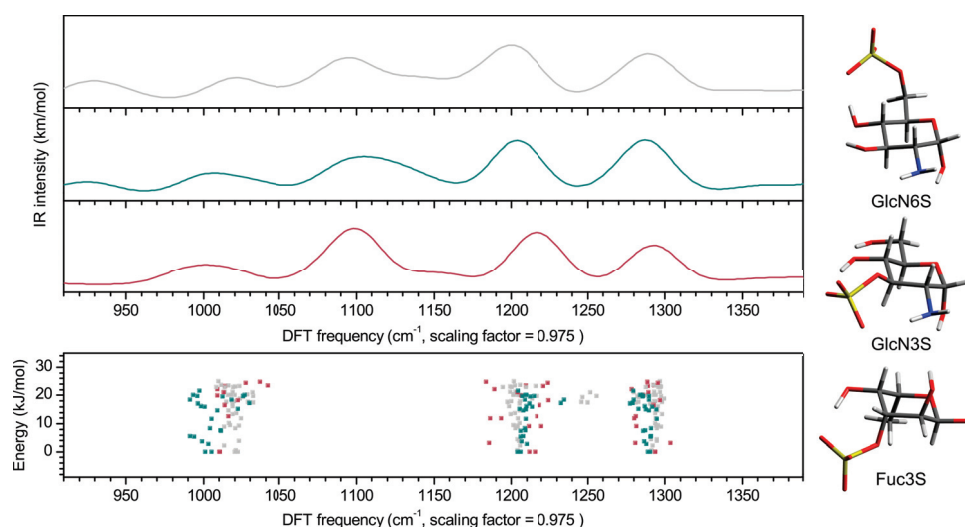


Figure S4: theoretical study of deprotonated species in the FEL range. Top panel: convoluted theoretical harmonic IR spectra of the lowest energy structure of deprotonated GlcN6S (grey line), GlcN3S (green line) and Fuc3S (red line) obtained with CAMB3LYP/6-311++G(2df,2pd). A scaling factor of 0.975 and a 20 cm^{-1} bandwidth convolution were used. The intense experimental doublet at 1200/1280 cm^{-1} is consistently observed in the calculated spectra, regardless of the molecule. These two intense features correspond to two asymmetric stretching modes of the sulfate group. The second experimental doublet at 1000/1060 cm^{-1} is also accounted for, with a pair of non-isomer dependant theoretical features corresponding to the symmetric SO_3 stretching mode and a CO elongation mode, respectively. Lower panel: scaled frequencies of the three the SO_3 modes of the stable conformations below 25 kJ/mol for deprotonated GlcN3S (green squares, 10 alpha anomers and 15 beta anomers), GlcN6S (grey squares, 22 alpha anomers and 22 beta anomers), Fuc3S (red squares, 7 alpha anomers and 8 beta anomers). These modes form three compact clouds around 1000, 1200 and 1300 cm^{-1} , which further confirms that the sulfate signature is poorly conformation-dependant.

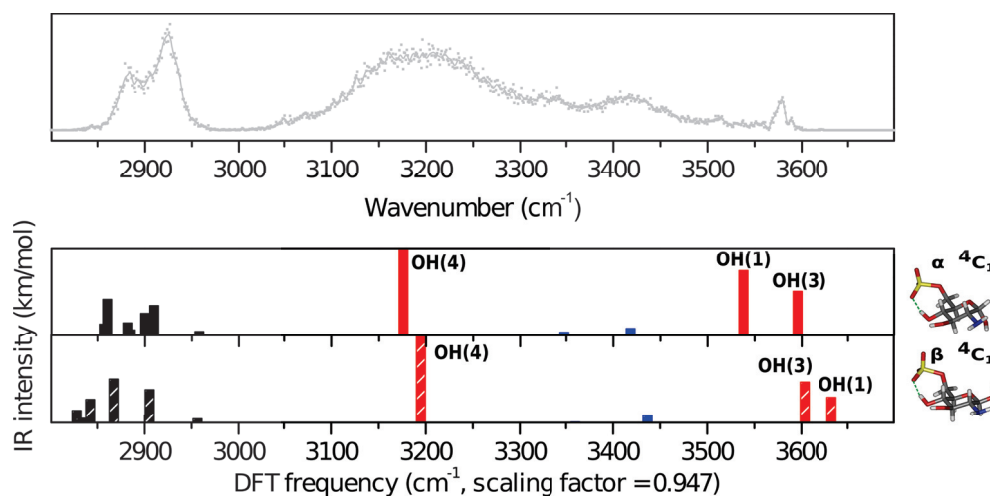


Figure S5: Top panel: IRMPD spectrum of deprotonated glucosamine 6-sulfate [GlcN6S - H]⁻. Lower panels: theoretical harmonic IR spectra of the lowest energy structure of deprotonated glucosamine 6-sulfate obtained with CAMB3LYP/6-311++G(2df,2pd). Solid bars represent the alpha anomer, dashed bars represent the beta anomer. A scaling factor of 0.947 was used for the 2800-3700 cm⁻¹ spectral range and a color code is used to highlight the mode analysis: NH's are indicated in blue, CH's are in black, carbohydrate OH's in red with individual labels indicating their position on the pyranose ring.

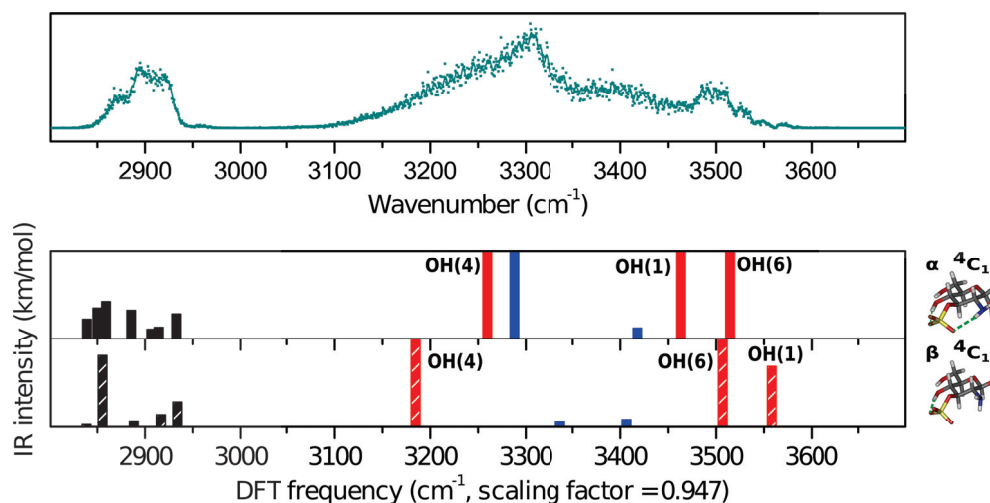


Figure S6: Top panel: IRMPD spectrum of deprotonated glucosamine 3-sulfate [GlcN3S - H]⁻. Lower panels: theoretical harmonic IR spectra of deprotonated glucosamine 3-sulfate obtained with CAMB3LYP/6-311++G(2df,2pd). Solid bars represent the alpha anomer, dashed bars represent the beta anomer. A scaling factor of 0.947 was used for the 2800-3700 cm⁻¹ spectral range and a color

S8

code is used to highlight the mode analysis: NH's are indicated in blue, CH's are in black, carbohydrate OH's in red with individual labels indicating their position on the pyranose ring.

REFERENCES

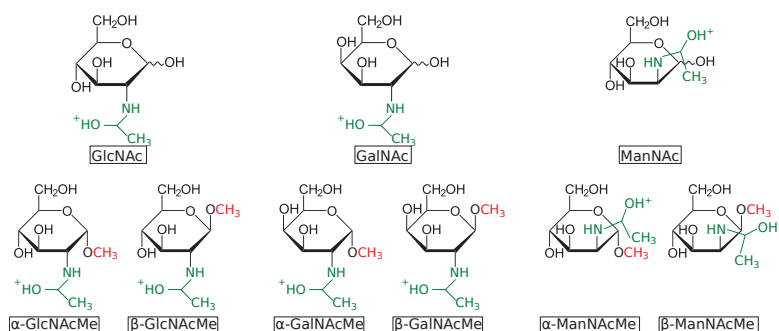
1. Beecher, C. N.; Larive, C. K. ^1H and ^{15}N NMR Characterization of the Amine Groups of Heparan Sulfate Related Glucosamine Monosaccharides in Aqueous Solution. *Anal. Chem.* **2015**, 150610095405007.
2. Schindler, B *et al.* Distinguishing Isobaric Phosphated and Sulfated Car-bohydrates by Coupling of Mass Spectrometry with Gas Phase Vibrational Spectroscopy. *Phys. Chem. Chem. Phys.* 2014, 16 (40), 22131–22138.

Conformational preferences of protonated *N*-acetylated hexosamines probed by Infrared multiple photon dissociation (IRMPD) spectroscopy and *ab initio* calculations[†]

L. BARNES^a, B. SCHINDLER^a, S. CHAMBERT^b, A-R. ALLOUCHE^a AND I. COMPAGNON^{ac}

Highlights

- First IRMPD study of protonated *N*-acetyl hexosamines.
- Disambiguation of HexNAc isomers, which could not be resolved by MS or MS/MS alone.
- Resolution of competitive protonation sites.
- Identification of the ring conformation and conformational heterogeneity.
- Analysis of the contribution of individual anomers in heterogeneous IRMPD spectra.



^a Institut Lumière Matière, Univ Lyon, Université Claude Bernard Lyon 1, CNRS, F-69622 Villeurbanne, France

^b Univ Lyon, INSA-Lyon, CNRS, Université Lyon 1, CPE Lyon, ICBMS, UMR 5246, Batiment Jules Verne, 20 avenue Albert Einstein, F-69621 Villeurbanne, France

^c Institut Universitaire de France, 103 Blvd St Michel, 75005 Paris, France. E-mail : isabelle.compagnon@univ-lyon1.fr

[†] Published, DOI : <https://doi.org/10.1016/j.ijms.2017.05.005>

6.1 Introduction

Structural analysis of carbohydrates in the gas phase was pioneered by the group of John Simons using cooled neutral jets and a double resonance UV/IR spectroscopic scheme [1]. It has since gained further momentum with IRMPD spectroscopy (InfraRed Multiple Photon Dissociation), which relies on a more frugal spectroscopic scheme and benefits from the performance of commercial mass spectrometers and the versatility of electrospray ions sources (ESI). By adding a spectroscopic dimension to Mass Spectrometric analysis, IRMPD allows the disambiguation of isomers and overcomes the intrinsic limitation of MS regarding the analysis of carbohydrates. Although the most abundant carbohydrate building blocks - hexoses - are neutral and do not readily ionize in ESI sources, several authors have reported IRMPD studies of mono- and disaccharides using metal adducts such as Lithium [2, 3, 4] or Rubidium [5]. On the other hand, monosaccharides can be naturally modified by a variety of functional modifications which contribute to modulate their charge in the biological environment. The structure of Lithium adducts of such functionalised monosaccharides have been reported [6, 7, 8]. In this context, we have previously reported the first IRMPD studies of protonated and deprotonated functionalized sugars, including hexuronic acids [9] and sulfated/phosphated amino sugars, which can be either protonated on the amino group or deprotonated on the sulfate/phosphate group [10, 11, 12]. Here we present the first IRMPD study of protonated *N*-acetylated hexosamines (GlcNAc, GalNAc and ManNAc) and their six methyl-blocked α and β anomers.

In section 6.2.1, we address the question of the protonation site. In section 6.2.2, we present a thorough analysis of the ring conformation. Finally, in section 6.2.3 we propose diagnostic spectroscopic features of the anomeric configuration.

6.1.1 Methods

6.1.1.1 Sample preparation

N-acetyl-D-glucopyranosyl amine (GlcNAc), *N*-acetyl-D-mannopyranosyl amine (ManNAc) and *N*-acetyl-D-galactopyranosyl amine (GalNAc) were purchased from Sigma Aldrich and used without further purification. Methyl- α -*N*-acetyl-D-galactopyranosyl amine (α GalNAcOMe) and methyl- β -*N*-acetyl-D-galactopyranosyl amine (β GalNAcOMe) were purchased from Carbosynth and used without further purification. Methyl- α -*N*-acetyl-D-glucopyranosyl amine (α GlcNAcOMe), methyl- α -*N*-acetyl-D-mannopyranosyl amine (α ManNAcOMe), methyl- β -*N*-acetyl-D-glucopyranosyl amine (β GlcNAcOMe) and methyl- β -*N*-acetyl-D-mannopyranosyl amine (β ManNAcOMe) were synthesized in house. They were prepared in water :methanol 50 :50 solutions at a concentration of 50 μ M for direct infusion and ionized in positive mode to produce protonated species.

6.1.1.2 IRMPD spectroscopy

IRMPD (InfraRed Multiple Photon Dissociation) spectroscopy was performed in a modified commercial quadrupolar ion trap (ThermoFinnigan LCQ classic) equipped with an electrospray source. The trapping electrode was drilled and an IR-transparent window was installed to allow irradiation of the ion cloud by the IR laser beam produced by a YAG-pumped OPO/OPA system (LaserVision) tunable in the 3 μ m region. Mass-selected ions

were isolated and irradiated for 700 ms, resulting in wavelength-dependent photofragmentation. The photofragmentation yield is calculated from the resulting photofragmentation spectrum using the following variation of the Beer-Lambert formula : $-\log(\frac{I_p}{I_p + \sum I_f})$, where I_p is the intensity of the precursor ion and I_f is the intensity of the photofragments. The IRMPD spectrum of the precursor ion is retrieved by monitoring its photofragmentation yield as a function of the wavelength in the 2700-3700 cm^{-1} spectral range. The data-points shown thorough the manuscript are averaged 3 times and a trend line (5-points Fourier Transform rolling averaging) is added to guide the eye. All experimental data are reported in Fig. SI1.

6.1.1.3 Computational methods

Molecular dynamic was used to explore the potential energy surface of the protonated ions with the PM7 [13] potential in OpenMopac [14] as described in a previous work [10]. One trajectory was calculated for each 1-*O*-methyl species and for each anomer of the non-methylated species. 1000 geometries were generated during each trajectory and optimized with PM7, yielding typically a few hundreds stable conformations after discarding the identical ones. The geometries were further optimized with two steps of Density Functional Theory (DFT) : firstly B3LYP [15, 16, 17]/6-31G*, then CAM-B3LYP [18]/6-311++G(2df,2pd) [19, 20, 21]. This resulted in the reduction of the number of stable geometries. The energies were further refined by single-point calculations with MP2 [22]/6-311++G(3df,3pd). At this stage, the 30 lowest energy conformers were selected for computing harmonic frequencies with CAM-B3LYP/6-311++G(2df,2pd). An empirical scaling factor of 0.947 was used for comparison with the experimental IRMPD spectra. The mode analysis was performed using Gabedit [23], Gaussian09 [24] was used for DFT calculations and Orca [25] was used for MP2 single-point energies.

6.2 Results and discussion

6.2.1 Identification of the protonation site

We have previously reported that protonation occurs on the amino group in glucosamine [10, 11]. In the case of *N*-acetyl hexosamines however, the amino function and the carbonyl function are competitive protonation sites.

To elucidate the protonation site we report the conformational exploration of GlcNAc protonated on each site as described in section 6.1.1.3. For protonation on the carbonyl function, molecular dynamics yielded 335 (474) structures for the α anomer (β anomer). 136 (136) stable conformations were obtained after DFT optimization at the B3LYP/6-31G* level of theory for the α anomer (β anomer). The absolute energy of the lowest energy conformer and the relative energies of the 30 lowest energy conformers are given in Table SI1 (α anomer) and Table SI2 (β anomer). For protonation on the amine function molecular dynamics yielded 476 (408) structures for the α anomer (β anomer). 74 (121) stable conformations were obtained after DFT optimization at the B3LYP/6-31G* level of theory for the α anomer (β anomer). The absolute energy of the lowest energy conformer and the relatives energies of the 30 lowest energy conformers are given in Table SI3 (α anomer) and Table SI4 (β anomer). It is found that the most stable conformation of α and β GlcNAc protonated on the carbonyl function have the lowest absolute MP2 energy :

they are 64 kJ/mol lower in energy than their counterparts protonated on the amine function.

The protonation on the carbonyl group raises an interesting question about the sharing of the proton with other hydroxyl group. This issue was also discussed for ortho-aminobenzoic acid [26] where the proton was involved in an hydrogen bond resulting in a large redshift of the corresponding stretching mode. In our case, the stretching mode of the carbonyl group and the proton (CO-H^+) has a low frequency : the scaled value is 2768 cm^{-1} , indicating a strong hydrogen bond with the O3H. An anharmonic calculation gives an even lower frequency showing the highly anharmonic nature of this bond. Unfortunately, this range of frequencies is not accessible in our IRMPD conditions. Hence the position of the stretching mode of CO-H^+ can not be confirmed experimentaly. However we have other qualitative means to assess this question. The calculation of the Mayer Bond Orders [27] (for the lowest energy conformation of the anomer α) indicates that the proton is shared. Indeed, for the hydrogen bond between O4H and O6H the Mayer bond order is about 0.94 for the bonding between O6 and its hydrogen compared to 0.77 for the bonding between the proton and the carbonyl oxygen. The Mayer bond order is 0.047 for the bond between O6 and the hydrogen of O4 while it is 0.16 between the proton and O3. This illustrates a strengthening of the hydrogen bond of the proton compared to typical hydrogen bonds at other location in the molecule. The computation of the non covalent interactions (NCI) via the reduced density gradient [28, 29] also support the fact that the hydrogen bond between the proton and the hydroxyl is strong . The computed NCI are shown in Fig. SI 17.

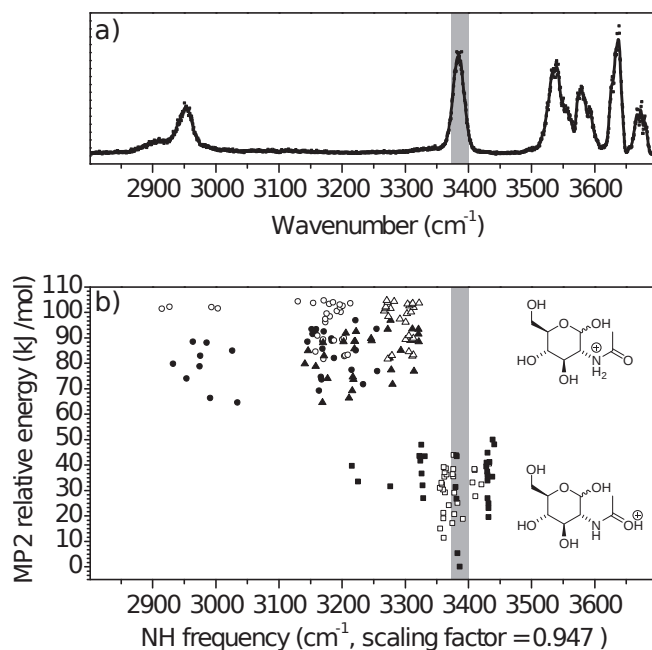


FIGURE 6.1 – (a) IRMPD spectrum of protonated GlcNAc. (b) scaled frequency of the NH stretching mode for the 30 lowest energy conformations of the α anomer (full squares) and β anomer (open squares) with a protonated carbonyl group, scaled frequencies of the two NH stretching modes (triangle : high frequency, circle : low frequency) for the 30 lowest energy conformations of the α anomer (full symbols) and β anomer (open symbols) with a protonated amine group. The position of the experimental band associated to the NH vibration is shaded in grey. The lowest energy structure of α GlcNAc protonated on the carbonyl function is taken as the global minimum.

Based on the energetic criterion, protonation of the carbonyl group appears highly favorable. To confirm this finding by experimental evidence, we further proceed to the mode analysis, in order to provide a robust spectroscopic diagnostic of the protonation site. The IRMPD spectrum of GlcNAc is shown in Fig. 6.1a. It displays four features above 3500 cm^{-1} which can be readily interpreted as OH stretching modes; an unresolved pattern of CH stretching modes around 2950 cm^{-1} ; and a strong NH stretching mode at 3384 cm^{-1} . Fig. 6.1b shows the calculated frequency of the NH stretching mode of the 30 lowest energy conformations of both anomers in the case of protonated carbonyl group (the theoretical spectra are shown in figures SI2 and SI3). They range from 3210 to 3450 cm^{-1} , depending on the conformation. However, it is remarkable that the distribution of NH frequencies for all stable structures below 25 kJ/mol is tightly centered around 3381 cm^{-1} with a deviation of 25.5 cm^{-1} , which perfectly matches the experimental pattern (shaded in grey in Fig. 6.1). Fig. 6.1c shows the calculated frequencies of the two NH stretching modes of the 30 lowest energy conformations of both anomers in the case of a protonated amine function (the theoretical spectra are shown in figures SI4 and SI5). These frequencies range below 3340 cm^{-1} regardless of the conformation and are incompatible with the experimental feature at 3384 cm^{-1} .

Based on both energetic and spectroscopic arguments, it appears clearly that GlcNAc is protonated on the carbonyl group. We find that the NH pattern proves to be a good diagnostic of the protonation site, however direct experimental observation of the protonated carbonyl was not possible due to its low frequency. In the next sections, we further proceed to the analysis of the ring conformation, assuming the same site of protonation for all *N*-acetyl hexosamine isomers. In each case the calculated frequency of the NH mode was found consistent with the experimental pattern, thus confirming the carbonyl protonation without repeating the complete energetic study.

6.2.2 Conformational analysis of 1-*O*-methyl *N*-acetyl hexosamines

In this section we further address the conformational preferences of the carbohydrate ring. In natural *N*-acetyl hexosamines samples, both anomers are present and contribute to the IRMPD spectrum. To remove this source of structural heterogeneity the conformational study was carried out on individual 1-*O*-methyl anomers, that is α GlcNAcOMe, β GlcNAcOMe, α GalNAcOMe, β GalNAcOMe, α ManNAcOMe and β ManNAcOMe. The molecular dynamics yielded 425, 458, 859, 1277, 314 and 509 structures, respectively. They converged into 143, 192, 248, 99, 137 and 152 stable conformers after DFT optimization. Their energies were calculated by MP2 and the 30 lowest energy conformations were selected for conformational analysis. Tables SI5 to SI10 summarize the energies and conformations obtained for each system.

For α GlcNAcOMe the lowest energy conformer is a 4C_1 chair stabilized by a network of hydrogen bonds between (Ac)O2H, O3H, O4H and O6H which are in equatorial position (Structure $\alpha_{Me}I$ in Fig. 6.2). The second lowest energy structure is a 1C_4 chair stabilized by a O3H $\cdots$ O6 hydrogen bond above the plane of the ring and a double coordination of the acetyl group (NH $\cdots$ O4 and (Ac)O2H $\cdots$ O1) below the plane (structure $\alpha_{Me}II$ in Fig. 6.2). Among the 30 lowest energy structures, a number of variations of these two structures are present. They include in particular 4C_1 and 1C_4 chairs with lesser hydrogen bond coordination (structures $\alpha_{Me}I'$ and $\alpha_{Me}II'$ in Fig. 6.2) and skew ring conformations, such as 1S_3 and 5S_1 (structures $\alpha_{Me}I''$ and $\alpha_{Me}II''$ in Fig. 6.2).

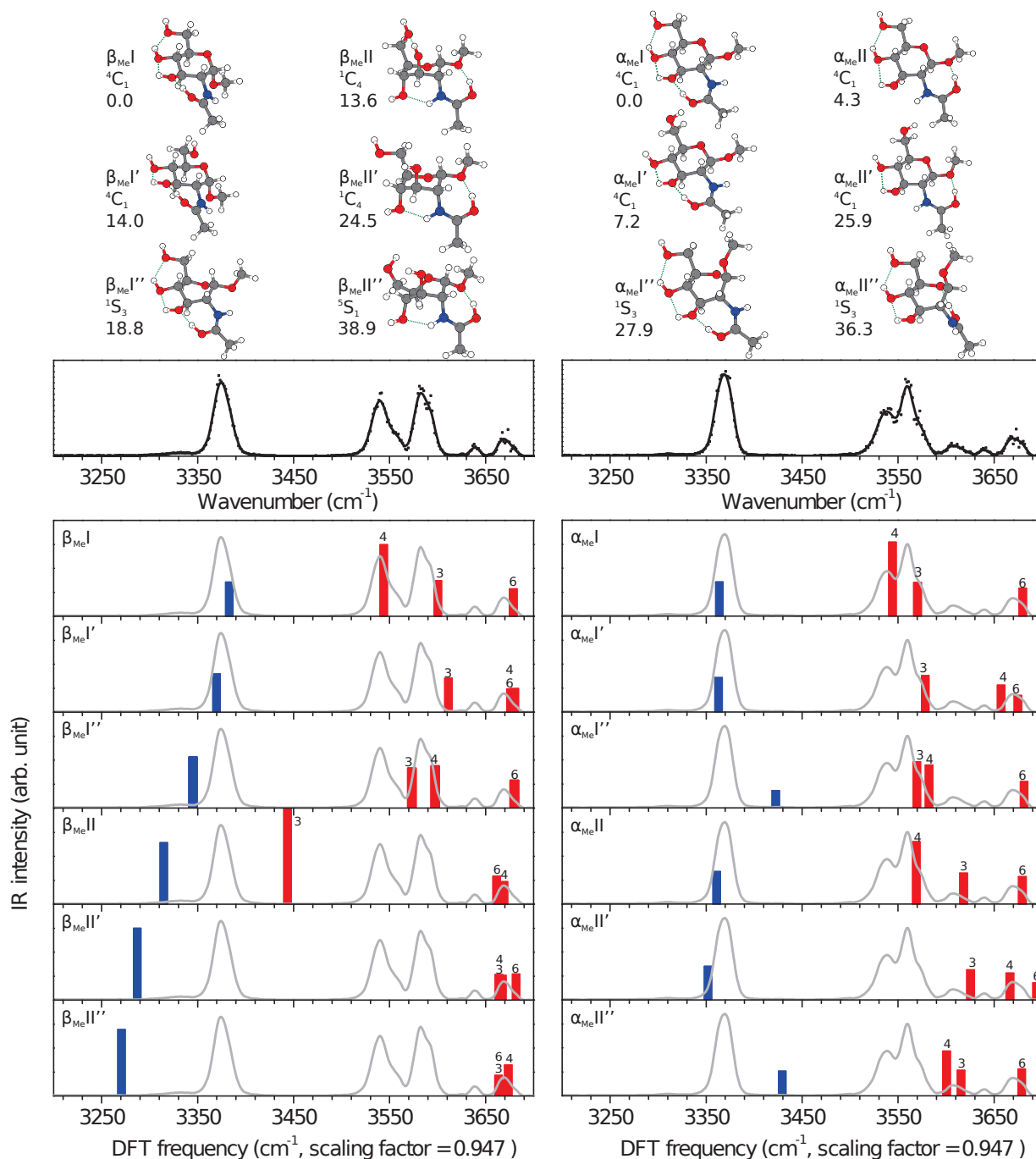


FIGURE 6.2 – Structures of 6 stable conformations of α GlcNAcOMe and their relative energy in kJ/mol calculated with MP2/6-311++G(3df,3pd). Theoretical harmonic IR spectra of these structures, between 3000 and 3700 cm^{-1} , obtained with CAM-B3LYP/6-311++G(2df,2pd) and scaled with a scaling factor of 0.947. For each IR spectrum, the experimental spectrum is plotted in gray for comparison. A color code is used to highlight the mode analysis : NH is indicated in blue, carbohydrate OH's in red with labels indicating their position on the pyranose ring.

For β GlcNAcOMe, the lowest energy conformer is also a 4C_1 chair stabilized by a maximal network of hydrogen bonds : (Ac)O2H $\cdots$ O3H $\cdots$ O4H $\cdots$ O6 and NH $\cdots$ O1 which are

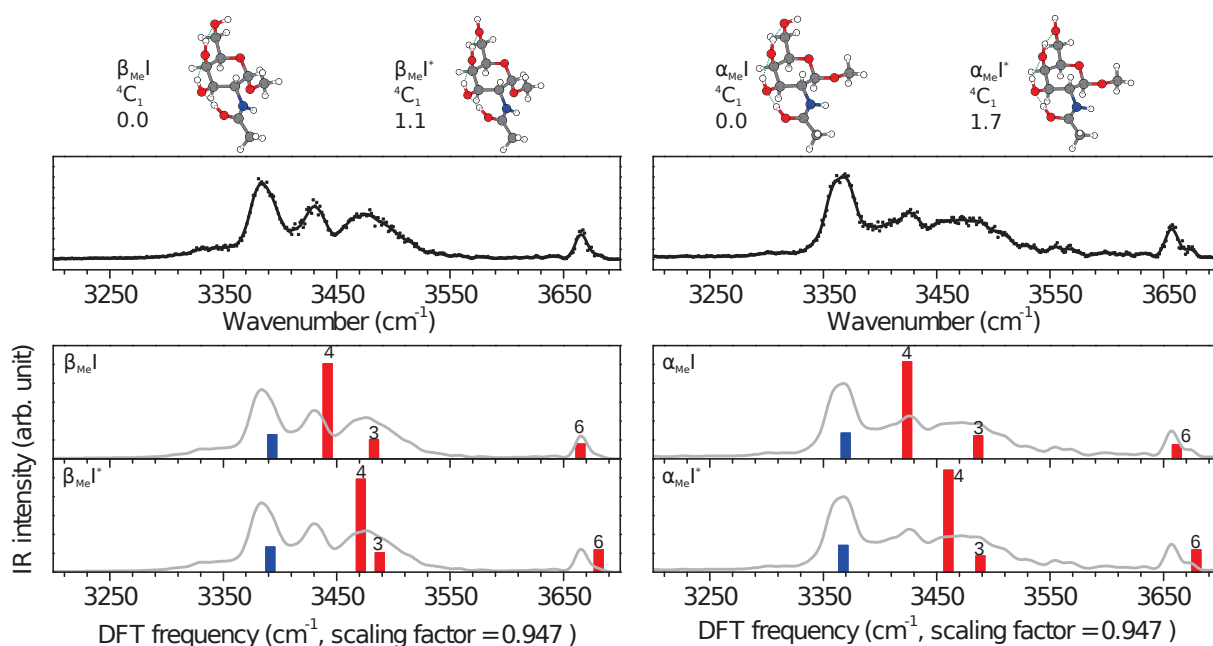


FIGURE 6.3 – Structures of the most stable conformations of α -GalNAcOMe (left panel) and β -GalNAcOMe (right panel) and their relative energy in kJ/mol calculated with MP2/6-311++G(3df,3pd). Experimental spectra of α -GalNAcOMe (middle left panel) and β -GalNAcOMe (middle right panel). Theoretical harmonic IR spectra of these structures, between 3000 and 3700 cm^{-1} , obtained with CAM-B3LYP/6-311++G(2df,2pd) and scaled with a scaling factor of 0.947. See Fig. 6.2 caption for color code.

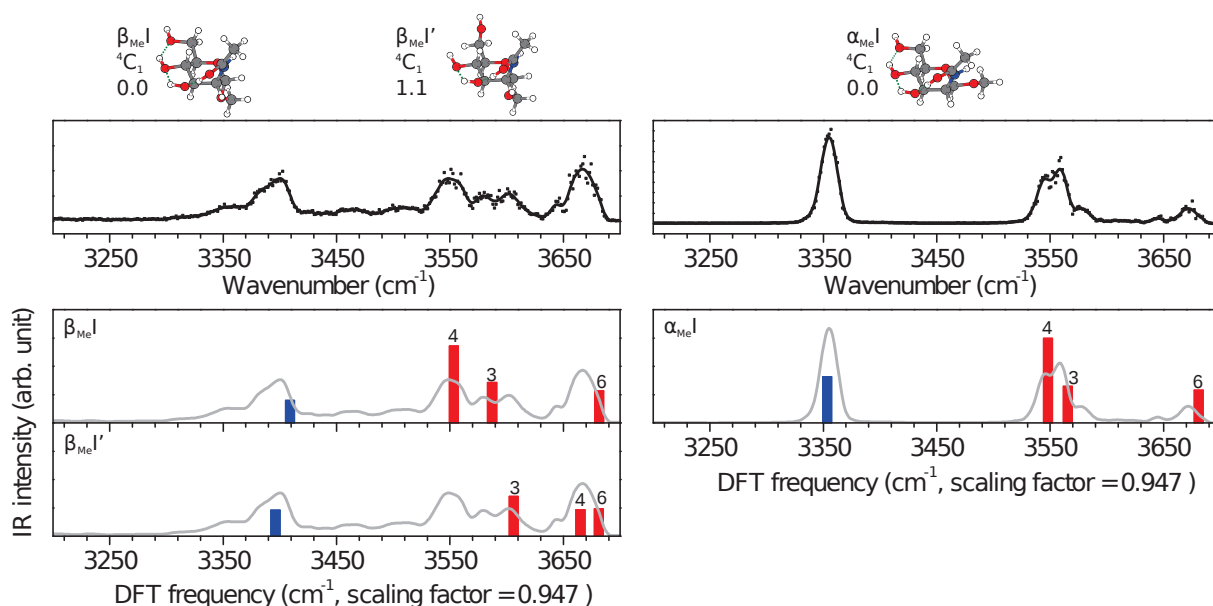


FIGURE 6.4 – Structures of the most stable conformations of α -ManNAcOMe (left panel) and β -ManNAcOMe (right panel) and their relative energy in kJ/mol calculated with MP2/6-311++G(3df,3pd). Experimental spectra of α -ManNAcOMe (middle left panel) and β -ManNAcOMe (middle right panel). Theoretical harmonic IR spectra of these structures, between 3000 and 3700 cm^{-1} , obtained with CAM-B3LYP/6-311++G(2df,2pd) and scaled with a scaling factor of 0.947. See Fig. 6.2 caption for color code.

all in equatorial position (structure $\beta_{Me}I$ in Fig. 6.2). Its variations (lower hydrogen bond coordinations and 1S_3 skew, respectively $\beta_{Me}I'$ and $\beta_{Me}I''$ in Fig. 6.2) are also present within the 30 lowest energy structures. As compared to the α anomer, the 1C_4 conformation is very unfavorable energetically due to the axial position of the OMe group above the plane (not shown). Finally, O3H and OMe in equatorial positions surround the acetyl group in a highly symmetrical environment. As a result, each β GlcNAcOMe conformer has a stable counterpart with an opposite orientation of the acetyl group ($\beta_{Me}II$, $\beta_{Me}II'$ and $\beta_{Me}II''$ in Fig. 6.2).

In order to further assess these conformations, their calculated IR spectra are compared to the IRMPD spectra in Fig. 6.2 and Fig. SI6. For α GlcNAcOMe the spectra of structures $\alpha_{Me}I$ and $\alpha_{Me}II$ show distinctive patterns of OH's and NH stretching modes : structure $\alpha_{Me}I$ matches the experimental spectrum whereas structure $\alpha_{Me}II$ (and its variants $\alpha_{Me}II'$ and $\alpha_{Me}II''$) can be unambiguously discarded. Comparison between the spectra of structure $\alpha_{Me}I$ and $\alpha_{Me}I'$ (which share the same 4C_1 conformation) confirms that the IR fingerprint is very sensitive diagnostic of the variations of the H-bonding network. Indeed, in $\alpha_{Me}I$ O4H is H-bonded to O6 and has a low frequency (3544 cm^{-1}) in excellent agreement with the experimental spectrum. In contrast, in $\alpha_{Me}I'$ O4H is free and has a higher frequency (3676 cm^{-1}). In $\alpha_{Me}I''$ the hydrogen bond network is preserved and the ring is distorted. Yet, its NH, O3H and O4H frequencies are significantly shifted from these of $\alpha_{Me}I$ which provides a direct diagnostic of the ring conformation.

For β GlcNAcOMe (see Fig. 6.2 and Fig. SI7) the spectrum of structure $\beta_{Me}I$ matches the main features of the experimental spectrum. The variants $\beta_{Me}I'$ and $\beta_{Me}I''$ show distinctive spectroscopic fingerprints and a lesser match with the experimental spectrum. Comparison of the spectra of $\beta_{Me}I$ and $\beta_{Me}II$ is very interesting : although they share the same ring conformation, the rotation of the acetyl group results in a significant blueshift of O3H and O4H stretching frequencies, which accounts for the small experimental feature at 3610 cm^{-1} .

To summarize, the spectroscopic fingerprint in the OH/NH stretching range offers a refined diagnostic of the carbohydrate structure, including the ring conformation (4C_1 vs. 1C_4), the network of stabilizing H bonds (full coordination vs. lower coordination) and the distortion of the ring (chair vs. skew). It is found that a single 4C_1 structure dominates the IRMPD spectrum of α GlcNAcOMe ($\alpha_{Me}I$), in agreement with the lowest energy structure obtained at the MP2 level. Two 4C_1 variants differing by the orientations of the acetyl group contributes to the spectrum of β GlcNAcOMe ($\beta_{Me}I$ and $\beta_{Me}II$), again corresponding to the two close energetic minima obtained at the MP2 level ($\Delta E = 4.3\text{ kJ/mol}$). Note that for both anomers, minor spectroscopic features are measured at 3340 and 3640 cm^{-1} and are not accounted for by the proposed structures. We have performed anharmonic calculations with a method previously developed [30, 31] using CAM-B3LYP/6-311++G(2df,2pd) as the harmonic level and MMFF94 [32] for the anharmonic part with HDCPT2 [33] to calculate the anharmonic frequencies for the lowest energy conformer of α GlcNAcOMe (fig. SI 18). No anharmonic mode (overtones or combination bands) can explain the non attributed experimental features observed at 3340 and 3640 cm^{-1} . Hence the presence of conformations with higher energies is probable. The IR spectra of the 30 lowest energy structures of α GlcNAcOMe and β GlcNAcOMe are shown in Figures SI6 and SI7 for completeness. Note also that the CH stretching pattern in the 2900 cm^{-1} region was measured but not used for structural analysis. The spectra recorded between 2700 and 3700 cm^{-1} are shown in Fig. SI1.

Based on the congruence of the energetic and spectroscopic criteria in the case of 1-*O*-

α GlcNAcMe				α GlcNAc			
$\alpha_{Me}I$	4C_1	full coordination	0.0	αI	4C_1	full coordination	0.0
β GlcNAcMe				β GlcNAc			
$\beta_{Me}I$	4C_1	full coordination	0.0	βI	4C_1	full coordination	0.0
$\beta_{Me}II$	4C_1	reverse Acetyl orientation	4.3	βII	4C_1	reverse Acetyl orientation	3.7
α GalNAcMe				α GalNAc			
$\alpha_{Me}I$	4C_1	full coordination	0.0	αI	4C_1	full coordination	0.0
$\alpha_{Me}I^*$	4C_1	rotation O6H	1.1	αI^*	4C_1	rotation O6H	0.8
β GalNAcMe				β GalNAc			
$\beta_{Me}I$	4C_1	full coordination	0.0	βI	4C_1	full coordination	0.5
$\beta_{Me}I^*$	4C_1	rotation O6H	1.7	βI^*	4C_1	rotation O6H	0.0
α ManNAcMe				α ManNAc			
$\alpha_{Me}I$	4C_1	full coordination	0.0	αI	4C_1	full coordination	0.0
$\alpha_{Me}I'$	4C_1	lower coordination	1.1	$\alpha I'$	4C_1	lower coordination	0.8
β ManNAcMe				β ManNAc			
$\beta_{Me}I$	4C_1	full coordination	0.0	βI	4C_1	full coordination	0.0

TABLE 6.1 – Conformations contributing to the IRMPD spectra [ring conformation/ structural variations/ relative energies in kJ/mol]

methyl GlcNAc, we can confidently proceed to the analysis of 1-*O*-methyl GalNAc and ManNAc.

Fig. 6.3 shows the IRMPD spectra of α GalNAcOMe (left panel) and β GalNAcOMe (right panel) together with the calculated IR spectra of their two lowest energy structures. The IR spectra of the 30 lowest energy structures of α GalNAcOMe and β GalNAcOMe are shown in Figures SI8 and SI9, respectively. For both anomers, the energetic minimum is a 4C_1 chairs stabilized by a network of hydrogen bonds between (Ac)O2H, O3H, O4H and O6 (structures $\alpha_{Me}I$ and $\beta_{Me}I$ in Fig. 6.3). The second minima ($\alpha_{Me}I^*$ and $\beta_{Me}I^*$, respectively) only differ from the lowest energy structure by the value of the H-C6-O6-H dihedral angle. The first minimum alone gives a good match for the main experimental features for both anomers. Nevertheless, the close proximity of the second minima ($\Delta E = 1.1$ kJ/mol and $\Delta E = 1.7$ kJ/mol for α GalNAcOMe and β GalNAcOMe, respectively), together with the noticeably broad experimental OH pattern, strongly suggests that both forms coexist.

For both anomers of 1-*O*-methyl ManNAc (Fig. 6.4), the most stable structure is a 4C_1 chair stabilized by a network of hydrogen bonds between (Ac)O2H, O3H, O4H and O6 ($\alpha_{Me}I$ and $\beta_{Me}I$). This first minimum alone is sufficient to account for the IRMPD spectrum of β ManNAcOMe. For α ManNAcOMe, the first minimum accounts for the main experimental features but the second minimum which shows a lower H-bond coordination ($\alpha_{Me}I'$) also contributes. None of the minima accounts however for the mild IR activity measured in the 3430-3510 cm^{-1} range and the minor contribution of higher energy structures is not excluded. The IR spectra of the 30 lowest energy structures of α ManNAcOMe and β ManNAcOMe are shown in Figures SI10 and SI11 for completeness.

Finally, the conformations contributing to the IRMPD spectra of the six 1-*O*-methyl *N*-acetyl hexosamines, which were elucidated using both energetic and spectroscopic criteria, are summarized in Table 6.1 and Fig SI16.

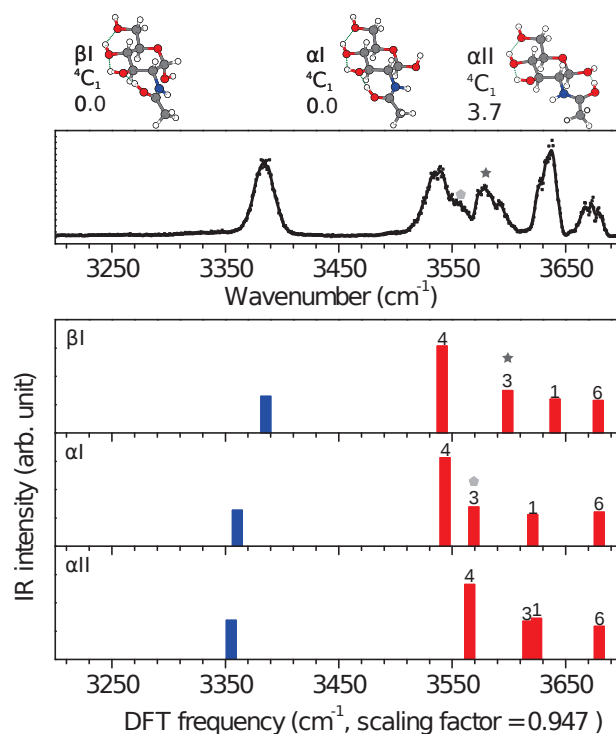


FIGURE 6.5 – Structures of the most stable conformations of α GlcNAc and β GlcNAc and their relative energy in kJ/mol calculated with MP2/6-311++G(3df,3pd). Experimental spectrum of GlcNAc. Theoretical harmonic IR spectra of these structures, between 3000 and 3700 cm^{-1} , obtained with CAM-B3LYP/6-311++G(2df,2pd) and scaled with a scaling factor of 0.947. Anomer diagnostic modes are marked with symbols (see text). See Fig. 6.2 caption for color code of the IR spectra.

6.2.3 Analysis of anomeric mixtures

In this last section, we address natural *N*-acetyl hexosamines in which both anomers are present simultaneously. Firstly, to assess the influence of the methyl group on the conformation, we now compare the lowest energy structures of the 1-*O*-methyl and non-methylated species. The conformational study was carried out on individual anomers, that is α GlcNAc, β GlcNAc, α GalNAc, β GalNAc, α ManNAc and β ManNAc. The molecular dynamics yielded 335, 474, 1612, 2161, 421 and 432 structures, respectively. They converged into 179, 166, 342, 356, 131 and 156 stable conformers after DFT optimization. Their energies were calculated by MP2 and the 30 lowest energy conformations were selected for conformational analysis. Tables SI1, SI2, SI11, SI12, SI13 and SI14 summarize the energies and conformations obtained for each system.

For α GlcNAc the lowest energy conformation is a 4C_1 chair (structure αI in Fig. 6.5) which corresponds exactly to the most stable structure for α GlcNAcOMe. For β GlcNAc, the two lowest energy conformations are 4C_1 chairs (structures βI and βII in Fig. 6.5) differing by the orientation of the acetyl group as for β GlcNAcOMe. For both α GalNAc and β GalNAc two 4C_1 chairs (structures αI and αI^* for α GalNAc and structures βI and βI^* for β GalNAc in Fig. 6.6) differing only by the value of the H-C6-O6-H dihedral angle are found at low energy, as for their 1-*O*-methyl counterparts. Similarly, for α ManNAc and β ManNAc the lowest energy conformations (structures αI and $\alpha I'$ for α ManNAc and structure βI for β ManNAc in Fig. 6.7) are equivalent to these found for α ManNAcOMe and β ManNAcOMe.

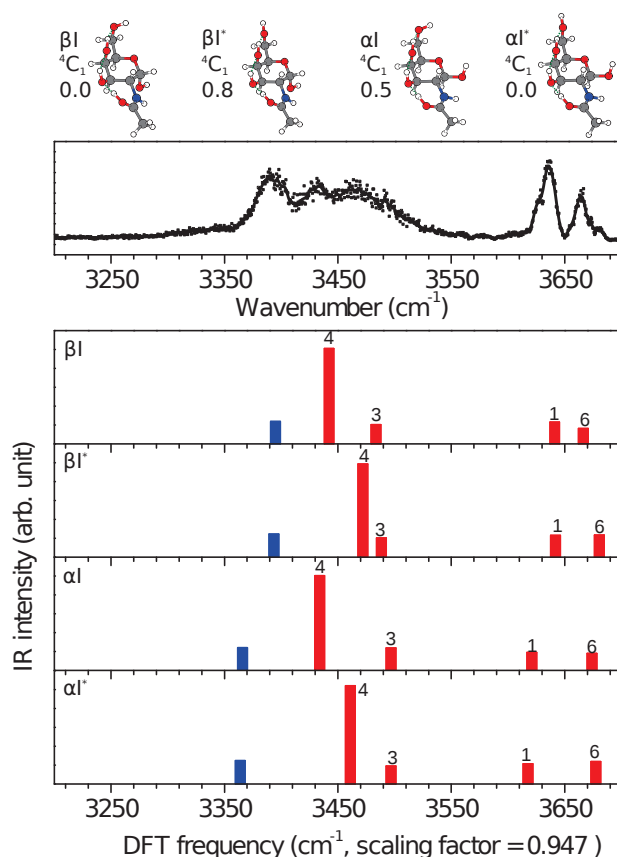


FIGURE 6.6 – Structures of the most stable conformations of α GalNAc and β GalNAc and their relative energy in kJ/mol calculated with MP2/6-311++G(3df,3pd). Experimental spectrum of GalNAc. Theoretical harmonic IR spectra of these structures, between 3000 and 3700 cm^{-1} , obtained with CAM-B3LYP/6-311++G(2df,2pd) and scaled with a scaling factor of 0.947. See Fig. 6.2 caption for color code of the IR spectra.

As shown in Table 6.1, the energy gap between the first and second lowest energy structures tends to be reduced for non-methylated species. In the case of β GalNAc, the order is of the two lowest energy conformers is reversed. Thus, methylation at the reducing end does not affect the conformational preference of protonated hexosamines. This validates that 1-*O*-methyl species are relevant models to address the more complex case of *N*-acetyl hexosamines anomeric mixtures.

For direct comparison with experiment, the spectra of the HexNAc conformers corresponding to the forms previously identified for their 1-*O*-methyl counterparts are reported in figures 6.5, 6.6 and 6.7 (additional theoretical spectra are reported in figures SI2, SI3, SI12, SI13, SI14 and SI15). Unsurprisingly, these spectra resemble these of the corresponding 1-*O*-methyl forms with one exception : a O1H stretching mode is now present between 3620 and 3650 cm^{-1} , matching the intense experimental feature at around 3640 cm^{-1} . In each case, the combination of the most stable α and β conformers is in agreement with the experimental spectra.

Finally, we attempt to identify anomer-diagnostic signatures in the theoretical spectra, and to retrieve their contribution in the experimental spectra. GlcNAc is shown in Fig. 6.5. Theory shows that the inversion of anomericity in the lowest energy structures (structures αI and βI) affects the frequency of the O3H mode by $\delta\nu = 30 \text{ cm}^{-1}$. Other frequencies are unchanged (or lesser than the experimental resolution). The α diagnostic frequency

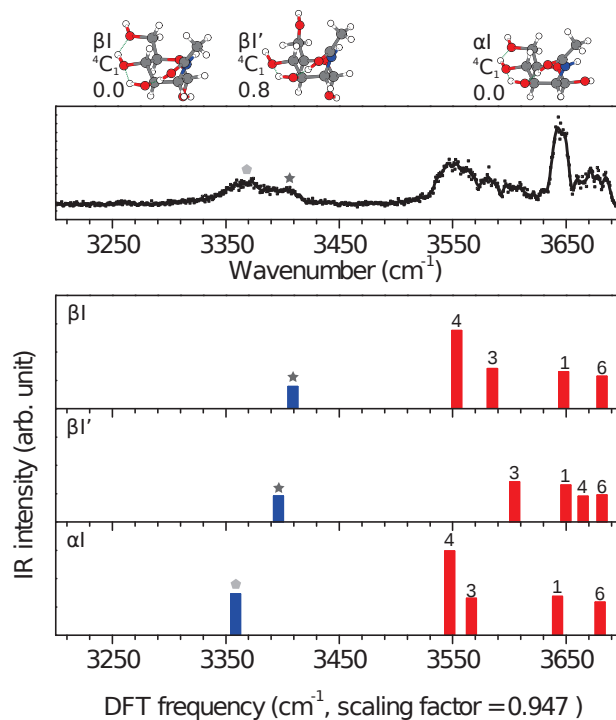


FIGURE 6.7 – Structures of the most stable conformations of α ManNAc and β ManNAc and their relative energy in kJ/mol calculated with MP2/6-311++G(3df,3pd). Experimental spectrum of ManNAc. Theoretical harmonic IR spectra of these structures, between 3000 and 3700 cm^{-1} , obtained with CAM-B3LYP/6-311++G(2df,2pd) and scaled with a scaling factor of 0.947. The anomeric-specific diagnostic is drawn on the IRMPD spectra in shaded grey. Anomer diagnostic modes are marked with symbols (see text). See Fig. 6.2 caption for color code of the IR spectra.

($\nu_{O3H}^{\alpha} = 3599 \text{ cm}^{-1}$, marked with a dark grey symbol) is clearly identified as the band at 3580 cm^{-1} in the experimental spectrum. The β diagnostic frequency ($\nu_{O3H}^{\beta} = 3569 \text{ cm}^{-1}$, marked with a light grey symbol) appears in the IRMPD spectrum at 3550 cm^{-1} , although partly unresolved from the O4H mode at 3540 cm^{-1} . Note that form βII , which was previously observed as a minor component of β GlcNAcOMe, is masked by the contributions of αI and βI . Indeed, its diagnostic O3H and O4H modes are masked by more dominant features (O1H and βI O3H, respectively). GalNAc is shown in Fig. 6.6. As in the previous section, the experimental spectrum is very broad around 3450 cm^{-1} , which was justified by the coexistence of two very close rotamers in the case of the 1-*O*-methyl species. Here, this unfortunately prevents the identification of anomeric-specific features. ManNAc is shown in Fig. 6.7. Theory shows that the anomeric configuration significantly affects the NH frequency. The α diagnostic NH frequency (marked with a dark grey symbol) is attributed to the band around 3400 cm^{-1} in the experimental spectrum. The β diagnostic frequency (marked with a light grey symbol) appears in the IRMPD spectrum at 3370 cm^{-1} . Anomeric contributions can thus be retrieved from heterogeneous IRMPD spectra in the case of GlcNAc (via the diagnostic frequency of the O3H mode) and ManNAc (via the diagnostic frequency of the NH mode).

6.3 Conclusion

In this work we have studied the structure of gas phase 1-*O*-methyl and non-methylated *N*-acetyl-D-hexosamines (GlcNAc, GalNAc and ManNAc) in their protonated charge state, using a combination of IRMPD spectroscopy and *ab-initio* calculations. Based on energetic and spectroscopic arguments built upon DFT calculations and IRMPD, we have shown that the protonation occurs on the carbonyl function for GlcNAc. For all other *N*-acetyl-D-hexosamine ions, we verified that the position of the diagnostic NH mode is always consistent with the protonation of the carbonyl function. Stable conformations of individual 1-*O*-methyl anomers of GlcNAc, GalNAc and ManNAc were proposed based on energetic criterion and validated by spectroscopy. While all isomers share a 4C_1 chair conformation of the carbohydrate ring, several structural variations were observed (coordination and orientation of the acetyl group). Interestingly, the structural heterogeneity depends on the epimer and on the anomeric configuration. The conformational preference however, is not affected by the methylation at the reducing end, as demonstrated in the last section. Instead, 1-*O*-methyl forms proved to be a relevant model to resolve the conformational and anomeric heterogeneity of natural *N*-acetyl-D-hexosamines. We finally proposed a spectroscopic diagnostic to identify the contributions of α and β anomers in the IRMPD spectra of GlcNAc and ManNAc. The capability of IRMPD spectroscopy to overcome the intrinsic limitation of Mass Spectrometry for the identification of carbohydrate isomers and anomers holds a great potential for the development of glycoanalysis.

Supplementary material

Online material includes IRMPD spectra in the 2800-3700 cm^{-1} spectral range; lists of low energy conformations and their calculated IR spectra.

Acknowledgement

This work was granted access to the HPC resources of the FLMSN, "Fédération Lyonnaise de Modélisation et Sciences Numériques", partner of EQUIPEX EQUIP@MESO and to the "Centre de calcul CC-IN2P3" at Villeurbanne, France. The authors are members of the Glycophysics Network (<http://www.glyms.univ-lyon1.fr>).

6.4 References

- [1] F. O. Talbot and J. P. Simons. Sugars in the gas phase : the spectroscopy and structure of jet-cooled phenyl A-D-glucopyranoside. *Phys. Chem. Chem. Phys.*, 4(15).
- [2] Oscar Hernandez, Samantha Isenberg, Vincent Steinmetz, Gary L. Glish, and Philippe Maitre. Probing mobility-selected saccharide isomers : Selective ion-molecule reactions and wavelength-specific IR activation. *J. Phys. Chem. A*, 119(23) :6057–6064, 2015.

-
- [3] Sarah E. Stefan and John R. Eyler. Differentiation of methyl-glucopyranoside anomers by infrared multiple photon dissociation with a tunable CO₂ laser. *Anal. Chem.*, 81(3) :1224–1227, 2009.
- [4] Nick C. Polfer, Jose J. Valle, David T. Moore, Jos Oomens, John R. Eyler, and Brad Bendiak. Differentiation of isomers by wavelength-tunable infrared multiple-photon dissociation-mass spectrometry : Application to glucose-containing disaccharides. *Anal. Chem.*, 78(3) :670–679, 2006.
- [5] Wright L. Pearson, Cesar Contreras, David Powell, Giel Berden, Jos Oomens, Brad Bendiak, and John R. Eyler. Differentiation of rubidiated methyl-D-glycoside stereoisomers by infrared multiple-photon dissociation spectroscopy in the O-H and C-H stretching regions. *J. Phys. Chem. B*, 119(41) :12970–12981, 2015.
- [6] Yanglan Tan, Ning Zhao, Jinfeng Liu, Pengfei Li, Corey N. Stedwell, Long Yu, and Nicolas C. Polfer. Vibrational Signatures of Isomeric Lithiated *N*-acetyl-D-hexosamines by Gas-Phase Infrared Multiple-Photon Dissociation (IRMPD) Spectroscopy. *J. Am. Soc. Mass Spectrom.*, pages 1–12, 2017.
- [7] Emilio B. Cagmat, Jan Szczepanski, Wright L. Pearson, David H. Powell, John R. Eyler, and Nick C. Polfer. Vibrational signatures of metal-chelated monosaccharide epimers : gas-phase infrared spectroscopy of Rb⁺-tagged glucuronic and iduronic acid. *Phys. Chem. Chem. Phys.*, 12 :3474–3479, 2010.
- [8] Cesar S. Contreras, Nicolas C. Polfer, Jos Oomens, Jeffrey D. Steill, Brad Bendiak, and John R. Eyler. On the path to glycan conformer identification : Gas-phase study of the anomers of methyl glycosides of *N*-acetyl-D-glucosamine and *N*-acetyl-D-galactosamine. *Int. J. Mass Spectrom.*, 330–332 :285–294, 2012.
- [9] Baptiste Schindler, Gina Renois-Predelus, Nassiba Bagdadi, Sihem Melizi, Loïc Barnes, Stéphane Chambert, Abdul-Rahman Allouche, and Isabelle Compagnon. MS/IR, a new MS-based hyphenated method for analysis of hexuronic acid epimers in glycosaminoglycans. *Glycoconj. J.*, 34(3) :421–425, 2017.
- [10] Baptiste Schindler, Janhavi Joshi, Abdul-Rahman Allouche, Daniel Simon, Stéphane Chambert, Vincent Brites, Marie-Pierre Gageot, and Isabelle Compagnon. Distinguishing isobaric phosphated and sulfated carbohydrates by coupling of mass spectrometry with gas phase vibrational spectroscopy. *Phys. Chem. Chem. Phys.*, 16 :22131–22138, 2014.
- [11] Loïc Barnes, Baptiste Schindler, Abdul-Rahman Allouche, Daniel Simon, Stéphane Chambert, Jos Oomens, and Isabelle Compagnon. Anharmonic simulations of the vibrational spectrum of sulfated compounds : application to the glycosaminoglycan fragment glucosamine 6-sulfate. *Phys. Chem. Chem. Phys.*, 17 :25705–25713, 2015.
- [12] B. Schindler, L. Barnes, C. J. Gray, S. Chambert, S. L. Flitsch, J. Oomens, R. Daniel, A. R. Allouche, and I. Compagnon. IRMPD spectroscopy sheds new (infrared) light on the sulfate pattern of carbohydrates. *J. Phys. Chem. A*, 121(10) :2114–2120, 2017.
- [13] James J. P. Stewart. Optimization of parameters for semiempirical methods VI : more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.*, 19(1) :1–32, 2012.
- [14] James J. P. Stewart. Mopac2012, 2012.

-
- [15] AD Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38 :3098–3100, 1988.
- [16] AD Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98 :5648–5652, 1993.
- [17] CT Lee, WT Yang, and RG Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37 :785–789, 1988.
- [18] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange-correlation functional using the coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.*, 393(1–3) :51–57, 2004.
- [19] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self-consistent molecular orbital methods. XII. further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.*, 56(5) :2257–2261, 1972.
- [20] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople. Self-consistent molecular orbital methods. XX. a basis set for correlated wave functions. *J. Chem. Phys.*, 72(1) :650–654, 1980.
- [21] Michael J. Frisch, John A. Pople, and J. Stephen Binkley. Self-consistent molecular orbital methods 25. supplementary functions for gaussian basis sets. *J. Chem. Phys.*, 80(7) :3265–3269, 1984.
- [22] Chr. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46 :618–622, 1934.
- [23] Abdul-Rahman Allouche. Gabedit-A graphical user interface for computational chemistry softwares. *J. Comput. Chem.*, 32(1) :174–182, 2011.
- [24] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision D.01, 2009. Gaussian Inc. Wallingford CT 2009.
- [25] Frank Neese. The ORCA program system. *WIREs Comput. Mol. Sci.*, 2 :73–78, 2012.
- [26] Adam P. Cismesia, Georgina R. Nicholls, and Nicolas C. Polfer. Amine vs. carboxylic acid protonation in ortho-, meta-, and para-aminobenzoic acid : An IRMPD spectroscopy study. *J. Mol. Spectrosc.*, 332 :79–85, 2017.

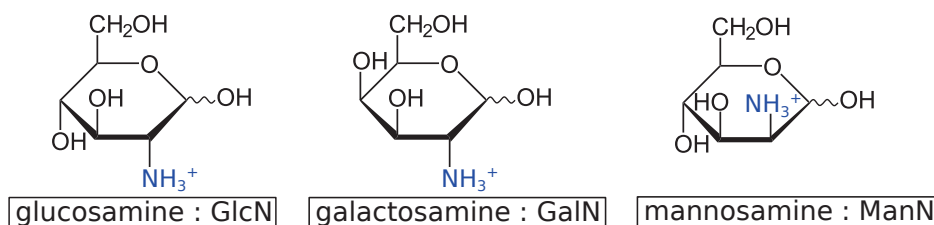
-
- [27] Adam J. Bridgeman, German Cavigliasso, Luke R. Ireland, and Joanne Rothery. The mayer bond order as a tool in inorganic chemistry. *J. Chem. Soc., Dalton Trans.*, pages 2095–2108, 2001.
- [28] Erin R. Johnson, Shahar Keinan, Paula Mori-Sánchez, Julia Contreras-García, Aron J. Cohen, and Weitao Yang. Revealing Noncovalent Interactions. *J. Am. Chem. Soc.*, 132(18) :6498–6506, 2010.
- [29] Julia Contreras-García, Erin R. Johnson, Shahar Keinan, Robin Chaudret, Jean-Philip Piquemal, David N. Beratan, and Weitao Yang. NCIPLOT : A Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput.*, 7(3) :625–632, 2011.
- [30] Loïc Barnes, Baptiste Schindler, Isabelle Compagnon, and Abdul-Rahman Allouche. Fast and accurate hybrid QM//MM approach for computing anharmonic corrections to vibrational frequencies. *J. Mol. Model.*, 22(11), 2016.
- [31] Loïc Barnes, Baptiste Schindler, Isabelle Compagnon, and Abdul-Rahman Allouche. iGVPT2. <https://sites.google.com/site/allouchear/igvpt2>, 2017.
- [32] Thomas A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.*, 17(5-6) :490–519, 1996.
- [33] Vincenzo Barone. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.*, 122(1) :014108, 2005.

The remarkable structural complexity of simple carbohydrate building blocks : conformational and anomeric heterogeneity of hexosamines resolved by IRMPD spectroscopy and *ab initio* calculations[†]

L. BARNES^a, B. SCHINDLER^a, S. CHAMBERT^b, A-R. ALLOUCHE^a AND I. COMPAGNON^{ac}

Highlights

- Assignment of the structures of gas phase glucosamine and galactosamine in their protonated charge state.
- Resolution of the anomeric and conformational signature.
- Developing and testing an approach to retrieve the relative contribution of anomers in a sample.
- Visualisation of the dynamic occuring in solution for galactosamine with IRMPD snapshots.



^a Institut Lumière Matière, Univ Lyon, Université Claude Bernard Lyon 1, CNRS, F-69622 Villeurbanne, France

^b Univ Lyon, INSA-Lyon, CNRS, Université Lyon 1, CPE Lyon, ICBMS, UMR 5246, Batiment Jules Verne, 20 avenue Albert Einstein, F-69621 Villeurbanne, France

^c Institut Universitaire de France, 103 Blvd St Michel, 75005 Paris, France. E-mail : isabelle.compagnon@univ-lyon1.fr

[†] To be submitted

7.1 Introduction

Carbohydrates play a key role in the living world. For example, they are part of the cell wall of the bacteria. They are also involved in the recognition process between cells. Hence, understanding correctly their structure is of utmost importance. Many groups have worked on analysing the structure of mono or disaccharides in the gas phase using metal adducts such as Lithium [1, 2, 3] or Rubidium [4, 5]. We have previously reported several studies of protonated [6, 7] or deprotonated [8] monosaccharides. A difficult question to answer in analysing structures of carbohydrates is the question about the relative population of conformers and anomers present in the sample. RMN can be used to answer this question but at the cost of a long period of purification or at the risk of damaging the functional modifications. Also it can only give the relative population of anomers. In this context, we present the first IRMPD study of protonated hexosamines (GlcN, GalN and ManN) and to the best of your knowledge, the first time an approach mixing IRMPD spectra and theoretical ones is used to retrieve the relative contribution of each conformer.

In section 7.3.1, we address the conformational preference of hexosamines and discuss about the conformation of monosaccharides studied in the literature. In section 7.3.2, we show by using simultaneously IRMPD spectra and theoretical spectra how we can access to the epimerization process occurring in solution for GalN. In section 7.3.3, we discuss the possibility to elucidate the conformational heterogeneity for GlcN by employing our approach.

7.2 Methods

7.2.1 IRMPD spectroscopy

IRMPD (InfraRed Multiple Photon Dissociation) spectroscopy is performed in a modified commercial quadrupolar ion trap (ThermoFinnigan LCQ classic) equipped with an electrospray source. The trapping electrode was drilled and an IR-transparent window was installed to allow irradiation of the ion cloud by the IR laser beam produced by a YAG-pumped OPO/OPA system (LaserVision) tunable in the 3 μm region. Mass-selected ions were isolated and irradiated for 700 ms, resulting in wavelength-dependent photofragmentation. The photofragmentation yield is calculated from the resulting photofragmentation spectrum using the following variation of the Beer-Lambert formula : $-\log(\frac{I_p}{I_p + \sum I_f})$, where I_p is the intensity of the precursor ion and I_f is the intensity of the photofragments. The IRMPD spectrum of the precursor ion is retrieved by monitoring its photofragmentation yield as a function of the wavelength in the 2700-3700 cm^{-1} spectral range. The data-points shown thorough the manuscript are averaged 3 times and a trend line (5-points Fourier Transform rolling averaging) is added to guide the eye.

7.2.2 Computational methods

Molecular dynamic was used to explore the potential energy surface of the protonated ions with the PM7 [9] potential in OpenMopac [10] as described in a previous work [6]. One trajectory was calculated for each anomer of the hexosamines. 1000 geometries were generated during each trajectory and optimized with PM7, yielding typically a few hundreds stable conformations after discarding the identical ones. The geometries

were further optimized with two steps of Density Functional Theory (DFT) : firstly B3LYP [11, 12, 13]/6-31G*, then CAM-B3LYP [14]/6-311++G(2df,2pd) [15, 16, 17]. This results in the reduction of the number of stable geometries. The energies are further refined by single-point calculations with MP2 [18]/6-311++G(3df,3pd). At this stage, the 20 lowest energy conformers were selected for computing harmonic frequencies with CAM-B3LYP/6-311++G(2df,2pd). An empirical scaling factor of 0.947 was used for comparison with the experimental IRMPD spectra. The mode analysis was performed using Gabedit [19], Gaussian09 [20] was used for all DFT calculations and Orca [21] for all MP2 single-point energies.

We have developed a tool to calculate the relative contribution of each anomer and/or conformer that contributes to the IRMPD spectrum. The total theoretical intensity is calculated thanks to the following formulae : $I_{tot} = \Lambda_{unit}[\gamma_1 I^{confo\ 1} + \gamma_2 I^{confo\ 2} + \dots]$. Λ_{unit} is the introduced coefficient to put IRMPD spectrum and IR one in the same unit. γ_i is the ratio of the i th conformer. $I^{confo\ i}$ is the calculated intensity for the i th conformer. It can be expressed as a sum of gaussians : $I^{confo\ i} = \sum_{i=0}^{Nm} I_i e^{-\log(2)(\frac{f_i-x}{\sigma})^2}$ with Nm being the number of modes considered, I_i the intensity of the i th mode and f_i its frequency and σ the full length half width (FLHW, here the value is fixed to 12 cm⁻¹). All the frequencies f_i are scaled by the scaling factor 0.947 for every conformer. We calculate the difference in intensity between IRMPD and theoretical spectra, $\Delta I = I_{tot} - I_{exp}$. Hence, it is possible to calculate the root-mean square deviation (RMSD), $RMSD = \sqrt{\frac{\sum \Delta I^2}{N_{points}}}$, with N_{points} being the number of points in the IRMPD spectrum. So lower this value is closer to the experiment the theory is. In optimizing the ratio γ of each conformer, we aim to reduce the RMSD. Note that all the modes corresponding to strongly bounded stretches are removed as the intensity of such modes are not well reproduced by calculations.

7.3 Results and discussion

7.3.1 Conformational preference of hexosamines

In this section, we explore the conformations of hexosamines (GlcN, GalN and ManN) to probe the heterogeneity of these compounds compared to existent literature on carbohydrates where mostly ⁴C₁ was found to be the predominant structure. In hexosamines samples, anomers α and β are present and both contribute to the IRMPD spectrum.

We have performed molecular dynamics exploration of the conformational energy landscape of α GlcN, β GlcN, α GalN, β GalN, α ManN and β ManN using the method described in Methods. The molecular dynamics yielded 79, 69, 87, 40, 108 and 114 structures, respectively. We selected the 20 lowest energy conformations obtained with B3LYP/6-31G* and optimized them with CAM-B3LYP/6-311++G(2df,2pd). The IR spectra of these selected structures were computed and the energy was refined by a MP2 single point calculation. For completeness IR spectra of the 20 structures of α GlcN, β GlcN, α GalN, β GalN, α ManN and β ManN are shown in Fig. A.1-A.6. The relative energies of the 20 lowest energy conformers of α GlcN, β GlcN, α GalN, β GalN, α ManN and β ManN are given in Tables A.1-A.6 .

For α GalN and β GalN, the lowest energy conformer is a ⁴C₁ chair with hydrogen bonds between O3H and O4H and between O4H and O6H.

For α GlcN, the lowest energy conformation is not a ⁴C₁ chair but a ¹C₄ chair with an

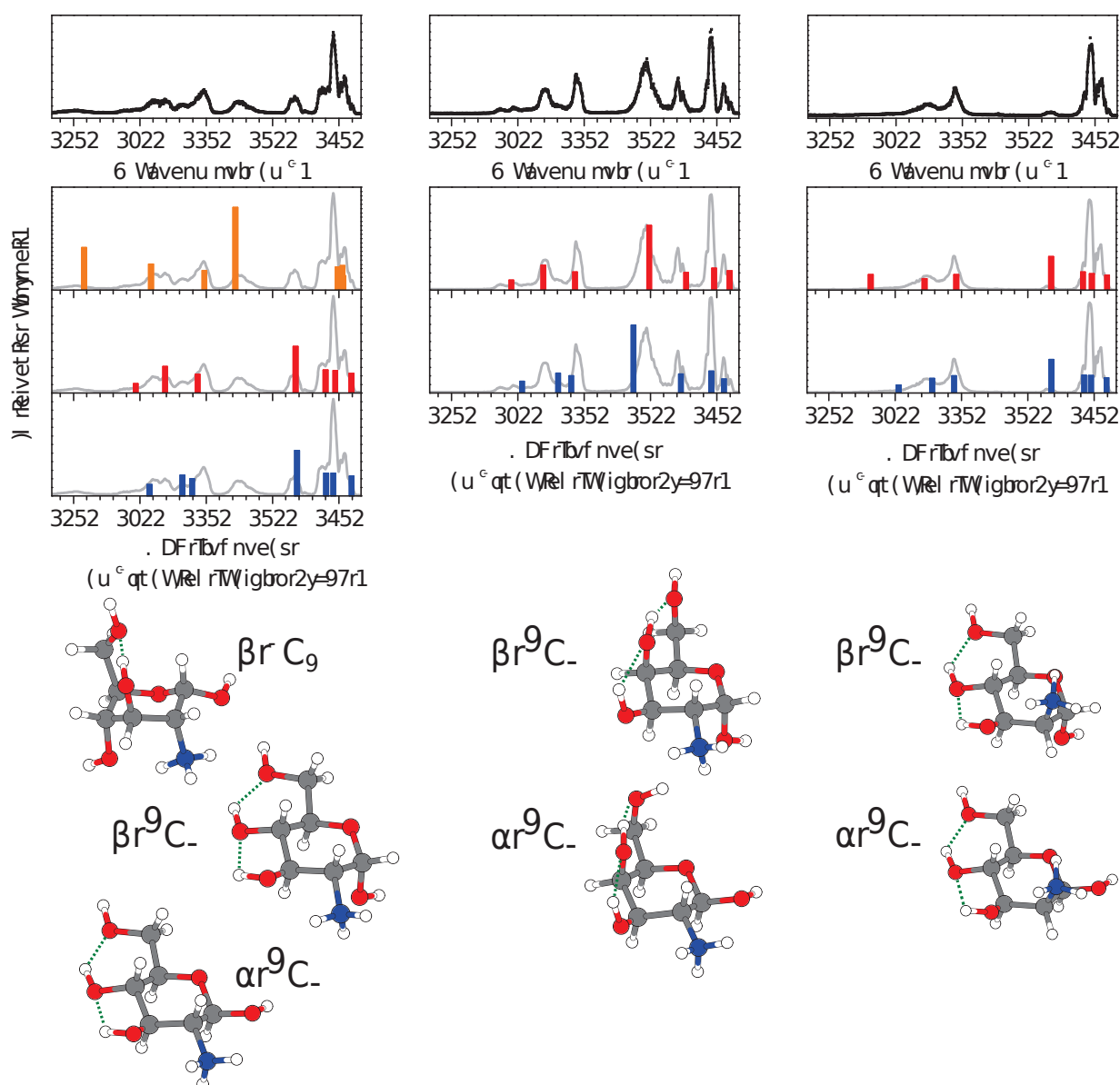


FIGURE 7.1 – Structures of the most stable conformations for α GlcN, the 4C_1 and 1C_4 chairs, and the most stable conformation of β GlcN, a 4C_1 chair (left panel), structures of the most stable conformations of α GalN and β GalN, 4C_1 chairs (middle panel) and structures of the 4C_1 conformations of α ManN and β ManN (right panel). Experimental spectra of GlcN, GalN and ManN on top of the figure (left, middle and left, respectively). Theoretical harmonic IR spectra of these structures, between 3000 and 3700 cm^{-1} , obtained with CAM-B3LYP/6-311++G(2df,2pd) and scaled with a scaling factor of 0.947. For each IR spectrum, the equivalent experimental spectrum is plotted in gray for comparison. The stretching modes of the 1C_4 of α GlcN are colored in orange, those of the 4C_1 chair for anomer α of each hexosamine are colored in red and those of the 4C_1 chair for anomer β of each hexosamine are colored in blue.

hydrogen bond between O3H and O6. The second lowest energy structure ($\Delta E = 7.1$ kJ/mol) is, however a 4C_1 chair with a network of hydrogen bonds between O3H, O4H and O6H. For β GlcN, the most stable conformer is a 4C_1 chair sharing the same hydrogen bonds than the 4C_1 of α GlcN.

For α ManN and β ManN, the minima are not 4C_1 chairs. In the case of α ManN, the lowest energy conformation is a 2H_1 half-chair with the amine group in radial position forming an hydrogen bond with O6. For the latter, β ManN, the lowest energy structure is a 3S_1 skew with an hydrogen bond between O3H and O6H. Regarding the 4C_1 chair, α ManN has a 4C_1 chair with a relative MP2 energy of $\Delta E = 2.9$ kJ/mol and presents a network of hydrogen bonds between O3H, O4H and O6H. As for β ManN, the second and third lowest energy conformations are two 4C_1 chairs. The first one is not fully coordinated contrary to the second one which has hydrogen bonds between O3H and O4H and between O4H and O6H.

The energetic criterion indicates that a 4C_1 chair is not always the most stable structure for these hexosamines.

Now we analysis the vibrational spectra of the structures cited before to check if other structures than the 4C_1 are required to explain the IRMPD spectra. Fig. 7.1 shows the spectra of the 4C_1 of α GalN and β GalN. They both contribute to the three experimental bands at around 3564, 3636 and 3670 cm^{-1} . The value of the O4H stretching mode is shifted between anomers but the experimental resolution is not sufficient to distinguish them. This stretching mode corresponds to the large experimental band (with a width of 35 cm^{-1} at half length) centered at 3491 cm^{-1} . The high asymmetric NH_3 stretching mode of α GalN and β GalN matches the experimental feature at 3335 cm^{-1} . The low asymmetric NH_3 stretching mode of α GalN reproduces the band at 3261 cm^{-1} while the one for β GalN reproduces the low intensity band at 3292 cm^{-1} . The NH_3 symmetric mode of both anomers, although a bit blue-shifted reproduces the two features at around 3161 and 3191 cm^{-1} . Energetic and spectroscopic evidences indicate that 4C_1 chairs are sufficient to explain the IRMPD spectrum of GalN.

The IR spectra of the 4C_1 of α GlcN and β GlcN (left panel) are shown in Fig. 7.1. The OH's stretching modes pattern is similar between the two chairs. Both 4C_1 chairs contribute to the experimental band at 3550 cm^{-1} (O4H stretching mode). However, the NH_3 pattern is tightened in the case of β GlcN. The NH_3 modes of the α 4C_1 match the IRMPD features at 3258 and the wide but low intensity around 3174 cm^{-1} . The NH_3 modes of the β 4C_1 match the experimental bands at 3230 and 3294 cm^{-1} . Both 4C_1 chairs present a NH_3 stretching mode corresponding to the shoulder around 3327 cm^{-1} . Taking into account only the two 4C_1 chairs is not sufficient to reproduce all the experimental features. Indeed, the bands at 3057, 3345 and 3428 cm^{-1} are not reproduced. Adding the contributions of the 1C_4 chair of α GlcN is necessary. The NH_3 symmetric, the high NH_3 asymmetric and the O3H stretching modes of the 1C_4 chair match the missing experimental bands (3057, 3345 and 3428 cm^{-1} , respectively). Thus, energetic and spectroscopic criteria support the fact that for GlcN, in top of the 4C_1 chairs, a 1C_4 chair is also present in the sample.

The IR spectra of the 4C_1 of α ManN and β ManN (right panel) are shown in Fig. 7.1. Regarding the OH's stretching modes, both chairs possess the same signature. They both had a O4H stretching mode matching the experimental band at 3548 cm^{-1} . No other close minima has a OH stretching that reproduces this feature. The two NH_3 asymmetric stretching modes are similar and match the bands at 3272 and 3334 cm^{-1} . The symmetric NH_3 stretching mode is different between the two anomers. But there is no match with the IRMPD spectrum. The intensities of the O4H stretching mode and of the symmetric NH_3 stretching mode indicate that 4C_1 are present but are not the most abundant in ManN. However, the resolution of the experimental spectrum and the similarity between IR spectra of the lowest energy conformers does not give us a clear picture of the mixture.

The spectroscopic analysis has brought into light anomer-specific signature for GlcN and

7.3. RESULTS AND DISCUSSION

hexoses					
Name	Charge state	Lowest structures	Level of theory	Attributed structures	References
α GlcMe	Rb ⁺	4C_1	B3LYP/6-31+G(d)/SDD	2 4C_1 (the 2 lowest)	[5]
β GlcMe	Rb ⁺	4C_1	B3LYP/6-31+G(d)/SDD	not said (first one plus others are possible) ?	[5]
α GalMe	Rb ⁺	4C_1	B3LYP/6-31+G(d)/SDD	not said (fourth one (1C_4) to explain a band	
β GalMe	Rb ⁺	4C_1	B3LYP/6-31+G(d)/SDD	low. struct 4C_1 second is a skew close in energy has to populate somehow	[5] [5] [5]
α GlcMe	H ₂ O-Li ⁺	4C_1	B3LYP/6-31G(d,p)	4C_1	[2]
β GlcMe	H ₂ O-Li ⁺	4C_1	B3LYP/6-31G(d,p)	4C_1	[2]
α ManMe	H ₂ O-Li ⁺	4C_1	B3LYP/6-31G(d,p)	4C_1	[2]
hexosamines					
Name	Charge state	Lowest structures	Level of theory	Attributed structures	References
GlcN-6P	H ⁺	4C_1 (α and β)	B3LYP/6-311+G(d)	4C_1 (α and β)	[6]
GlcN-6S	H ⁺	4C_1 (α and β)	B3LYP/6-311+G(d)	4C_1 (α and β)	[6]
GlcN-3S	H ⁺	4C_1 (α and β)	CAM-B3LYP/6-311++G(2df,2pd)	4C_1 (α and β)	[8]
GlcN-6S	H ⁻	4C_1 (α and β)	CAM-B3LYP/6-311++G(2df,2pd)	4C_1 (α and β)	[8]
GlcN-3S	H ⁻	4C_1 (α and β)	CAM-B3LYP/6-311++G(2df,2pd)	4C_1 (α and β)	[8]
GlcN	H ⁺	1C_4 (α), 4C_1 (β)	CAM-B3LYP/6-311++G(2df,2pd)	1C_4 , 4C_1 for α and 4C_1 for β	this work
GalN	H ⁺	4C_1 (α and β)	CAM-B3LYP/6-311++G(2df,2pd)	4C_1 (α and β)	this work
ManN	H ⁺	2H_1 (α), 3S_1 (β)	CAM-B3LYP/6-311++G(2df,2pd)	complex but 4C_1 may be present	this work
N-acetyl hexosamines					
Name	Charge state	Lowest structures	Level of theory	Attributed structures	References
α GlcNAcMe	H ⁺	4C_1	CAM-B3LYP/6-311++G(2df,2pd)	4C_1	[7]
β GlcNAcMe	H ⁺	4C_1	CAM-B3LYP/6-311++G(2df,2pd)	2 4C_1 rotation acetyl group	[7]
α GalNAcMe	H ⁺	4C_1	CAM-B3LYP/6-311++G(2df,2pd)	2 4C_1	[7]
β GalNAcMe	H ⁺	4C_1	CAM-B3LYP/6-311++G(2df,2pd)	2 4C_1	[7]
α ManNAcMe	H ⁺	4C_1	CAM-B3LYP/6-311++G(2df,2pd)	2 4C_1 + other structures	[7]
β ManNAcMe	H ⁺	4C_1	CAM-B3LYP/6-311++G(2df,2pd)	1 4C_1	[7]
GlcNAc	H ⁺	4C_1 (α and β)	CAM-B3LYP/6-311++G(2df,2pd)	4C_1 (α and β)	[7]
GalNAc	H ⁺	4C_1 (α and β)	CAM-B3LYP/6-311++G(2df,2pd)	4C_1 (α and β)	[7]
ManNAc	H ⁺	4C_1 (α and β)	CAM-B3LYP/6-311++G(2df,2pd)	4C_1 (α and β)	[7]
α GlcNAcMe	Li ⁺	4C_1	B3LYP/6-311++G(d,p)	4C_1	[1]
β GlcNAcMe	Li ⁺	B_{O3}	B3LYP/6-311++G(d,p)	B_{O3} + chair	[1]
α GalNAcMe	Li ⁺	4C_1	B3LYP/6-311++G(d,p)	4C_1	[1]
β GalNAcMe	Li ⁺	4C_1	B3LYP/6-311++G(d,p)	4C_1	[1]
GlcNAc	Li ⁺	4C_1 (α and β)	B3LYP/6-31+G(d,p)	4C_1 (α and β)	[3]
GalNAc	Li ⁺	4C_1 (α and β)	B3LYP/6-31+G(d,p)	4C_1 (only α ?)	[3]
ManNAc	Li ⁺	1C_4 (α and β)	B3LYP/6-31+G(d,p)	1C_4 (β predominant)	[3]
uronic acid					
Name	Charge state	Lowest structures	Level of theory	Attributed structures	References
GlcA	Rb ⁺	4C_1 (α and β)	B3LYP/6-31+G(d,p)/SDD	4C_1 (α , β major contrib.)	[4]
IdoA	Rb ⁺	1C_4 (α and β)	B3LYP/6-31+G(d,p)/SDD	1C_4 second α and first β	[4]

TABLE 7.1 – Summarizing table of all structural analysis of carbohydrates in the gas phase

GalN. We will sum it up in the following. For GalN the bands located at 3161 and 3261 cm^{-1} characterize the anomer α and the bands at 3191 and 3292 cm^{-1} are diagnostic of the anomer β . All these bands are attributed to NH_3 stretching modes. Concerning GlcN, there are five experimental features specific to the anomer α : 2 from the 4C_1 chair, which are the broad band around 3174 and the band at 3258 cm^{-1} both corresponding to NH_3 stretching modes; and 3 from the 1C_4 chair, the low intensity band at 3057 (NH_3 stretching mode), the bands at 3345 (NH_3 stretching mode) and 3428 cm^{-1} (O_3H stretching mode). Only the band at 3294 cm^{-1} which is a NH_3 stretching mode is specific to the anomer β . Unfortunately, for ManN, it was not possible to highlight any anomeric specificity. In Table 7.1, we report all the stable structures found by DFT calculations and the structures validated by IR spectroscopy of gas phase ions of carbohydrates in the literature. We separate the studied carbohydrates in four categories, the first one corresponds to the hexoses which have their methyl-blocked α and β protonated with a rubidium cation or with the complex $\text{H}_2\text{O-Li}^+$. The second are the hexosamines or hexosamines substituted either by a sulfate function on the carbon in position 3 or 6 in a protonated or deprotonated charge state or a phosphate on the carbon in position 6. The third category is composed of the *N*-acetyl hexosamines, 1-O-methylated or not and protonated or lithiated. The iduronic and glucuronic acid protonated with a rubidium cation constitute the forth category. We have shown for GalN that only the lowest energy

structure of anomer α and the lowest energy structure of anomer β , 2 4C_1 , are sufficient to reproduce the IRMPD spectrum. For other cases of Gal from Table 7.1, the predicted lowest energy structure for α and β is always a 4C_1 chair. The attributed structures are almost always a 4C_1 chair except for $\text{Rb}^+-\alpha\text{GalMe}$ where a 1C_4 chair is needed to account for some features of the experimental spectrum. For Glc and his derivatives, apart from $\text{Li}^+\beta\text{GlcNAcMe}$ where a boat B_{O3} was predicted as the minimum, a 4C_1 conformation was always the lowest energy structure. Regarding the attributed structures, it is almost everytime a 4C_1 chair. In some cases, in top of the anomeric heterogeneity, several conformers of the same anomer contribute. For $\text{Rb}^+-\alpha\text{GlcMe}$, two 4C_1 conformers need to be taken into account to match the experimental spectrum. For the particular case of $\text{Li}^+\beta\text{GlcNAcMe}$, the stable boat B_{O3} and a chair contribute to the spectrum. For $\text{Rb}^+-\beta\text{GlcMe}$, several 4C_1 conformers may be present in the sample. And finally for $\text{H}^+\beta\text{GlcNAcMe}$, we showed that two 4C_1 chairs are required to account for the bands and the broadening of the IRMPD spectrum. In the case of GlcA, the heterogeneity is lesser, the single structure of anomer β is predominant in the mixture. In this work, we established that for αGlcN , a 1C_4 chair is the most stable structure while for βGlcN it is a 4C_1 chair. We also demonstrated the heterogeneity of conformations for αGlcN . Indeed, in addition to the presence of the 1C_4 chair, the second lowest energy structure, *i.e.* a 4C_1 chair, is required to explain all the features of the IRMPD spectrum. The most stable structure (4C_1) predicted for βGlcN is also present in the mixture. The lowest energy structures for both anomers of ManN are not 4C_1 chairs and are not even 1C_4 chairs like Li^+ManNAc . For $\text{Rb}^+-\alpha\text{ManMe}$, for both anomers of $\text{H}^+\text{ManNAcMe}$ and for both anomers of H^+ManNAc , 4C_1 conformers are the lowest energy structures. The close proximity of several conformers and their similar IR signature for both anomers of ManN does not allow a clear attribution of the structures. But, we proved that 4C_1 are not the major constituents but are present in the sample. For Li^+ManNAc , the attributed structures are the 1C_4 chairs and the anomer β was shown to be predominant. For $\text{H}^+\alpha\text{ManNAcMe}$, an other 4C_1 has to be considered to best match the IRMPD spectrum. For all other cases, the 4C_1 predicted as the lowest energy conformation is the attributed structure to the experimental spectrum.

7.3.2 Capture of the dynamic in solution by IRMPD snapshots

In the previous section, we pointed out the complex heterogeneity in anomers and in conformers. This hererogeneity could make the analysis of the experimental spectrum a bit tedious. In addition, when comparing separately each conformers and anomers to the experimental spectrum, the population are never taken into account. In this section we evaluate a theoretical tool described in Methods to have access to the population of anomers in GalN. The RMN experiments revealed that for GalN in solution a new equilibrium is reached. Indeed, at the beginning of the experiment the ratio between anomer α and anomer β is 7/93. After five hours, the relative contribution of β has collapsed to 53%, almost divided by 2, while the population of anomer α was multiplied by 7. After eight hours, the new equilibrium is reached, the ratio is 56/44 and will not change (an other point was taken after 3 days and the same ratio was found). Here, we see that in solution an epimerization occurs that leads to a change in the equilibrium configuration. We measured from the same sample of GalN (prepared in solid phase), several IRMPD spectra at different times (measured every forty minutes). In order to retrieve the relative population between anomer α and anomer β we used our proposed theoretical method described in Methods. Fig 7.2 shows the fitted relative contribution of anomer β for each IRMPD spectrum in function of time. The curve of the calculated

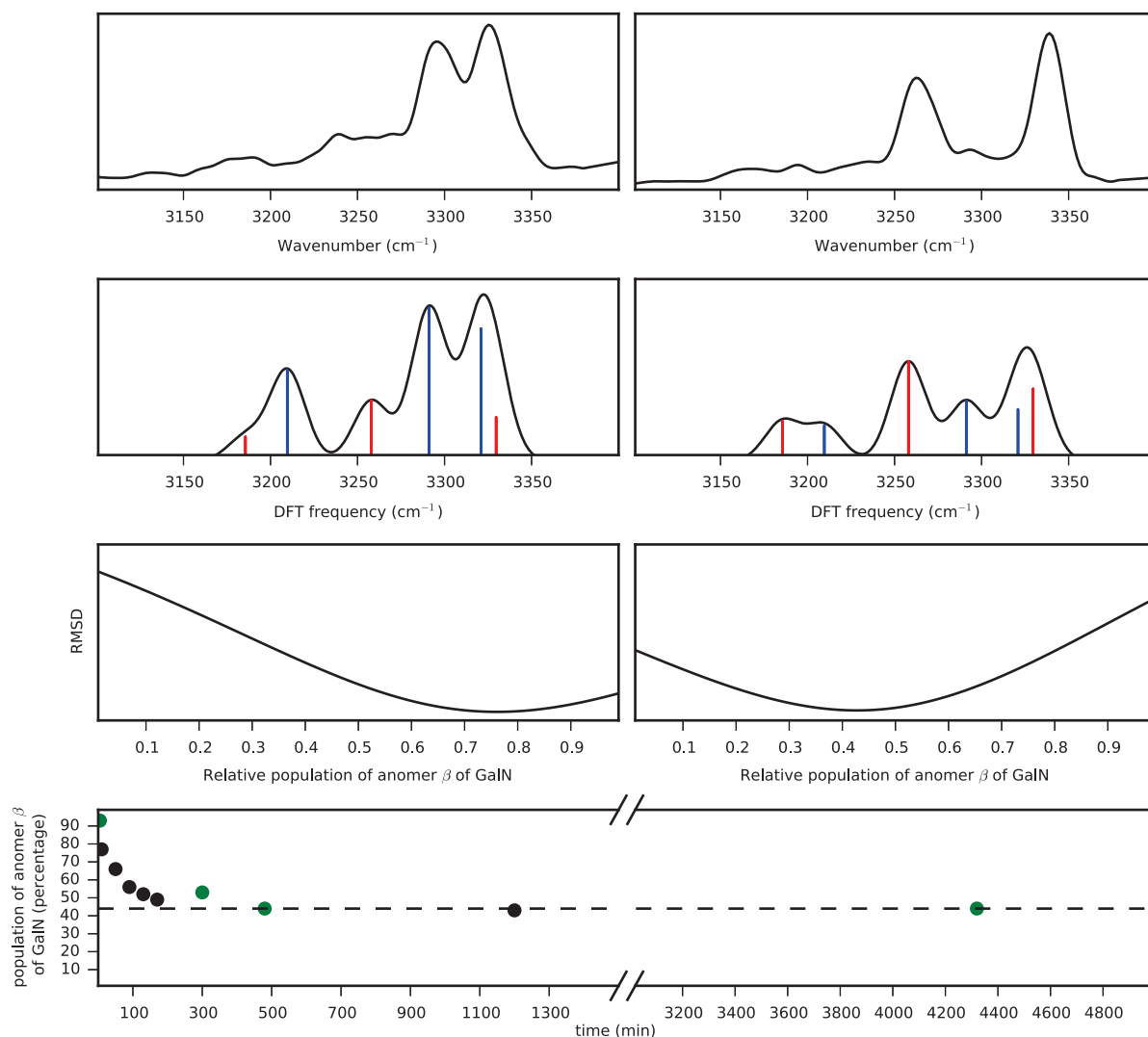


FIGURE 7.2 – First three rows are the IRMPD spectrum, the total theoretical spectrum obtained with anomeric populations corresponding to the minimum of the RMSD curve (third panel) and the evolution of the RMSD with the relative population of β GalN, respectively. Left side corresponds to the measurement of the IRMPD spectrum after 10 min and its corresponding theoretical spectrum and RMSD curve and right side corresponds to a measure after 1200 min. Bottom panel : Evolution in time of the relative contribution of β GalN. Black circles for relative contribution calculated by our approach and green circles for relative contribution obtained by RMN.

population of anomer β has the same trend as RMN. It starts with a high value of the population of anomer β and reach an asymptotic value after a certain amount of time. At the beginning, we found a population of anomer β equal to 78% and at the end it is 42%. In view of the above, using IRMPD we are able to probe the dynamic occurring in solution for GalN. IRMPD spectra are like taken pictures of the epimerization process in this case or simply pictures of the instantaneous configuration in solution in general.

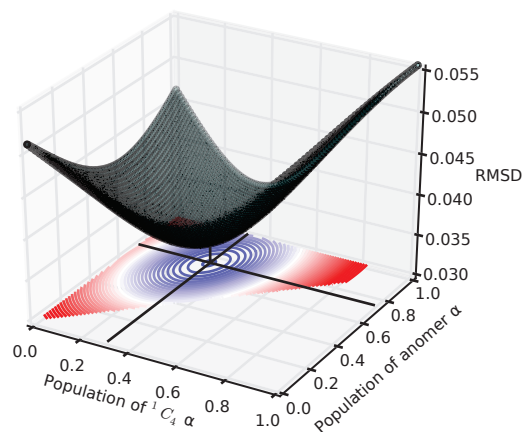


FIGURE 7.3 – Evolution of the RMSD in function of the population of the 1C_4 chair of anomer α and the total population of anomer α .

7.3.3 Elucidation of conformational heterogeneity

We have previously shown that combining IRMPD with a technique to calculate the relative population of anomers allows us to follow the dynamic in the solution for GalN. For GlcN, we selected the lowest energy 4C_1 of both anomers plus the lowest energy structure of α GlcN, a 1C_4 chair to explain the IRMPD spectrum. RMN experiment gives 69% of anomer α and 31% of anomer β . But this kind of experiment gives no information about the population of conformers. We used IRMPD and our technique to have an idea about the relative contribution of each conformer. To do that we let the population of the α 1C_4 chair and of the two 4C_1 chairs vary between 0 and 1 with a step of 0.01 such that the total population is always one. We evaluated the convoluted theoretical spectrum to the IRMPD one for each set of population ($\rho_{1C_4}^\alpha, \rho_{4C_1}^\alpha, \rho_{4C_1}^\beta$). All the calculated RMSD are plotted in Fig 7.3. The curve has a minimum at (0.31, 0.41, 0.028). It corresponds to 72% of anomer α and 28% of anomer β which is close to RMN results. Fig. 7.4 shows the IRMPD spectrum of GlcN and the total theoretical spectrum obtained with relative contribution of each conformer from the minimum in the RMSD curve in the 3100-3400 cm^{-1} region. The theoretical spectrum perfectly matches the IRMPD one. Regarding the conformations of α GlcN, the calculation predicts that the 4C_1 is more populated than the 1C_4 with 41% and 31%, respectively. However, the energetic criteria highly favored the 1C_4 . This fact is not contradictory with our present result. Indeed, in solution we have a certain mixing of each compound and this mixing is conserved in the electron spray ionization process. The energies are not necessarily the same in solution and in gas phase. Using IR and IRMPD spectra, it seems to be possible to have informations on the composition of the sample.

7.4 Conclusions

In this work we have studied the structures of gas phase hexosamines (GlcN, GalN and ManN) in their protonated charge state, using a combination of IRMPD spectroscopy and *ab initio* calculations. Based on energetic and spectroscopic criteria, we have proposed stable conformations for GlcN et GalN. For ManN, we have pointed out the required

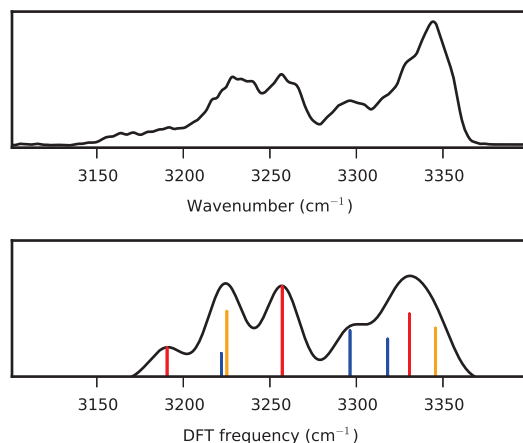


FIGURE 7.4 – IRMPD spectrum of GlcN (top panel) and the total theoretical spectrum (bottom panel) obtained with conformational and anomeric populations corresponding to the minimum in the RMSD curve (see fig. 7.3).

presence of 4C_1 chairs to explain some features of the corresponding IRMPD spectrum. But we could not conclude about the stable conformations for this hexosamine. We have tested our original approach to gain access to the relative contributions of conformers and anomers in a sample. Using the IRMPD spectrum and the theoretical spectra of the lowest energy structures, this approach has reproduced the epimerization process occurring in solution for GalN. Measuring a IRMPD spectrum is like taking a picture of the configuration of the sample in solution. With this in mind, we have elucidated the conformational heterogeneity of GlcN. The relative anomeric population obtained with our approach is consistent with RMN data. Our results suggest that the 4C_1 chair of the anomer α is more populated than the 1C_4 chair (41% and 31%, respectively). This method shows promising results to retrieve the relative population of the constituents in a sample.

Acknowledgement

This work was granted access to the HPC resources of the FLMSN, "Fédération Lyonnaise de Modélisation et Sciences Numériques", partner of EQUIPEX EQUIP@MESO and to the "Centre de calcul CC-IN2P3" at Villeurbanne, France. The authors are members of the Glycophysics Network (<http://www.glyms.univ-lyon1.fr>).

7.4.1 References

- [1] Cesar S. Contreras, Nicolas C. Polfer, Jos Oomens, Jeffrey D. Steill, Brad Bendiak, and John R. Eyler. On the path to glycan conformer identification : Gas-phase study of the anomers of methyl glycosides of *N*-acetyl-D-glucosamine and *N*-acetyl-D-galactosamine. *Int. J. Mass Spectrom.*, 330–332 :285–294, 2012.
- [2] Oscar Hernandez, Samantha Isenberg, Vincent Steinmetz, Gary L. Glish, and Philippe Maitre. Probing mobility-selected saccharide isomers : Selective ion-molecule

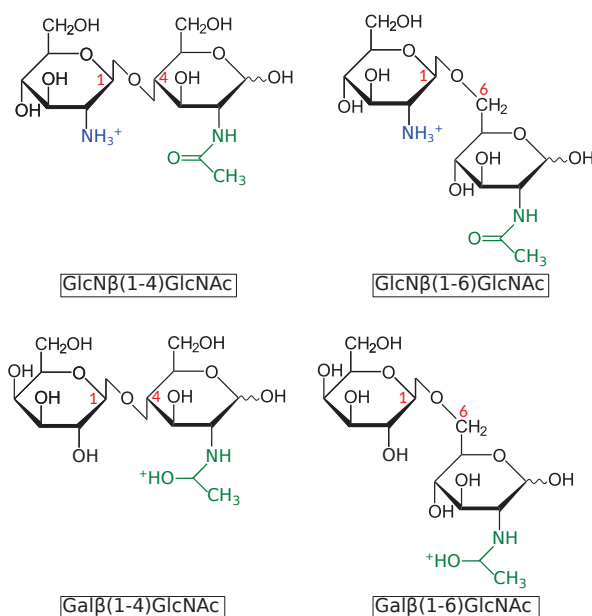
-
- reactions and wavelength-specific IR activation. *J. Phys. Chem. A*, 119(23) :6057–6064, 2015.
- [3] Yanglan Tan, Ning Zhao, Jinfeng Liu, Pengfei Li, Corey N. Stedwell, Long Yu, and Nicolas C. Polfer. Vibrational Signatures of Isomeric Lithiated *N*-acetyl-D-hexosamines by Gas-Phase Infrared Multiple-Photon Dissociation (IRMPD) Spectroscopy. *J. Am. Soc. Mass Spectrom.*, pages 1–12, 2017.
- [4] Emilio B. Cagmat, Jan Szczepanski, Wright L. Pearson, David H. Powell, John R. Eyler, and Nick C. Polfer. Vibrational signatures of metal-chelated monosaccharide epimers : gas-phase infrared spectroscopy of Rb^+ -tagged glucuronic and iduronic acid. *Phys. Chem. Chem. Phys.*, 12 :3474–3479, 2010.
- [5] Wright L. Pearson, Cesar Contreras, David Powell, Giel Berden, Jos Oomens, Brad Bendiak, and John R. Eyler. Differentiation of rubidiated methyl-D-glycoside stereoisomers by infrared multiple-photon dissociation spectroscopy in the O-H and C-H stretching regions. *J. Phys. Chem. B*, 119(41) :12970–12981, 2015.
- [6] Baptiste Schindler, Janhavi Joshi, Abdul-Rahman Allouche, Daniel Simon, Stephane Chambert, Vincent Brites, Marie-Pierre Gageot, and Isabelle Compagnon. Distinguishing isobaric phosphated and sulfated carbohydrates by coupling of mass spectrometry with gas phase vibrational spectroscopy. *Phys. Chem. Chem. Phys.*, 16 :22131–22138, 2014.
- [7] Loïc Barnes, Baptiste Schindler, Stéphane Chambert, Abdul-Rahman Allouche, and Isabelle Compagnon. Conformational preferences of protonated N-acetylated hexosamines probed by InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy and ab initio calculations. *Int. J. Mass Spectrom.*, 421 :116–123, 2017.
- [8] B. Schindler, L. Barnes, C. J. Gray, S. Chambert, S. L. Flitsch, J. Oomens, R. Daniel, A. R. Allouche, and I. Compagnon. IRMPD Spectroscopy Sheds New (Infrared) Light on the Sulfate Pattern of Carbohydrates. *J. Phys. Chem. A*, 121(10) :2114–2120, 2017.
- [9] James J. P. Stewart. Optimization of parameters for semiempirical methods VI : more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.*, 19(1) :1–32, 2012.
- [10] James J. P. Stewart. Mopac2012, 2012.
- [11] AD Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38 :3098–3100, 1988.
- [12] AD Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98 :5648–5652, 1993.
- [13] CT Lee, WT Yang, and RG Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37 :785–789, 1988.
- [14] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange-correlation functional using the coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.*, 393(1–3) :51–57, 2004.

-
- [15] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.*, 56(5) :2257–2261, 1972.
- [16] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.*, 72(1) :650–654, 1980.
- [17] Michael J. Frisch, John A. Pople, and J. Stephen Binkley. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.*, 80(7) :3265–3269, 1984.
- [18] Chr. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46 :618–622, 1934.
- [19] Abdul-Rahman Allouche. Gabedit-A graphical user interface for computational chemistry softwares. *J. Comput. Chem.*, 32(1) :174–182, 2011.
- [20] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision D.01, 2009. Gaussian Inc. Wallingford CT 2009.
- [21] Frank Neese. The ORCA program system. *WIREs Comput. Mol. Sc.*, 2 :73–78, 2012.

Étude de la liaison osidique

Highlights

- L'étude des modes de vibration de la liaison osidique s'avère inadaptée pour caractériser cette dernière.
- L'étude des modes d'élongation NH et OH permet de caractériser la liaison osidique.
- L'énergie relative pour les disaccharides est très sensible au niveau de théorie.
- Les effets anharmoniques ne sont pas toujours bien pris en compte.



8.1 Introduction

La liaison osidique a un rôle important dans l'organisation structurale en 3D des oligosaccharides. La position et la stéréochimie de la liaison peuvent conduire à former des hélices, des chaînes sans ou avec ramifications influençant ainsi les interactions entre les éléments de l'oligosaccharide ainsi qu'avec son environnement. La stéréochimie de la liaison osidique est importante par exemple pour la dégradation des saccharides par l'organisme humain. En effet, les enzymes humaines ne sont capables de dégrader que les liaisons osidiques en configuration α . Les oligosaccharides possédant des liaisons osidiques en configuration β ne peuvent être pratiquement dégrader que par des enzymes bactériennes. Pour caractériser les oligosaccharides, il est donc important d'identifier, en plus des constituants de l'oligosaccharide, la liaison entre ces constituants.

La liaison osidique s'effectue entre le groupement hydroxyle en position 1 d'un monosaccharide et n'importe quel autre groupement hydroxyle d'un autre monosaccharide. Il y a deux approches possibles afin d'étudier la position de la liaison osidique. La première est de recourir à l'analyse de la signature directe de la liaison, c'est-à-dire les modes de vibrations locaux C-O-C de cette liaison. Si la fréquence des modes de vibration C-O-C dépend de la position de la liaison alors on pourra caractériser la liaison. L'autre approche est une approche indirecte et consiste à étudier les modes d'élongations OH et NH. Les liaisons hydrogènes formées ne seront pas forcément identiques selon la position de la liaison osidique indiquant ainsi la position de la liaison.

Nous testerons la première approche sur un exemple : les disaccharides GlcN β (1 $\rightarrow$ 4)GlcNAc α et GlcN β (1 $\rightarrow$ 6)GlcNAc α . Nous étudierons ensuite la seconde approche sur l'exemple précédent ainsi que sur les disaccharides Gal β (1 $\rightarrow$ 6)GlcNAc et Gal β (1 $\rightarrow$ 4)GlcNAc. Le but c'est de caractériser la position de la liaison osidique pour ces disaccharides et de développer la base de données sur les disaccharides.

8.2 Signature directe de la liaison osidique

Dans cette partie nous allons discuter de l'approche reposant sur la signature directe de la liaison osidique, c'est-à-dire des modes d'élongation et de déformation C-O-C. Nous allons montrer que cette approche ne fonctionne pas parce que les modes de vibrations C-O-C ne sont pas localisés (voir figure 8.1). Cela rend difficile leur identification et leur comparaison avec les modes de vibrations C-O-C d'un autre disaccharide. La table 8.1 et la figure 8.2 montrent les valeurs pour les modes de déformation de la liaison C-O-C pour les structures les plus stables des anomères de GlcN β (1 $\rightarrow$ 4)GlcNAc et de GlcN β (1 $\rightarrow$ 6)GlcNAc ainsi que leurs spectres théoriques dans la zone 600-1700 cm^{-1} . Les modes indiqués dans la table 8.1 sont ceux des structures I_α et I_β de la table 8.2. La table 8.1 montre que le mode de déformation C-O-C le plus intense pour GlcN β (1 $\rightarrow$ 4)GlcNAc α possède une fréquence (662 cm^{-1}) très proche de la fréquence du mode de déformation pour GlcN β (1 $\rightarrow$ 6)GlcNAc α (667 cm^{-1}) ainsi que pour GlcN β (1 $\rightarrow$ 6)GlcNAc β (664 cm^{-1}). Le mode le moins intense pour GlcN β (1 $\rightarrow$ 4)GlcNAc α a quant à lui une fréquence (716 cm^{-1}) similaire aux fréquences (714 et 728 cm^{-1}) des modes de déformation pour GlcN β (1 $\rightarrow$ 4)GlcNAc β . On voit donc que la signature directe de la position de la liaison osidique n'est pas claire. La signature spectroscopique directe de la liaison osidique ne permet donc pas d'identifier la position de la liaison.

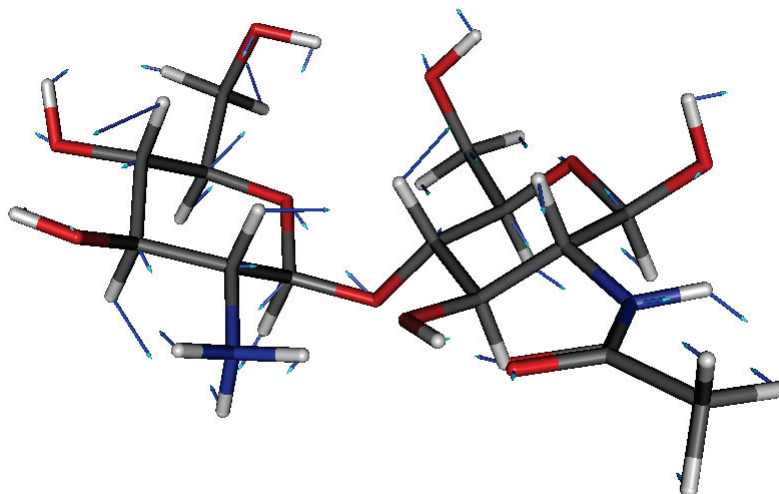


FIGURE 8.1 – Exemple d’un mode de déformation de la liaison osidique non local pour une conformation de l’anomère β de GlcN β (1 $\rightarrow$ 4)GlcNAc

GlcN β (1 $\rightarrow$ 4)GlcNAc α		GlcN β (1 $\rightarrow$ 4)GlcNAc β	
51.6	662	5.6	714
6.2	716	0.7	728
GlcN β (1 $\rightarrow$ 6)GlcNAc α		GlcN β (1 $\rightarrow$ 6)GlcNAc β	
15.5	667	15.4	664

TABLE 8.1 – Intensités et fréquences des modes de déformation de la liaison C-O-C avec CAM-B3LYP/6-311++G(2df,2pd) pour les structures les plus stables des anomères des disaccharides GlcN β (1 $\rightarrow$ 4)GlcNAc et GlcN β (1 $\rightarrow$ 6)GlcNAc [intensité en km.mol/ fréquence corrigée par le scaling factor 0,975 en cm $^{-1}$]

8.3 Signature indirecte de la liaison osidique

8.3.1 Analyse des disaccharides GlcN β (1 $\rightarrow$ 6)GlcNAc et GlcN β (1 $\rightarrow$ 4)GlcNAc

Dans cette partie, nous explorons la seconde approche qui repose sur l’analyse des modes d’élongation OH et NH. Nous allons montrer sur l’exemple de GlcN β (1 $\rightarrow$ 4)GlcNAc et GlcN β (1 $\rightarrow$ 6)GlcNAc que cette approche permet de déterminer la position de la liaison osidique. Le spectre expérimental a été mesuré pour le disaccharide GlcN β (1 $\rightarrow$ 4)GlcNAc mais pas pour GlcN β (1 $\rightarrow$ 6)GlcNAc. En effet, le disaccharide GlcN β (1 $\rightarrow$ 6)GlcNAc n’est pas présent naturellement et est difficilement synthétisable.

À partir du séquençage d’un tétrasaccharide, le spectre IRMPD a pu être mesuré pour l’anomère β de GlcN β (1 $\rightarrow$ 4)GlcNAc [1]. En effet, la liaison osidique du disaccharide GlcN β (1 $\rightarrow$ 4)GlcNAc avec un autre disaccharide est en configuration β dans le tétrasaccharide séquençé et le séquençage a conservé cette stéréochimie. Pour discuter la signature indirecte de la position de la liaison osidique, nous allons étudier uniquement l’anomère β des disaccharides GlcN β (1 $\rightarrow$ 4)GlcNAc et GlcN β (1 $\rightarrow$ 6)GlcNAc.

La figure 8.3 présente les spectres dans la zone OPO de la structure de plus basse énergie (structure I $_{\beta}$ dans la table 8.2) de l’anomère β de GlcN β (1 $\rightarrow$ 4)GlcNAc et de

GlcNβ(1$\rightarrow$4)GlcNAcα				GlcNβ(1$\rightarrow$4)GlcNAcβ			
I $_{\alpha}$	0	0	0	I $_{\beta}$	0	0	0
II $_{\alpha}$	29,7	25	27,4	II $_{\beta}$	8,5	13	11
GlcNβ(1$\rightarrow$6)GlcNAcα				GlcNβ(1$\rightarrow$6)GlcNAcβ			
I $_{\alpha}$	0	0	0	I $_{\beta}$	40,7	0	0
II $_{\alpha}$	6,2	5,3	4,6	II $_{\beta}$	0	28,2	26,7
II $_{\alpha}$	9,7	10,1	9,4	III $_{\beta}$	3,5	30,4	29,3
				IV $_{\beta}$	5,2	32,6	31,4

TABLE 8.2 – Énergies relatives des structures de plus basse énergie pour GlcN β (1 $\rightarrow$ 4)GlcNAc et GlcN β (1 $\rightarrow$ 6)GlcNAc [énergie du point zéro CAM-B3LYP/6-311++G(2df,2pd) / énergie électronique MP2/6-311++G(3df,3pd) / correction du point zéro CAM-B3LYP/6-311++G(2df,2pd) + énergie MP2/6-311++G(3df,3pd)]

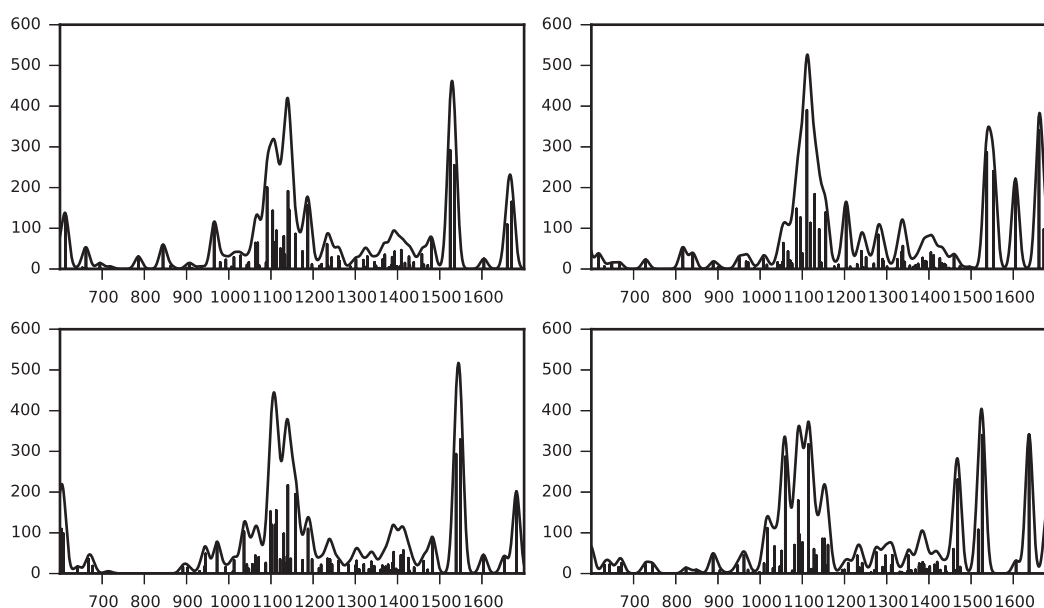


FIGURE 8.2 – Spectres théoriques obtenus avec CAM-B3LYP/6-311++G(2df,2pd) pour les conformères les plus stables des anomères α (haut) et β (bas) de GlcN β (1 $\rightarrow$ 4)GlcNAc (partie gauche) et de GlcN β (1 $\rightarrow$ 6)GlcNAc (partie droite) dans la zone 600-1700 cm^{-1}

GlcN β (1 $\rightarrow$ 6)GlcNAc.

Les spectres théoriques des deux disaccharides sont différents. En effet, la bande autour de 3510 cm^{-1} sur le spectre de l'anomère β de GlcN β (1 $\rightarrow$ 4)GlcNAc n'est pas présente sur le spectre de l'anomère β de GlcN β (1 $\rightarrow$ 6)GlcNAc. Ce mode correspond à l'élongation de la liaison hydrogène O6H-O6H pour l'anomère β de GlcN β (1 $\rightarrow$ 4)GlcNAc. Ce mode n'existe évidemment pas pour l'autre disaccharide car l'un des hydroxyles en position intervient dans la liaison osidique. Cette constatation conforte l'idée que la signature indirecte permet d'identifier la position de la liaison osidique.

Nous allons maintenant vérifier que la structure de plus basse énergie pour l'anomère β de GlcN β (1 $\rightarrow$ 4)GlcNAc reproduit bien le spectre IRMPD. Les bandes expérimentales situées à 3272 et 3335 cm^{-1} correspondent aux modes d'élongation NH3 $^{+}$ de GlcN β (1 $\rightarrow$ 4)GlcNAc β . Le mode d'élongation NH reproduit la bande expérimentale à 3430 cm^{-1} . La bande large centrée à 3548 cm^{-1} n'est pas parfaitement reproduite par

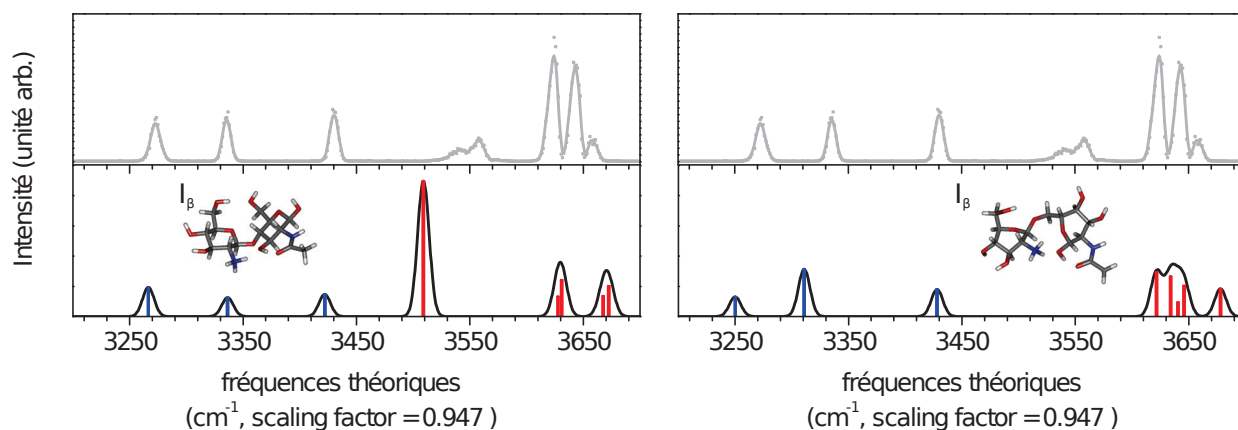


FIGURE 8.3 – Comparaison des spectres théoriques de la structure I_β de l'anomère β de GlcN β (1 $\rightarrow$ 4)GlcNAc (bas-gauche) et de la structure I_β de l'anomère β de GlcN β (1 $\rightarrow$ 6)GlcNAc (bas-droite) par rapport au spectre expérimental de GlcN β (1 $\rightarrow$ 4)GlcNAc (en haut). Le code couleur indique les modes d'élongation NH 3^+ et NH en bleu et les modes d'élongation OH en rouge.

le mode d'élongation O6H-O6H. Cependant ce mode d'élongation implique que la liaison hydrogène entre les deux monosaccharides est fortement anharmonique. La distance de la liaison hydrogène est de 1,94 Å alors que pour des liaisons plus faibles la distance est supérieure ou égale à 2,2-2,4 Å. Une autre quantité permettant de juger de la force de la liaison est l'ordre de liaison de Mayer [2]. Pour cette liaison hydrogène, on a un ordre de liaison de 0,056 contre 0,02 pour une autre liaison hydrogène entre deux hydroxyles. Cette anharmonicité pourrait expliquer le décalage entre la théorie et l'expérience et l'élargissement de la bande expérimentale. Enfin pour les modes d'élongation OH les plus libres (autour de 3670 cm^{-1}), on observe un décalage autour de 15 cm^{-1} avec le spectre IRMPD. Pour les deux bandes situées à 3624 et 3644 cm^{-1} il y a deux modes OH libres couplés dont les fréquences sont très proches qui pourraient expliquer ces bandes. Pour la bande à 3644 cm^{-1} l'écart avec la fréquence du mode O3H (du monosaccharide protoné) est autour de 11 cm^{-1} et la bande à 3624 cm^{-1} correspond au mode O1H.

En ce qui concerne le disaccharide GlcN β (1 $\rightarrow$ 6)GlcNAc, le spectre théorique de la structure la plus stable (selon des calculs MP2) ne reproduit pas le spectre expérimental. En effet, les fréquences des modes d'élongation NH 3^+ sont décalées de 20 à 25 cm^{-1} par rapport aux bandes expérimentales correspondantes. Cependant le mode d'élongation NH reproduit correctement la bande expérimentale à 3430 cm^{-1} . En revanche, il n'y a pas de modes théoriques compatibles avec la bande large centrée à 3548 cm^{-1} qui avait été attribuée à une liaison hydrogène entre les deux hydroxyles en position 6 pour le GlcN β (1 $\rightarrow$ 4)GlcNAc. Les structures de plus hautes énergies (à plus de 28 kJ/mol en MP2) présentent un mode d'élongation O4H-O6H (d'un même monosaccharide) à cet emplacement mais elles restent très hautes en énergie. Pour la zone au-delà de 3600 cm^{-1} , les deux bandes intenses sont bien reproduites mais l'écart avec la bande peu intense est un peu plus grande que pour GlcN β (1 $\rightarrow$ 4)GlcNAc avec une valeur de 20 cm^{-1} .

Cet exemple nous montre clairement que la signature indirecte spectroscopique de la liaison osidique peut permettre de connaître la position de la liaison.

Nous avons montré pour certains monosaccharides que l'on pouvait identifier les anomères à partir d'une signature spectroscopique particulière. Dans cette partie, nous allons montrer que c'est aussi possible pour le disaccharide GlcN β (1 $\rightarrow$ 4)GlcNAc.

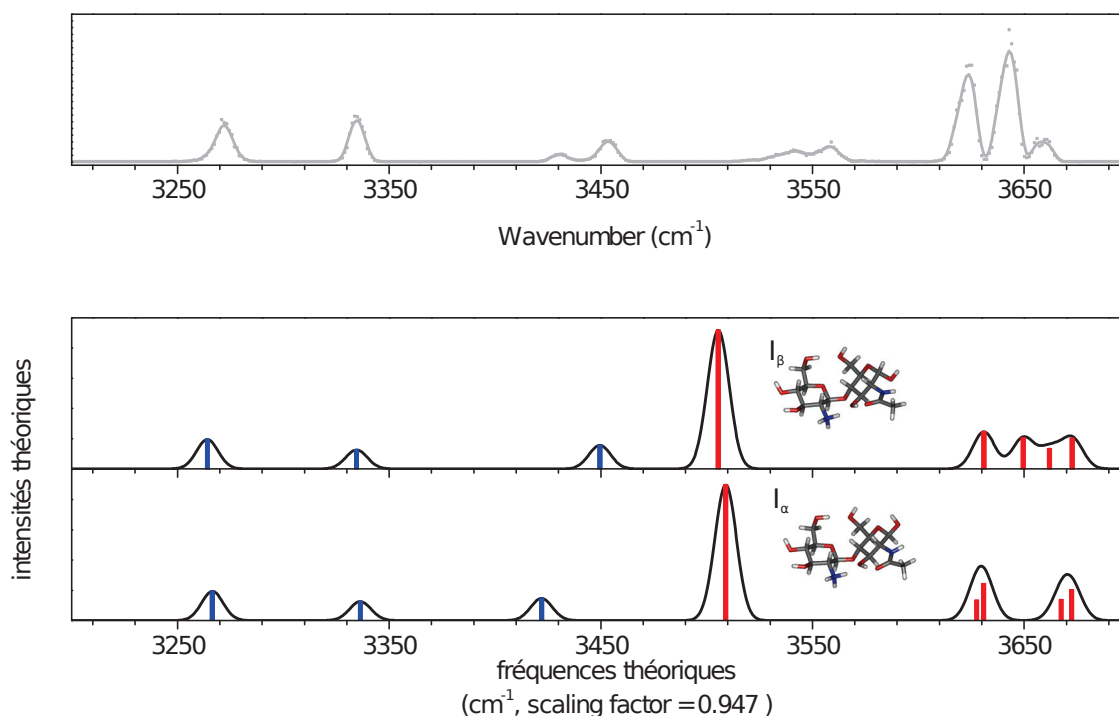


FIGURE 8.4 – Comparaison des spectres théoriques des structures I_α (milieu) et I_β (bas) de $\text{GlcN}\beta(1\rightarrow4)\text{GlcNAc}$ par rapport au spectre expérimental de $\text{GlcN}\beta(1\rightarrow4)\text{GlcNAc}$ (spectre IRMPD en gris clair). Le code couleur indique les modes d'élongation NH_3^+ et NH en bleu et les modes d'élongation OH en rouge.

Le spectre IRMPD a été mesuré pour le mélange anomérique de $\text{GlcN}\beta(1\rightarrow4)\text{GlcNAc}$ [1]. La figure 8.4 montre le spectre IRMPD avec le spectre théorique de la structure de plus basse énergie de l'anomère α et de l'anomère β . La plus importante différence entre les deux spectres théoriques et la fréquence du mode d'élongation NH . Alors que pour l'anomère β la fréquence est égale à 3422 cm^{-1} , la fréquence pour l'anomère α vaut 3450 cm^{-1} . Ce décalage permet d'expliquer le doublet expérimental dont les bandes sont situées à 3430 et 3453 cm^{-1} . Autrement dit ce mode de vibration est une signature de l'anomère.

8.3.2 Analyse des disaccharides $\text{Gal}\beta(1\rightarrow6)\text{GlcNAc}$ et $\text{Gal}\beta(1\rightarrow4)\text{GlcNAc}$

La figure 8.5 montre que les spectres expérimentaux de $\text{Gal}\beta(1\rightarrow6)\text{GlcNAc}$ et de $\text{Gal}\beta(1\rightarrow4)\text{GlcNAc}$ sont différents. En effet, le spectre de $\text{Gal}\beta(1\rightarrow6)\text{GlcNAc}$ présente un motif peu résolu de 3000 à 3550 cm^{-1} avec une bande large et plus intense centrée autour de 3440 cm^{-1} . Le spectre de $\text{Gal}\beta(1\rightarrow4)\text{GlcNAc}$, lui possède un motif peu résolu de 3000 à 3250 cm^{-1} mais présente une bande bien résolue autour de 3400 cm^{-1} . De plus, le spectre IRMPD de $\text{Gal}\beta(1\rightarrow6)\text{GlcNAc}$ présente une bande intense autour de 3600 cm^{-1} alors que sur le spectre de $\text{Gal}\beta(1\rightarrow4)\text{GlcNAc}$ on voit des bandes peu intenses.

Nous avons effectué les recherches de conformation pour les anomères de $\text{Gal}\beta(1\rightarrow6)\text{GlcNAc}$ et de $\text{Gal}\beta(1\rightarrow4)\text{GlcNAc}$ en supposant que la protonation a lieu sur l'atome d'oxygène du groupement acétyl. Les tables 8.3 et 8.4 donnent les énergies relatives des conformères les plus stables pour les anomères de $\text{Gal}\beta(1\rightarrow6)\text{GlcNAc}$ et de $\text{Gal}\beta(1\rightarrow4)\text{GlcNAc}$. A part pour l'anomère α de $\text{Gal}\beta(1\rightarrow6)\text{GlcNAc}$, les énergies relatives

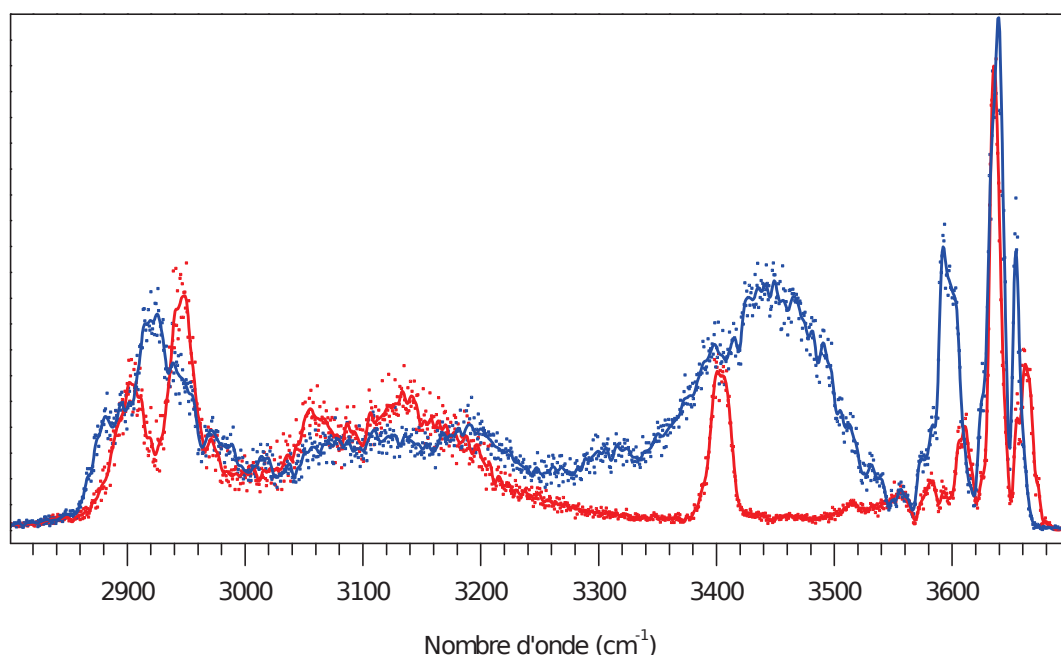


FIGURE 8.5 – Spectres IRMPD des disaccharides $\text{Gal}\beta(1\rightarrow6)\text{GlcNAc}$ (en bleu) et $\text{Gal}\beta(1\rightarrow4)\text{GlcNAc}$ (en rouge)

varient énormément entre CAM-B3LYP et MP2 (avec ou sans la correction vibrationnelle de CAM-B3LYP) avec un écart maximal de plus de 1000 kJ/mol. En revanche, les énergies relatives obtenues par MP2 avec ou sans la correction vibrationnelle changent peu, seulement de quelques kJ/mol au maximum. Ces grandes variations des énergies relatives peuvent vouloir dire que l'approximation consistant à calculer l'énergie MP2 en utilisant la géométrie optimisée avec CAM-B3LYP n'est pas adaptée aux disaccharides dont l'énergie est beaucoup plus sensible aux variations de géométrie. Il faudrait réoptimiser les géométries CAM-B3LYP avec une méthode plus précise comme par exemple B2PLYP ainsi que calculer les fréquences. Cependant, optimiser la géométrie et calculer les fréquences est difficilement faisable pour tous les conformères avec la méthode B2PLYP. Les énergies relatives indiquent aussi que plusieurs conformères peuvent être présents dans l'échantillon compliquant ainsi le problème.

Nous allons maintenant étudier le spectre infrarouge théorique des conformères de plus basse énergie obtenue avec MP2. La figure 8.6 montre le spectre expérimental de $\text{Gal}\beta(1\rightarrow6)\text{GlcNAc}$ ainsi que la conformation de l'anomère α et les deux conformations les plus stables pour l'anomère β . Étant donné la multitude de modes OH libres présents dans la zone de 3580 à 3700 cm^{-1} , les contributions aux bandes expérimentales de cette zone sont multiples. Ainsi l'attribution des modes est difficile. Cependant, globalement cette zone est reproduite correctement par les spectres théoriques. Le reste du spectre est très peu résolu par conséquent l'identification des bandes expérimentales est impossible. Par exemple, le mode NH (3340 cm^{-1} pour l'anomère α et 3360 et 3390 cm^{-1} pour les deux conformères de plus basse énergie de l'anomère β) ne peut être clairement associé à un motif expérimental. Néanmoins, pour l'anomère α , le mode d'élongation NH est un peu plus lié que pour l'anomère β . En effet, la distance entre l'hydrogène et un oxygène accepteur est de 1,996 Å contre 2,2 Å ou plus pour l'anomère β (voir table 8.5). Cette différence dans la liaison hydrogène pourrait signifier que le mode NH de l'anomère α participe à l'élargissement des bandes expérimentales. Le mode d'élongation O1H $\cdots$ O6H très lié (avec une distance H $\cdots$ O de 1,631 Å) de l'anomère α et les modes d'élongation

Gal β (1 $\rightarrow$ 6)GlcNAc α				Gal β (1 $\rightarrow$ 6)GlcNAc β			
I $_{\alpha}$	0,00	0,00	0,00	I $_{\beta}$	0,00	0,00	0,00
II $_{\alpha}$	10,03	15,31	14,42	II $_{\beta}$	15,46	5,19	10,30
III $_{\alpha}$	11,23	19,88	16,88	III $_{\beta}$	-34,74	36,52	38,54
IV $_{\alpha}$	7,52	22,39	22,14	IV $_{\beta}$	-29,64	40,50	42,99
V $_{\alpha}$	29,44	30,52	28,29	V $_{\beta}$	-29,14	40,38	43,34
VI $_{\alpha}$	24,71	30,67	29,79	VI $_{\beta}$	-10,22	44,57	47,30
VII $_{\alpha}$	29,69	31,86	29,98	VII $_{\beta}$	-22,37	48,09	49,98
VIII $_{\alpha}$	30,55	33,06	33,93	VIII $_{\beta}$	1043,19	53,19	58,49
IX $_{\alpha}$	32,60	33,55	34,62	IX $_{\beta}$	-6,02	58,32	60,54
X $_{\alpha}$	28,88	38,86	34,75	X $_{\beta}$	2,17	63,09	63,35

TABLE 8.3 – Énergies relatives des 10 conformations les plus stables au niveau MP2 des anomères α et β de Gal β (1 $\rightarrow$ 6)GlcNAc. Les énergies relatives indiquées sont l'énergie relative corrigée avec l'énergie de point zéro calculée avec CAM-B3LYP/6-311++G(2df,2pd), l'énergie relative obtenue avec MP2/6-311++G(3df,3pd) et l'énergie relative obtenue avec MP2/6-311++G(3df,3pd) corrigée avec l'énergie du point zéro de CAM-B3LYP/6-311++G(2df,2pd). Les énergies relatives sont données en kJ/mol.

Gal β (1 $\rightarrow$ 4)GlcNAc α				Gal β (1 $\rightarrow$ 4)GlcNAc β			
I $_{\alpha}$	0,00	0,00	0,00	I $_{\beta}$	0,00	0,00	0,00
II $_{\alpha}$	12,61	2,88	1,57	II $_{\beta}$	33,98	-1,19	2,55
III $_{\alpha}$	-17,55	18,08	23,58	III $_{\beta}$	6,72	7,17	8,86
IV $_{\alpha}$	-22,47	35,80	35,82	IV $_{\beta}$	9,29	11,42	11,03
V $_{\alpha}$	-20,60	36,50	36,17	V $_{\beta}$	14,32	8,20	12,23
VI $_{\alpha}$	-23,58	34,61	36,52	VI $_{\beta}$	19,38	11,76	13,92
VII $_{\alpha}$	-23,38	35,60	37,15	VII $_{\beta}$	13,78	14,57	14,59
VIII $_{\alpha}$	-22,09	37,47	37,60	VIII $_{\beta}$	15,51	12,02	14,64
IX $_{\alpha}$	-20,18	39,60	39,63	IX $_{\beta}$	12,96	12,61	15,21
X $_{\alpha}$	-3,76	41,50	42,77	X $_{\beta}$	16,71	9,90	15,47

TABLE 8.4 – Énergies relatives des 10 conformations les plus stables au niveau MP2 des anomères α et β de Gal β (1 $\rightarrow$ 4)GlcNAc. Les énergies relatives indiquées sont l'énergie relative corrigée avec l'énergie de point zéro calculée avec CAM-B3LYP/6-311++G(2df,2pd), l'énergie relative obtenue avec MP2/6-311++G(3df,3pd) et l'énergie relative obtenue avec MP2/6-311++G(3df,3pd) corrigée avec l'énergie du point zéro de CAM-B3LYP/6-311++G(2df,2pd). Les énergies relatives sont données en kJ/mol.

O3H $\cdots$ O6H très liés (distance de 1,64 Å) couplés aux modes d'élongation CH du second conformère de l'anomère β pourraient expliquer le motif non résolu de 3000 à 3250 cm^{-1} . L'anomère α présente deux modes OH liés, l'élongation de O3H en liaison hydrogène avec l'oxygène de la liaison osidique et le mode d'élongation O6H $\cdots$ O4H dont les fréquences sont respectivement, 3480 et 3500 cm^{-1} . La distance entre l'hydrogène et l'oxygène accepteur est de 1,849 et 1,885 Å, respectivement. La conformation la plus stable de l'anomère β ne possède pas de mode OH lié dans cette zone contrairement à la seconde conformation qui présente 3 modes OH liés. Ces derniers sont les modes O6H $\cdots$ O4H, O4H $\cdots$ O3H et le mode d'élongation O1H en liaison hydrogène avec l'oxygène de la liaison osidique ayant pour fréquences respectives, 3500, 3540 et 3510 cm^{-1} . La distance entre l'hydrogène et l'oxygène accepteur est de 1,928, 1,88 et 1,976 Å, respectivement. Ces modes OH liés

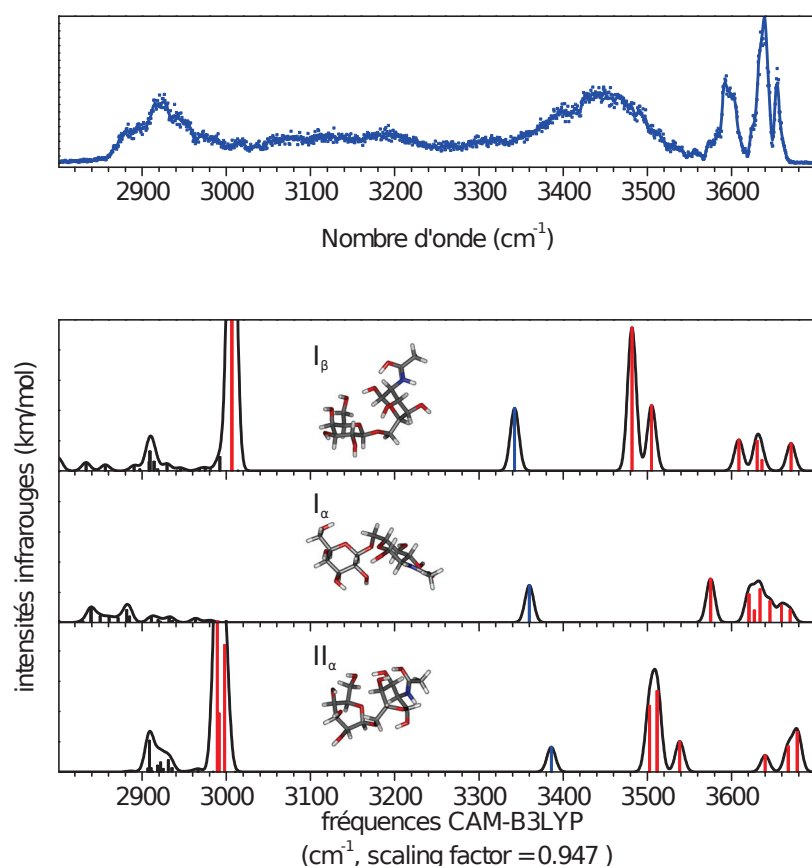


FIGURE 8.6 – Comparaison des spectres théoriques de la structure I_α de l'anomère α et des structures I_β et II_β de l'anomère β de $\text{Gal}\beta(1\rightarrow6)\text{GlcNAc}$ par rapport au spectre IRMPD (haut). Le code couleur indique les modes d'élongation OH en rouge, le mode d'élongation NH en bleu et les modes d'élongation CH en noir. Les spectres théoriques sont convolués en utilisant une largeur à mi-hauteur de 6 cm^{-1} et les fréquences sont scalées par un scaling factor de 0,947.

dont la fréquence est autour de 3500 cm^{-1} pourraient expliquer la bande large et intense expérimentalement.

La figure 8.7 montre le spectre expérimental de $\text{Gal}\beta(1\rightarrow4)\text{GlcNAc}$ ainsi que les deux conformations les plus stables de l'anomère α et de l'anomère β de $\text{Gal}\beta(1\rightarrow4)\text{GlcNAc}$. De même que pour $\text{Gal}\beta(1\rightarrow6)\text{GlcNAc}$, du fait de la complexité de la surface d'énergie potentielle (2 conformères de chaque anomère) il est difficile d'attribuer les bandes expérimentales dans la zone de 3620 à 3700 cm^{-1} . En revanche, seules les deux conformères β contribuent au motif expérimental entre 3580 et 3620 cm^{-1} alors que pour le disaccharide précédent les contributions provenaient des deux anomères. Le conformère de plus basse énergie pour l'anomère α et le second conformère de l'anomère β présente un mode OH situé à 3460 et 3480 cm^{-1} , respectivement. Or expérimentalement il n'y pas de bande. Les distances pour ces modes sont de l'ordre de $1,81\text{ \AA}$, l'anharmonicité de ce mode est probablement importante et les calculs statiques avec un scaling factor ne sont pas suffisants. De plus, pour le second conformère de l'anomère β , il y a un mode très proche du mode d'élongation NH, un mode OH lié. On s'attendrait alors à avoir une bande large sur le spectre expérimental or on a une bande bien résolue et fine. Cette structure n'est peut-être pas présente dans l'échantillon. Le mode NH des conformères de l'anomère α reproduit bien la bande résolue à 3400 cm^{-1} . Pour ce qui est du mode NH des conformères

Galβ(1→6)GlcNAcα				Galβ(1→4)GlcNAcα					
Iα	O6H...O4H	3505	1,885	Iα	O4H...O3H	3532	1,913		
	O3H...O(liaison osidique)	3482	1,849		O6H...O4H	3454	1,812		
	O1H...O6H	3007	1,631		O3H...O6H	2870	1,613		
	NH...O	3343	1,996		NH...O	3401	2,208		
				IIα	O4H...O3H	3541	1,942		
					O3H...O4H	2944	1,61		
					NH...O	3403	2,201		
Galβ(1→6)GlcNAcβ				Galβ(1→4)GlcNAcβ					
Iβ	NH...O	3360	2,20	Iβ	O3H...O(cycle)	3212	1,72		
				NH...O				3371	2,211
IIβ	O6H...O4H	3503	1,928	IIβ	O4H...O6H	3478	1,818		
	O4H...O3H	3539	1,88		O3H...O4H	3376	1,784		
	O1H...O(liaison osidique)	3511	1,976		O3H...O3H	2978	1,584		
	O3H...O6H	2989	1,64		NH...O	3366	2,18		
	NH...O	3385	2,476						

TABLE 8.5 – Tableau récapitulatif des modes d'élongation OH liés et NH ainsi que les distances entre l'hydrogène et l'atome accepteur pour les structures de plus basse énergie des anomères de Gal β (1 $\rightarrow$ 6)GlcNAc et Gal β (1 $\rightarrow$ 4)GlcNAc [Structure/Type de mode/fréquence scalée en cm $^{-1}$ / distance en Å]

de l'anomère β , la fréquence de ce mode est plus basse de 20 à 30 cm $^{-1}$. Les conformères α ont tous deux un mode OH lié autour de 3530-3540 cm $^{-1}$ (distance de plus de 1,91 Å) qui pourrait correspondre à la bande expérimentale peu intense entre 3520 et 3570 cm $^{-1}$. Enfin, la zone entre 3000 et 3250 cm $^{-1}$, peu résolue, peut s'expliquer par l'existence de modes d'élongation OH très fortement liés dont la distance entre l'hydrogène et l'oxygène accepteur est de l'ordre de 1,6 Å. Aucune des conformations α ou β de Gal β (1 $\rightarrow$ 4)GlcNAc n'a de modes d'élongation OH dont l'hydroxyle est en liaison hydrogène avec l'oxygène de la liaison osidique. C'est peut-être cela qui explique la différence du motif expérimental dans la zone 3400-3600 cm $^{-1}$ entre Gal β (1 $\rightarrow$ 4)GlcNAc et Gal β (1 $\rightarrow$ 6)GlcNAc.

8.4 Conclusion

L'approche consistant à étudier la signature directe de la liaison osidique, c'est-à-dire les modes de vibrations C-O-C ne permet pas de connaître la position de cette liaison osidique. En revanche, nous avons montré que l'analyse de la signature indirecte constituée des modes d'élongation OH et NH permettait de connaître la position à laquelle s'effectue la liaison entre deux monosaccharides. Cette approche a été utilisée avec succès sur les disaccharides GlcN β (1 $\rightarrow$ 4)GlcNAc et GlcN β (1 $\rightarrow$ 6)GlcNAc. Cette approche a aussi été utilisée sur les disaccharides Gal β (1 $\rightarrow$ 6)GlcNAc et Gal β (1 $\rightarrow$ 4)GlcNAc. Bien que l'on puisse discuter qualitativement des disaccharides Gal β (1 $\rightarrow$ 6)GlcNAc et Gal β (1 $\rightarrow$ 4)GlcNAc et de leurs différences, la résolution expérimentale ne permet pas de confirmer les structures de plus basse énergie. Afin d'augmenter la résolution expérimentale, il faudrait que les mesures expérimentales soient menées à basse température. Ce genre d'approche est explorée par Rizzo et al [3]. qui arrivent à obtenir des spectres expérimentaux bien résolus pour une température de 13 K. Dans cette expérience, les disaccharides sont protonés par

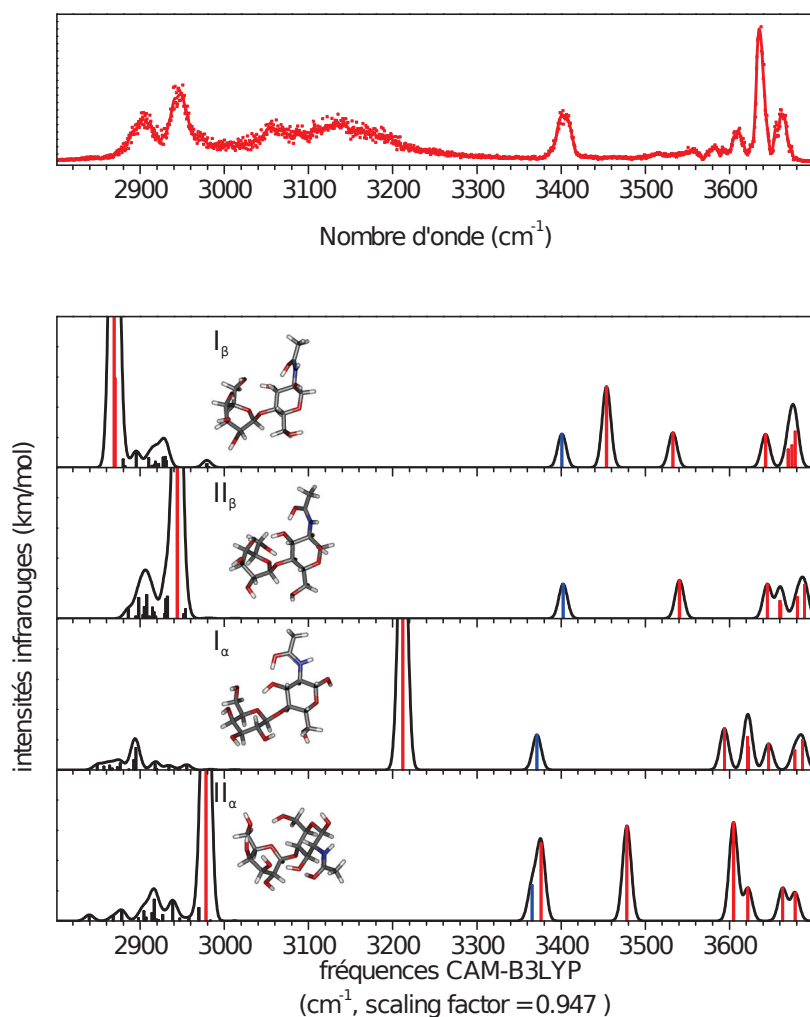


FIGURE 8.7 – Comparaison des spectres théoriques des structures I_α et II_α de l'anomère α et des structures I_β et II_β de l'anomère β de $\text{Gal}\beta(1\rightarrow4)\text{GlcNAc}$ par rapport au spectre IRMPD (haut). Le code couleur indique les modes d'élongation OH en rouge, le mode d'élongation NH en bleu et les modes d'élongation CH en noir. Les spectres théoriques sont convolués en utilisant une largeur à mi-hauteur de 6 cm^{-1} et les fréquences sont scalées par un scaling factor de 0,947.

l'ajout d'atomes de sodium sous forme de cations. La position des bandes est modifiée par la présence des ions sodium ainsi nos résultats théoriques ne peuvent pas être comparés aux résultats de Rizzo. Du point de vue théorique, il faudrait pouvoir prendre en compte les effets de température afin de comparer les spectres théoriques aux spectres IRMPD. Entre autre, il faut faire de la dynamique moléculaire. Or la dynamique moléculaire AIMD [4] est limitée aux fonctionnelles GGA comme BLYP. Les fonctionnelles hybrides comme B3LYP ne sont pas accessibles en terme de temps de calcul. Même avec la fonctionnelle BLYP, un calcul de dynamique moléculaire pour un disaccharide contenant une cinquantaine d'atomes reste très difficile surtout que nous ne sommes pas certains de nos structures de plus basse énergie. Il existe des approches de dynamique moléculaire reposant sur l'emploi du champ de force AMOEBA comme potentiel. Les types d'atomes, les moments multipolaires et les autres paramètres sont obtenus à partir d'un calcul statique de fréquences et un calcul du potentiel électrostatique [5]. La dynamique moléculaire est alors effectuée à partir de ce nouveau potentiel. Ce pourrait être un bon compromis

entre précision et temps de calcul pour pouvoir explorer les effets de température sur les spectres théoriques des disaccharides.

8.5 Références

- [1] Baptiste Schindler. *Caractérisation structurale et séquençage de carbohydrates par spectroscopie infrarouge intégrée à la spectrométrie de masse*. PhD thesis, Université Claude Bernard Lyon 1, 2016.
- [2] Adam J. Bridgeman, German Cavigliasso, Luke R. Ireland, and Joanne Rothery. The Mayer bond order as a tool in inorganic chemistry. *J. Chem. Soc., Dalton Trans.*, pages 2095–2108, 2001.
- [3] Chiara Masellis, Neelam Khanal, Michael Z. Kamrath, David E. Clemmer, and Thomas R. Rizzo. Cryogenic vibrational spectroscopy provides unique fingerprints for glycan identification. *J. Am. Soc. Mass Spectrom.*, 2017.
- [4] Marie-Pierre Gageot. Theoretical spectroscopy of floppy peptides at room temperature. A DFTMD perspective : gas and aqueous phase. *Phys. Chem. Chem. Phys.*, 12 :3336–3359, 2010.
- [5] Johnny C. Wu, Gaurav Chattree, and Pengyu Ren. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor Chem Acc*, 131(3) :1138, 2012.

Quatrième partie

Conclusion et perspectives

8.6 Résumé des travaux

Ces travaux de thèse s'inscrivent dans la volonté de développer une méthode permettant l'étude des oligosaccharides, de pouvoir les séquencer et connaître leur structure afin de mieux appréhender leur rôle dans la nature. L'alliance des calculs théoriques à la spectroscopie d'action permet d'obtenir des informations détaillées sur leur nature et leur structure.

Nous avons appliqué cette approche sur des monosaccharides sulfatés dans deux états de charge, protoné et déprotoné. Notre recherche conformationnelle a permis d'obtenir les structures de plus basse énergie dont le spectre infrarouge explique le spectre IRMPD pour les monosaccharides sulfatés protonés. Ces structures sont toutes des chaises 4C_1 . Nous avons observé que les spectres expérimentaux et théoriques étaient bien différents selon la position du groupement sulfate. Des signatures distinctes sont visibles, notamment le mode d'élongation O3H (présent dans le glucosamine-6 sulfate et pas dans le glucosamine-3 sulfate) et le mode d'élongation O6H (c'est l'inverse). Pour les monosaccharides sulfatés dans leur état déprotoné, le spectre IRMPD est différent dans la zone OPO (2800-3800 cm^{-1}). Même si les spectres théoriques ne reproduisent pas l'élargissement des bandes expérimentales, les spectres théoriques du glucosamine-6 sulfate et du glucosamine-3 sulfate déprotonés sont bien différents.

Nous avons étudié des monosaccharides protonés non méthylés et méthylés en position 1. À l'aide de critères énergétiques et spectroscopiques pour un grand nombre de conformations obtenues à partir de notre recherche conformationnelle, nous avons montré que la protonation s'effectuait sur la fonction carboxylique. Le mode de vibration caractéristique de cette protonation est le mode d'élongation N-H et sa signature spectroscopique est présente dans le spectre expérimental pour tous les monosaccharides méthylés et non méthylés. La théorie a montré que les structures de plus basse énergie présentes dans l'échantillon étaient toutes des chaises 4C_1 . Pour certains épimères et anomères les structures en chaise 4C_1 étaient légèrement différentes. Certaines chaises 4C_1 n'avaient pas la même coordination ou bien le groupement acétyl avait tourné pour former une nouvelle liaison hydrogène. De plus, nous avons déterminé des signatures permettant d'identifier les anomères α et β pour le N-acétyl glucosamine et le N-acétyl mannosamine. En revanche, cela n'a pas été possible pour le N-acétyl galactosamine parce que la résolution expérimentale n'était pas suffisante.

Pour terminer sur les monosaccharides, nous avons étudié 3 hexosamines protonées. Contrairement aux monosaccharides précédents, les structures de plus basse énergie obtenues ne sont pas toujours des structures en chaise 4C_1 . La recherche conformationnelle a montré que la surface d'énergie potentielle était plus complexe que celle des monosaccharides étudiés auparavant et elle a montré une plus grande hétérogénéité de structures. Pour le glucosamine, en plus des structures 4C_1 pour les anomères α et β , l'anomère α présente une structure en chaise 1C_4 . Ces trois structures doivent être prises en compte pour reproduire le spectre expérimental. Pour la mannosamine, les structures de plus basse énergie ne sont pas des structures en chaise 4C_1 ou 1C_4 . Cependant, pour la mannosamine, une bande expérimentale ne peut être expliquée que par la présence de structures 4C_1 dans l'échantillon. Pour la galactosamine, les structures les plus stables sont des chaises 4C_1 . La confrontation des spectres théoriques aux spectres expérimentaux a permis d'identifier des signatures pour les anomères du glucosamine et du galactosamine. De plus des signatures spécifiques aux conformations ont été trouvées pour les deux conformations (chaises 4C_1 et 1C_4) de l'anomère α du glucosamine et pour les structures en chaise 4C_1 de la mannosamine.

Nous avons pour la première fois exploré une approche pour déterminer les contributions relatives des structures contribuant au spectre IRMPD. Notre approche a permis de reproduire la dynamique (mutarotation) en solution observée pour la galactosamine par des mesures de RMN. Les pourcentages obtenus pour la glucosamine par cette méthode sont en bon accord avec les mesures RMN. Notre approche montre qu'en utilisant des spectres infrarouges et IRMPD il est possible d'obtenir des informations sur la composition de l'échantillon.

Pour caractériser un oligosaccharide, il est important de caractériser la liaison osidique en plus de ses constituants. Nous avons donc exploré dans cette thèse la possibilité de caractériser cette liaison par notre approche IRMPD/calculs *ab initio*. Notre étude a montré que l'approche reposant sur la signature directe ne permet pas une telle caractérisation. En revanche, nous avons montré qu'en étudiant les modes d'élongation OH et NH, autrement dit la signature indirecte nous pouvions obtenir des informations sur la liaison osidique et notamment sa position. Cependant pour certains cas la résolution expérimentale ne permet pas de confirmer les structures de plus basse énergie ce qui empêche la caractérisation de la liaison osidique.

8.7 Perspectives

Le travail effectué au cours de cette thèse devra être poursuivi afin de compléter la base de données des spectres infrarouges et des structures de monosaccharides et de disaccharides en vue du séquençage d'oligosaccharides. On pourra étudier des mono et disaccharides présentant d'autres groupements chimiques ou bien situés à d'autres endroits de la molécule. Pour les disaccharides, il faut étudier d'autres types de liaisons osidiques. L'approche développée pour l'analyse de la composition du mélange pourra être étendue à d'autres monosaccharides afin de vérifier la validité de cette méthode. En ce qui concerne les disaccharides, les mesures RMN seront difficilement réalisables voire impossibles et donc nous ne pourrons pas confronter les résultats de notre méthode à ceux de la RMN. Il serait intéressant de chercher à améliorer la qualité des énergies. En effet, nous avons vu pour les disaccharides que les énergies relatives variaient beaucoup selon le niveau de théorie et qu'en plus les spectres des structures de plus basse énergie avec la méthode MP2 ne reproduisaient pas forcément les spectres expérimentaux. On pourra faire des calculs avec la fonctionnelle double hybride B2PLYP et une assez grande base afin d'améliorer ces énergies relatives. Il faudra pour cela optimiser la géométrie et calculer les fréquences avec cette même fonctionnelle pour toutes les conformations possibles. Cependant, cela coûtera extrêmement cher en terme de temps de calcul. Au vu de la résolution expérimentale et à moins que de nouvelles mesures soient effectuées à plus basse température, un travail supplémentaire est à fournir pour améliorer la qualité des spectres théoriques pour des monosaccharides dans un état déprotoné mais aussi pour des disaccharides. En effet, les calculs statiques ne sont pas suffisants pour reproduire les effets anharmoniques importants pour des monosaccharides déprotonés et pour certains disaccharides. Des calculs de dynamique moléculaire pourraient être une solution mais la dynamique moléculaire avec des fonctionnelles hybrides comme B3LYP est impossible. Il faudra trouver un compromis entre précision et temps de calcul. Des calculs de dynamique moléculaire avec des champs de force comme AMOEBA paraissent tout indiqué. Il faudra s'assurer cependant que les fréquences harmoniques que l'on peut obtenir avec ces champs de force soient plus précises que celles obtenues par BLYP.

Annexe

Annexe **A**

Informations supplémentaires pour le chapitre 7

TABLE A.1 – Twenty lowest energy conformers of protonated α GlcN at MP2 level. Energies include : the relative ZPE-corrected energy calculated with CAM-B3LYP/6-311++G(2df.2pd), the relative energy calculated with MP2/6-311++G(3df.3pd) and the relative ZPE-corrected energy calculated with MP2/6-311++G(3df.3pd) (the ZPE correction was calculated at the CAM-B3LYP/6-311++G(2df.2pd) level of theory). The corresponding absolute energy for the lowest energy structure is indicated in parentheses in Hartree. Relative energies are given in kJ/mol.

Name	Structure	Relative CAM-B3LYP ZPE-corrected energy	Relative MP2 energy	Relative MP2 ZPE-corrected energy
g1	1C_4	0.00	0.00	0.00
g23	4C_1	3.12	7.10	6.12
g6	1C_4	12.41	11.47	10.98
g12	1C_4	10.97	11.98	9.68
g8	4C_1	9.20	12.85	10.84
g11	${}^3\text{-}OB$	14.81	19.32	17.61
g14	OS_2	18.91	19.76	18.15
g13	OS_2	21.53	19.85	19.99
g10	1C_4	18.77	20.23	18.60
g15	1C_4	21.50	22.67	19.43
g19	4C_1	17.38	23.24	20.13
g16	4C_1	21.38	25.29	22.62
g18	2SO	20.44	26.02	23.51
g17	4C_1	23.00	26.69	24.28
g21	2SO	26.13	30.71	28.67
g20	4C_1	25.71	30.96	28.06
g29	1S_3	30.41	35.69	33.45
g24	${}^3\text{-}OB$	34.02	36.08	34.08
g27	${}^3\text{-}OB$	35.89	39.31	37.28
g32	2SO	35.86	41.10	38.44

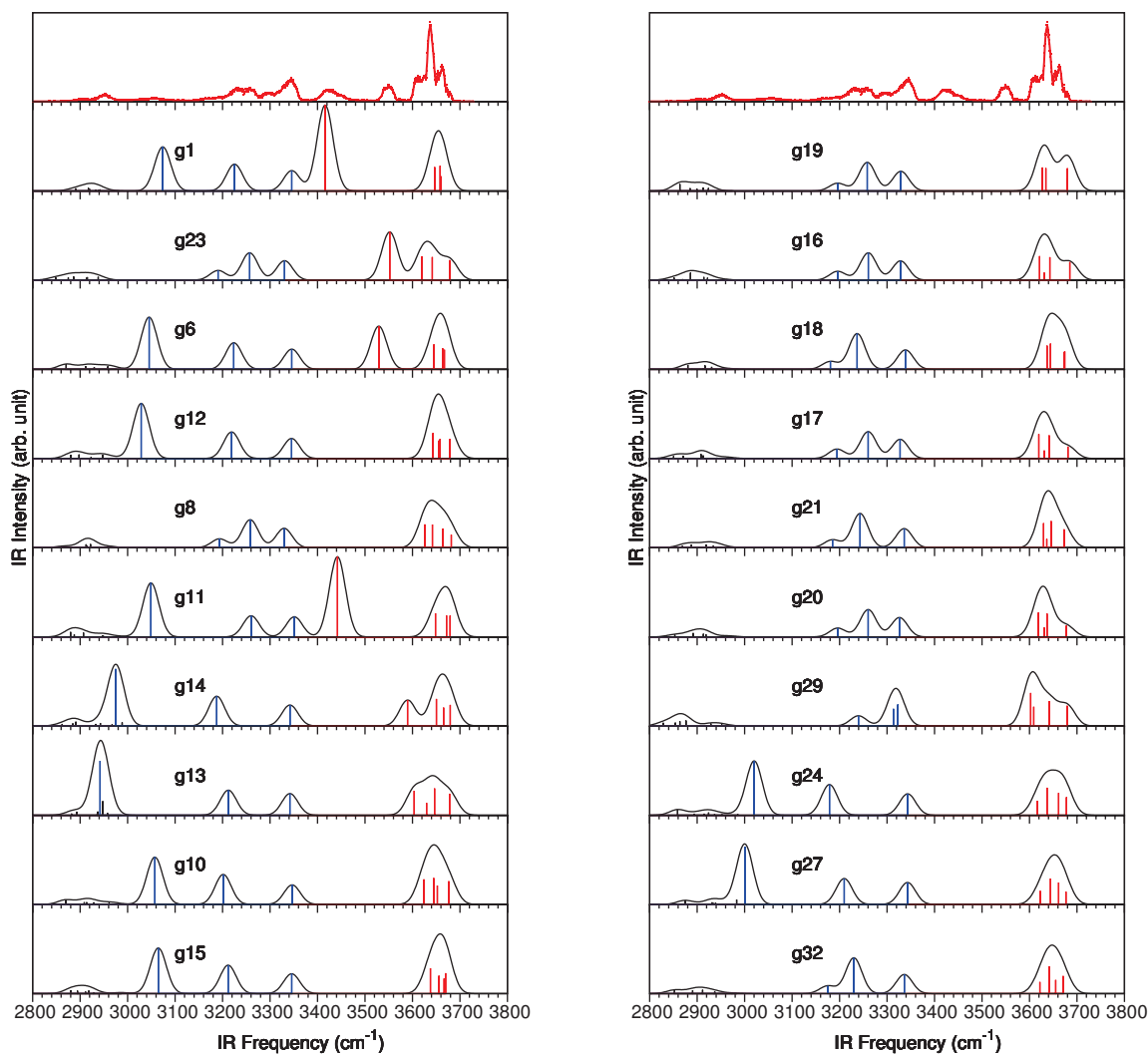


FIGURE A.1 – IR spectra of the 20 lowest energy structures of α GlcN protonated on the amine group. The IR spectra are calculated at the CAM-B3LYP/6-311++G(2df,2pd) level of theory and scaled with a scaling factor of 0.947. They are convoluted (black line) with a gaussian profile and a full width-half maximum of 10 cm^{-1} . The stretching modes are color coded : NH in blue, OH's carbohydrate in red, CH in black. The IRMPD spectrum is shown in the top panel (red line) for comparison.

TABLE A.2 – Twenty lowest energy conformers of protonated β GlcN at MP2 level. Energies include : the relative ZPE-corrected energy calculated with CAM-B3LYP/6-311++G(2df.2pd), the relative energy calculated with MP2/6-311++G(3df.3pd) and the relative ZPE-corrected energy calculated with MP2/6-311++G(3df.3pd) (the ZPE correction was calculated at the CAM-B3LYP/6-311++G(2df.2pd) level of theory). The corresponding absolute energy for the lowest energy structure is indicated in parentheses in Hartree. Relative energies are given in kJ/mol.

Name	Structure	Relative CAM-B3LYP ZPE-corrected energy	Relative MP2 energy	Relative MP2 ZPE-corrected energy
g1	4C_1	0.00	0.00	0.00
g3	4C_1	6.79	6.36	5.33
g18	4C_1	14.68	16.75	14.42
g5	1C_4	25.29	17.43	20.84
g11	3S_1	20.95	17.58	17.33
g4	1C_4	26.68	18.53	22.06
g8	1C_4	31.24	19.26	22.45
g7	1C_4	31.29	19.92	23.08
g15	4C_1	19.92	20.48	18.05
g12	2S_O	24.71	20.90	22.67
g13	4C_1	21.15	20.91	19.33
g19	5S_1	27.26	23.08	24.25
g14	3S_1	26.15	23.58	23.11
g27	1C_4	30.45	23.68	24.34
g17	2S_O	29.42	24.73	26.87
g21	4C_1	23.65	25.03	22.95
g30	${}^3-OB$	34.02	28.42	28.78
g32	2S_O	33.28	32.10	31.57
g33	B_{3-O}	34.17	34.67	34.83
g37	4C_1	32.54	34.98	31.94

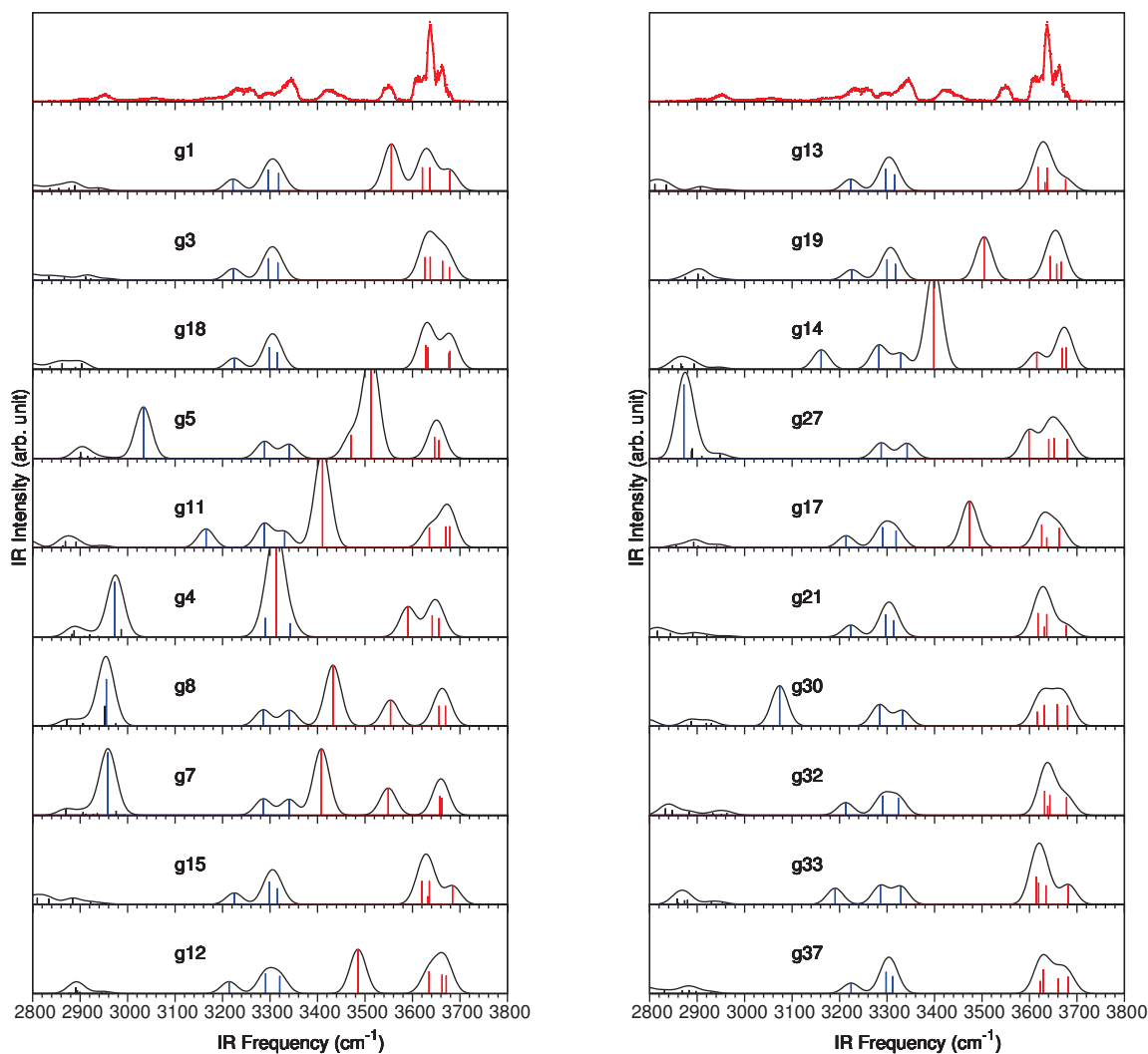


FIGURE A.2 – IR spectra of the 20 lowest energy structures of β GlcN protonated on the amine group. The IR spectra are calculated at the CAM-B3LYP/6-311++G(2df,2pd) level of theory and scaled with a scaling factor of 0.947. They are convoluted (black line) with a gaussian profile and a full width-half maximum of 10 cm^{-1} . The stretching modes are color coded : NH in blue, OH's carbohydrate in red, CH in black. The IRMPD spectrum is shown in the top panel (red line) for comparison.

TABLE A.3 – Twenty lowest energy conformers of protonated α GalN at MP2 level. Energies include : the relative ZPE-corrected energy calculated with CAM-B3LYP/6-311++G(2df.2pd), the relative energy calculated with MP2/6-311++G(3df.3pd) and the relative ZPE-corrected energy calculated with MP2/6-311++G(3df.3pd) (the ZPE correction was calculated at the CAM-B3LYP/6-311++G(2df.2pd) level of theory). The corresponding absolute energy for the lowest energy structure is indicated in parentheses in Hartree. Relative energies are given in kJ/mol.

Name	Structure	Relative CAM-B3LYP ZPE-corrected energy	Relative MP2 energy	Relative MP2 ZPE-corrected energy
g2	4C_1	0.00	0.00	0.00
g1	4C_1	1.00	0.16	0.61
g3	4C_1	5.98	5.18	5.24
g7	4C_1	17.14	19.00	16.27
g5	4C_1	18.49	19.99	18.59
g11	4C_1	17.61	20.14	17.90
g37	1S_3	27.59	26.34	26.39
g10	1S_3	26.62	26.57	26.22
g12	4C_1	24.36	26.58	24.11
g6	2S_O	28.11	26.74	28.22
g13	4C_1	25.14	27.31	24.54
g26	1S_3	32.51	31.74	30.52
g18	4C_1	32.42	33.17	31.91
g15	2S_O	33.22	33.68	33.14
g17	1C_4	37.53	35.19	35.39
g19	B^{3-O}	36.94	37.07	36.33
g20	1C_4	43.79	43.29	42.01
g21	5S_1	46.99	44.99	45.87
g27	1S_3	44.27	46.31	43.59
g31	1S_3	47.19	49.31	46.62

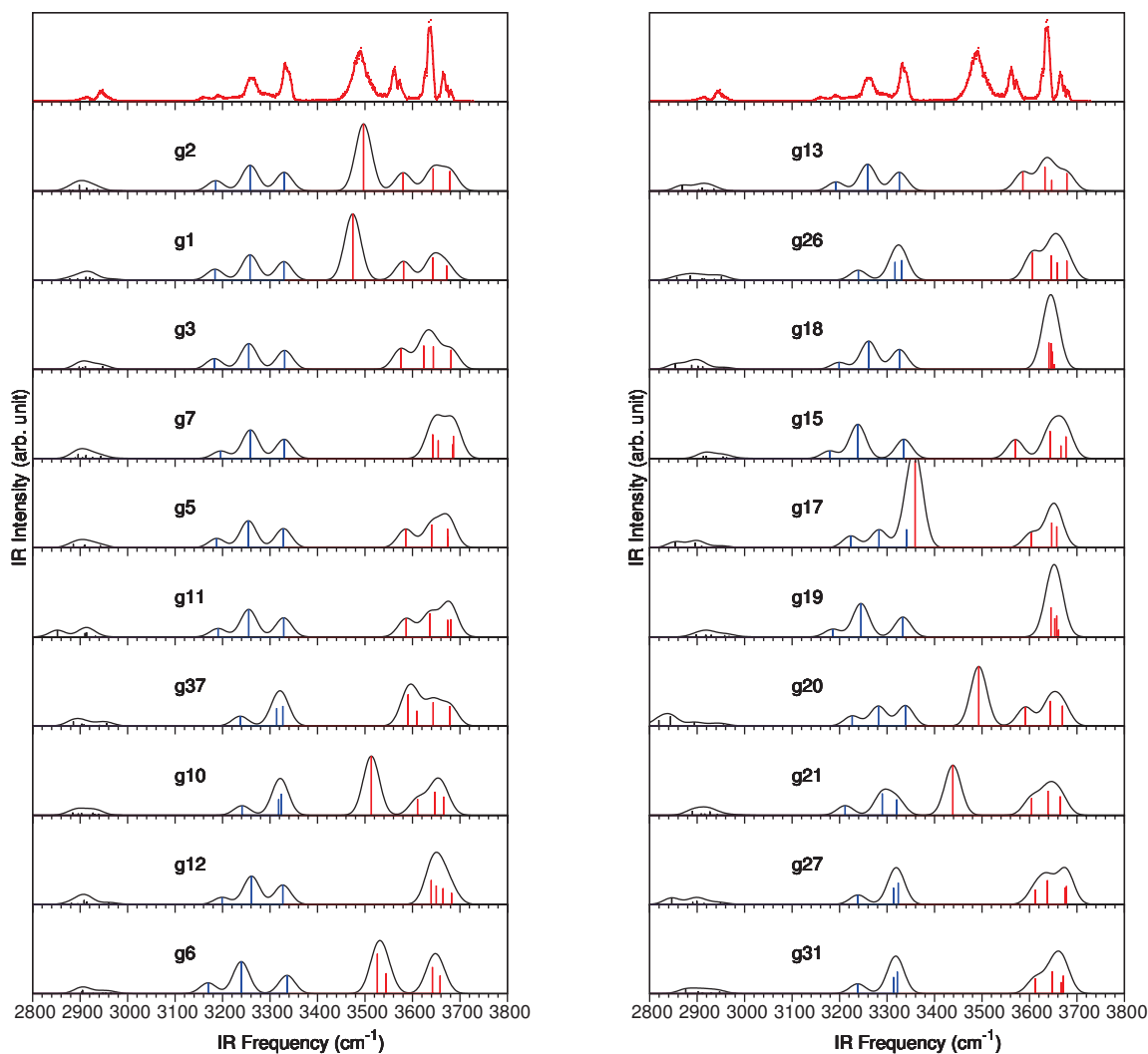


FIGURE A.3 – IR spectra of the 20 lowest energy structures of α GalN protonated on the amine group. The IR spectra are calculated at the CAM-B3LYP/6-311++G(2df,2pd) level of theory and scaled with a scaling factor of 0.947. They are convoluted (black line) with a gaussian profile and a full width-half maximum of 10 cm^{-1} . The stretching modes are color coded : NH in blue, OH's carbohydrate in red, CH in black. The IRMPD spectrum is shown in the top panel (red line) for comparison.

TABLE A.4 – Eighteen lowest energy conformers of protonated β GalN at MP2 level. Energies include : the relative ZPE-corrected energy calculated with CAM-B3LYP/6-311++G(2df.2pd), the relative energy calculated with MP2/6-311++G(3df.3pd) and the relative ZPE-corrected energy calculated with MP2/6-311++G(3df.3pd) (the ZPE correction was calculated at the CAM-B3LYP/6-311++G(2df.2pd) level of theory). The corresponding absolute energy for the lowest energy structure is indicated in parentheses in Hartree. Relative energies are given in kJ/mol.

Name	Structure	Relative CAM-B3LYP ZPE-corrected energy	Relative MP2 energy	Relative MP2 ZPE-corrected energy
g1	4C_1	0.00	0.00	0.00
g2	4C_1	-0.46	0.87	0.00
g3	4C_1	3.93	4.16	3.71
g10	4C_1	17.01	19.69	15.95
g5	4C_1	17.11	19.80	17.44
g7	4C_1	19.29	20.91	18.27
g12	4C_1	24.78	26.90	24.41
g14	4C_1	23.81	27.17	23.51
g8	2S_O	29.58	28.70	29.94
g6	2S_O	34.63	29.51	32.91
g18	1S_3	46.11	39.79	42.11
g15	2S_O	41.38	40.77	41.16
g16	5S_1	42.65	41.78	42.22
g17	2S_O	46.51	46.23	44.95
g19	2S_O	54.40	53.14	52.28
g20	2S_O	53.57	53.77	52.96
g21	B_{3-O}	56.45	56.43	56.03
g22	1S_3	64.05	60.53	60.89

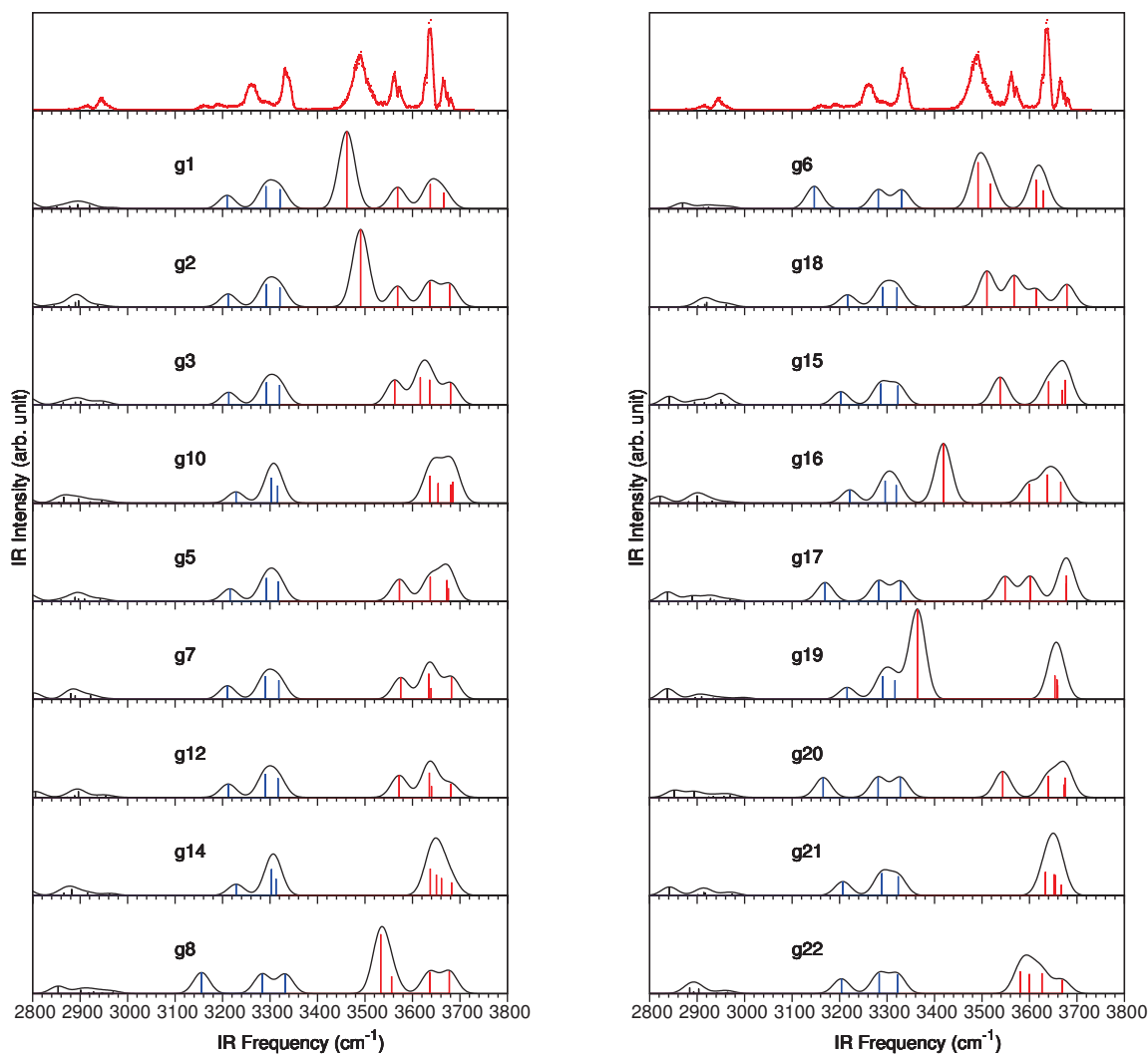


FIGURE A.4 – IR spectra of the 18 lowest energy structures of β GalN protonated on the amine group. The IR spectra are calculated at the CAM-B3LYP/6-311++G(2df,2pd) level of theory and scaled with a scaling factor of 0.947. They are convoluted (black line) with a gaussian profile and a full width-half maximum of 10 cm^{-1} . The stretching modes are color coded : NH in blue, OH's carbohydrate in red, CH in black. The IRMPD spectrum is shown in the top panel (red line) for comparison.

TABLE A.5 – Twenty lowest energy conformers of protonated α ManN at MP2 level. Energies include : the relative ZPE-corrected energy calculated with CAM-B3LYP/6-311++G(2df.2pd), the relative energy calculated with MP2/6-311++G(3df.3pd) and the relative ZPE-corrected energy calculated with MP2/6-311++G(3df.3pd) (the ZPE correction was calculated at the CAM-B3LYP/6-311++G(2df.2pd) level of theory). The corresponding absolute energy for the lowest energy structure is indicated in parentheses in Hartree. Relative energies are given in kJ/mol.

Name	Structure	Relative CAM-B3LYP ZPE-corrected energy	Relative MP2 energy	Relative MP2 ZPE-corrected energy
g4	2H_1	0.00	0.00	0.00
g7	2H_1	0.62	0.23	0.43
g8	1C_4	0.73	1.95	3.48
g16	E_1	2.84	2.67	4.28
g23	4C_1	1.08	2.88	3.50
g5	1C_4	4.64	6.82	7.70
g21	E_1	10.81	9.96	11.77
g14	E_1	13.62	13.45	14.59
g11	E_1	18.10	14.89	18.09
g12	B_{1-4}	20.72	15.27	20.00
g13	3S_1	15.65	16.69	17.59
g17	B_{3-O}	11.36	16.78	16.89
g18	B_{3-O}	14.04	17.94	18.96
g19	B_{3-O}	13.43	18.76	18.90
g20	${}^3-O B$	16.93	20.53	20.18
g27	4C_1	16.99	20.54	19.21
g29	1C_4	21.63	25.22	24.04
g43	4C_1	22.95	25.40	24.65
g31	${}^O S_2$	20.44	25.74	26.07
g32	1S_5	22.37	28.39	27.74

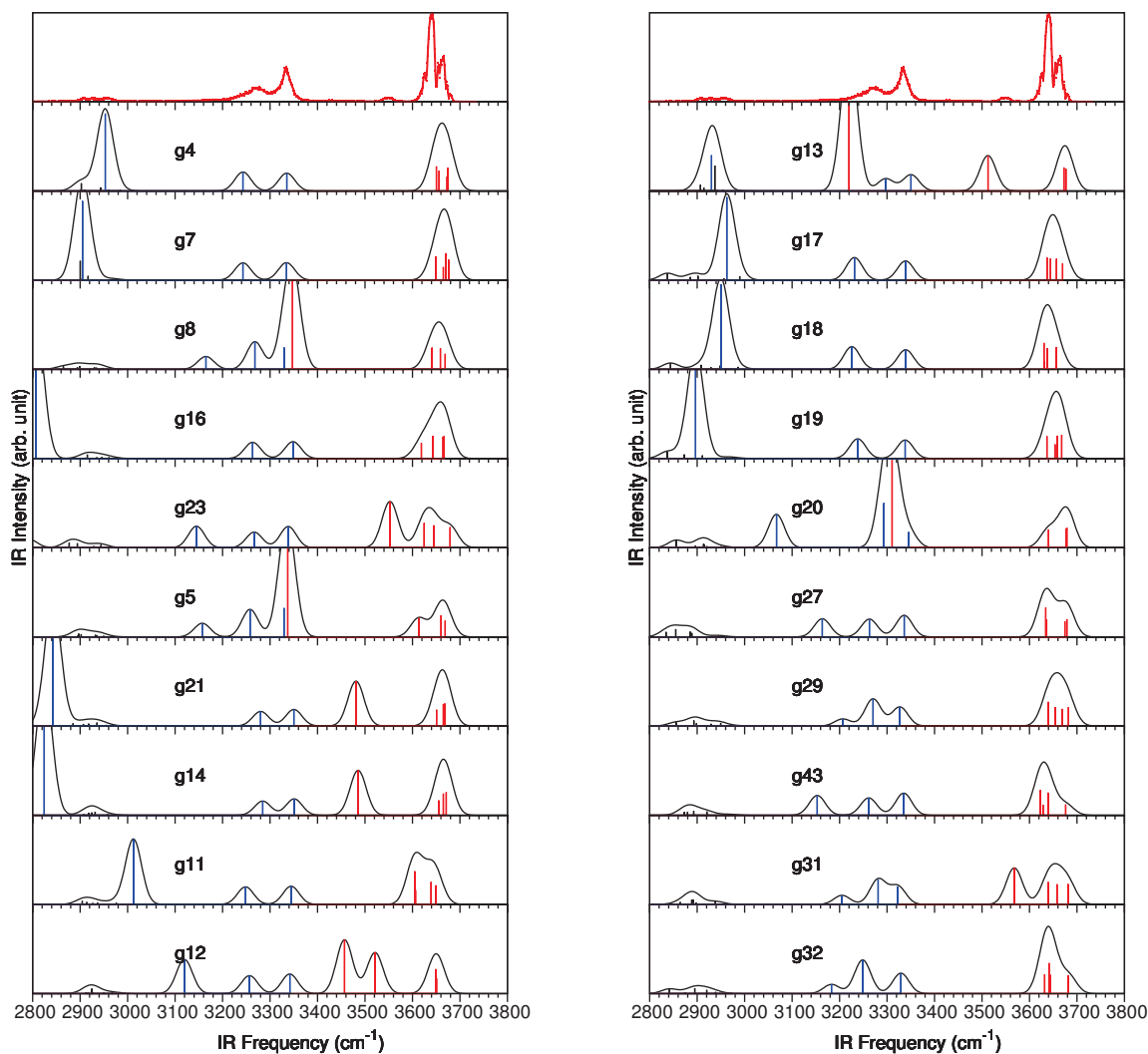


FIGURE A.5 – IR spectra of the 20 lowest energy structures of α ManN protonated on the amine group. The IR spectra are calculated at the CAM-B3LYP/6-311++G(2df,2pd) level of theory and scaled with a scaling factor of 0.947. They are convoluted (black line) with a gaussian profile and a full width-half maximum of 10 cm^{-1} . The stretching modes are color coded : NH in blue, OH's carbohydrate in red, CH in black. The IRMPD spectrum is shown in the top panel (red line) for comparison.

TABLE A.6 – Twenty lowest energy conformers of protonated β ManN at MP2 level. Energies include : the relative ZPE-corrected energy calculated with CAM-B3LYP/6-311++G(2df.2pd), the relative energy calculated with MP2/6-311++G(3df.3pd) and the relative ZPE-corrected energy calculated with MP2/6-311++G(3df.3pd) (the ZPE correction was calculated at the CAM-B3LYP/6-311++G(2df.2pd) level of theory). The corresponding absolute energy for the lowest energy structure is indicated in parentheses in Hartree. Relative energies are given in kJ/mol.

Name	Structure	Relative CAM-B3LYP ZPE-corrected energy	Relative MP2 energy	Relative MP2 ZPE-corrected energy
g1	3S_1	0.00	0.00	0.00
g9	4C_1	1.54	0.03	-1.02
g2	4C_1	0.61	0.95	0.75
g4	B_{1-4}	8.87	5.31	5.92
g7	B_{1-4}	11.91	5.52	7.90
g17	B_{1-4}	6.58	5.95	5.32
g40	2H_1	7.92	7.44	6.01
g10	2S_O	8.55	9.07	9.72
g3	1S_3	8.36	9.10	9.80
g16	2H_1	13.52	11.60	10.99
g14	2S_O	11.64	12.92	12.81
g6	1S_3	14.08	14.52	15.44
g25	4C_1	15.53	17.23	15.50
g20	1C_4	25.60	22.44	24.97
g18	1S_3	21.22	23.46	24.24
g19	1C_4	24.29	23.74	24.47
g23	4C_1	24.16	25.24	23.79
g26	4C_1	29.51	30.76	29.35
g28	$^3-O B$	32.46	31.72	31.20
g32	3S_1	37.48	40.77	38.68

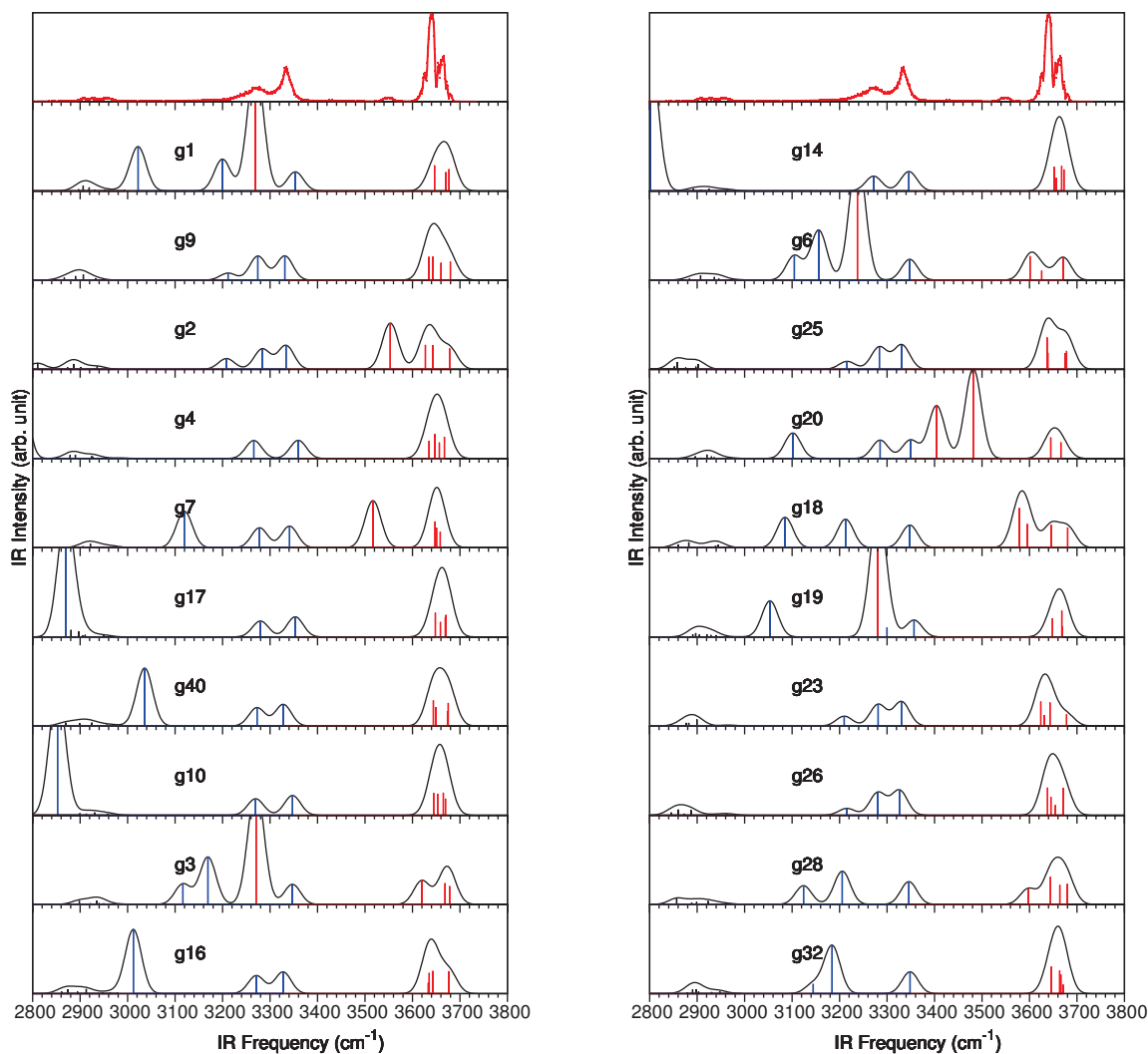


FIGURE A.6 – IR spectra of the 20 lowest energy structures of β ManN protonated on the amine group. The IR spectra are calculated at the CAM-B3LYP/6-311++G(2df,2pd) level of theory and scaled with a scaling factor of 0.947. They are convoluted (black line) with a gaussian profile and a full width-half maximum of 10 cm^{-1} . The stretching modes are color coded : NH in blue, OH's carbohydrate in red, CH in black. The IRMPD spectrum is shown in the top panel (red line) for comparison.

Liste d'articles

1. Baptiste Schindler, **Loïc Barnes**, Gina Renois, Christopher Gray, Abdul-Rahman Allouche, Stéphane Chambert, Sebastien Fort, Sabine Flitsch, Claire Loison and Isabelle Compagnon. Anomeric memory of the glycosidic bond upon fragmentation, and its consequences for carbohydrate sequencing, *Nature Communications*, 8, 973, 2017.
2. **Loïc Barnes**, Baptiste Schindler, Stéphane Chambert, Abdul-Rahman Allouche, and Isabelle Compagnon. The remarkable structural complexity of simple carbohydrate building blocks : conformational and anomeric heterogeneity of hexosamines resolved by IRMPD spectroscopy and ab initio calculations (being drafted).
3. **Loïc Barnes**, Baptiste Schindler, Stéphane Chambert, Abdul-Rahman Allouche, and Isabelle Compagnon. Conformational preferences of protonated N-acetylated hexosamines probed by infrared multiple photon dissociation (IRMPD) spectroscopy and ab initio calculations. *International Journal of Mass Spectrometry*, 421 :116-123, 2017.
4. B. Schindler, **L. Barnes**, C. J. Gray, S. Chambert, S. L. Flitsch, J. Oomens, R. Daniel, A. R. Allouche, and I. Compagnon. IRMPD spectroscopy sheds new (infrared) light on the sulfate pattern of carbohydrates. *The Journal of Physical Chemistry A*, 121(10) :2114-2120, 2017.
5. Baptiste Schindler, Gina Renois-Predelus, Nassiba Bagdadi, Sihem Melizi, **Loïc Barnes**, Stéphane Chambert, Abdul-Rahman Allouche, and Isabelle Compagnon. MS/IR, a new MS-based hyphenated method for analysis of hexuronic acid epimers in glycosaminoglycans. *Glycoconjugate Journal*, 34(3) :421-425, 2017.
6. **Loïc Barnes**, Baptiste Schindler, Isabelle Compagnon, and Abdul-Rahman Allouche. Fast and accurate hybrid QM//MM approach for computing anharmonic corrections to vibrational frequencies. *J. Mol. Model.* 22 : 285, 2016.
7. **Loïc Barnes**, Baptiste Schindler, Abdul-Rahman Allouche, Daniel Simon, Stéphane Chambert, Jos Oomens, and Isabelle Compagnon. Anharmonic simulations of the vibrational spectrum of sulfated compounds : application to the glycosaminoglycan fragment glucosamine 6-sulfate. *Phys. Chem. Chem. Phys.*, 17 :25705-25713, 2015.

Étude théorique des saccharides : structures et spectres infrarouges

Notre travail s'effectue dans le but de fournir une méthode pour permettre le séquençage des oligosaccharides, des biopolymères jouant un rôle clé dans le domaine de la santé comme dans certains domaines industriels tels que l'énergie ou l'alimentation. Les méthodes utilisées pour séquencer les protéines sont limitées pour l'étude des oligosaccharides. En revanche, nous montrons que l'alliance entre la méthode expérimentale de spectroscopie d'action (IRMPD) et les calculs théoriques est une méthode prometteuse pour le séquençage des oligosaccharides.

Notre méthode théorique consiste à obtenir des minima de la surface d'énergie potentielle. Les spectres infrarouges de ces minima sont calculés avec la DFT et en utilisant une grande base. Les spectres calculés sont alors comparés aux spectres expérimentaux afin d'obtenir des informations sur la structure des constituants ou la composition du mélange dans l'échantillon.

Notre approche alliant les calculs théoriques aux mesures IRMPD s'est avérée particulièrement efficace pour obtenir les structures des mono ou disaccharides présents dans l'échantillon. De plus elle permet d'identifier des signatures spécifiques à des anomères ou à des conformères.

Theoretical study of saccharides : structures and infrared spectra

Our work aims to develop a method to allow the sequencing of oligosaccharides, which are biopolymers playing a key role in the field of Health as in several industrial fields such as energy or nutrition. The methods used for sequencing proteins are limited in the study of oligosaccharides. However, we show that the combination of action spectroscopy experimental method (IRMPD) and theoretical calculations is a promising method to sequence oligosaccharides.

Our theoretical method consist of obtaining minima of the potential energy surface. The infrared spectra of these minima are computed with DFT and a large basis. The calculated spectra are then compared to the experimental ones to gain informations on the structure of the compounds or on the composition of the mixture in the sample.

Our approach combining theoretical computations to IRMPD measurements is particularly effective to obtain the structures of mono and disaccharides present in the sample. Moreover, it allows to identify specific signatures of anomers and conformers.