

THESE DE DOCTORAT DE
l'Université de Pau et des Pays de l'Adour
École doctorale des Sciences Exactes et leurs Applications
Spécialité Chimie Physique

Présentée par

Sandrine THICOÏPÉ

Pour obtenir le grade de DOCTEUR de l'UPPA

**Développement et Application de modèles dépendant et
indépendant du temps pour l'étude des propriétés
spectroscopiques de composés d'intérêt biologique**

soutenue le 10 décembre 2012 devant le jury composé de :

M. Xavier ASSFELD Université de Lorraine *Président*
Mme Gilberte CHAMBAUD Université de Marne la Vallée *Rapporteuse*
M. Eric PERPETE Université de Namur *Rapporteur*
Mme Isabelle BARAILLE Université de Pau *Examinatrice*
M. Claude POUCHAN Université de Pau *Directeur de thèse*
M. Philippe CARBONNIERE Université de Pau *Co-directeur de thèse*

Remerciements

Ah les remerciements... C'est la dernière partie à rédiger, et ce n'est pas forcément la plus simple... Je vais tenter de faire en sorte de n'oublier personne, et je m'excuse par avance si c'est le cas.

Je tiens tout d'abord à remercier sincèrement mon directeur de thèse Claude POUCHAN, pour m'avoir fait confiance après mon M2R et pouvoir offert l'opportunité de découvrir le milieu de la recherche. J'en profite aussi pour y associer Philippe CARBONNIERE, qui m'a patiemment encadrée durant ces trois années.

Je tiens aussi à remercier tous les membres du jury d'avoir accepté de juger ce travail malgré leurs emplois du temps surchargés, et l'IPREM, en particulier l'ECP, de m'avoir accueillie.

Quand j'aurai plus de recul sur ces trois années, je repenserai sans aucun doute aussi à mes collègues de bureau et aux fous rires partagés, pas toujours autour de problématiques scientifiques, je l'avoue! Je pense à Sharky, Hustst (ils se reconnaîtront...) qui sont devenus tour à tour fleuristes, sherpas et puis re-fleuristes... Je pense aussi à Pierre, parti montrer ces talents de docteur sous d'autres cieux et à Khaled, même si on n'a pas eu l'occasion de partager longtemps notre bureau. Enfin, je joins à tout ce petit monde Claire et Hélène, dont j'apprécie la bonne humeur et le naturel.

Mes années de doctorat sont aussi indissociables de l'enseignement, qui m'a permis de me sortir la tête du guidon. J'ai pu travailler et découvrir des personnes de mon laboratoire ou d'autres différents. J'aimerais en particulier remercier Julie et Hélène, dont l'accueil chaleureux m'a souvent ressourcée lors des TP. Je vais citer en vrac Didier BEGUE, Germain VALLVERDU, Eve PERE, Delphine FLAHAUT, Pierre MOCHO, Laurent OUERDANE qui m'ont tous permis, à un moment ou à un autre et à leur façon, d'avancer.

J'ai aussi une pensée pour Michel Gelize, sans qui le travail sur les aérosols de nitrate n'aurait pas été possible.

Enfin, je remercie ma mère et ma sœur, sans qui tout ça n'aurait pas été possible. Je sais que le stress peut me rendre grognon, alors merci d'avoir supporté cela et d'avoir été toujours là pour me soutenir et m'encourager pendant ces trois années, malgré des contextes parfois tout aussi durs pour vous.

Table des matières

1	Introduction	10
2	Rappels théoriques	13
I	Généralités	14
II	Les différentes méthodes de détermination de la structure électronique	15
II.1	Méthode Hartree-Fock	15
II.2	Méthodes semi-empiriques	19
II.3	Méthodes post-HF	20
II.4	Théorie de la fonctionnelle de la densité	23
II.5	Bases de fonctions atomiques	28
III	Algorithmes de recherche de structure géométrique	30
III.1	Méthodes locales	31
III.2	Méthodes non locales	33
IV	Problème vibrationnel	36
IV.1	Traitement indépendant du temps : méthode statique	36
IV.2	Traitement dépendant du temps : méthode dynamique (DM)	44
IV.3	Conditions de calcul	51
	Bibliographie	52
3	Étude de molécules hydratées : cas d'un système inorganique, les aérosols de nitrate	58
I	Introduction	59
I.1	Formation et fonction des aérosols	59
I.2	Position du problème	60
II	Conditions calculatoires	61
II.1	Choix de la méthode	62

II.2	Choix de la base	63
II.3	Choix du modèle structural	63
III	Propriétés vibrationnelles des aérosols	67
III.1	Publication	67
III.2	Compléments	87
IV	Conclusion	88
	Bibliographie	90
4	Étude de molécules hydratées : cas d'un système organique, structures géométriques des bases d'acides nucléiques	92
I	Introduction	93
I.1	Présentation générale de l'ADN et de l'ARN	93
I.2	Hydratation de l'ADN	95
I.3	Position du problème	96
II	Étude bibliographique	97
II.1	Cas de l'uracile	97
II.2	Cas de la thymine	97
II.3	Cas de la cytosine	98
II.4	Cas de l'adénine	98
II.5	Cas de la guanine	98
III	Publication	99
IV	Conclusion	126
	Bibliographie	127
5	Étude de molécules hydratées : cas d'un système organique, spectres vibrationnels des bases d'acides nucléiques	130
I	Choix du modèle de représentation des bases hydratées	131
I.1	Cas de la cytosine	131
I.2	Un cas comparable à la cytosine : étude de la thymine	137
I.3	Validation sur les autres bases dans la zone $1000-1800\text{cm}^{-1}$	141
II	Choix du traitement vibrationnel	146
II.1	Cas de petites molécules azotées	147
II.2	Cas des bases d'acides nucléiques : zone $1000-1800\text{cm}^{-1}$	148
III	Conclusion et Perspectives	149

Bibliographie	150
6 Conclusion	152
A Nombres d'onde anharmoniques de la cytosine isolée et en solution aqueuse déterminés à un niveau B3LYP/6-31+G(d,p)	155
B Nombres d'onde anharmoniques de l'uracile isolée et en solution aqueuse déterminés à un niveau B3LYP/6-31+G(d,p)	158
C Nombres d'onde anharmoniques de la thymine isolée et en solution aqueuse déterminés à un niveau B3LYP/6-31+G(d,p)	161
D Nombres d'onde anharmoniques de la guanine isolée et en solution aqueuse déterminés à un niveau B3LYP/6-31+G(d,p)	164
E Nombres d'onde anharmoniques de l'adénine isolée et en solution aqueuse déterminés à un niveau B3LYP/6-31+G(d,p)	168
F Publication sur la cytosine	171

Table des figures

2.1	Principe de la méthode VCI-P	43
3.1	Description des modes internes des aérosols de nitrate étudiés	60
3.2	Valeurs des nombres d'onde relatifs aux 5 modes de vibration de l'ion nitrate en fonction de n	61
3.3	Écarts (en cm^{-1}) par rapport aux fréquences harmoniques de NO_3^- calculées au niveau CCSD(T)/aug-cc-pVTZ pour différents niveaux calculatoires	62
3.4	Structures les plus stables adoptées par les systèmes $(\text{NaNO}_3, n\text{H}_2\text{O})$ au niveau B3LYP/6-31+G(d,p) obtenues à l'issue d'une dynamique moléculaire à 200K	64
3.5	Fréquences harmoniques (en cm^{-1}) des structures les plus stables de la forme $(\text{NaNO}_3, n\text{H}_2\text{O})$ issues d'une thermalisation à 200K, et à un niveau B3LYP/6-31+G(d,p)	64
3.6	Structures les plus stables adoptées par les systèmes $(\text{NaNO}_3, 1\text{H}_2\text{O})_x$ avec $x=1$ à 4 au niveau B3LYP/6-31+G(d,p)	65
3.7	Écarts (en cm^{-1}) entre les fréquences anharmoniques obtenues à un niveau B3LYP/6-31+G(d,p) pour chaque structure la plus stable de la forme $(\text{NaNO}_3, 1\text{H}_2\text{O})_x$ avec $x=1$ à 4 et leurs contreparties expérimentales en fonction du mode considéré	66
3.8	Bandes spectrales correspondant aux modes 3A et 3B pour $n=1$ à 4 dans le cadre de la méthode B3LYP/6-311+G(d,p) appliquée au modèle $(\text{NaNO}_3, n\text{H}_2\text{O})_3$	88
3.9	Nombres d'onde anharmoniques (en cm^{-1}) obtenus au niveau B3LYP/6-311+G(d,p) pour les 5 modes caractéristiques de l'unité NO_3^- de $\text{NaNO}_3, n\text{H}_2\text{O}$ en fonction de n (triangle) avec leurs incertitudes (en pointillés) - comparaison avec les données expérimentales (rond)	89
4.1	Molécules d'ADN (A,T,G,C) et d'ARN (U) ainsi que la numérotation associée aux atomes	94
4.2	Représentation schématique de l'ADN	94

5.1	Paramètres géométriques de la cytosine obtenus au niveau B3LYP/6-31+G(d,p) selon le modèle de représentation de la solvation considéré	132
5.2	Fréquences anharmoniques DM entre 100 et 1800cm ⁻¹ obtenues pour différents modèles représentant l'uracile en solution aqueuse à un niveau B3LYP/6-31+G(d,p) - Calculs des écarts (Δ) entre théorie et expérience	133
5.3	Fréquences anharmoniques VPT2 entre 100 et 1800cm ⁻¹ obtenues pour différents modèles représentant la thymine en solution aqueuse à un niveau B3LYP/6-31+G(d,p) - Calculs des écarts (Δ) entre théorie et expérience	138
5.4	Écarts en cm ⁻¹ entre les valeurs harmoniques et anharmoniques obtenues pour la thymine isolée et hydratée dans la zone 1000-1800cm ⁻¹ aux niveaux VPT2 et DM dans le cadre des méthodes B3LYP et B3LYP-D/6-31+G(d,p)- Comparaison avec les données expérimentales (<i>suite : page suivante</i>)	139
5.5	Fréquences anharmoniques VPT2 et DM de la méthanimine et du pyrrole obtenues avec un calcul de structure électronique réalisé à un niveau B3LYP/6-31+G(d,p)	147
5.6	Écarts moyens obtenus (en cm ⁻¹) entre les nombres d'onde anharmoniques théoriques (DM ou VPT2) et leurs contreparties expérimentales dans la zone 1000-1800cm ⁻¹ pour les différentes bases d'acides nucléiques isolées ou en solution aqueuse à un niveau B3LYP/6-31+G(d,p)	149

Chapitre 1

Introduction

NOM THICOÏPÉ

PRÉNOM Sandrine

SUJET Développement et application des modèles dépendant et indépendant du temps pour l'étude des propriétés spectroscopiques de composés d'intérêt biologique

L'étude spectroscopique des molécules, en particulier IR et Raman, donne lieu à de nombreuses expériences afin de comprendre les propriétés des systèmes étudiés. Cependant, leur complexité constitue un frein à la bonne interprétation des spectres, d'où la nécessité de mener des études théoriques en parallèle. Les premières études de ce type concernant les propriétés spectroscopiques ont été basées sur le calcul des champs de force harmoniques pour des raisons pratiques. Toutefois, afin d'évaluer avec précision les fréquences de transitions vibrationnelles, la prise en compte de l'anharmonicité est indispensable. L'obtention *a priori* de ces données n'est pas aisée pour des systèmes de grande taille tels que les molécules d'intérêt biologique et les molécules solvatées. Dans ce dernier cas, la qualité des résultats dépend de la bonne représentation de la structure géométrique considérée. La recherche de modèles géométriques et le traitement théorique mis en jeu pour comprendre et expliquer les observations spectrales expérimentales ont donc été au centre de ce travail de thèse. La présentation générale des systèmes étudiés, à savoir les aérosols de nitrate et les bases d'acides nucléiques isolées et microhydratées, a fait l'objet de deux parties introductives au début de chaque chapitre les concernant : y sont rassemblés l'analyse des données bibliographiques antérieures et les enjeux de notre travail.

Les méthodes théoriques utilisées au cours de ce travail ont été décrites dans une première partie. Ce chapitre donne un aperçu des méthodes calculatoires que nous avons utilisées en insistant davantage sur les méthodes et les algorithmes mis au point au laboratoire tels que GSAM, qui permet de déterminer les structures géométriques des clusters moléculaires, et VCI-P pour traiter des modes de vibrations dans l'hypothèse anharmonique.

La deuxième partie propose l'élaboration d'un modèle géométrique représentant les aérosols de nitrate à partir des résultats issus du code GSAM, utilisé jusqu'à maintenant avec succès pour décrire les agrégats atomiques, et appliqué ici à la détermination géométrique des clusters moléculaires. Le but ultime de ce travail est de reproduire avec la meilleure précision possible les fréquences de vibration de ces systèmes en fonction de leurs taux d'humidité relative. Ce travail vient d'être soumis à publication. L'étude des bases d'acides nucléiques a fait l'objet des deux parties suivantes de notre travail. Dans un premier temps, le chapitre quatre est consacré à une étude géométrique complète des bases d'acides nucléiques microhy-

dratées (pour un nombre de molécules d'eau compris entre 1 et 5). Cette étude a été menée à l'aide de calculs DFT en utilisant deux niveaux calculatoires B3LYP et B3LYP-D afin de s'assurer de l'influence des termes dispersifs pour traduire l'existence des liaisons hydrogène. Cette étude, qui utilise aussi le code GSAM, fait l'objet d'un article invité à *PCCP* dans un numéro spécial consacré aux clusters.

Enfin, le cinquième chapitre s'intéresse au calcul des propriétés vibrationnelles des molécules de cytosine et de thymine isolées et hydratées afin d'établir les modèles géométriques et les conditions calculatoires optimales permettant de décrire au mieux les données vibrationnelles expérimentales. Pour ce faire, différentes méthodes de détermination de structures électroniques (B3LYP et B3LYP-D), et de traitement vibrationnel (VPT2 et DM) ont été testées dans la zone de signature vibrationnelle, située entre 1000 et 1800cm⁻¹. Ces conditions permettent ensuite de proposer dans cette même région spectrale les modes attendus pour les autres bases d'acides nucléiques.

MOTS CLÉS *aérosol de nitrate, base d'acide nucléique, hydratation, GSAM, modèle géométrique, modèle théorique, vibration, B3LYP, B3LYP-D, VPT2, dynamique moléculaire*

Ce travail a été réalisé à : IPREM - Equipe de Chimie Physique - UMR5254

Hélioparc - 2, avenue du Président Angot

64053 PAU Cedex 09

... et a donné lieu à : 1 communication orale, 4 posters et 3 publications dont une effective, deux soumises

- "Vibrational analysis beyond the harmonicity from *ab initio* molecular dynamics" Carbonnière P., Thicoïpé S., Very T., Assfeld X. *IJQC* (2012)112, 9, 2221-2230
- "A theoretical strategy to build structural models of microhydrated inorganic systems for the knowledge of their vibrational properties. The case of the hydrated nitrate aerosols" Thicoïpé S., Carbonnière P., Pouchan C. *J. Phys. Chem.*, en révision
- "The use of the GSAM approach for the structural investigation of low lying isomers of molecular clusters from density functional theory-based potential energy surfaces : the structures of microhydrated nucleic acid bases" Thicoïpé S., Carbonnière P., Pouchan C. *PCCP*, en révision

Chapitre 2

Rappels théoriques

Sommaire

I	Généralités	14
II	Les différentes méthodes de détermination de la structure électronique	15
II.1	Méthode Hartree-Fock	15
II.2	Méthodes semi-empiriques	19
II.3	Méthodes post-HF	20
II.4	Théorie de la fonctionnelle de la densité	23
II.5	Bases de fonctions atomiques	28
III	Algorithmes de recherche de structure géométrique	30
III.1	Méthodes locales	31
III.2	Méthodes non locales	33
IV	Problème vibrationnel	36
IV.1	Traitement indépendant du temps : méthode statique	36
IV.2	Traitement dépendant du temps : méthode dynamique (DM)	44
IV.3	Conditions de calcul	51
	Bibliographie	52

I Généralités

Les méthodes de chimie quantique visent en premier lieu à la détermination de la fonction d'onde du système polyélectronique moléculaire que l'on veut étudier. Cela nécessite de résoudre l'équation de Schrödinger correspondante dans le cadre des états stationnaires, c'est-à-dire pour une énergie totale constante :

$$H\Psi(r, R) = E\Psi(r, R) \quad (\text{I.1})$$

Ψ désigne la fonction d'onde indépendante du temps et (r, R) représentent respectivement l'ensemble des coordonnées électroniques et nucléaires. Quant à H qui constitue l'hamiltonien indépendant du temps, il prend la forme générale suivante :

$$H = T_n + T_e + V_{ee} + V_{nn} + V_{ne} \quad (\text{I.2})$$

où T_n et T_e sont les opérateurs associés respectivement à l'énergie cinétique des noyaux et des électrons. V_{nn} , V_{ee} et V_{ne} représentent les potentiels coulombiens : les deux premiers sont des termes répulsifs tandis que le dernier renvoie à l'attraction noyaux-électrons. En unités atomiques et pour un système à n électrons et M noyaux, l'hamiltonien s'écrit :

$$H = - \sum_{A=1}^M \frac{\nabla_A^2}{2M_A} - \frac{1}{2} \sum_{i=1}^n \nabla_i^2 - \sum_{i=1}^n \sum_{A=1}^M \frac{Z_A}{R_{i,A}} + \sum_{i=1}^n \sum_{j < i} \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B < A} \frac{Z_A Z_B}{R_{AB}} \quad (\text{I.3})$$

Les indices A et B sont associés aux noyaux de masse M_A et M_B et de charge Z_A et Z_B , i et j désignant les électrons.

La présence du terme de répulsion électronique dans l'expression du hamiltonien rend impossible la résolution de l'équation de Schrödinger, sauf dans le cas d'un système monoélectronique. Cela implique donc de faire une approximation pour dépasser cet obstacle à l'étude de systèmes électroniques complexes. On considère alors que la masse des noyaux est supérieure à celle des électrons. Ainsi, les mouvements électroniques peuvent être découplés de ceux des noyaux : les électrons se déplacent plus rapidement que les noyaux fixes. C'est l'approximation de Born-Oppenheimer [1]. Les distances entre les noyaux sont alors traitées comme des paramètres à part entière, le terme relatif à la répulsion coulombienne noyau-noyau devenant une constante. De plus, le terme T_n étant négligeable, les termes restants de l'équation initiale constituent l'hamiltonien électronique H_e décrivant le mouvement des n électrons dans le champ des M noyaux fixes :

$$H_e = -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 - \sum_{i=1}^n \sum_{A=1}^M \frac{Z_A}{R_{i,A}} + \sum_{i=1}^n \sum_{j<i}^n \frac{1}{r_{ij}} \quad (\text{I.4})$$

Ainsi, la solution de l'équation de Schrödinger électronique $H_e(r, R)\Psi_e(r, R) = E(R)\Psi_e(r, R)$ est la fonction d'onde $\Psi_e(r, R)$ qui décrit le mouvement des électrons, les paramètres nucléaires étant constants. Pour la déterminer, le terme de répulsion interélectronique fait l'objet d'approximations et différentes hypothèses peuvent être émises, répertoriées sous différentes appellations. La méthode Hartree-Fock, basée sur l'approximation des particules indépendantes, constitue un premier niveau de calcul. Ensuite, d'autres méthodes prenant en compte la corrélation électronique, type post-HF ou des approches basées sur la théorie de la fonctionnelle de la densité, seront présentées.

II Les différentes méthodes de détermination de la structure électronique

II.1 Méthode Hartree-Fock

L'équation de Schrödinger n'étant pas directement résoluble analytiquement, on cherche une solution approchée telle que la fonction d'onde Ψ soit un produit de fonctions monoélectroniques χ_i [2, 3]. L'expression la plus simple pour cette fonction d'onde prend la forme d'un déterminant de Slater construit sur des spinorbitales orthonormées :

$$\Psi = |\chi_1 \chi_2 \dots \chi_n| \quad (\text{II.1})$$

avec $\langle \chi_i | \chi_j \rangle = \delta_{ij}$. La répulsion coulombienne électronique n'y est pas spécifiquement prise en compte.

II.1.1 Champ auto-cohérent (SCF)

Les équations de Roothan décrites un peu plus loin sont résolues de façon itérative par la procédure du champ auto-cohérent (SCF pour Self-Consistent Field). Elle est présentée dans le détail ci-dessous.

Les éléments F_{jk} de la matrice de Fock F peuvent être développés comme suit :

$$F_{jk} = H_{jk}^c + G_{jk} = \sum_{r=1}^m \sum_{s=1}^m P_{rs} \left[(jk|rs) - \frac{1}{2}(jr|ks) \right] \quad (\text{II.2})$$

avec $H_{jk}^c = \langle \chi_j | H^c | \chi_k \rangle$ et P_{rs} l'élément rs de la matrice densité totale du système P. L'introduction de la matrice X d'orthogonalisation de la base d'OA (définie telle que $X^+ S X = I$ avec X^+ matrice adjointe de X), associée au remplacement de C par $X C'$ permet de ramener la résolution des équations de Roothan à un problème aux valeurs propres : $F' C'_i = \epsilon_i C'_i$.

La résolution de l'équation ci-dessus s'effectue par la diagonalisation de la matrice F' dépendante des coefficients recherchés C_i . La méthode de résolution auto-cohérente est alors employée. Elle comprend plusieurs étapes :

1. Calcul des termes S_{jk} , H_{jk}^c et $(jk|rs)$;
2. Estimation de la matrice densité P ;
3. Construction de la matrice de Fock $F = H_{jk}^c + G_{jk}$;
4. Expression de F dans la base orthogonale : $F' = X^+ F' X$ avec $X = S^{-\frac{1}{2}}$;
5. Diagonalisation de F' ;
6. Transformation des OM dans la base non-orthogonale : $C = X C'$;
7. Calcul de la nouvelle matrice densité P et calcul de l'énergie $E = \frac{1}{2} Tr [P(H^c + F)]$;
8. Test de convergence sur E et/ou P .

Au premier cycle SCF, P est estimée souvent en négligeant par exemple les interactions électrons/électrons, ou bien encore par un calcul semi-empirique peu coûteux. A la fin de chaque cycle (étape 8), on compare la variation de la matrice P par rapport au cycle précédent à une valeur seuil fixée dès le départ. Si elle est inférieure à cette limite, la procédure SCF est stoppée ; sinon, elle se poursuit jusqu'à obtention de cette condition.

II.1.2 Equation de Hartree-Fock

L'énergie électronique associée à la fonction Ψ de l'équation (I.4) est égale à :

$$E = \langle \Psi | H_e \Psi \rangle \quad (\text{II.3})$$

On note H_c l'hamiltonien de coeur, qui correspond à l'opérateur monoélectronique décrivant un électron dans le champ des noyaux M :

$$H^c(i) = -\frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{R_{i,A}} \quad (\text{II.4})$$

L'hamiltonien électronique devient alors en fonction de H_c :

$$H_e = \sum_{i=1}^n H^c(i) + \sum_{i=1}^n \sum_{j < i}^n \frac{1}{r_{ij}} \quad (\text{II.5})$$

L'énergie électronique peut s'exprimer de la façon suivante d'après (II.2) :

$$E = \sum_{i=1}^n H_{ii}^c + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\langle \chi_i(1) \chi_j(2) | \frac{1}{r_{12}} \chi_i(1) \chi_j(2) \rangle - \langle \chi_i(1) \chi_j(2) | \frac{1}{r_{12}} \chi_j(1) \chi_i(2) \rangle) \quad (\text{II.6})$$

avec $H_{ii}^c = \langle \chi_i(1) | H^c(1) | \chi_i(1) \rangle$ qui représente l'intégrale monoélectronique de coeur.

L'application du principe variationnel impose que l'ensemble des spinorbitales χ_i est optimal quand l'énergie E est minimale. Cette condition revient à résoudre l'équation de Hartree-Fock :

$$F\chi_i = e_i\chi_i$$

où F est l'opérateur de Fock s'exprimant comme :

$$F = H^c(1) + \sum_{j=1}^n (J_j(1) - K_j(1)) \quad (\text{II.7})$$

$$J_j(1)\chi_i(1) = \chi_i(1) \int_{\tau_2} \chi_j^*(2)\chi_j(2) \frac{1}{r_{12}} d\tau_2 / K_j(1)\chi_i(1) = \chi_j(1) \int_{\tau_2} \chi_j^*(2)\chi_j(2) \frac{1}{r_{12}} d\tau_2 \quad (\text{II.8})$$

J_j et K_j sont appelés respectivement opérateurs de Coulomb et d'échange car le premier prend en compte la répulsion coulombienne entre les électrons et le second traduit la modification de l'énergie due aux effets de corrélation de spin. Les solutions de ces équations sont déterminées via un processus itératif du type méthode du champ cohérent (notée SCF). Deux cas de figure sont possibles : soit on se place dans le cadre du modèle HF avec contrainte de spin (formalisme RHF pour Restricted Hartree-Fock), soit dans le cadre du même modèle, mais sans contrainte de spin (UHF pour Unrestricted Hartree-Fock).

II.1.3 Fonction d'onde restreinte et non-restreinte (RHF/UHF)

Les développements précédents n'imposent aucune hypothèse sur les spinorbitales χ_i autre que la condition d'orthogonalité.

Dans le cadre du formalisme RHF, on impose à chaque spinorbitale de systèmes à couches fermées de représenter deux électrons de spin opposés. Autrement dit, deux spinorbitales consécutives doivent avoir la même fonction d'espace, mais leur fonction de spin doit être différente. La fonction d'onde polyélectronique d'un système possédant un nombre pair d'électrons est donnée par le déterminant de Slater suivant : $\Psi = |\chi_1 \bar{\chi}_1 \dots \chi_{\frac{n}{2}} \bar{\chi}_{\frac{n}{2}}|$

L'opérateur de Fock à couches complètes s'écrit alors de la façon suivante :

$$F = H^c(1) + \sum_{j=1}^{n/2} (2J_j(1) - K_j(1)) \quad (\text{II.9})$$

avec, en notant désormais χ_i l'orbitale moléculaire i :

$$J_j(1)\chi_i(1) = \chi_i(1) \int_{\tau_2} \chi_j^*(2)\chi_j(2) \frac{1}{r_{12}} d\tau_2 \quad (\text{II.10})$$

$$K_j(1)\chi_i(1) = \chi_j(1) \int_{\tau_2} \chi_j^*(2)\chi_i(2) \frac{1}{r_{12}} d\tau_2 \quad (\text{II.11})$$

On obtient alors un nouveau système d'équations exprimées en fonction des orbitales moléculaires (OM) et non plus des spinorbitales. D'après la méthode LCAO (pour Linear Combination of Atomic Orbitals), chaque OM peut être développée par une combinaison linéaire d'OA $\chi_i = \sum_{p=1}^m c_{pi} \phi_p$ et les équations de Hartree-Fock initiales conduisent aux équations de Roothan[4] suivantes :

$$\sum_{p=1}^{n/2} F_{jp} c_{pi} = \epsilon_i \sum_{p=1}^{n/2} S_{jp} c_{pi} \quad (\text{II.12})$$

avec $S_{jp} = \langle \phi_j | \phi_p \rangle$ l'intégrale de recouvrement entre les deux OA.

Les coefficients sont optimisés lors de la procédure d'itération assurant l'auto-cohérence, qui porte sur l'opérateur de Fock et l'énergie.

Quant au formalisme UHF qui correspond notamment au traitement de systèmes à couches ouvertes, il constitue souvent la meilleure approximation. En effet, si la distance interatomique devient infinie (cas de la dissociation), la fonction d'onde du système tend vers le produit antisymétrisé des deux orbitales atomiques qui s'éloignent. En d'autres termes, elle ne tend pas vers la fonction biélectronique imposée par le formalisme précédent. Chaque orbitale n'est plus occupée que par un seul électron et la fonction polyélectronique devient un déterminant de Slater construit sur deux ensembles orthonormés de spinorbitales. Le premier est associé aux électrons α ($\chi_1^\alpha \chi_2^\alpha \dots \chi_t^\alpha$) et le second aux électrons β ($\chi_1^\beta \chi_2^\beta \dots \chi_u^\beta$) et $t > u$). L'énergie totale du système est telle que :

$$E = \sum_{i=1}^{t+u} H_{ii} + \frac{1}{2} \left[\sum_{i=1}^{t+u} \sum_{j=1}^{t+u} J_{ij} - \sum_{i=1}^t \sum_{j=1}^t K_{ij}^\alpha - \sum_{i=1}^u \sum_{j=1}^u K_{ij}^\beta \right] \quad (\text{II.13})$$

Après applications du principe variationnel et de la méthode LCAO, on obtient les équations de Pople et Nesbet[5, 6] :

$$\sum_{p=1}^{n_\alpha} F_{jp}^\alpha c_{pi}^\alpha = \epsilon_i^\alpha \sum_{p=1}^{n_\alpha} S_{jp} c_{pi}^\alpha \quad (\text{II.14})$$

$$\sum_{p=1}^{n_\beta} F_{jp}^\beta c_{pi}^\beta = \epsilon_i^\beta \sum_{p=1}^{n_\beta} S_{jp} c_{pi}^\beta \quad (\text{II.15})$$

Ces équations doivent être résolues en parallèle et de façon itérative car chaque opérateur de Fock dépend à la fois des coefficients relatifs à α et β . L'énergie électronique peut alors s'écrire :

$$E = \sum_{i=1}^{t+u} H_{ii} + \frac{1}{2} \left[\sum_{i=1}^t \epsilon_i^\alpha + \sum_{i=1}^u \epsilon_i^\beta \right] \quad (\text{II.16})$$

Le formalisme UHF fournit de meilleurs résultats que la méthode RHF. Mais l'approximation orbitale fait que les états de spin sont mal définis. Cela peut entraîner des défauts irréversibles dans la résolution du problème électronique. Globalement, l'énergie SCF totale obtenue sera d'autant plus proche de la

limite HF que la base d'OA utilisée sera étendue, cette limite n'étant atteinte que dans le cas d'une base complète. Les propriétés moléculaires sont alors bien décrites dans ces conditions. En revanche, il a été établi que les énergies de réaction, associées à des formations ou des ruptures de liaison, ne sont pas bien reproduites. Les écarts entre théorie et expérience sont imputables au modèle sur lequel est basée la méthode HF. C'est pourquoi d'autres méthodes prenant en compte la corrélation électronique ont été établies et seront décrites dans les paragraphes II.3 et II.4 de ce chapitre.

II.2 Méthodes semi-empiriques

Lors de la résolution des équations HF, l'étape la plus problématique est le calcul des intégrales biélectroniques intervenant dans la matrice de Fock F . En effet, si n est la dimension de la base, le nombre d'intégrales à calculer atteint n^4 , ce qui limite la taille des systèmes étudiés. La plupart du temps, seules les orbitales de la couche de valence sont prises en compte, les orbitales de cœur étant gelées. Les effets induits par celles-ci, type polarisation ou corrélation, sont alors inclus dans la paramétrisation. Selon la nature des approximations utilisées pour s'affranchir du coût calculatoire, on distingue deux variantes :

- les méthodes pour lesquelles l'hamiltonien est une somme d'opérateurs hamiltoniens monoélectroniques non explicités : on peut citer la méthode de Hückel étendu qui ne sera pas développée ici ;
- les méthodes dérivant des équations de Roothan. Ces dernières sont associées à l'approximation du recouvrement différentiel nul, qui consiste à négliger les recouvrements différentiels entre deux orbitales atomiques dès lors qu'elles sont différentes (méthode CNDO[7]) ou qu'elles sont portées par des atomes de nature différente (méthode NDDO[8]).

Autrement dit, la méthode CNDO (Complete Neglected of Differential Overlap) néglige toutes les intégrales biélectroniques qui dépendent du recouvrement et des densités de charge entre des atomes différents, ce qui en fait la méthode semi-empirique la plus simple. Une de ses limites réside dans le traitement de systèmes sous forme d'états triplets ou de radicaux libres, pour lesquels l'énergie d'échange est élevée, alors que la méthode CNDO implique la nullité des intégrales d'échange. C'est pour combler cette lacune qu'une autre méthode appelée INDO[9] (Intermediate Neglected of Differential Overlap) a été développée. Seules les intégrales monocentriques sont intégrées au calcul. Des versions modifiées de la méthode ont été ensuite proposées, telle que MINDO/1,2,3[10], et ont été paramétrées dans le but de permettre une évaluation fiable des énergies de formation des molécules. L'approche NDDO (Neglect of Diatomic Differential Overlap) est encore plus complète puisque seuls les recouvrements diatomiques différentiels sont négligés. Plus précisément, si on considère les couples d'orbitales pq et rs centrés respectivement sur un atome A et sur un atome B de la molécule étudiée, ne sont conservés que

les intégrales biélectroniques de la forme $(pp|qq)$, $(pq|pq)$, $(pp|rs)$. Suivant le mode d'évaluation empirique des termes non négligés, plusieurs techniques dérivant de cette méthode sont répertoriées. L'approche MNDO[11] (Modified Neglected of Diatomic Overlap) qui est adaptée au calcul de grandeurs telles que la géométrie moléculaire, les moments dipolaires et les constantes de force, traduit mal le phénomène de liaison par pont hydrogène et ne permet pas de traiter les métaux de transition ou les systèmes présentant une forte conjugaison. C'est pourquoi sa paramétrisation a été modifiée et de nouvelles méthodes sont apparues telles que AM1[12] ou PM6[13, 14]. Cette dernière méthode a été réécrite récemment sur la base d'un nouveau paramétrage plus précis, appliqué à tous les principaux éléments et aux métaux de transition et tiré de PM3 et PM5. PM6 sera appliquée dans ce travail à une étude géométrique.

II.3 Méthodes post-HF

On appelle énergie de corrélation la grandeur correspondant à la différence entre l'énergie exacte et l'énergie HF du système étudié. Cet écart provient de la structure de la fonction d'onde totale qui ne s'annule pas lorsque deux électrons de spin opposés sont situés au même point de l'espace. Par conséquent, les répulsions électroniques sont surestimées. La simple prise en compte de la corrélation existant entre le mouvement des divers électrons permet de s'affranchir de ce problème. Plusieurs classes de théories ont été proposées :

- * les théories de perturbation (MPn);
- * l'intégration de configuration (CI);
- * la théorie de l'espace actif (CAS,CASPT);
- * les clusters couplés (CC);
- * le champ multiconfigurationnel auto-cohérent (MCSCF);
- * la configuration d'intégration multi-référence (MRCI).

II.3.1 Théorie des perturbations

Il existe deux formes majeures de cette théorie. La première a été mise au point par Rayleigh et Schrödinger. On considère un opérateur hamiltonien non perturbé H_0 -dont on connaît les fonctions et les valeurs propres associées- auquel on ajoute une petite perturbation de la forme λV :

$$H = H_0 + \lambda V \tag{II.17}$$

On peut noter que si la perturbation est négligeable devant H_0 , les énergies et les fonctions d'onde totales ne doivent pas être différentes de celles résultant du problème non perturbé. Les solutions exactes de $H(\lambda)$ peuvent être développées en série de Taylor par rapport à λ :

$$\Psi(\lambda) = \Psi_i^{(0)} + \sum_{n=1}^{\infty} \lambda^n \Psi_i^{(n)} \quad (\text{II.18})$$

$$E_i(\lambda) = E_i^{(0)} + \sum_{n=1}^{\infty} \lambda^n E_i^{(n)} \quad (\text{II.19})$$

$E_i^{(n)}$ et $\Psi_i^{(n)}$ constituent les corrections d'ordre n respectivement sur l'énergie et sur la fonction d'onde. Dans le cas de petites perturbations, un nombre de termes relativement réduit peut suffire à obtenir une convergence des résultats.

La théorie de Moller-Plesset[15] repose sur la méthode de Rayleigh-Schrödinger qui est appliquée à l'opérateur hamiltonien de Hartree-Fock décrit précédemment avec une valeur de $\lambda = 1$. On exprime l'hamiltonien non perturbé H_0 comme une somme d'opérateurs monoélectroniques de Fock $F(i)$. Il vaut alors : $H_0 = \sum_i F(i)$. On en déduit la valeur du paramètre λV .

$$\lambda V = H_e - H_0 = \sum_{j>i} \frac{1}{r_{ij}} - \sum_{j=1}^n (2J_j - K_j) \quad (\text{II.20})$$

L'énergie de perturbation à l'ordre zéro $E_0^{(0)}$ correspond à la valeur propre de l'équation de Schrödinger non perturbée. L'énergie d'ordre 1 $E_0^{(1)}$ dont une expression est $\langle \Psi_{(0)} | V \Psi_{(0)} \rangle$ devient, après développement,

$$E_0^{(1)} = -\frac{1}{2} \sum_{a,b} (\langle \chi_a(1)\chi_b(2) | \frac{1}{r_{12}} \chi_a(1)\chi_b(2) \rangle - \langle \chi_a(1)\chi_b(2) | \frac{1}{r_{12}} \chi_b(1)\chi_a(2) \rangle) \quad (\text{II.21})$$

La somme $E_0^{(0)} + E_0^{(1)}$ est égale à l'énergie de Hartree-Fock : on en déduit que c'est la correction de l'énergie à l'ordre deux $E_0^{(2)}$ qui apporte la première correction à l'énergie HF. Elle est égale à :

$$E_0^{(2)} = \sum_n \frac{|\Psi_0^{(0)} | V \Psi_n^{(0)}|^2}{E_0^{(0)} - E_n^{(0)}} \quad (\text{II.22})$$

Le théorème de Brillouin impose que seules les doubles excitations interviennent. L'énergie globale représentée par la somme $E_{HF} + E_0^{(2)}$ est l'énergie MP2. Si l'on prend en compte les doubles corrections au deuxième ordre, ou bien les corrections simples, triples et quadruples, on utilise respectivement les méthodes MP3 et MP4. Une des limites de la méthode MPn réside dans le traitement des cas de quasi-dégénérescence, même à des ordres élevés. C'est pourquoi d'autres approches ont été envisagées pour pallier ce défaut.

II.3.2 Méthodes d'interaction de configurations CI

Dans le cadre de ces méthodes, la fonction d'onde décrivant le système est développée sous la forme d'une combinaison linéaire de déterminants de Slater. Ces derniers sont construits en excitant un ou

plusieurs électrons des spinorbitales occupées vers les spinorbitales virtuelles.

$$\Psi = \sum_k c_k \Psi_k \tag{II.23}$$

Ψ_k désignent les déterminants qui décrivent successivement l'état fondamental, un état monoexcité etc... L'application du principe variationnel permet de déterminer les coefficients c_k et par suite l'énergie du système. Si le nombre de déterminants k considéré est infini, on obtient la valeur exacte de l'énergie et on se place dans le cadre de la méthode d'interaction de configuration complète (notée FCI pour Full Configuration Interaction). Cependant, la diagonalisation complète est très coûteuse pour des systèmes électroniques complexes. On préférera alors limiter le nombre de déterminants : seules les excitations dont l'ordre sera inférieur à un certain seuil seront prises en compte. Ainsi, la méthode CI tronquée la plus utilisée est celle dans laquelle sont conservées les mono et double excitations (méthode SDCI). L'inclusion des triples (SDTCI) ou bien des quadruples (SDTQCI) excitations alourdit considérablement les calculs. Ces méthodes ne sont donc pas utilisées préférentiellement pour l'étude des systèmes de grande taille.

Il existe des variantes à la méthode CI. La méthode MCSCF (pour MultiConfiguration SCF) considère une fonction analogue à celle employée dans le cadre de la méthode CI. En plus de l'optimisation des coefficients c_k , les coefficients des orbitales atomiques intervenant dans l'écriture des OM sont eux aussi optimisés via le principe variationnel. On peut citer parmi cette famille de méthodes la méthode CASSCF (pour Complete Active Space SCF), pour laquelle les OM prises en compte dans l'écriture des déterminants de Slater sont divisées en deux catégories. D'une part, il y a les OM inactives, c'est-à-dire celles possédant deux électrons ou bien celles qui sont inoccupées pour n'importe quelle configuration et d'autre part, celles qui constituent l'espace actif, pour lesquelles une CI complète est effectuée. Enfin, la méthode MRCI (pour MultiReference CI) combine les deux méthodes décrites auparavant.

II.3.3 Méthode "Coupled Cluster"

Cette méthode est, au même titre que la théorie des perturbations, non variationnelle. Comme dans le cadre des méthodes CI, il s'agit de déterminer les coefficients c_k . Pour cela, on représente la fonction d'onde par un opérateur exponentiel e^T agissant sur une fonction d'onde de référence $|0\rangle$ (ici, on prend la fonction d'onde de Hartree-Fock). Le développement de Taylor de l'exponentielle associée à l'explicitation de T en tant que somme d'opérateurs d'excitation individuels T_i conduit à l'expression de la fonction d'onde Ψ_{cc} suivante :

$$|\Psi_{cc}\rangle = \{1 + (T_1 + T_2 + T_3 + \dots) + \frac{1}{2!}(T_1 + T_2 + T_3 + \dots)^2 + \dots\}|0\rangle \tag{II.24}$$

Le développement de T est souvent stoppée aux quadruples excitations car la contribution des ordres

supérieurs est faible. La classification des méthodes de cluster couplé dépend de la troncature appliquée à T . Ainsi, si seules les excitations sont intégrées à l'expression de T , on parle de méthodes CCS. Lorsqu'on y ajoute le terme T_2 relatif aux doubles excitations, on se place dans le cadre de la méthode CCSD. La méthode la plus communément utilisée est CCSD(T), ce qui signifie que les contributions des triples excitations (T) sont calculées via la théorie des perturbations MP4 et ajoutées aux simples et doubles excitations totalement prises en compte dans le calcul. En conclusion, les méthodes CC convergent plus rapidement que les méthodes CI vers le résultat FCI. C'est pourquoi le niveau CCSD(T) s'avère être une des techniques les plus efficaces pour le calcul de corrélation, malgré un certain coût calculatoire.

Lorsque les systèmes étudiés sont de très grande taille, une alternative à ces méthodes est la théorie de la fonctionnelle de la densité (DFT=Density Functional Theory). Son principe repose, à la différence des méthodes post-HF qui approchent la fonction d'onde, sur l'approximation de l'opérateur hamiltonien.

II.4 Théorie de la fonctionnelle de la densité

II.4.1 Bases de la théorie

Elle repose sur le calcul exact de la densité électronique associée à la fonction d'onde via la relation suivante :

$$\rho(r) = |\Psi(r)|^2 \quad (\text{II.25})$$

L'expression de l'énergie totale du système en tant que fonctionnelle de la densité a été envisagée la première fois par Thomas, Fermi et Wigner[16, 17, 18]. C'est en 1964 que Hohenberg et Kohn ont proposé une formulation exacte de ce modèle[19] sous la forme de deux théorèmes. Le premier démontre l'existence d'une fonctionnelle de la densité ρ , tandis que le second précise que le principe variationnel s'applique bien à cette grandeur.

Si le nombre d'électrons est connu, au même titre que la forme du potentiel d'attraction noyaux-électrons (ou potentiel externe V_{ext}), l'hamiltonien est déterminé. Il en est de même pour la fonction d'onde de l'état fondamental et la densité électronique qui lui est associée. C'est la réciproque qui a été démontrée par Hohenberg et Kohn. Le premier théorème de HK stipule ainsi que pour un état fondamental non dégénéré, la densité électronique détermine le potentiel externe. L'énergie du système s'exprime alors de la façon suivante, avec F une fonctionnelle de la densité :

$$E(\rho) = F(\rho) + \int \rho(r)V_{ext}(r)dr \quad (\text{II.26})$$

On peut noter que F étant universelle, ce théorème d'existence s'applique à tous les systèmes quantiques (solides, gaz d'électrons, molécules).

Le second théorème de HK indique que toute densité électronique approchée d'un état fondamental non dégénéré conduit à une borne supérieure de l'énergie exacte du système étudié. L'énergie de l'état fondamental pourra alors être obtenue en appliquant le principe variationnel à $\rho(r)$ par minimisation de l'énergie totale, de manière analogue à $\Psi(r)$ dans le cadre des méthodes ab initio, semi-empiriques notamment.

Globalement, l'énergie électronique prend la forme ci-dessous en fonction de la densité ρ :

$$E(\rho) = T(\rho) + V_{ne}(\rho) + V_{ee}(\rho) \quad (\text{II.27})$$

où $T(\rho)$ représente l'énergie cinétique des électrons, $V_{ne}(\rho)$ l'énergie d'attraction noyaux-électrons et $V_{ee}(\rho)$ l'énergie d'interaction électronique.

Le terme $V_{ne}(\rho)$ s'écrit en fonction de ρ :

$$V_{ne} = \sum_{A=1}^M \int \frac{Z_A \rho(r)}{|R_A - r|} dr \quad (\text{II.28})$$

En revanche, le terme $V_{ee}(\rho)$ pose problème, comme c'était déjà le cas dans les méthodes dites de fonction d'onde, rendant impossible la résolution directe de l'équation de Schrödinger électronique. Une approximation, permettant le calcul du minimum de la fonctionnelle énergie et par suite celui de la densité électronique et de l'énergie, a donc été proposée.

II.4.2 Equations de Kohn et Sham

Kohn et Sham[20] ont substitué le hamiltonien décrivant un système de n particules en interaction par un hamiltonien de référence renvoyant à n particules qui n'interagissent pas entre elles. La densité électronique de ce dernier système est considérée comme identique à celle du système réel. L'opérateur monoélectronique de Kohn et Sham prend la forme :

$$H_{KS} = -\frac{1}{2} \nabla^2 + \int \frac{\rho(r) - \rho(r')}{|r - r'|} dr' + V_{xc}(\rho) + V_{ext}(r) = -\frac{1}{2} \nabla^2 + V_{eff}(r) \quad (\text{II.29})$$

V_{xc} est appelé potentiel d'échange-corrélation et est défini par :

$$V_{xc}(\rho) = \frac{\partial E_{xc}(\rho(r))}{\partial \rho(r)} \quad (\text{II.30})$$

Afin de pouvoir minimiser l'énergie du système, on introduit les orbitales de Kohn et Sham $\psi(i)$, qui vont permettre de résoudre de manière itérative le jeu d'équations suivant :

$$H_{KS}\psi(i) = \epsilon_i\psi(i) \quad (\text{II.31})$$

La méthode de résolution est identique à celle utilisée pour les équations HF : on emploie le processus SCF. L'énergie totale du système est alors donnée par $E_{KS}^{tot}(\rho) = T_s(\rho) + J(\rho) + E_{xc}(\rho) + \int V_{ext}(r)\rho(r)dr$.

Les différents termes correspondent respectivement à l'énergie cinétique des électrons, sans interaction, pris comme système de référence, à l'énergie d'interaction coulombienne entre les électrons, à l'énergie d'échange et de corrélation et enfin, au terme d'interaction avec le potentiel externe.

D'après Hohenberg et Kohn, $E_{KS}^{tot}(\rho)$ doit être équivalente à l'énergie totale du système $E_{KS}^{r\acute{e}el(\rho)}$ qui est égale à :

$$E_{KS}^{r\acute{e}el(\rho)} = T_{r\acute{e}el}(\rho) + V_{ee}(\rho) + \int V_{ext}(r)\rho(r)dr \quad (\text{II.32})$$

On en déduit une expression de $E_{xc}(\rho)$ qui vaut $E_{xc}(\rho) = T_{r\acute{e}el} - T_s(\rho) + V_{ee}(\rho) - J(\rho)$.

La différence constituée par les deux premiers termes correspond à la correction de l'énergie cinétique apportée par l'interaction des électrons, tandis que celle relative aux deux derniers englobe les corrections non classiques à la répulsion électronique.

L'énergie du système polyélectronique est la somme de plusieurs termes s'exprimant comme des fonctionnelles de la densité ρ . Cependant, les expressions analytiques de certains d'entre eux, dont E_{xc} , sont inconnues. Différentes approximations sont alors possibles pour le déterminer.

II.4.3 Différents types de fonctionnelles

L'énergie d'échange-corrélation peut être décomposée en deux termes : $E_{xc}(\rho) = E_x(\rho) + E_c(\rho)$, dans lequel $E_x(\rho)$ constitue le terme d'échange pur et $E_c(\rho)$ l'énergie de corrélation pure. Les approximations faites sur chacun de ces deux termes simultanément ou indépendamment l'un de l'autre ont donné naissance à trois grandes familles de fonctionnelles, regroupées sous les appellations suivantes : LDA, GGA et hybride.

II.4.3.1 Approximation de la densité locale (LDA) Elle constitue l'approximation la plus grossière et est basée sur le modèle du gaz uniforme d'électrons, pour lequel $E_x(\rho)$ et $E_c(\rho)$ sont connus. Ainsi, l'énergie d'échange-corrélation, fonction de la seule densité électronique, s'exprime telle que :

$$E_{xc}^{LDA} = \int \rho(r)E_{xc}(r)dV \quad (\text{II.33})$$

avec $E_{xc}(r)$ pouvant être décomposé comme ci-dessus. $E_x(\rho)$ a été proposé exactement par Dirac[21] et $E_c(\rho)$ peut adopter notamment, dans la cadre du modèle du gaz d'électrons uniforme, la forme analytique présentée par Vosko, Wilk et Nusair en 1980[22] et notée généralement VWN.

Les inhomogénéités de la densité électronique n'étant pas prises en compte, certains problèmes peuvent être rencontrés, notamment pour les calculs d'énergie de liaison ou de distances de systèmes

possédant des électrons non appariés. Une alternative est de considérer l'approche UHF sans contrainte de spin, mais elle est trop aléatoire. Par conséquent, une approximation plus complexe a été faite dans le cadre du modèle GGA.

II.4.3.2 Approximation du gradient généralisé (GGA) La densité électronique n'étant que rarement uniforme dans un élément dV , une autre génération de fonctionnelles, introduisant l'hypothèse d'un gradient sur la densité, a été créée. La fonctionnelle dépend de ρ , ainsi que de $\nabla\rho$ et son expression devient :

$$E_{xc}^{GGA}(\rho) = A_x \int \rho^{\frac{4}{3}} F^{GGA} dr^3 \quad (\text{II.34})$$

La fonctionnelle d'échange-corrélation reste, malgré cette nouvelle dépendance, semblable à la fonctionnelle LDA. Mais il en existe de très nombreuses variantes. Comme potentiel d'échange, on peut citer celles de Perdew et Wang, PW86 et PW91[23, 24], ainsi que celle de Becke, B88, ajustée sur l'énergie d'échange des gaz rares[25]. Les principales fonctionnelles de corrélation associées sont, par ordre de popularité décroissante, celles de Lee, Yang et Par[26] et celle de Perdew et Wang tirée de VWN[27].

II.4.3.3 Fonctionnelles hybrides L'approximation hybride repose au départ sur la connexion adiabatique[28], qui consiste à exprimer l'énergie d'échange-corrélation en fonction d'un paramètre continu λ traduisant le couplage entre les électrons. $\lambda = 1$ indique qu'il y a interaction entre eux, tandis que $\lambda = 0$ signifie que les électrons n'interagissent pas. Ainsi, E_{xc} s'écrit :

$$E_{xc} = \int_0^1 E_{xc,\lambda} d\lambda = \int_0^1 (\langle \Psi_\lambda | V_{ee} \Psi_\lambda \rangle - J(\rho)) d\lambda \quad (\text{II.35})$$

Pour déterminer E_{xc} , on introduit une part d'échange calculable via la théorie de HF ; il correspond au cas $\lambda = 0$. Lorsque $\lambda = 1$, on retrouve une expression de la fonctionnelle assimilable à GGA. Ainsi, l'énergie d'échange-corrélation devient :

$$E_{xc}^{hyb} = aE_x^{HF} + (1-a)E_x^{DFT} \quad (\text{II.36})$$

Les paramètres fonctions de a sont des constantes à déterminer. Ces derniers donnent accès au rapport $\frac{HF}{DFT}$ qui permet d'engendrer les différentes familles de fonctionnelles hybrides. Parmi elles, la plus utilisée, notamment dans cette thèse, est celle proposée par Becke, B3LYP[29]. Trois fonctionnelles d'échange différentes (E_x^{LDA} , $E_x^{HFexact}$ et E_x^{B88}) interviennent dans l'expression de E_{xc} , au même titre que deux fonctionnelles de corrélation LYP et VWN :

$$E_{xc}^{B3LYP} = (1-a)E_x^{LDA} + aE_x^{HFexact} + b(E_x^{B88} - E_x^{LDA}) + E_c^{VWN} + c(E_c^{LYP} - E_c^{VWN}) \quad (\text{II.37})$$

a, b et c valent respectivement 0.2, 0.7 et 0.8. Ils ont été déterminés de façon empirique pour reproduire aux plus près des résultats expérimentaux. Une autre fonctionnelle qui sera utilisée lors de ce travail est PBE0[30], dont la fonctionnelle présente 25% d'échange HF dans une fonctionnelle GGA de type PBE. A la différence de B3LYP, la part d'échange HF n'a pas été choisie empiriquement, mais sur la base d'arguments de la théorie perturbative. Ces méthodes, au même titre que certains résultats expérimentaux, ont servi de référence pour l'élaboration d'autres méthodes beaucoup moins coûteuses en temps et très faciles à mettre en œuvre : les méthodes semi-empiriques. Cependant, elles sont tributaires des composés qui ont servi à l'étalonnage.

II.4.3.4 Autres types d'approximations On peut citer notamment l'approximation de type méta-GGA[31, 32, 33]. Afin de prendre mieux en compte l'inhomogénéité de la densité, la fonctionnelle dépend alors de $\rho(r)$, de son gradient $\nabla\rho(r)$, mais aussi du laplacien de la densité $\nabla^2\rho(r)$. Elle nécessite de plus l'introduction de la notion de densité d'énergie cinétique $\tau(r)$, qui dépend des orbitales occupées issues d'un calcul Kohn et Sham :

$$\tau(r) = \frac{1}{2} \sum_{i=1}^N |\nabla\Psi(i)|^2 \quad (\text{II.38})$$

L'énergie d'échange-corrélation s'exprime alors :

$$E_{xc}^{meta-GGA}(\rho) = \int f(\rho, \nabla\rho, \tau, \Delta\rho) dr \quad (\text{II.39})$$

D'autres méthodes plus récentes prennent en compte des effets plus fins, dont les interactions à longue portée (méthode CAMB3LYP[34]) ou bien les effets de dispersion (méthode DFT-D[35]). La caractéristique principale de l'approximation CAMB3LYP réside dans l'énergie d'échange, majoritaire, qui est divisée en deux parties : une partie est relative à la longue portée (LC), la seconde à la courte portée (CC). Globalement, l'optimisation des paramètres donne lieu à l'expression suivante :

$$E_{xc}^{CAMB3LYP} = (0.19E_x^{HF} + 0.81E_x^{B88})_{CC} + (0.65E_x^{HF} + 0.35E_x^{B88})_{LC} + 0.19E_c^{LYP} + 0.81E_c^{VWN} \quad (\text{II.40})$$

Les méthodes du type DFT-D traduisent mieux les interactions de Van der Waals, dont l'importance est non négligeable pour les larges systèmes tels que les systèmes hydratés. Les phénomènes de dispersion sont pris en compte via un terme correctif E_{disp} sur l'énergie DFT de Kohn et Sham qui adopte la forme suivante, suivant le formalisme proposé par Wu et Wang[36] :

$$E_{disp} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} - \dots \quad (\text{II.41})$$

Suivant le niveau de troncature utilisé, il existe différentes variantes : la troncature à C_6 constitue la DFT-D2, le niveau C_8 la DFT-D3 etc... Dans le cadre de la première citée, E_{disp} s'exprime ainsi :

$$E_{disp} = -s_6 \sum_{i,j>1}^{Nat} \frac{C_{ij}^6}{R_{ij}^6} f_{damp}(R_{ij}) \quad (\text{II.42})$$

s_6 est une constante qui dépend de la fonctionnelle à laquelle on a ajouté la correction dispersive. Ainsi, pour la méthode B3LYP-D, s_6^{B3LYP} vaut 1.05. La fonction f_{damp} permet d'activer la correction au-delà d'une distance caractérisée par la somme des deux rayons de Van der Waals de i et j .

II.5 Bases de fonctions atomiques

Il existe deux grands types de bases atomiques pour exprimer les OM :

- * les fonctions de Slater, plus communément appelées STO (pour Slater-Type Orbitals) ;
- * les fonctions gaussiennes, ou GTO (pour Gaussian-Type Orbitals).

II.5.1 Fonctions STO et GTO

Les fonctions de Slater adoptent la forme suivante (en coordonnées sphériques) :

$$\phi_{\zeta,n,l,m}(r, \theta, \varphi) = N r^{n-1} e^{-\zeta r} Y_{l,m}(\theta, \varphi) \quad (\text{II.43})$$

$Y_{l,m}$ représente la partie harmonique sphérique et N est un facteur de normalisation. Ces fonctions ont pour avantage de bien décrire l'occupation électronique au voisinage du noyau, ainsi qu'à grande distance : elles sont donc appropriées pour les calculs atomiques. Mais elles restent très peu utilisées pour l'étude de systèmes moléculaires car elles engendrent des calculs d'intégrales biélectroniques trop complexes.

En ce qui concernent les fonctions gaussiennes, elles s'expriment comme :

$$g_{\zeta,n,l,m}(r, \theta, \varphi) = N r^{2n-2-l} e^{-\zeta r^2} Y_{l,m}(\theta, \varphi) \quad (\text{II.44})$$

Elles possèdent une propriété remarquable : le produit de deux gaussiennes $g_A(\zeta_A)$ et $g_B(\zeta_B)$ centrées sur deux atomes différents A et B donnent une autre gaussienne $g_C(\zeta_C)$, située entre les atomes A et B, avec $\zeta_C = \zeta_A + \zeta_B$. Cela a pour conséquence d'autoriser un calcul plus simple des intégrales biélectroniques à quatre centres qui se ramènent à des calculs d'intégrales à deux centres. Cependant, leur comportement à courte et longue distances du noyau n'est pas bon (dérivée nulle et décroissance trop rapide respectivement). Globalement, la précision des résultats dépend du nombre de gaussiennes utilisées pour développer chaque orbitale.

Les deux descriptions présentées ci-dessus apparaissent complémentaires : c'est pourquoi Huzinaga[37] envisagea de remplacer une STO par une combinaison linéaire de N gaussiennes dites primitives (notées STO-NG, et qualifiées de bases minimales). Les exposants ζ sont ajustés pour permettre de reproduire les fonctions de type Slater. Ainsi, on obtient un comportement physique cohérent vis à vis du noyau et le calcul des intégrales est favorisé. Dans un second temps, des jeux de coefficients et d'exposants de gaussiennes décrivant des fonctions de Slater ont été proposés par Pople notamment[38, 39, 40]. D'autres types de bases sont alors apparues, respectant la structure du modèle en couche :

- * les bases double, triple ou quadruple zeta (notées DZP etc...) : chaque OA est représentée respectivement par 2, 3 ou 4 combinaisons de gaussiennes ;
- * les bases split-valence double zeta, triple zeta, quadruple zeta (notées n-n'n"G) avec n désignant le nombre de gaussiennes de la couche interne et n'n" stipulant le nombre de gaussiennes utilisée dans chaque couche de valence.

Plus ces bases sont grandes, plus le coût calculatoire est élevé. Pour qu'elles gagnent en flexibilité, c'est-à-dire pour qu'elles traduisent au mieux les distorsions dues à l'environnement, des fonctions de polarisation peuvent être ajoutées.

II.5.2 Fonctions de polarisation

Elles correspondent à des orbitales inoccupées dans l'atome à l'état fondamental. Elles prennent en compte les déformations des orbitales atomiques de valence lors de la formation de la molécule. Différentes valeurs des exposants de ces fonctions ont été proposées[41, 42, 43, 44]. La présence des fonctions de polarisation est généralement notifiée par les lettres "(d,p)", ou bien "*" dans la nomenclature de Pople : dans ce cas, les fonctions "d" concernent tous les atomes, sauf l'hydrogène sur lequel sont des fonctions de type "p". La notation "***" implique que tous les atomes, sans exception, sont concernés par l'extension de la base. Dunning a, quant à lui, mis au point des jeux de bases de type split-valence[45, 46] incluant des fonctions de polarisation dont les exposants et les coefficients ont été optimisés à partir de calculs post-HF. Ces bases sont notées cc-pVDZ (pour Correlation Consistent Polarized Valence Double Zeta), cc-pVTZ, cc-VQZ pour les plus communes.

Pour obtenir une bonne description de l'espace situé au-delà des orbitales de valence, il est nécessaire d'étoffer encore la base atomique utilisée. On fait intervenir des fonctions diffuses.

II.5.3 Fonctions diffuses

Elles possèdent des exposants ζ très faibles, ce qui leur permet de recouvrir l'espace diffus situé à grande distance du noyau. Elles sont nécessaires à l'étude des anions, de systèmes présentant des liaisons

hydrogène ou bien possédant de faibles énergies de liaison. Différentes méthodes de détermination de leurs exposants ont été proposées. On peut citer la minimisation de l'énergie des anions, ou encore l'extrapolation. De manière générale, l'exposant d'une diffuse est plus petit que le plus faible exposant d'une fonction de valence. Dans la nomenclature de Pople, une fonction diffuse est désignée par le symbole "+", tandis que les bases de Dunning incluent le préfixe "aug".

II.5.4 Erreur de superposition de base

L'évaluation de l'énergie exacte implique l'utilisation d'une base infinie de fonctions -en plus d'un calcul FCI-. Or dans la pratique, un nombre fini de fonctions est utilisé : cela provoque un abaissement artificiel de l'énergie. C'est ce qu'on nomme l'erreur de superposition de base (ou BSSE pour Basis Set Superposition Error). Plus concrètement, si on appelle x et y les fonctions de base servant à décrire les atomes des deux systèmes X et Y en interaction, le système global devient XY et est associé à la base xy . Si les premières citées se superposent, l'énergie de liaison E_l s'exprime telle que, la notation * faisant référence à la géométrie du complexe :

$$E_l = E(XY)_{xy}^* - E(X)_x - E(Y)_y \quad (\text{II.45})$$

Le déséquilibre provoqué par une description des nuages électroniques du système global et de chacun des deux fragments avec des bases x et y différentes induit l'erreur de superposition de base. Elle peut être corrigée par la méthode dite de contreponds (Counterpoise Correction)[47]. L'énergie corrigée du complexe devient :

$$E_l^{corr} = E_l - [E(X)_{xy}^* + E(Y)_{xy}^* - E(X)_x^* - E(Y)_y^*] \quad (\text{II.46})$$

La correction de superposition de base peut se faire de deux manières : a priori, en chaque point de la courbe d'énergie potentielle, ou a posteriori, c'est-à-dire uniquement au minimum de la courbe d'énergie potentielle. Enfin, sa quantification dépend de la méthode de calcul utilisée. Elle est faible dans le cadre des méthodes DFT, mais devient de plus en plus prononcée lorsqu'il s'agit de méthodes de type MP2 ou Coupled Cluster. Certains traitements ab initio permettent d'éliminer directement cette erreur à partir des équations de Roothan modifiées[48].

III Algorithmes de recherche de structure géométrique

La méthode de calcul de l'énergie du système considéré étant choisie, il est maintenant nécessaire de déterminer la géométrie optimale adoptée par cette structure à partir de laquelle le calcul de propriétés

sera mené. La géométrie optimale correspond au minimum global de la surface d'énergie potentielle, c'est-à-dire au point de la SEP pour lequel le gradient est nul et le hessien est positif dans toutes les directions. Différents types de méthodes peuvent être utilisées, suivant le mode d'approximation de la fonction employée : on parle alors de méthodes locales ou non locales.

III.1 Méthodes locales

La recherche d'un minimum local se fait dans ce cadre-là au moyen d'algorithmes regroupés en deux classes : les méthodes déterministes et probabilistes. Les premières citées sont sous-divisées en deux catégories, selon si la recherche fait intervenir ou non un calcul de gradient. La convergence sera alors atteinte plus ou moins rapidement.

III.1.1 Méthode du Simplex

Cet algorithme a été proposé par Dantzig en 1947. L'idée de départ consiste à entourer le minimum recherché par un simplexe, celui-ci étant un ensemble de $(N+1)$ points qui entoure le minimum (N étant le nombre de dimensions). La recherche du point optimal se fait via une méthode itérative par un nombre minimal d'expériences. Au départ, on force le simplexe à s'écartier du point initial grâce à une série de déplacements mettant en jeu des opérations de symétrie. Cette méthode étant, telle quelle, longue à mettre en œuvre et pouvant conduire à un faux minimum, Nelder et Mead y ont apporté des modifications[49]. La localisation devient plus rapide et plus précise car deux autres opérations ont été incluses dans la procédure de recherche : l'expansion et la contraction. La méthode du Simplex a l'avantage de ne pas nécessiter le calcul du gradient ou du hessien. En revanche, elle dépend énormément du choix du point de départ, et elle n'est pas très efficace si le minimum recherché se trouve sur un bord de la SEP ou bien si la dimension N devient trop élevée. C'est pourquoi d'autres méthodes, certes plus lourdes en matière de calculs mais plus précises, ont été mises au point. Elles reposent sur la détermination du gradient et sont qualifiées de méthodes à direction de descente.

III.1.2 Méthodes à direction de descente

Trois méthodes principales seront présentées ici : la méthode du gradient, la méthode du gradient conjugué ainsi que la méthode de Newton, plus complexe puisqu'elle nécessite aussi le calcul du hessien. C'est la détermination du gradient en un point donné qui fournit la direction de la descente.

III.1.2.1 Méthode du gradient Cette méthode est la plus simple afin d'accéder au minimum recherché au moyen d'un calcul de gradient. La distance entre un point x_k et le point suivant x_{k+1}

est calculé en fonction de la valeur du gradient g_k et d'un pas fixé au préalable λ (qualifié de pas de descente).

$$x_{k+1} = x_k - \lambda g_k \quad (\text{III.1})$$

Cette méthode a l'avantage d'être facile à mettre en œuvre, mais les conditions de convergence sont assez lourdes, d'où une certaine lenteur associée : plus le pas est grand, plus la convergence sera rapide, mais une valeur trop importante peut engendrer une divergence.

III.1.2.2 Méthode du gradient conjugué Au même titre que la méthode du gradient, cette méthode est itérative, mais elle utilise un algorithme pour l'optimisation de la direction de recherche. Celle-ci est déterminée à partir du gradient au point x_k , mais aussi en fonction de celui calculé au point précédent x_{k-1} . Pour la première itération, la direction choisie correspond à la valeur négative du gradient : $d_0 = -g_0$. Afin de diminuer le nombre d'itérations, le point x_{k+1} est choisi en déterminant le minimum de la fonction dans la direction précédente :

$$x_{k+1} = x_k + \alpha_k d_{k-1} \quad (\text{III.2})$$

avec $d_k = -g_k + p_k d_{k-1}$. Il faut noter que la direction de la descente d_k est désormais différente de $-g_k$, elle est corrigée par un terme associé à la direction précédente. Suivant la procédure utilisée pour déterminer le paramètre p_k , on distingue différentes variantes de la méthode du gradient conjugué, dont notamment :

- * la méthode de Fletcher-Reeves
- * la méthode de Polak-Ribière

Cette méthode est excellente si la fonction étudiée est quadratique. En effet si elle est en N dimensions, seules N étapes de calcul sont nécessaires. Sinon, son efficacité dépend de la valeur de N . Globalement, la convergence vers le minimum se fait très rapidement. L'évaluation du hessien peut être nécessaire pour déterminer précisément le pas de descente, mais ce n'est pas indispensable, contrairement à la méthode de Newton présentée ci-après.

III.1.2.3 Méthode de Newton La méthode de Newton n'est pas une méthode d'optimisation à proprement dite. Elle permet de résoudre des équations non linéaires de la forme $F(x) = 0$: dans le cas de la recherche d'un minimum local, cette équation devient : $gradF = 0$. Sa résolution permet l'obtention des points critiques de la fonction, qui ne sont pas forcément que des minima. Le point x_{k+1} est alors défini tel que :

$$x_{k+1} = x_k - h_k^{-1} g_k \tag{III.3}$$

où h_k est le hessien de la fonction au point x_k et $-h_k^{-1}g_k$ constitue la direction de descente. La calcul du hessien h_k est l'étape la plus coûteuse de la méthode : son approximation permettrait donc de gagner en temps de calcul. C'est pourquoi les méthodes de quasi-Newton ont été proposées. Les méthodes dites "de région de confiance" s'attachent à rendre le processus calculatoire peu dépendant du choix du point de départ. Par exemple, les méthode BFGS, DFP ou SR1 proposent de calculer h_{k+1} en fonction de h_k grâce à des formules algébriques. On peut aussi conserver h_k durant plusieurs itérations et l'actualiser périodiquement. Une autre méthode de quasi-Newton est celle de Levenberg Marquardt, plus complexe car couplée avec une règle de recherche linéaire.

En conclusion, l'efficacité des méthodes déterministes dépend de la complexité de la fonction (et donc de la surface) étudiée. Elles convergent vers un unique minimum, et ne peuvent en aucun cas renseigner de manière automatique sur tous les minima locaux présents. Compte tenu des avantages et inconvénients de chacune, une méthode du type gradient conjugué avec estimation du hessien semble constituer un bon compromis entre efficacité et temps de calcul. C'est d'ailleurs à cette famille qu'appartient l'algorithme de Berny[50] que nous utiliserons plus tard via le logiciel GAUSSIAN09[51]. La recherche du minimum global implique aussi l'utilisation d'autres méthodes. C'est pourquoi des méthodes de recherche de minimum non locales sont envisagées.

III.2 Méthodes non locales

III.2.1 Présentation générale

La détermination de toutes les structures adoptées par un système devient vite fastidieux en utilisant des méthodes locales telles que celles évoquées auparavant. De plus, certains minima locaux peuvent être omis en raison du choix aléatoire du point de départ, ce qui ne permet pas d'isoler de façon quasi-certaine le minimum global de la SEP. Les méthodes non locales autorisent ce type de recherche. On en distingue trois types majeurs.

1. Méthode de Monte Carlo

Elle consiste à explorer la surface d'énergie potentielle par des déplacements aléatoires des atomes. Sa formulation initiale a été proposée par Métropolis[52]. La simulation du système étudié est réalisée dans un bain thermique à une température T ; un processus de modification aléatoire des coordonnées atomiques est alors initié. Si à l'étape n l'itéré vaut $X_n = x$, on regarde la valeur de l'énergie E pour un point y voisin de x (choisi aléatoirement). Si $E(y) < E(x)$, x n'est pas adapté et on choisit $X_{n+1} = y$. Dans le cas contraire, $X_{n+1} = x$ apparaît plus judicieux. Mais il existe

le risque d'être déjà dans un minimum local $X_n = x$. Pour s'affranchir de ce problème, on pose $X_{n+1} = y$ si $E(y) - E(x) < 0$ ou $X_{n+1} = x$ dans le cas contraire etc...

2. Le recuit simulé

L'algorithme est basé sur le processus physique de refroidissement d'un liquide ou de cristallisation d'un métal. Il utilise la dynamique de Métropolis à des températures T_n décroissantes. La première étape consiste à appliquer cet algorithme à la température élevée T_0 . Dans un second temps, la température est abaissée de ΔT ($= T_n - T_{n+1}$) et l'algorithme de Métropolis est alors appliqué à cette nouvelle température. Cette étape est répétée jusqu'à ce que les déplacements d'atomes ne soient plus marqués, ce qui correspond souvent à une température très basse. Lors du refroidissement, le système devient de plus en plus ordonné, jusqu'à atteindre l'état de plus basse énergie, c'est-à-dire le minimum global ;

3. La dynamique moléculaire

La méthode de dynamique moléculaire la plus connue pour l'exploration de la surface d'énergie potentielle est la dynamique Born-Oppenheimer[53]. L'énergie du système est déterminée de manière quantique à chaque pas de temps de la dynamique. Les forces sont quant à elles calculées classiquement à partir du théorème de Hellmann-Feynman :

$$F_I(t) = \int \frac{\partial V(r, R(t))}{\partial R_I} \quad (\text{III.4})$$

avec V la fonction potentielle du système considéré. La résolution itérative des équations du mouvement de Newton, qui se fait notamment au moyen de l'algorithme de Verlet[54], provoque alors une nouvelle configuration du système notée $R(t, \partial t)$ caractérisées par la position, la vitesse et la force appliquée sur chaque atome i à $t + \partial t$:

$$r_i(t + \partial t) = r_i(t) + \partial t v_i(t) + \frac{\partial t^2}{2m_i} f_i(t) \quad (\text{III.5})$$

$$v_i(t + \partial t) = v_i(t) + \frac{\partial t}{2m_i} [f_i(t) + f_i(t + \partial t)] \quad (\text{III.6})$$

$$f_i(t + \partial t) = \frac{\partial V}{\partial r_i(t + \partial t)} \quad (\text{III.7})$$

Ce moyen d'exploration de la SEP dépend du choix du pas de la dynamique. Il implique aussi un coût calculatoire élevé, le calcul de l'énergie nécessitant une diagonalisation de matrice. Afin

de remédier à ce problème, Car-Parinello[55] ont développé une nouvelle approche basée sur la propagation de la fonction d'onde des noyaux. Elle sera présentée ultérieurement, au même titre que la dynamique Oppenheimer qui sera approfondie, dans le cadre du traitement vibrationnel. Le choix de la température est crucial lors d'une dynamique. En effet, si elle est trop importante, le système tendra vers un état dissociatif, tandis que l'abaisser de manière excessive peut ne pas permettre de franchir certaines barrières énergétiques.

Toutes ces méthodes non locales ont un défaut majeur : elles sont longues à mettre en œuvre et dépendent du choix de la structure initiale qui lorsqu'elle est éloignée du point recherché, allonge le temps de calcul. Les influences conjugués du pas et/ou de la température peuvent aussi avoir des conséquences sur le minimum obtenu, qui ne sera pas forcément global. Une nouvelle approche s'affranchissant de ces inconvénients a été mise au point : c'est l'approche GSAM[56].

III.2.2 Algorithme GSAM

Cet algorithme, créé au sein du laboratoire, est toujours en cours d'évolution et appliqué à des systèmes de plus en plus variés[57, 58, 59]. Il comprend trois étapes :

- * la génération
- * la sélection
- * le processus de ratissage et d'optimisation

III.2.2.1 La génération Il s'agit de générer un jeu de structures initiales. C'est le point crucial qui conditionne la fiabilité de la méthode. Deux approches complémentaires ont été mises au point : la génération aléatoire et la génération sphéroïdale, qui ne sera pas détaillée ici car elle ne sera pas utilisée par la suite. La génération aléatoire est basée sur la procédure d'Orland[60, 61]. On génère aléatoirement, atome par atome, un ensemble de configurations dans un volume prédéfini et obéissant à une distribution boltzmannienne. Parmi les structures générées, celles qui sont proches d'une forme symétrique sont symétrisées. Cependant, ce mode de génération ne permet pas d'obtenir des structures hautement symétriques telles que les sphères. La contrainte du volume permet d'obtenir des structures non dissociées, ayant un sens physique, à la différence des autres méthodes décrites précédemment. De plus, cette méthode ne nécessite pas d'hypothèse de départ.

III.2.2.2 La sélection Parmi les structures générées, les plus représentatives sont sélectionnées. Il faut donc notamment supprimer les doublons. Pour ce faire, on applique un mode de sélection a priori, et non a posteriori comme dans le cadre des méthodes non locales où le calcul de l'énergie est nécessaire avant d'établir un quelconque choix. Deux structures semblables devant conduire au même minimum de

la SEP, l'une des deux peut être écartée en se basant sur la métrique de Minkowski[62]. Si on note Q le vecteur propriété qui contient l'ensemble des distances atomiques, la différence topologique D_{ij} entre les structures i et j s'écrit :

$$D_{ij} = \sum_{\alpha} \frac{|Q_{i\alpha} - Q_{j\alpha}|^2}{\max_{kl} |Q_{k\alpha} - Q_{l\alpha}|^2}^{\frac{1}{2}} \quad (\text{III.8})$$

Si $D_{ij} > D_{seuil}$ (valeur seuil fixée au préalable), les deux structures i et j sont conservées. Dans le cas contraire, j est rejetée.

III.2.2.3 Le processus de ratissage On réalise m tranches de cycles d'optimisation successives de n_{opt} cycles. Puis on rejette à l'issue de ces derniers les structures possédant une énergie relative $E_{rel} > \Delta E_m$, où ΔE_m est réajustée au fur et à mesure que le nombre de cycles d'optimisation effectués augmente. De plus, si on appelle i et j les formes géométriquement et énergétiquement identiques à l'issue d'une tranche de cycles d'optimisation, la structure i sera éliminée selon le critère suivant : $\Delta E_{ij} < 0.0002 a.u.$ et $D_{ij} < 1\%$. Les m tranches d'optimisation étant terminées, on calcule les probabilités de Boltzmann à 298K -la température d'étude de non systèmes- : les structures possédant une probabilité non négligeable, c'est-à-dire supérieure à 5%, sont alors pleinement relaxées et des calculs vibrationnels seront réalisés à partir de celles-ci. Les paramètres considérés lors de l'application de GSAM sont : $m=5$; $\Delta E_m = 6, 4, 3, 2eV$ à l'issue respectivement des 1^e, 2^e, 3^e et 4^e tranches d'optimisation ; $n_{opt} = 200$.

IV Problème vibrationnel

IV.1 Traitement indépendant du temps : méthode statique

IV.1.1 Fonction potentielle - traitement harmonique

L'approximation de Born-Oppenheimer a permis d'établir que la fonction d'onde totale pouvait être séparée en deux facteurs : une fonction d'onde électronique et une fonction d'onde nucléaire. L'équation de Schrödinger nucléaire s'écrit :

$$H_n \Psi_n(R) = E_n \Psi_n \quad (\text{IV.1})$$

avec $H_n = T_n + V_{él}$. $V_{él}$ et E_n représentent respectivement l'énergie des noyaux et la fonction potentielle nucléaire. Les déplacements des noyaux par rapport à la position d'équilibre des noyaux (notée 0) étant faibles, la fonction potentielle V peut être développée en une série de Taylor de ces déplacements Δr :

$$V = V_0 + \sum_{i=1}^{3N} \left(\frac{\partial V}{\partial \Delta r_i} \right)_0 \Delta r_i + \frac{1}{2} \sum_{ij} \left(\frac{\partial^2 V}{\partial \Delta r_i \partial \Delta r_j} \right)_0 \Delta r_i \Delta r_j + \dots \quad (\text{IV.2})$$

Les Δr_i constituent les coordonnées de déplacements nucléaires, N étant le nombre d'atomes de la molécule considérée. La position d'équilibre étant la plus stable, le potentiel est minimum à la position d'équilibre $V_0 = 0$ et $\left(\frac{\partial V}{\partial \Delta r_i} \right)_0 = 0$. Le premier terme non-nul de l'équation (précédente) est donc quadratique :

$$V = \frac{1}{2} \sum_{ij} \left(\frac{\partial^2 V}{\partial \Delta r_i \partial \Delta r_j} \right)_0 \Delta r_i \Delta r_j + \dots \quad (\text{IV.3})$$

La détermination des constantes du développement permet alors de calculer la forme du potentiel au voisinage de la position d'équilibre. Il reste donc à expliciter l'énergie cinétique T_n et le potentiel V en fonction des déplacements et des vitesses dans un état électronique donné. Pour cela, on résout l'équation suivante : $(T_n + V)\Psi_n = E_n\Psi_n$. L'énergie cinétique des noyaux peut se décomposer en la somme de plusieurs termes, fonctions de deux référentiels distincts, un repère fixe et un repère mobile lié à la molécule, permettant de définir le mouvement global de celle-ci par rapport au trièdre fixe :

- * l'énergie de translation de la molécule dans le repère fixe ;
- * l'énergie cinétique des noyaux dans le repère mobile : c'est cette grandeur qui, associée au potentiel V , correspond à l'énergie cinétique de vibration ;
- * l'énergie de rotation ;
- * les interactions rotation–vibration, translation–vibration ainsi que translation–rotation.

Compte tenu des 6 relations d'Eckart, on peut établir que :

$$T_n = T_{vib} + T_{trans} + T_{rot} + T_{rovib} \quad (\text{IV.4})$$

Ainsi, dans le cas général, les vibrations sont caractérisées par $(3N-6)$ coordonnées de déplacement indépendantes, tandis que la rotation est caractérisée par 3 coordonnées définissant l'orientation du repère mobile par rapport au repère fixe. Les 3 coordonnées de translation définissent quant à elles la position du centre de gravité. En revanche, la rotation des molécules linéaires n'étant caractérisée que par deux angles, elles ne comptent que $(3N-5)$ coordonnées de vibration.

L'équation de Schrödinger nucléaire s'écrit alors d'après (IV.4) :

$$(T_{trans} + T)\Psi_{trans}\Psi_{rovib} = E\Psi_{trans}\Psi_{rovib} \quad (\text{IV.5})$$

avec $T = T_{vib} + T_{rot} + T_{rovib}$. L'absence d'interaction entre la translation et les autres mouvements moléculaires permet d'exprimer Ψ comme le produit $\Psi_{trans}\Psi$. De plus, T et V sont invariants par translation. On obtient :

$$(T + V)\Psi = E_{rovib}\Psi \quad (IV.6)$$

avec $E_{rovib} = E - E_{trans}$ et $T_{trans}\Psi_{trans} = E_{trans}\Psi_{trans}$.

Afin de simplifier encore le problème, on peut considérer que le couplage entre rotation et vibration est négligeable si la molécule étudiée est un rotateur. Cela impose que $T_n = T_{rot} + T_{vib}$ et $\Psi_{rovib} = \Psi_{rot}\Psi_{vib}$. L'énergie rotationnelle pour un niveau vibrationnel donné peut alors être déterminée en résolvant l'équation : $H_{rot}\Psi_{rot} = E_{rot}\Psi_{rot}$. H_{rot} est connu et prend la forme suivante :

$$H_{rot} = \frac{J_a^2}{2I_a} + \frac{J_b^2}{2I_b} + \frac{J_c^2}{2I_c} \quad (IV.7)$$

en notant a, b et c les axes d'inertie de la molécule, J_α les projections du moment angulaire total et I_α les moments d'inertie associés tels que $I_\alpha = \sum_i^N m_i r_i^{\alpha 2}$ avec m_i la masse de l'atome i et $r_i^{\alpha 2}$ sa distance de l'axe d'inertie correspondant.

La résolution de l'équation vibrationnelle peut se faire dans différents systèmes de coordonnées : ici, les résultats sont présentés en fonction des coordonnées cartésiennes de déplacement pondérées (s_i) telles que $s_i = \sqrt{m_i(x_i - (x_i)_0)}$ où les x_i renvoient aux coordonnées cartésiennes de chaque atome i.

On obtient notamment :

$$V = \frac{1}{2} \sum_{ij}^{3N} \left(\frac{\partial^2 V}{\partial s_i \partial s_j} \right)_0 s_i s_j + \dots \quad (IV.8)$$

Il existe une transformation des coordonnées cartésiennes pondérées permettant d'exprimer H_{vib} comme :

$$H_{vib} = T_{vib} + V = \frac{1}{2} \sum_{k=1}^{3N-6} (\dot{q}_k^2 + \lambda_k q_k^2) + \dots \quad (IV.9)$$

avec $q_k = \sum_{j=1}^{3N} L_{kj} s_j$, L_{kj} est une matrice de dimension $(3N-6) \times 3N$ et q_k constituent les coordonnées normales.

Par conséquent, l'équation de Schrödinger vibrationnelle s'écrit dans ce même système de coordonnées :

$$-\frac{\hbar^2}{2} \sum_{i=1}^{3N-6} \frac{\partial^2}{\partial q_i^2} \Psi_{vib} = V \Psi_{vib} + E_{vib} \Psi_{vib} \quad (IV.10)$$

L'énergie potentielle est déduite du calcul de la structure électronique effectuée grâce aux méthodes présentées dans la partie II.2. La fonction potentielle est telle que :

$$\begin{aligned}
 V = \frac{1}{2} \sum_{i=1} \frac{\partial^2 V}{\partial q_i^2} q_i^2 + \frac{1}{3!} \sum_{i=1} \sum_{j=1 \leq i} \sum_{k=1 \leq j} \frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k} q_i q_j q_k \\
 + \frac{1}{4!} \sum_{i=1} \sum_{j=1 \leq i} \sum_{k=1 \leq j} \sum_{l=1 \leq k} \frac{\partial^4 V}{\partial q_i \partial q_j \partial q_k \partial q_l} q_i q_j q_k q_l + \dots \quad (\text{IV.11})
 \end{aligned}$$

La troncature à l'ordre 2 constitue l'approximation harmonique du potentiel. La considération d'un ordre supérieur -généralement l'ordre 4- permet la prise en compte des effets anharmoniques. Les dérivées d'ordre 2, 3 et 4 sont respectivement les constantes de force quadratiques, cubiques et quartiques.

Dans le cadre harmonique, la résolution de l'équation (IV.10) ne pose pas de problème : il suffit de diagonaliser le hessien (dérivée seconde) de dimension $3N \times 3N$. Les $(3N-6)$ valeurs et vecteurs propres associés conduisent aux fréquences harmoniques ω_i et aux coordonnées normales q_i . Les fonctions d'onde vibrationnelles Ψ_{vib} se présentent sous la forme du produit des fonctions d'onde d'oscillateurs harmoniques ϕ_i :

$$\Psi_{vib} = \prod_i^{3N-6} \phi_i(q_i, \omega_i) \quad (\text{IV.12})$$

avec $\phi_i(q_i, \omega_i) = \frac{\sqrt{\omega}}{\sqrt{\pi}} H_n(q_i, \omega_i) \exp^{-\frac{\omega_i q_i^2}{2}}$. L'énergie de l'état vibrationnel s'exprime comme :

$$E_{vib}^i = \prod_i \left(\nu_i + \frac{1}{2} \right) \omega_i \quad (\text{IV.13})$$

Lorsque $\nu_i = 0$, E_{vib} constitue l'énergie du niveau fondamental : on la nomme ZPE (pour Zero Point Energy). L'énergie totale du système est alors telle que $E_{tot} = ZPE + E_{él}$.

La considération des termes d'ordre supérieur dans l'équation (IV.11) est plus complexe : elle nécessite des approximations. Les méthodes de traitement de l'anharmonicité sont analogues à celles utilisées pour l'étude de la structure électronique.

IV.1.2 Traitement anharmonique

Le potentiel dont l'expression est donnée ci-dessus permet de résoudre analytiquement l'équation de Schrödinger vibrationnelle qui repose ici sur le hamiltonien de Watson[63] :

$$\begin{aligned}
 H = hc \left[\frac{1}{2} \sum_i \omega_i p_i^2 + \frac{1}{2} \sum_i \omega_i q_i^2 + \sum_{i,j,k} k_{ijk} q_i q_j q_k \right. \\
 \left. + \sum_{ijkl} k_{ijkl} q_i q_j q_k q_l + \sum_{\alpha} B_{\alpha} P_{\alpha}^2 \right] \quad (\text{IV.14})
 \end{aligned}$$

Les termes entre parenthèses correspondent au hamiltonien d'ordre zéro, p_i est le moment conjugué de q_i . $\sum B_\alpha P_\alpha^2$ constitue le terme de Coriolis du second ordre[64].

Quatre grandes méthodes de résolution de cette équation seront présentées, par ordre croissant de complexité : VSCF, VPT2, VCI et VCC.

IV.1.2.1 VSCF L'approximation VSCF[65, 66] est basée sur une expression de la fonction d'onde en tant que produit de fonctions monodimensionnelles de type Hartree, N étant le nombre de modes de vibration :

$$\Psi_n(q_1, \dots, q_N) = \prod_{j=1}^N \chi_j^{(n)}(q_j) \quad (\text{IV.15})$$

Ce qui conduit aux équations SCF monomodes suivantes :

$$\left[-\frac{1}{2} \frac{\partial^2}{\partial q_j^2} + V_j^{(n)}(q_j) \right] \chi_j^{(n)}(q_j) = \epsilon_n \chi_j^{(n)}(q_j) \quad (\text{IV.16})$$

$V_j^{(n)}(q_j)$ est le potentiel effectif pour le mode q_j et s'exprime ainsi :

$$V_j^{(n)}(q_j) = \left\langle \prod_{i \neq j}^N \chi_i^{(n)}(q_i) \middle| V(q_1, \dots, q_N) \middle| \prod_{i \neq j}^N \chi_i^{(n)}(q_i) \right\rangle \quad (\text{IV.17})$$

Les équations précédentes sont résolues via la procédure SCF présentée auparavant. L'approximation VSCF fournit l'énergie totale suivante :

$$E_n = \sum_{j=1}^N \epsilon_j^{(n)} - (N-1) \left\langle \prod_{j=1}^N \chi_j^{(n)}(q_j) \middle| V(q_1, \dots, q_N) \middle| \prod_{j=1}^N \chi_j^{(n)}(q_j) \right\rangle \quad (\text{IV.18})$$

La principale difficulté réside dans l'équation (IV.17). L'évaluation du potentiel effectif $V_j^{(n)}(q_j)$ nécessite le calcul des intégrales multidimensionnelles. Plusieurs approches ont été proposées[67] :

- * développement de la fonction potentielle en termes de modes normaux ;
- * inclusion des interactions entre deux modes normaux dans le développement du potentiel ;
- * intégration des intégrales multidimensionnelles multidimensionnelles par la méthode Monte-Carlo.

Les solutions qui résultent de l'équation (IV.17) peuvent être corrigées par les effets de corrélation entre les modes de vibration à l'aide de la théorie des perturbations du second ordre (méthode CC-VSCF). Le potentiel du système est alors la somme de termes séparables (les termes monomodes) et des termes de couplage croisés[67] :

$$V(q_1, \dots, q_N) = \sum_j^N V_{eff}(q_j) + \sum_i^{N-1} \sum_{j>i}^N V_{ij}^{coup}(q_i, q_j) \quad (\text{IV.19})$$

Les équations VSCF peuvent être résolues grâce à des méthodes reposant sur une grille de points. Une forme analytique du potentiel n'est pas forcément nécessaire, ce qui constitue un avantage compte tenu de l'importance de sa qualité dans les résultats obtenus.

IV.1.2.2 VPT2 Les fréquences anharmoniques peuvent aussi être déterminées au moyen de la théorie des perturbations en termes de coordonnées normales. La construction des constantes de force anharmoniques et leur utilisation dans une évaluation du second ordre perturbatif des paramètres ro-vibrationnels a d'ailleurs été implémentée dans le logiciel GAUSSIAN09[68]. L'hamiltonien d'ordre zéro correspondant au hamiltonien de l'oscillateur harmonique, l'anharmonicité peut être considérée comme une perturbation. On a alors : $H_{VPT2} = H_0 + H_{anh}$. H_{anh} s'exprime comme :

$$H_{anh} = \sum_{ijk} \left(\frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k} \right)_0 q_i q_j q_k + \sum_{ijkl} \left(\frac{\partial^4 V}{\partial q_i \partial q_j \partial q_k \partial q_l} \right)_0 q_i q_j q_k q_l \quad (IV.20)$$

$$= \sum_{ijk} k_{ijk} q_i q_j q_k + \sum_{ijkl} k_{ijkl} q_i q_j q_k q_l \quad (IV.21)$$

Les dérivées d'ordre 3 et 4 sont évaluées par différenciation numérique des hessiens analytiques, pour de petits incréments de la géométrie d'équilibre le long d'une coordonnée normale. Cela nécessite le calcul de $6N-11$ hessiens en notant toujours N le nombre de modes de vibration de la molécule. Ainsi, les constantes cubiques et quartiques ont la forme suivante :

$$k_{ijk} = \frac{1}{3} \left(\frac{\omega_{jk}(\delta q_i) - \omega_{jk}(-\delta q_i)}{2\delta q_i} + \frac{\omega_{ki}(\delta q_k) - \omega_{ki}(-\delta q_j)}{2\delta q_j} + \frac{\omega_{ij}(\delta q_k) - \omega_{ij}(-\delta q_k)}{2\delta q_k} \right) \quad (IV.22)$$

$$k_{ijkk} = \frac{\omega_{ij}(\delta q_k) + \omega_{ij}(-\delta q_k) - 2\omega_{ij}(0)}{2\delta q_k^2} \quad (IV.23)$$

$$k_{iikk} = \frac{1}{2} \left(\frac{\omega_{ii}(\delta q_k) + \omega_{ii}(-\delta q_k) - 2\omega_{ii}(0)}{\delta q_k^2} + \frac{\omega_{kk}(\delta q_i) + \omega_{kk}(-\delta q_i) - 2\omega_{kk}(0)}{\delta q_i^2} \right) \quad (IV.24)$$

L'énergie vibrationnelle obtenue dans le cadre de la théorie perturbative du second ordre s'écrit :

$$E_n^{VPT2} = \chi_0 + \sum_i \omega_i \left(n_i + \frac{1}{2} \right) + \sum_i \sum_{j < i} \chi_{ij} \left(n_i + \frac{1}{2} \right) \left(n_j + \frac{1}{2} \right) = \dots \quad (IV.25)$$

avec ω les fréquences harmoniques issues du hamiltonien d'ordre 0 et χ désignent les contributions anharmoniques fonctions des constantes cubiques et quartiques k_{ijk} et k_{ijkl} [68, 69].

Cette méthode possède deux défauts majeurs : elle traite mal les résonances et reproduit mal les couplages, les surestimant souvent.

IV.1.2.3 VCI La contribution perturbative surestimant les valeurs anharmoniques, on préférera réaliser des calculs variationnels VCI[70, 71, 72, 73, 74][75, 76, 77, 78, 79], plus coûteux. Ils reposent sur la diagonalisation de la représentation du hamiltonien vibrationnel dans la base des fonctions propres du hamiltonien non perturbé H_0 . Elle contient les fonctions d'onde étudiées (sous-espace S_0) et leurs multiexcitations (sous-espace S_1).

Dans l'approche VSCF, l'expression de la fonction d'onde globale sous la forme d'un produit de fonctions de type Hartree n'autorise pas la prise en compte des interactions explicites intermodes. La méthode VCI permet d'inclure les excitations des orbitales occupées vers des orbitales virtuelles, qui correspondent aux niveaux inoccupés en VSCF. Elles sont générées par des opérateurs d'excitation τ_μ . L'espace des configurations global s'écrit alors : $|\Psi_i\rangle, \tau_\mu |\Psi_i\rangle$. L'indice μ fournit l'information nécessaire à la description de l'excitation (états de départ et d'arrivée). L'ensemble des excitations peut être désigné par des produits d'excitation mono-modes du type $a_{a^m}^{m+} a_i^m$ (en considérant une excitation d'un niveau i^m occupé vers un niveau a^m virtuel). La paramétrisation VCI de la fonction d'onde consiste en un développement linéaire adoptant la forme suivante pour les N modes du système :

$$|VCI\rangle = C_i |\Psi_i\rangle + \sum_{\mu} C_{\mu} \tau_{\mu} |\Psi_i\rangle \quad (\text{IV.26})$$

Les différentes valeurs des paramètres sont déterminées à partir de critères variationnels aboutissant aux équations aux valeurs propres usuelles CI.

La prise en compte de toutes les excitations μ permet d'obtenir la fonction d'onde FVCI (pour Full-VCI). Elle constitue la fonction d'onde exacte pour un hamiltonien donné et la différence entre les énergies VSCF et FVCI est l'énergie de corrélation. Ainsi, plus la dimension de la matrice à diagonaliser est grande, plus la solution exacte est approchée et plus le nombre de couplages considérés entre états vibrationnels est grand. Sachant qu'on ne peut pas diagonaliser une matrice de taille infinie, une approximation pour l'étude de gros systèmes a été mise au point : on limite le nombre de produits de type Hartree. L'approche VCI tronquée permet d'inclure seulement une part de l'espace S_1 , et une partie de l'énergie de corrélation. Le moyen pour introduire cette approximation sur la fonction FVCI est de tronquer directement le développement correspondant à un certain niveau d'excitation, souvent, quadruple. Deux autres options sont envisageables pour rendre le calcul accessible : on peut considérer soit un critère énergétique (toutes les configurations existantes comprises entre des valeurs d'énergie choisies au préalable sont générées), soit un critère lié à la fonction potentielle (le niveau de troncature de celle-ci impose le degré d'excitation possible).

Malgré cela, le défaut majeur de cette approche réside dans le temps de calcul très élevé. En effet, la vitesse de la diagonalisation de la matrice représentative du hamiltonien évolue comme N^3 . Pour tirer

le meilleur parti des avantages des méthodes VCI et VPT2, des méthodes dites mixtes de "variation-perturbation" ont été mises au point et permettent de limiter substantiellement la taille de la matrice à diagonaliser sans perte notable de précision. C'est notamment le cas de la méthode VCI-P.

IV.1.2.4 VCI-P La méthode VCI-P[80, 81] consiste à calculer état par état l'énergie des niveaux vibrationnels anharmoniques par une méthode basée sur le principe variationnel. Cela nécessite la diagonalisation de $3N - 5$ matrices CI (où N est le nombre d'atomes), chacune d'entre elles étant construite avec les interactions les plus fortes qui décrivent l'état considéré. La contributions des configurations les moins pertinentes, c'est-à-dire qui décrivent très faiblement l'état, sont évalués par une approche perturbative au second ordre, comme le montre la figure 2.1 ci-dessous.

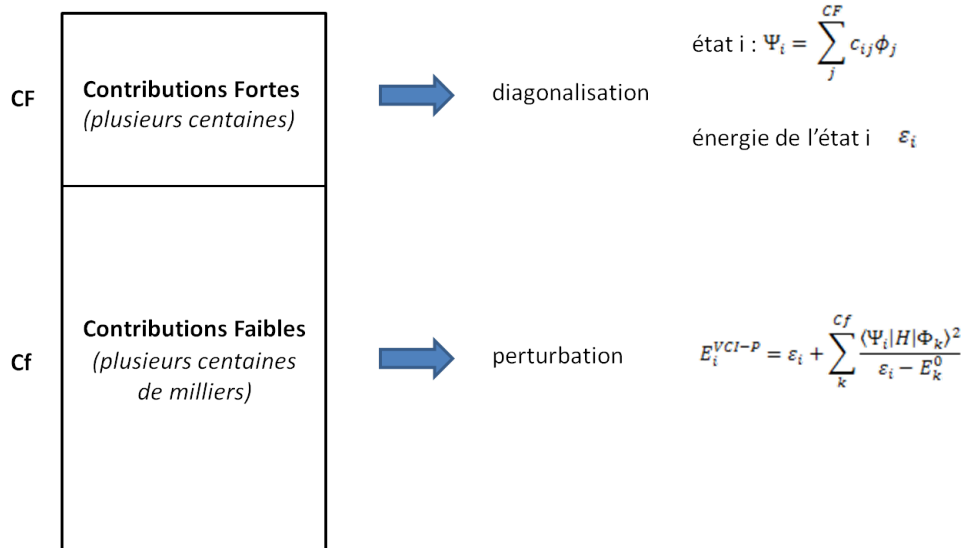


FIGURE 2.1 – Principe de la méthode VCI-P

IV.1.2.5 VCC Il a déjà été noté dans le cadre des présentations des méthodes de calcul de la structure électronique que l'approche "Coupled Cluster" était la plus précise, mais aussi la plus gourmande en temps de calcul. La méthode CC vibrationnelle (VCC)[82] repose sur l'équation suivante :

$$|VCC\rangle = e^T |\Psi_i\rangle \tag{IV.27}$$

où $|\Phi_i\rangle$ est la fonction d'onde de référence (dans l'état VSCF) et T l'opérateur de clusterisation, identique à celui présenté précédemment. Il peut être décomposé sous la forme d'opérateurs décrivant

tous les excitations du système étudié :

$$T = \sum_{\mu} t_{\mu} T_{\mu} \quad (\text{IV.28})$$

avec t_{μ} désignant l'amplitude et T_{μ} l'opérateur de l'excitation représentée par l'indice μ .

Les équations (IV.27 et IV.28) conduisent à une expression de l'équation de Schrödinger vibrationnelle suivante :

$$e^{-T} H e^T |\Psi_i\rangle = E_{VCC} |\Psi_i\rangle \quad (\text{IV.29})$$

La projection de chaque excitation sur l'état de référence VSCF fournit un jeu d'équations non linéaires :

$$0 = \langle \mu | e^{-T} H e^T |\Psi_i\rangle \quad (\text{IV.30})$$

La projection sur l'état VSCF conduit à une valeur de l'énergie E_{VCC} de :

$$E_{VCC} = \langle \Psi_i | e^{-T} H e^T |\Psi_i\rangle \quad (\text{IV.31})$$

Cette dernière ne peut être évaluée qu'une fois que les équations (IV.30) auront été résolues. Au même titre que pour l'approche VCI, une troncature de T s'impose à la 4e excitation.

Cette méthode possède enfin l'avantage d'être extensive par la taille : elle rend bien compte des évolutions de l'énergie, de la fonction d'onde ainsi que des propriétés en fonction de la taille du système.

Toutes ces méthodes nécessitent une forme analytique du potentiel, dont la justesse conditionne la qualité des résultats vibrationnels obtenus.

Pour s'affranchir de cet obstacle, on peut employer des méthodes de dynamique moléculaire. Elles consistent à simuler une trajectoire durant un temps de simulation défini à une température donnée. Le système étudié est alors libre de se mouvoir en fonction des forces qui l'animent. A partir de la trajectoire établie, une signature vibrationnelle qui rend compte de l'ensemble des conformations visitées peut être déterminée.

IV.2 Traitement dépendant du temps : méthode dynamique (DM)

Les observations expérimentales paraissent plus faciles à reproduire dans le cadre de la dynamique moléculaire que dans le contexte statique en raison de trois facteurs : l'importance de la température ainsi que de l'effet entropique, et la diversité des temps propres des techniques de spectroscopie. Ces effets peuvent être introduits de différentes manières, notamment via des méthodes de dynamique ab

initio, pour lesquelles la dynamique des noyaux est traitée classiquement et les noyaux sont considérés de façon quantique.

IV.2.1 Trajectoires

IV.2.1.1 Cadre général Le point de départ de la DM repose sur l'équation de Schrödinger dépendante du temps :

$$i\hbar \frac{\partial}{\partial t} \Psi(r_i, R_I, t) = H \Psi(r_i, R_I, t) \quad (\text{IV.32})$$

avec H le hamiltonien présenté auparavant.

On introduit alors une approximation sur la fonction d'onde, qui consiste à séparer les contributions électroniques Φ et nucléaires χ . Celle-ci est différente de l'hypothèse de Born et Oppenheimer proposée dans le cas statique, puisqu'un facteur de phase noté E_e intervient :

$$E_e = \int \Phi^*(r_i, t) \chi^*(R_I, t) H_e \Phi(r_i, t) \chi(R_I, t) dr dR \quad (\text{IV.33})$$

On aboutit aux équations suivantes :

$$i\hbar \frac{\partial \Phi}{\partial t} = - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 \Phi + \left[\int \chi^*(R_I, t) V_{ne} \chi(R_I, t) dR \right] \Phi \quad (\text{IV.34})$$

$$i\hbar \frac{\partial \chi}{\partial t} = - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 \chi + \left[\int \Phi^*(r_i, t) H_e \Phi(r_i, t) dr \right] \chi \quad (\text{IV.35})$$

Les électrons et les noyaux traités de manière quantique évoluent ainsi dans un potentiel dépendant du temps.

La seconde étape consiste à considérer les noyaux comme des particules ponctuelles. La fonction d'onde nucléaire χ adopte la forme suivante :

$$\chi(R_I, t) = A(R_I, t) \exp\left(\frac{i}{\hbar} S(R_I, t)\right) \quad (\text{IV.36})$$

où A est le facteur d'amplitude et S le facteur de phase qui sont des réels[83]. L'équation (IV.34) donne alors lieu à deux égalités, fonctions respectivement de A et de S :

$$\frac{\partial A}{\partial t} + \sum_I \frac{1}{M_I} (\nabla_I A) (\nabla_I S) + \frac{A}{2} \nabla_i^2 S = 0 \quad (\text{IV.37})$$

$$\frac{\partial S}{\partial t} + \sum_I \frac{1}{2M_I} (\nabla_I S)^2 + \int \Phi^* H_e \Phi dr = \hbar^2 \sum_I \frac{1}{2M_I} \frac{\nabla_i^2 A}{A} \quad (\text{IV.38})$$

C'est la seconde citée qui permet d'accéder à l'équation classique du mouvement décrite ci-dessous, avec $P_I = \nabla_I S$:

$$\frac{dP_I}{dt} = -\nabla_I \int \Phi^* H_e \Phi dr \quad (\text{IV.39})$$

soit

$$M_I \ddot{R}_I = -\nabla_I \int \Phi^* H_e \Phi dr = -\nabla_I V_e(R_I) \quad (\text{IV.40})$$

On constate que les noyaux évoluent classiquement dans le potentiel V_e créé par les électrons, qui dépend de la position instantanée des noyaux. En ce qui concerne l'équation (IV.32) qui fait intervenir la fonction d'onde nucléaire au même titre que son homologue électronique, on impose une limite classique qui donne lieu à l'équation d'évolution électronique suivante :

$$i\hbar \frac{\partial \Phi}{\partial t} = H_e(r_i, R_I) \Phi(r_i, R_I) \quad (\text{IV.41})$$

D'après cette dernière, les électrons évoluent de façon auto-cohérente avec les noyaux. Le hamiltonien et la fonction d'onde électronique dépendent des positions instantanées des noyaux de part la forme du potentiel d'interaction électrons/noyaux V_{ne} . A partir de ce constat, plusieurs méthodes de résolution du problème dynamique sont envisageables.

IV.2.1.2 Dynamique d'Ehrenfest Les équations (IV.40) et (IV.41) peuvent être résolues simultanément en considérant la fonction d'onde électronique Φ comme une combinaison linéaire de plusieurs états Φ_k :

$$\Phi(r_i, R_I, t) = \sum_k^{\infty} c_k(t) \Phi_k(r_i, R_I) \quad (\text{IV.42})$$

Il est commun d'adopter comme formes de Φ_k la base issue de la résolution de l'équation de Schrödinger indépendante du temps $H_e \Phi_k = E_k \Phi_k$. De plus, la restriction à un seul état électronique dans la forme développée de Φ permet d'obtenir les équations d'évolution ci-dessous :

$$M_I \ddot{R}_I(t) = -\nabla_I \langle \Phi_0 | H_e | \Phi_0 \rangle \quad (\text{IV.43})$$

$$i\hbar \frac{\partial \Phi_0}{\partial t} = H_e \Phi_0 \quad (\text{IV.44})$$

Le dynamique d'Ehrenfest[84] est la première approche dite "en vol". Cependant, elle n'est presque pas utilisée car l'échelle et le pas de temps dictés par la dynamique des électrons engendrent des coûts calculatoires trop élevés. D'autres approches ont été proposées.

IV.2.1.3 Dynamique moléculaire "Born-Oppenheimer" (BOMD) Comme cela a déjà été évoqué dans la partie III, cette méthode[85, 86] consiste à résoudre dans le cadre statique le problème de la structure électronique à chaque pas de la simulation pour un jeu de coordonnées nucléaires correspondant à leurs positions instantanées. Autrement dit, les noyaux suivent une trajectoire classique et la structure électronique est déduite de la résolution de l'équation de Schrödinger indépendante du temps. Les équations permettant de traduire ces hypothèses prennent la forme suivante, pour l'état fondamental Φ_0 :

$$M_I \ddot{R}_I = -\nabla_I \min \langle \Phi_0 | H_e | \Phi_0 \rangle \quad (\text{IV.45})$$

$$H_e \Phi_0 = E_0 \Phi_0 \quad (\text{IV.46})$$

Dans le cadre de l'approximation Hartree-Fock, la minimisation contrainte de l'énergie par rapport aux orbitales, soit $\min \langle \Phi_0 | H_e^{HF} | \Phi_0 \rangle_{\langle \psi_i | \psi_j \rangle = \delta_{ij}}$, équivaut, dans le formalisme de Lagrange, à $L = -\langle \Phi_0 | H_e^{HF} | \Phi_0 \rangle + \sum_{ij} \Lambda_{ij} (\langle \psi_i | \psi_j \rangle - \delta_{ij})$ si $\Phi_0 = \det(\psi_i)$. Λ_{ij} représentent les multiplicateurs de Lagrange associés aux contraintes d'orthonormalisation. On a alors : $\frac{\delta L}{\delta \psi_i^*} = 0$, ce qui conduit aux équations HF usuelles $H_e^{HF} \psi_i = \sum_j \Lambda_{ij} \psi_j$. Les équations (IV.45) et (IV.46) peuvent s'écrire :

$$M_I \ddot{R}_I = -\nabla_I \min \langle \Phi_0 | H_e^{HF} | \Phi_0 \rangle_{\langle \psi_i | \psi_j \rangle = \delta_{ij}} \quad (\text{IV.47})$$

$$H_e^{HF} \psi_i = \epsilon_i \psi_i \quad (\text{IV.48})$$

Ce jeu d'égalités peut aussi être obtenu dans le cadre de la DFT où H_e^{HF} sera remplacé par H_e^{KS} .

On notera que l'énergie électronique doit être minimisée à chaque pas de la simulation, ce qui constitue une différence majeure avec la DM d'Ehrenfest. En effet, la minimisation de l'énergie par la fonction d'onde de départ suffit à sa minimisation naturelle par la fonction d'onde propagée. Une autre différence remarquable est que la BOMD ne nécessite pas d'évolution temporelle explicite de la fonction d'onde. Quant au pas de temps, il doit être important car l'échelle repose sur les noyaux. Ces trois inconvénients, surtout le premier cité, font que cette méthode de simulation est très gourmande en temps de calcul. C'est pourquoi d'autres approches ont été suggérées, dites DM de "Lagrange Etendu", pour lesquelles on exploite la différence entre les échelles de temps relatifs aux noyaux (mouvement lent) et aux électrons (mouvement rapide).

IV.2.1.4 Dynamique moléculaire "Lagrange Etendu" (ELMD) Suivant l'expression du Lagrangien considéré, deux principales approches sont utilisées : la DM dite de Car-Parinello (CPMD)[55]

ou bien celle reposant sur le formalisme "Atom-centered Density Matrix Propagation" (ADMP)[87, 88, 89].

IV.2.1.4.1 CPMD On considère le Lagrangien suivant :

$$L_{CP} = \frac{1}{2} \sum_I M_I \dot{R}_I^2 - \langle \Phi_0 | H_e | \Phi_0 \rangle + \frac{1}{2} \sum_i \mu_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle + \sum_{ij} (\langle \psi_i | \psi_j \rangle - \delta_{ij}) \quad (IV.49)$$

Le premier terme fait référence à l'énergie cinétique des noyaux, le second à l'énergie potentielle électronique, le troisième à l'énergie cinétique des OM et le dernier pose la condition d'orthonormalisation de celles-ci.

Les équations d'Euler-Lagrange qui s'écrivent de manière générale :

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{R}_I} = \frac{\partial L}{\partial R_I} \quad (IV.50)$$

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\psi}_I^*} = \frac{\partial L}{\partial \psi_I^*} \quad (IV.51)$$

permettent d'obtenir les équations du mouvement suivantes :

$$M_I \frac{\partial^2 R_I(t)}{\partial t^2} = - \frac{\partial}{\partial R_I} \langle \Phi_0 | H_e | \Phi_0 \rangle + \frac{\partial}{\partial R_I} \sum_{ij} \Lambda_{ij} (\langle \psi_i | \psi_j \rangle - \delta_{ij}) \quad (IV.52)$$

$$\mu_i \frac{\partial^2 \psi_i(t)}{\partial t^2} = - H_e \psi_i + \sum_j \Lambda_{ij} \psi_j \quad (IV.53)$$

On peut ainsi dire que les noyaux évoluent au cours du temps à une certaine température proportionnelle au terme $\sum_I M_I \dot{R}_I^2$, tandis que les électrons seraient soumis à une température fictive associée au terme $\mu \sum_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle$.

La CPMD propose une évolution temporelle explicite de la fonction d'onde et ne nécessite pas de minimiser le problème électronique à chaque pas puisque la fonction d'onde est propagée au cours du temps. Le pas de temps est raisonnablement grand, son choix étant dicté par la valeur de la masse μ_i . Généralement, l'approche de Car-Parinello emploie les orbitales de Kohn et Sham développées sur une base d'ondes planes.

IV.2.1.4.2 ADMP Une des étapes les plus lourdes de la CPMD réside dans le calcul itératif du terme fonction de Λ_{ij} . Pour s'affranchir de cet inconvénient, une expression du Lagrangien sans cette contrainte a été proposée. Elle prend la forme suivante, sous forme matricielle :

$$L_{ADMP} = \frac{1}{2} Tr(V^T M V) + \frac{1}{2} \mu Tr(\dot{P} \dot{P}) - E(R, P) - Tr(\Lambda(P P - P)) \quad (IV.54)$$

où R , V et M renvoient respectivement aux positions nucléaires, aux vitesses et aux masses, P représentant la matrice densité. μ constitue toujours la masse fictive associée aux degrés de liberté électroniques. Elle permet de compenser l'erreur implicite due à la propagation de la structure électronique par rapport à sa convergence totale. Les équations du mouvement peuvent se développer ainsi :

$$M \frac{d^2 R}{dt^2} = - \frac{\partial E}{\partial R} |_P \quad (\text{IV.55})$$

$$\frac{d^2 P}{dt^2} = -\mu^{-\frac{1}{2}} \left(\frac{\partial E}{\partial P} |_R + \Lambda P + P \Lambda - \Lambda \right) \mu^{-\frac{1}{2}} \quad (\text{IV.56})$$

Dans le cadre de cette méthode, la structure électronique, représentée par la matrice densité, et les noyaux traités classiquement sont propagées simultanément en ajustant les échelles de temps relatives aux noyaux et aux électrons.

L'efficacité et le justesse de ce modèle dépend du choix de la masse fictive. De plus, il a été noté que l'écart entre les trajectoires ADMP et BOMD était directement proportionnel à cette grandeur[87]. Son choix doit être tel que les oscillations de la densité soient d'un ordre de grandeur plus important que les mouvements de plus haute fréquence affichés par les noyaux. Plusieurs avantages de cette méthode sont répertoriés :

- * la liberté de traiter rigoureusement tous les électrons du système ;
- * l'utilisation de pas de temps raisonnablement importants au travers de petites valeurs de masses fictives ;
- * le large choix de fonctionnelles d'échange-corrélation (incluant la DFT hybride) ;
- * la possibilité de traiter des systèmes chargés ;
- * le contrôle en "temps de vol" de la déviation par rapport à la surface Born-Oppenheimer.

Cette approche est implémentée dans GAUSSIAN09 et sera utilisée pour la détermination des spectres vibrationnels de nos systèmes d'étude. Leur obtention à partir de la trajectoire issue des calculs ci-dessus est détaillée dans la partie suivante.

IV.2.2 Transformées de Fourier de données temporelles

Le calcul des spectres vibrationnels à partir d'une trajectoire de DM est basé sur la théorie de la réponse linéaire. La fonction d'autocorrélation $C_A(t)$ d'une grandeur A permet d'exprimer la corrélation entre les valeurs $A(t)$ et $A(t_0)$, t_0 désignant l'instant initial. Elle s'exprime telle que : $C_A(t) = \langle A(t_0).A(t_0 + t) \rangle$. Elle constitue une valeur moyenne sur l'ensemble statistique. Pour améliorer les résultats, plusieurs origines de temps peuvent être considérées. $C_A(t)$ peut alors s'écrire :

$$C_A(t) = \frac{1}{\tau_{max}} \frac{1}{N} \sum_{j=1}^{\tau_{max}} \sum_{i=1}^N A_i(\tau_j) \cdot A_i(\tau_j + t) \quad (IV.57)$$

où τ_{max} représente le nombre d'origines temporelles τ_j choisies et N le nombre d'entités. Suivant la forme de A_i , différentes données spectrales peuvent être obtenues.

IV.2.2.1 A_i =vitesses atomiques pondérées Les vitesses atomiques pondérées \hat{v}_i sont égales à $\hat{v}_i = \sqrt{m_i} v_i$ avec m_i et v_i respectivement la masse et la vitesse cartésienne de l'atome i . La transformée de Fourier de la fonction d'autocorrélation $C_v(t)$ avec $C_v(t)$ déduite de l'équation (IV.57) en prenant A_i comme \hat{v}_i permet d'obtenir la densité d'états vibrationnels du système étudié, appelés spectres DOS (pour Density Of States)[90, 91] :

$$DOS(\omega) = \frac{1}{B} \int e^{i\omega t} C_v(t) dt \quad (IV.58)$$

avec B un facteur de normalisation.

IV.2.2.2 A_i =moment dipolaire M La transformée de Fourier de la fonction d'autocorrélation des moments dipolaires permet de déterminer le spectre IR. Le coefficient d'absorption $I(\omega)$ de la radiation IR s'écrit[92, 93, 94] :

$$I(\omega) = \frac{4\pi^2\omega^2}{3cnk_B T} \int e^{i\omega t} \langle M_0 \cdot M_t \rangle dt \quad (IV.59)$$

avec c la vitesse de la lumière et n l'indice de réfraction du milieu. Le moment dipolaire est défini à partir des positions r_i et des charges q_i des particules : $M_t = \sum_{i=1}^N q_i r_i(t)$. Sa considération directe pour l'établissement du spectre peut poser des problèmes, notamment en cas de conditions périodiques, pour lesquelles M est discontinu. On préférera alors calculer le spectre IR à partir des dérivées du moment dipolaire $\dot{M}_t = \sum_{i=1}^N q_i v_i(t)$:

$$I(\omega) = \frac{4\pi^2\omega^2}{3cnk_B T} \int e^{i\omega t} \langle \dot{M}_0 \cdot \dot{M}_t \rangle dt \quad (IV.60)$$

IV.2.2.3 Attribution des bandes spectrales L'utilisation de cette approche fournit l'image théorique d'un spectre IR dont l'attribution des bandes n'est pas triviale. Pour identifier une à une les bandes fondamentales (et même indirectement certaines bandes harmoniques ou de combinaison), on considérera les transformées de Fourier des fonctions d'autocorrélation des vitesses projetées chacune le long d'un mode normal[95, 96]. Chacun d'entre eux peut être calculé en diagonalisant le hessien, fonction des positions atomiques : $H_{ij} = \frac{1}{\sqrt{m_i m_j}} \frac{\partial^2 E}{\partial r_i \partial r_j}$. Les vecteurs propres de cette matrice représentent alors les modes normaux. A la différence de l'approche décrite précédemment, on obtiendra (3N-6) spectres

DOS, dont la position du pic principal correspondra à la valeur de la transition fondamentale provenant du mode normal sur lesquelles les vitesses temporelles ont été projetées :

$$\hat{V}_j(\omega) = e^{i\omega t} \langle V_{j_0} \cdot V_{j_t} \rangle dt \quad (\text{IV.61})$$

où $\hat{V}_j(\omega)$ est la vitesse selon la coordonnée normale j . Les vecteurs propres correspondants aux modes normaux sont extraits des simulations de DM en diagonalisant la matrice de covariance des vitesses pondérées par les masses, notées K_{ij}^v , et définies par :

$$K_{ij}^v = \frac{1}{2} \langle \sqrt{m_i m_j} v_i v_j \rangle \quad (\text{IV.62})$$

Les valeurs propres associées λ_{ij}^v représentent l'énergie cinétique moyenne sur chaque mode normal. De manière analogue, les vitesses peuvent être projetées sur les coordonnées internes[97]. Le pic principal correspond alors à la transition vibrationnelle pour laquelle la dite coordonnée interne est la plus sollicitée. Son principe général est similaire à un calcul d'orbitales localisées déterminées à partir des orbitales SCF obtenus à un niveau Hartree-Fock. Dans les deux cas, une entité (soit l'OM, soit le mode normal), qui est délocalisé sur toute la molécule, est transformée en son équivalent local. Concrètement, les fréquences correspondant aux coordonnées normales deviennent des fréquences "intrinsèques", qui représentent la somme des contributions de tous les modes normaux de vibration à une coordonnée interne spécifique.

Il arrive que ces modes de projection ne suffisent pas à obtenir des valeurs cohérentes pour les nombres d'onde correspondants à certains modes de vibration, notamment pour les stretching NH ou CH au-delà de 2800cm^{-1} . Afin de comprendre les phénomènes mis en jeu, une IVR (pour Intramolecular Vibrational energy Redistribution) peut être menée. Cela consiste à mettre initialement une certaine quantité d'énergie sur le vecteur déplacement du mode à analyser. Au cours du temps, des couplages intermodes peuvent apparaître et favoriser la propagation de l'énergie de départ sur ces modes. Ce phénomène de transfert énergétique peut entraîner des valeurs d'intensité aberrantes pour certains modes et engendrer des problèmes indirects dans l'attribution de bandes spectrales.

Ces méthodes peuvent être appliquées grâce au code VAMD implémenté au laboratoire[98].

IV.3 Conditions de calcul

Dans le cadre de la détermination statique des propriétés de structure électronique, différentes méthodes ont été testées sur nos systèmes d'étude, à savoir les aérosols de nitrate et les bases d'ADN microhydratées.

Tout d'abord, des calculs basés sur la DFT vont être menés, de part leur faible coût calculatoire par rapport aux méthodes post-HF et leur cohérence communément constatée par rapport aux données

expérimentales. En particulier, la méthode B3LYP est connue pour fournir des résultats qui approchent à 10cm^{-1} les données expérimentales sur les systèmes organiques[99] et radicalaires[100]. Quant à la méthode PBE0, sa fiabilité a été validée sur des systèmes similaires aux oxydes d'azote, tels que les oxydes de phosphore[101]. La première citée possède son analogue prenant en compte les interactions à longue portée, à savoir CAMB3LYP. Elle sera testée au même titre que les méthodes précédentes car nos systèmes possèdent des liaisons hydrogène, ainsi que des interactions métal/ligand.

D'autres méthodes développées récemment permettent elles aussi de reproduire l'influence des différentes interactions pouvant entrer en jeu au sein même d'un composé. On peut citer celles du type B3LYP-D. Les méthodes prenant en compte un haut niveau de corrélation électronique fournissent certes des résultats équivalents, mais elles sont beaucoup plus coûteuses en temps et sont donc adaptées à l'étude de petites molécules.

La méthode MP2 constitue elle aussi un niveau de calcul envisageable car les résultats issus de son application ont déjà été meilleurs que ceux observés en B3LYP sur des systèmes organiques tels que la glycine[102]. Les calculs réalisés à un niveau CCSD(T) seront des références pour le choix de la méthode utilisée, son application ne pouvant être généralisée en raison de sa trop lourde mise en œuvre de part la taille des systèmes d'intérêt.

Quant aux bases, deux grands types de bases seront associées à toutes ces méthodes : 6-31+G(d,p), 6-311+G(d,p) d'une part, et aug-cc-pVTZ d'autre part.

En ce qui concerne la DM, les forces temporelles seront déterminées par la méthode choisie pour l'étude statique, afin de pouvoir établir une comparaison directe entre ces deux familles de résultats. D'autres paramètres ont une importance particulière pour le calcul vibrationnel dynamique. Le temps total de simulation sera pris égal à 10ps, ce qui permet de bien reproduire les mouvements de grande amplitude (modes mous, de bas nombre d'onde). Le pas sera de 0.2fs, pour traduire au mieux les mouvements de faible amplitude, très rapprochés. La température de la simulation sera imposée par la valeur de l'énergie cinétique initialement insufflée au système. Cette dernière vaut : $\frac{3}{2}(N-1)k_B T$ où N est le nombre d'atomes et T la température prise égale à 2 fois la température réelle d'étude, la structure de base étant la forme optimisée à 0K. Le dernier facteur à évaluer est la masse fictive. Sa valeur est fixée à 1000. Cette valeur est celle prise par défaut dans le logiciel GAUSSIAN09, car elle permet un bon compromis entre rapidité de la DM et le traitement des liaisons labiles. Enfin, la production de spectres se réalisera dans l'espace microcanonique NVE.

Bibliographie

- [1] Born, M. ; Oppenheimer, R. Annalen der Physik 1927, 389, 457.

-
- [2] Hartree, D. R. *Mathematical Proceedings of the Cambridge Philosophical Society* 1928, 24, 89.
- [3] Fock, V. *Zeitschrift für Physik A Hadrons and Nuclei* 1930, 61, 126.
- [4] Roothaan, C. C. J. *Reviews of Modern Physics* 1951, 23, 69.
- [5] Pople, J. A.; Nesbet, R. K. *J Chem Phys* 1954, 22, 571.
- [6] Berthier, G. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences* 1954, 238, 91.
- [7] Pople, J. A.; Santry, D. P.; Segal, G. A. *J Chem Phys* 1965, 43, S129.
- [8] Pople, J. A.; Segal, G. A. *J Chem Phys* 1965, 43, S136.
- [9] Pople, J. A.; Beveridge, D. L.; Dobosh, P. A. *J Chem Phys* 1967, 47, 2026.
- [10] Bingham, R. C.; Dewar, M. J. S.; Lo, D. H. *J Am Chem Soc* 1975, 97, 1285.
- [11] Dewar, M. J. S.; Thiel, W. *J Am Chem Soc* 1977, 99, 4899.
- [12] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J Am Chem Soc* 1985, 107, 3902.
- [13] Stewart, J. J. P. *J of Molecular Modeling* 2007, 13, 1173.
- [14] Stewart, J. J. P. *J of Molecular Modeling* 2008, 14, 499.
- [15] Møller, C.; Plesset, M. S. *Phys Rev* 1934, 46, 618.
- [16] Fermi, E. *Rend Accad Naz Lincei* 1927, 6, 602.
- [17] Thomas, L. H. *Mathematical Proceedings of the Cambridge Philosophical Society* 1927, 23, 542.
- [18] Wigner, E. *Phys Rev* 1934, 46, 1002.
- [19] Hohenberg, P.; Kohn, W. *Phys Rev* 1964, 136, B864.
- [20] Kohn, W.; Sham, L. J. *Phys Rev* 1965, 140, A1133.
- [21] Dirac, P. a. M. *Mathematical Proceedings of the Cambridge Philosophical Society* 1930, 26, 376.
- [22] Vosko, S.; Wilk, L.; Nusair, M. *Canadian Journal of Physics* 1980, 58, 1200.
- [23] Perdew, J. P.; Wang, Y. *Electronic Structure of Solids*; P. Ziesche and H. Eschrig : Berlin, 1991.
- [24] Perdew, J. P.; Yue, W. *Phys Rev B* 1986, 33, 8800.
- [25] Becke, A. D. *Phys Rev A* 1988, 38, 3098.
- [26] Lee, C.; Yang, W.; Parr, R. G. *Phys Rev B* 1988, 37, 785.
- [27] Perdew, J. P.; Wang, Y. *Phys Rev B* 1992, 45, 13244.
- [28] Harris, J. *Phys Rev A* 1984, 29, 1648.
- [29] Becke, A. D. *J Chem Phys* 1993, 98, 5648.
- [30] Adamo, C.; Barone, V. *J Chem Phys* 1999, 110, 6158.

- [31] Ghosh, S. K.; Parr, R. G. *Phys Rev A* 1986, 34, 785.
- [32] Becke, A. D.; Roussel, M. R. *Phys Rev A* 1989, 39, 3761.
- [33] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys Rev Lett* 2003, 91, 146401.
- [34] Yanai, T.; Tew, D. P.; Handy, N. C. *Chem Phys Lett* 2004, 393, 51.
- [35] Grimme, S. *Journal of Computational Chemistry* 2006, 27, 1787.
- [36] Wu, Q.; Yang, W. *The Journal of Chemical Physics* 2002, 116, 515.
- [37] Huzinaga, S. *J Chem Phys* 1977, 67, 5973.
- [38] Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J Chem Phys* 1970, 52, 5001.
- [39] Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J Chem Phys* 1969, 51, 2657.
- [40] Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. *J Chem Phys* 1970, 52, 2769.
- [41] Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J Chem Phys* 1982, 77, 3654.
- [42] Hariharan, P. C.; Pople, J. A. *Theoretical Chemistry Accounts : Theory, Computation, and Modeling (Theoretica Chimica Acta)* 1973, 28, 213.
- [43] Roos, B.; Siegbahn, P. *Theoretical Chemistry Accounts : Theory, Computation, and Modeling (Theoretica Chimica Acta)* 1970, 17, 199.
- [44] Dunning, T.; Hay, J. In *Modern Theoretical Chemistry*; vol. 3 Schaefer III H.F.: New York, 1977.
- [45] Woon, D. E.; Thom H. Dunning, J. *J Chem Phys* 1993, 98, 1358.
- [46] Dunning, T. H. *J Chem Phys* 1989, 90, 1007.
- [47] Boys, S.; Bernardi, F. *Molecular Physics* 1970, 19, 553.
- [48] Gianinetti, E.; Raimondi, M.; Tornaghi, E. *IJQC* 1996, 60, 157.
- [49] Nelder, J. A.; Mead, R. *The Computer Journal* 1965, 7, 308.
- [50] Schlegel, H. B. *J of Computational Chemistry* 1982, 3, 214.
- [51] Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.; Izmaylov, A.; Bloino, J.; Zheng, G.; Sonnenberg, J.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr, J.; Peralta, J.; Ogliaro, F.; Bearpark, M.; Heyd, J.; Brothers, E.; Kudin, K.; Staroverov, V.; Kobayashi, R.; Normand, J.; Krishnan Raghavachari; Rendell, A.; Burant, J.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, E.; Cross, J.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.; Yasyev, O.; Austin, A.;

- Cammi, R. ; Pomelli, C. ; Ochterski, J. ; Martin, R. ; Morokuma, K. ; Zakrzewski, V. ; Voth, G. A. ; Salvador, P. ; Dannenberg, J. ; Dapprich, S. ; Daniels, A. D. ; Farkas, O. ; Foresman, J. ; Ortiz, J. ; Cioslowski, J. ; Fox, D. Gaussian 09, Revision A.1 ; Gaussian Inc., Wallingford CT, 2009.
- [52] Metropolis, N. ; Rosenbluth, A. W. ; Rosenbluth, M. N. ; Teller, A. H. ; Teller, E. J Chem Phys 1953, 21, 1087.
- [53] Leforestier, C. J Chem Phys 1978, 68, 4406.
- [54] Swope, W. C. ; Andersen, H. C. ; Berens, P. H. ; Wilson, K. R. J Chem Phys 1982, 76, 637.
- [55] Car, R. ; Parrinello, M. Phys Rev Lett 1985, 55, 2471.
- [56] Marchal, R. ; Carbonnière, P. ; Pouchan, C. J Chem Phys 2009, 131, 114105.
- [57] Marchal, R. ; Carbonnière, P. ; Bégué, D. ; Pouchan, C. Chem Phys Lett 2008, 453, 49.
- [58] Karamanis, P. ; Marchal, R. ; Carbonnière, P. ; Pouchan, C. J Chem Phys 2011, 135, 044511.
- [59] Marchal, R. ; Carbonnière, P. ; Pouchan, C. Computational and Theoretical Chemistry 2012, 990, 100.
- [60] Garel, T. ; Niel, J. ; Orland, H. ; Smith, J. ; Velikson, B. J Chim Phys 1990, 23, 2479.
- [61] Garel, T. ; Orland, H. Journal of Physics A : Mathematical and General 1990, 23, L621.
- [62] Einstein, A. ; Minkowski, H. Dover publications 1952, , 73.
- [63] Watson, J. K. Molecular Physics 1968, 15, 479.
- [64] Carbonniere, P. ; Barone, V. Chem Phys Lett 2004, 392, 365.
- [65] Ratner, M. A. ; Gerber, R. B. J Phys Chem 1986, 90, 20.
- [66] Bowman, J. M. Accounts of Chemical Research 1986, 19, 202.
- [67] Jung, J. O. ; Gerber, R. B. J Chem Phys 1996, 105, 10332.
- [68] Barone, V. J Chem Phys 2004, 122, 014108.
- [69] Miller, W. H. ; Hernandez, R. ; Handy, N. C. ; Jayatilaka, D. ; Willetts, A. Chem Phys Lett 1990, 172, 62 .
- [70] Bowman, J. M. ; Carter, S. ; Huang, X. International Reviews in Physical Chemistry 2003, 22, 533.
- [71] Carter, S. ; Sharma, A. R. ; Bowman, J. M. ; Rosmus, P. ; Tarroni, R. J Chem Phys 2009, 131, 224106.
- [72] Cassam-Chenai, P. ; Liévin, J. J of Computational Chemistry 2006, 27, 627.
- [73] Dunn, K. M. ; Boggs, J. E. ; Pulay, P. J Chem Phys 1986, 85, 5838.
- [74] Ribeiro, F. ; Iung, C. ; Leforestier, C. J Chem Phys 2005, 123, 054106.

-
- [75] Bégué, D. ; Pouchan, C. AIP Conference Proceedings 2007, 963, 14.
- [76] Bégué, D. ; Labeguerie, P. ; Zhang-Negrerie, D. Y. ; Avramopoulos, A. ; Serrano-Andres, L. ; Papadopoulos, M. G. PCCP 2010, 12, 13746.
- [77] Marchal, R. ; Bégué, D. ; Pouchan, C. ; Guillemin, J.-C. ; Benidar, A. Chem Phys Lett 2009, 480, 31 .
- [78] Bégué, D. ; Gohaud, N. ; Pouchan, C. ; Cassam-Chenai, P. ; Liévin, J. J Chem Phys 2007, 127, 164115.
- [79] Gohaud, N. ; Begue, D. ; Darrigan, C. ; Pouchan, C. Journal of Computational Chemistry 2005, 26, 743.
- [80] Pouchan, C. ; Zaki, K. J Chem Phys 1997, 107, 342.
- [81] Carbonnière, P. ; Dargelos, A. ; Pouchan, C. Theoretical Chemistry Accounts 2009, 125, 543.
- [82] Christiansen, O. J Chem Phys 2004, 120, 2149.
- [83] Caligiana, A. ; Aquilanti, V. ; Burcl, R. ; Handy, N. C. ; Tew, D. P. Chem Phys Lett 2003, 369, 335 .
- [84] Ehrenfest, P. Z Physik 1927, 45, 455.
- [85] Helgaker, T. ; Uggerud, E. ; Jensen, H. J. A. Chem Phys Lett 1990, 173, 145 .
- [86] Uggerud, E. ; Helgaker, T. J Am Chem Soc 1992, 114, 4265.
- [87] Schlegel, H. B. ; Iyengar, S. S. ; Li, X. ; Millam, J. M. ; Voth, G. A. ; Scuseria, G. E. ; Frisch, M. J. J Chem Phys 2002, 117, 8694.
- [88] Iyengar, S. S. ; Schlegel, H. B. ; Millam, J. M. ; A. Voth, G. ; Scuseria, G. E. ; Frisch, M. J. J Chem Phys 2001, 115, 10291.
- [89] Schlegel, H. B. ; Millam, J. M. ; Iyengar, S. S. ; Voth, G. A. ; Daniels, A. D. ; Scuseria, G. E. ; Frisch, M. J. J Chem Phys 2001, 114, 9758.
- [90] Berens, P. H. ; Mackay, D. H. J. ; White, G. M. ; Wilson, K. R. J Chem Phys 1983, 79, 2375.
- [91] Jobic, H. ; Smirnov, K. S. ; Bougeard, D. Chem Phys Lett 2001, 344, 147 .
- [92] McQuarrie, D. Statistical Mechanics ;, 1976.
- [93] Gordon, R. Advances in Magnetic Resonance 1968, 3, 1.
- [94] Berens, P. H. ; Wilson, K. R. J Chem Phys 1981, 74, 4872.
- [95] Schmitz, M. ; Tavan, P. J Chem Phys 2004, 121, 12233.
- [96] Schmitz, M. ; Tavan, P. J Chem Phys 2004, 121, 12247.

- [97] Boatz, J. A. ; Gordon, M. S. *J Phys Chem* 1989, 93, 1819.
- [98] Carbonniere, P. ; Dargelos, A. ; Ciofini, I. ; Adamo, C. ; Pouchan, C. *PCCP* 2009, 11, 4375.
- [99] Carbonniere, P. ; Barone, V. *Chem Phys Lett* 2004, 399, 226.
- [100] Barone, V. ; Carbonniere, P. ; Pouchan, C. *J Chem Phys* 2005, 122, 224308.
- [101] Carbonnière, P. ; Pouchan, C. *Chem Phys Lett* 2008, 462, 169.
- [102] Chaban, G. ; Gerber, R. *Theoretical Chemistry Accounts : Theory, Computation, and Modeling (Theoretica Chimica Acta)* 2008, 120, 273.

Chapitre 3

Étude de molécules hydratées : cas d'un système inorganique, les aérosols de nitrate

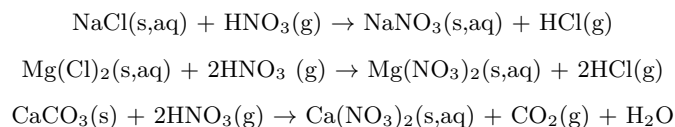
Sommaire

I	Introduction	59
I.1	Formation et fonction des aérosols	59
I.2	Position du problème	60
II	Conditions calculatoires	61
II.1	Choix de la méthode	62
II.2	Choix de la base	63
II.3	Choix du modèle structural	63
III	Propriétés vibrationnelles des aérosols	67
III.1	Publication	67
III.2	Compléments	87
IV	Conclusion	88
	Bibliographie	90

I Introduction

I.1 Formation et fonction des aérosols

Les aérosols sont un des principaux composants de l'atmosphère terrestre. Ce sont des particules en suspension dans l'air, de taille micronique ou submicronique[1], qui sont d'origine naturelle (poussières désertiques, cendres volcaniques, embruns marins) ou bien anthropique (fumées des industries, transport, poussières issues de feux agricoles)[2]. Elles se forment à partir de poussières minérales ou de sels marins qui ont la propriété d'interagir avec des oxydes d'azote tels que NO_2 , HNO_3 et N_2O_5 . Ainsi, s'obtiennent des aérosols de nitrate, dont NaNO_3 , $\text{Mg}(\text{NO}_3)_2$, $\text{Ca}(\text{NO}_3)_2$, ou encore $(\text{NH}_4)\text{NO}_3$ suivant les réactions ci-dessous[2] :



La quantité d'aérosols de nitrate, plus importante au-dessus des zones polluées[3], et plus grande à l'intérieur des terres qu'au dessus des océans, va augmenter d'ici à 2030[4]. Cela entraînera une augmentation de l'effet dit « parasol » via leur capacité à diffuser les rayons solaires UV[2], qui sera annihilé par l'effet de serre induit par la production croissante d'autres molécules[4]. Par ailleurs, ces composés toxiques pour l'homme, participent à la cancérogenèse du poumon[3].

Une autre propriété remarquable des aérosols, et plus particulièrement des aérosols de nitrate, est qu'ils président à la formation des nuages[2, 5]. Cette propension dépend d'une part de l'hygroscopicité du complexe inorganique et d'autre part de la nature de l'espèce cationique associée[2]. A ce sujet, l'augmentation de son hygroscopicité amplifie ce processus[6]. Il a été observé que pour un fort taux d'humidité relative ambiant (RH) dans l'atmosphère, les particules d'aérosols de nitrate deviennent plus hygroscopiques et réfléchissent 2 à 3 fois plus de rayonnement que ces mêmes particules sèches[7, 8]. La modification des propriétés d'absorption Infrarouge (IR) de l'aérosol en fonction du RH s'observe alors facilement par le déplacement des bandes IR caractéristiques du sel et de l'eau liée au cation[6]. De nombreuses études expérimentales ont été menées à ce sujet sur les différents types d'aérosols de nitrate cités en début de paragraphe (NaNO_3 , $\text{Mg}(\text{NO}_3)_2$, $\text{Ca}(\text{NO}_3)_2$, NH_4NO_3) pour plusieurs valeurs du taux d'humidité relative[2, 6, 9, 10, 11, 12].

I.2 Position du problème

En particulier, les études expérimentales menées sur les aérosols de nitrate NaNO_3 ont été réalisées principalement pour des taux d'humidité relative compris entre 20 et 80%. Il a été établi que de telles valeurs de RH correspondaient à un nombre de molécules autour du nitrate de sodium variant de 0.65 à 6.1. Cependant, aucune étude théorique permettant de mieux appréhender les variations des propriétés spectroscopiques observées n'a été effectuée sur ces aérosols. En outre, les effets d'anharmonicité, de solvation et dynamique qui constituent leur signature vibrationnelle n'ont pas été étudiés. Les structures géométriques des conformations les plus probables à la température stratosphérique (200K) n'ont pas été non plus recherchées, alors qu'elles président à la détermination de ces propriétés. L'objectif de l'étude présentée ci-dessous est de déterminer les propriétés microscopiques à l'origine du comportement vibrationnel observé pour nos systèmes. La démarche sera double : elle nécessite l'identification par le calcul des formes les plus stables adoptées par ces clusters en fonction du taux d'humidité, puis un calcul des propriétés vibrationnelles correspondantes viendra compléter les résultats expérimentaux développés dans la littérature. L'application théorique se limitera à l'analyse des aérosols de la forme $\text{NaNO}_3 \cdot n\text{H}_2\text{O}$, avec n un entier variant de 1 à 4.

Concrètement, l'objectif initial est de reproduire au mieux, par le calcul, les déplacements des bandes IR relatives aux cinq modes de vibration caractéristiques de NO_3^- constatées avec l'augmentation du taux d'humidité et donc de la valeur de n . Ces cinq modes, dont une représentation est disponible sur la figure 3.1, sont décrits ci-dessous :

- * le stretching symétrique ν_1 ;
- * les stretching asymétriques ν_{3A} et ν_{3B} ;
- * les bending asymétriques ν_{4A} et ν_{4B} .

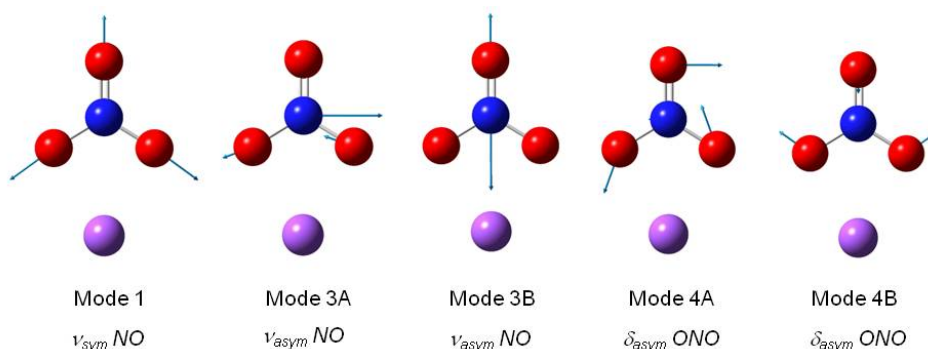


FIGURE 3.1 – Description des modes internes des aérosols de nitrate étudiés

Il faut noter que l'intensité très faible du mode ν_2 , relatif à la déformation hors du plan de NO_3^- ,

ne permet pas son étude en fonction du taux d'humidité. Les valeurs relevées expérimentalement[2, 10] pour chaque mode actif en fonction du nombre n de molécules d'eau associées à NaNO_3 sont reportées dans le tableau 3.2.

Mode	$n=0.8(1)$	$n=1.7(1.9)$	$n=3.3(3)$	$n=4.3(4)$
1	1058	1055	1053	1050
3A	1367	1360	1353	1353
3B	1427	1425	1423	1409
4A	717	717	717	720
4B	728	728	728	728

FIGURE 3.2 – Valeurs des nombres d'onde relatifs aux 5 modes de vibration de l'ion nitrate en fonction de n

Les incertitudes ont été évaluées à 10cm^{-1} sur les nombres d'onde, et à 0.12 sur le nombre de molécules d'eau agrégées autour de NaNO_3 .

II Conditions calculatoires

Les résultats présentés ci-dessous font suite aux travaux réalisés par Gibson[2] et Li[10]. L'objectif a été de tenter de reproduire par une approche théorique les évolutions expérimentales observées pour les nombres d'onde relatifs à cinq modes de vibration de l'ion nitrate NO_3^- au sein d'un environnement particulier : un aérosol de nitrate dont le taux d'humidité varie. Ce dernier est directement dépendant du nombre de molécules d'eau n nécessaire à la bonne représentation de l'aérosol par un modèle géométrique du type $(\text{NaNO}_3, n\text{H}_2\text{O})$ proposé dans la littérature. Dans ce travail, n est un entier compris entre 1 et 4. La justesse de nos résultats va être conditionnée par la qualité de nos modèles théoriques. D'une part, le modèle géométrique considéré doit être adapté à l'étude des aérosols de nitrate et doit permettre de retranscrire réellement le comportement des ions nitrate en présence de cations Na^+ et surtout d'une quantité variable d'eau. D'autre part, la qualité des résultats vibrationnels dépend du modèle calculatoire choisi : pour se faire, le choix de la méthode ainsi que de la base est primordial.

II.1 Choix de la méthode

Avant d'envisager une quelconque étude structurale ou vibrationnelle permettant de traduire, voire conforter, les données expérimentales concernant les aérosols de nitrate pour 4 taux d'hydratation différents, il est nécessaire de déterminer les conditions calculatoires les plus adaptées. Des tests préliminaires ont été effectués sur l'ion NO_3^- .

Des calculs des fréquences harmoniques de l'ion NO_3^- ont été menés à différents niveaux théoriques afin d'être comparés à des calculs de référence réalisés au niveau CCSD(T)/aug-cc-pVTZ grâce au logiciel C4[13]. Pour se faire, les méthodes B3LYP, PBE0, CAMB3LYP et MP2 ont été associées aux bases 6-31+G(d,p), 6-311+G(d,p), aug-cc-pVTZ. Les écarts entre les valeurs harmoniques calculées pour toutes ces méthodes et les valeurs CCSD(T) sont reportées dans le tableau 3.3. Il faut noter que la symétrie de cet ion étant D_{3h} , les modes 3A et 3B d'une part, et 4A et 4B d'autre part, sont dégénérés.

Method	ω_1	$\omega_{3A} = \omega_{3B}$	$\omega_{4A} = \omega_{4B}$
CCSD(T)/aug-cc-pVTZ	1046	1372	703
CCSD(T)/6-311G(d)	+7	+67	+11
B3LYP/6-31+G(d,p)	+24	+30	-6
B3LYP/6-311+G(d,p)	+20	+6	+6
B3LYP/aug-cc-pVTZ	+16	-8	+4
MP2/6-31+G(d,p)	+21	+131	-7
MP2/6-311+G(d,p)	+33	+127	+16
MP2/aug-cc-pVTZ	+20	+102	+11
CAMB3LYP/6-31+G(d,p)	+68	+89	+12
CAMB3LYP/6-311+G(d,p)	+63	+64	+24
PBE0/6-311+G(d,p)	+71	+85	+29
PBE0/aug-cc-pVTZ	+64	+69	+24

FIGURE 3.3 – Écarts (en cm^{-1}) par rapport aux fréquences harmoniques de NO_3^- calculées au niveau CCSD(T)/aug-cc-pVTZ pour différents niveaux calculatoires

On constate que l'écart moyen par rapport aux valeurs CCSD(T)/aug-cc-pVTZ est le plus faible pour les méthodes B3LYP : il est de 13cm^{-1} , contre 52, 53 et 57cm^{-1} respectivement pour les niveaux MP2, CAMB3LYP et PBE0. Ce sont surtout les modes de stretching asymétriques qui sont mal traduits

par ces dernières méthodes pour lesquelles l'écart moyen atteint 120, 77, 77 cm^{-1} . On en déduit que la méthode alliant bonne description de l'ion nitrate et coût calculatoire raisonnable est B3LYP, associée soit à la base 6-31+G(d,p), soit à la base 6-311+G(d,p).

II.2 Choix de la base

En ce qui concerne le choix de la base, l'étude du tableau précédent nous montre que les nombres d'onde des modes de bending sont du même ordre de grandeur que les valeurs de référence pour les 3 bases testées. La sélection se fait alors à partir des modes 1 et 3. Plus la base est étendue, plus les fréquences associées au mode 1 sont proches des valeurs CCSD(T). Quant aux modes 3, la petite base 6-31+G(d,p) les reproduit mal (30 cm^{-1} d'écart contre 6 ou -8 cm^{-1} pour les deux autres). Par conséquent, si l'on s'en tient à ses seuls résultats, la base 6-311+G(d,p) paraît la mieux adaptée à l'étude de NO_3^- . Cependant, la base 6-31+G(d,p) a été considérée pour les études qualitatives. Son coût calculatoire est moins important que celui de 6-311+G(d,p) : la différence entre les deux sera d'autant moins négligeable que l'hydratation, et donc la taille du système, augmentera. Ensuite, l'association de la méthode B3LYP avec la base 6-31+G(d,p) a déjà été validée pour le calcul de structure électronique de systèmes tels que H_2CO , H_2CS , H_2CNH ou C_2H_4 , ainsi que l'uracile ou le pyrrole[14, 15]. Enfin, la prise en compte d'orbitales diffuses, qui sont adaptées pour l'étude des anions, est identique pour les deux bases, 6-31+G(d,p) et 6-311+G(d,p).

II.3 Choix du modèle structural

II.3.1 Application aux systèmes $\text{NaNO}_3, n\text{H}_2\text{O}$: niveau harmonique

Dans un premier temps, une approche théorique basique a été envisagée. Des calculs harmoniques sur les structures les plus stables adoptées par nos aérosols représentés par $\text{NaNO}_3, n\text{H}_2\text{O}$ avec $n=1$ à 4 ont été réalisés. Les structures les plus probables pour chaque aérosol ont été obtenues à l'issue d'une simulation de dynamique moléculaire via une étape de thermalisation à 200K. Les structures les plus stables ont été répertoriées dans la figure 3.4.

Le calcul des propriétés vibrationnelles de chacune de ces formes à un niveau harmonique et dans la cadre de la méthode B3LYP/6-31+G(d,p) a abouti aux résultats du tableau 3.5.

Tout d'abord, on peut remarquer les fréquences différentes obtenues pour les modes 3A, 3B et 4A, 4B. Elle provient de la levée de dégénérescence des modes 3 et 4 qui se produit lorsque l'ion nitrate passe de la symétrie D_{3h} à la symétrie C_{2v} en présence de son contre-ion Na^+ . D'un point de vue qualitatif, les évolutions observées expérimentalement pour chacun des modes ne sont pas reproduites. Quantitativement, les résultats ne sont pas meilleurs puisque des écarts atteignant 40, 60, 137 cm^{-1} pour

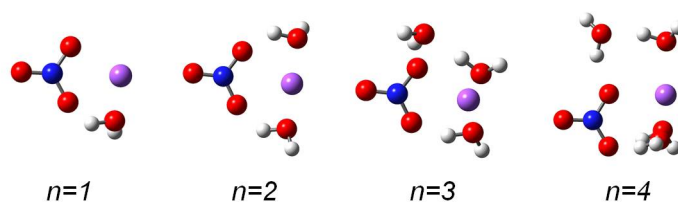


FIGURE 3.4 – Structures les plus stables adoptées par les systèmes $(\text{NaNO}_3, n\text{H}_2\text{O})$ au niveau B3LYP/6-31+G(d,p) obtenues à l'issue d'une dynamique moléculaire à 200K

n	1	2	3	4
ω_1	1054	1083	1053	1090
ω_{3A}	1307	1342	1321	1366
ω_{3B}	1556	1540	1560	1528
ω_{4A}	701	718	695	699
ω_{4B}	723	736	725	731

FIGURE 3.5 – Fréquences harmoniques (en cm^{-1}) des structures les plus stables de la forme $(\text{NaNO}_3, n\text{H}_2\text{O})$ issues d'une thermalisation à 200K, et à un niveau B3LYP/6-31+G(d,p)

les modes 1, 3A et 3B respectivement ont été relevés. La non prise en compte de l'anharmonicité ne suffit pas à expliquer de telles différences. La pertinence de la méthode de calcul a été re-vérifiée en menant ce calcul harmonique pour $n=1$ aux niveaux CAMB3LYP/6-31+G(d,p) et B3LYP+BSSE/6-31+G(d,p), dont l'influence peut jouer un rôle au sein de l'aérosol. Les valeurs obtenues sont respectivement : 1129, 1477, 1547, 747, 761 cm^{-1} et 1083, 1417, 1488, 734 et 739 cm^{-1} . On en conclut que la méthode de calcul n'est pas en cause puisque les premiers résultats sont encore plus éloignés et que les seconds engendrent des écarts entre expérience et théorie de l'ordre de 60 cm^{-1} , inexplicables encore par la seule considération de l'anharmonicité. Le modèle basique consistant à simuler l'environnement humide autour de NaNO_3 par n molécules d'eau n'est sans doute pas représentatif du milieu. De plus, plusieurs formes stables peuvent être représentatives de l'aérosol dans des conditions fixées. Ceci fera l'objet de développements dans la partie II. Enfin, les expériences ont été réalisées en laboratoire, à 298K : tous les calculs devront être considérés à cette même température. Au préalable, la méthode de calcul de l'anharmonicité la plus adaptée à l'étude de nos aérosols sera déterminée.

II.3.2 Influence du modèle anharmonique

Il s'agit ici de comparer les résultats anharmoniques obtenus à partir des structures précédentes avec les deux méthodes VPT2 et VCI-P. Cela permettra d'évaluer indirectement la force des couplages anharmoniques mettant en jeu les modes du soluté dans les aérosols de nitrate. Les écarts par rapport aux valeurs VCI-P obtenus par le logiciel Gaussian sont reportées dans le tableau reporté dans la publication placée en fin de chapitre.

Les écarts extrêmes n'excèdent pas 10cm^{-1} , ce qui signifie que les couplages soluté/soluté et soluté/solvant ne sont pas très prononcés. L'anharmonicité sera donc considérée via la méthode VPT2, moins coûteuse en temps.

Le modèle théorique étant totalement établi pour l'étude vibrationnelle des aérosols de nitrate, on peut maintenant s'attacher à déterminer le modèle géométrique représentatif de ces systèmes.

II.3.3 Étude du système $(\text{NaNO}_3, 1\text{H}_2\text{O})_x$ avec $x=1$ à 4

Le modèle $\text{NaNO}_3, n\text{H}_2\text{O}$ ne fournissant pas des résultats vibrationnels qualitativement satisfaisants, il doit être reconsidéré afin de reproduire le comportement réel des aérosols dans le milieu humide. Des systèmes sous la forme de clusters d'agrégats tels que $(\text{NaNO}_3, n\text{H}_2\text{O})_x$ ont été envisagés, avec x un entier de 1 à 4. Le choix du modèle se fera à partir de l'étude pour $n=1$.

L'algorithme GSAM a permis de générer, à partir des motifs NaNO_3 et H_2O , 200 structures préoptimisées en PM6, pour lesquelles les probabilités de Boltzmann %P ont été déterminées à 298K. Leurs calculs sont basés sur les valeurs d'énergie corrigées par le ZPVE (Zero Point Vibrational Energy). Les formes dont %P est supérieur à 5% ont été optimisées au niveau B3LYP/6-31+G(d,p) : les plus stables pour chaque valeur de x sont reportées sur la figure 3.6.

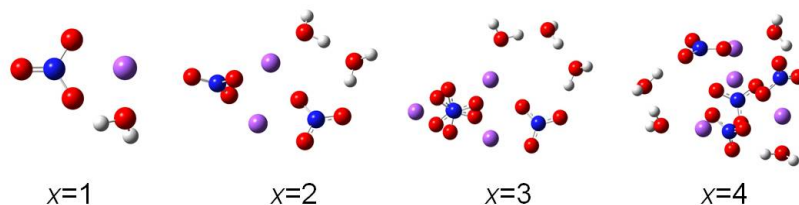


FIGURE 3.6 – Structures les plus stables adoptées par les systèmes $(\text{NaNO}_3, 1\text{H}_2\text{O})_x$ avec $x=1$ à 4 au niveau B3LYP/6-31+G(d,p)

D'un point de vue géométrique, on constate que les molécules ont tendance à s'agréger dans un même plan, puis lorsque x devient plus important, leur organisation se fait dans l'espace autour d'interactions

Na-NO₃ et de liaisons hydrogène. Le nombre de ces dernières au sein du système augmente avec le nombre de motifs NaNO₃,nH₂O. En revanche, leur longueur moyenne qui diminue jusqu'à $x=3$ repart à la hausse pour $x=4$ (1.86Å contre 1.72Å), ce qui peut signifier une plus grande instabilité.

A partir de cette étude structurale, un calcul anharmonique des fréquences correspondantes a été mené au même niveau calculatoire. La figure 3.7 rapporte les écarts entre ces valeurs et leurs contreparties expérimentales pour chacun des modes étudiés.

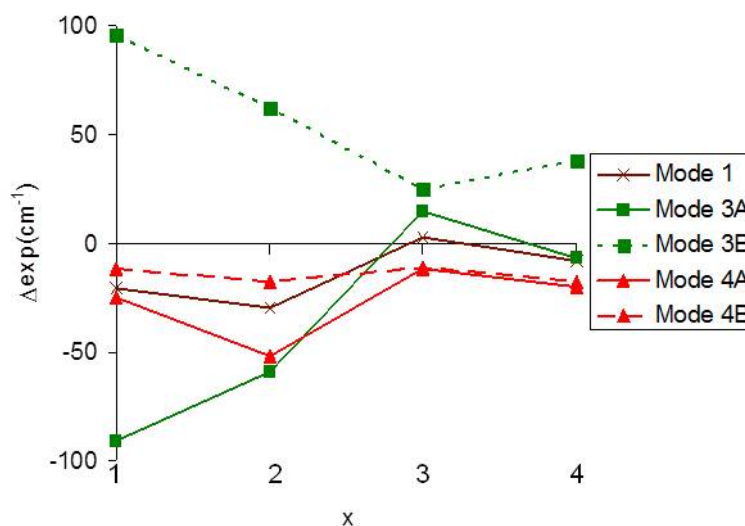


FIGURE 3.7 – Écarts (en cm^{-1}) entre les fréquences anharmoniques obtenues à un niveau B3LYP/6-31+G(d,p) pour chaque structure la plus stable de la forme $(\text{NaNO}_3,1\text{H}_2\text{O})_x$ avec $x=1$ à 4 et leurs contreparties expérimentales en fonction du mode considéré

Lorsque le nombre de motifs NaNO₃,1H₂O augmente, l'écart expérience/théorie diminue jusqu'à $x=3$. Pour $x=4$, il se stabilise : on observe donc une convergence. Pour $x=3$, il atteint au maximum 25cm^{-1} pour le mode 3B, alors qu'il était de 96cm^{-1} pour $x=1$. Ainsi, la forme $(\text{NaNO}_3,1\text{H}_2\text{O})_3$ sera considérée pour la suite, en faisant l'hypothèse que son comportement est similaire pour toutes les valeurs de n.

La procédure GSAM a été appliquée et 1000 structures ont été générées au départ pour $x=3$ et $n=2, 3$ et 4 afin de considérer l'augmentation de la taille du cluster. Après préoptimisation au niveau PM6 et la suppression des "doublons", il reste environ 40% de structures représentatives pour chaque valeur de n. L'optimisation de celles-ci avec la méthode B3LYP/6-31+G(d,p) permet de répertorier les formes les plus probables (respectant $\%P < 5\%$). On obtient ainsi 2 ou 3 structures caractéristiques pour chaque

aérosol pour $n=1$, 2, 4 et $n=3$ respectivement. Elles sont présentées dans la publication située à la fin du chapitre, associées à leurs valeurs de %P. On remarque un étalement plus prononcé des molécules d'eau et de NaNO_3 avec l'augmentation de la taille du système. Ce phénomène est mis en évidence par l'évolution moyenne de certains paramètres géométriques caractéristiques de ces clusters, dont par exemple les distances Na-N ainsi que Na-O. Les longueurs Na-N augmentent avec le taux d'humidité : pour $n=1$, elle vaut 2.77\AA pour atteindre 2.86\AA pour $n=4$. Quant aux secondes Na-O, le nombre d'entre elles inférieures à 3\AA diminue avec une plus grande hydratation. Ceci confirme l'hypothèse d'une séparation de charges de plus en plus marquée avec l'augmentation du nombre de molécules d'eau. L'étude vibrationnelle sur ces aérosols de nitrate hydratés sera conduite sur les structures ainsi déterminées.

III Propriétés vibrationnelles des aérosols

III.1 Publication

Il s'agit de déterminer les fréquences de vibration de chaque aérosol pour une valeur de n donnée. La propriété globale sera obtenue en pondérant les intensités des bandes relatives à chaque forme représentative par le pourcentage de Boltzmann correspondant. Pour chaque mode d'intérêt, les données sont ajustées par une lorentzienne dont la largeur à mi-hauteur est prise de telle sorte qu'un pic se révèle. Le maximum correspondra alors au nombre d'onde de la vibration considérée. Un spectre vibrationnel contient deux informations majeures : les valeurs des nombres d'onde ainsi que les intensités des bandes vibrationnelles. L'objectif principal de notre étude concerne ici les premières citées : les résultats finaux sont présentés dans la publication insérée ci-après.

A theoretical strategy to build structural models of microhydrated inorganic systems for the knowledge of their vibrational properties. The case of the hydrated nitrate aerosols.

Sandrine Thicoipe, Philippe Carbonniere* and Claude Pouchan

Groupe de Chimie Théorique et Réactivité, IPREM/ECP UMR CNRS 5254, Université de Pau et des Pays de l'Adour, F-64000 Pau, France.

*corresponding author: philippe.carbonniere@univ-pau.fr

Abstract

This study provides theoretical anharmonic calculations for microhydrated NaNO_3 –labeled $(\text{NaNO}_3, n\text{H}_2\text{O})_x$ – with a water-to-solute ratio (n) ranging from 1 to 4. A representative geometrical model of these forms was first investigated by simulating molecular clusters as $(\text{NaNO}_3, 1\text{H}_2\text{O})_x$ with $x=1$ to 4. The comparison between the calculated time independent anharmonic frequencies using the B3LYP/6-31+G(d,p) method and their experimental counterparts led to the choice of a supercluster model. Time independent anharmonic calculations were performed at the B3LYP/6-311+G(d,p) level of theory from the most probable structures of $(\text{NaNO}_3, n\text{H}_2\text{O})_3$ supercluster determined by using our global search algorithm we developed recently (GSAM code). The quality of the structural model is illustrated by comparing their corresponding anharmonic vibrational signatures with those obtained from IR experiences. An average deviation of 15 cm^{-1} is observed between the two series for the modes of the solute in these molecular systems ranging from 24 to 51 atoms.

Introduction

The computation of the vibrational signatures of a molecular system relies on the proper location of its associated nuclear configuration which corresponds to a particular minimum (or a particular set of minima) within a potential energy surface (PES).

In small and semi-rigid systems, quantum mechanical direct approach is able to describe accurately the PES^{1,2} which is often characterized by one minimum, while non-local methods are more appropriate for complex systems which have many degrees of freedom and large number of minima.

Even for larger and floppy carbon-based compounds exhibiting *a priori* well characterized minima, the optimized geometries of the corresponding lowest energy equilibrium conformations may be obtained from local search algorithms. The (an)harmonic vibrational signatures are then computed from time independent (or static) approaches or time dependent (or dynamic) methods by starting from the geometries obtained at 0K or at finite temperature after a step of thermalization^{3,4,5}.

For the vast majority of molecular systems for which the overall of physically reasonable conformer structures cannot be guessed, global search algorithms are used. Thus, the proper evaluation of their vibrational properties is related to the proper exploration to their complex potential energy landscapes. This is generally performed in this field from a classical molecular dynamics (MD) simulation in which the forces acting on the nuclei during their motions are computed from Quantum Mechanics (QM) or QM/Molecular Mechanics (QM/MM) methods⁶.

While the MD is by essence the best choice to explore exhaustively the PES, its efficiency is hampered by the height of the energetic barriers connecting the minima. The metadynamics avoids the drawback by adding a positive Gaussian potential around the space explored in order to discourage the system to come back to the previous point⁷.

We have recently developed an effective algorithm called GSAM⁸ (Global Search Algorithm of Minima exploration) for finding the most stable isomers of clusters. This method is based on the sampling of the most different topological forms before any electronic structure computations allowing the generation of a quite small initial set of promising structures.

Such an investigation has already been carried out on microhydrated cytosine⁹ with a number of water molecules ranging from 1 to 5 to build a simple and computationally cheap solvent model able to reproduce reliably the experimental infrared spectrum of its aqueous form.

This scheme is used in the present paper to build a structural model aiming to reproduce the anharmonic vibrational properties of more polarized molecular systems.

We chose to treat the case of nitrate aerosol which has been extensively studied experimentally, namely in reason of its hygroscopic property that makes the molecule involved in cloud condensation^{10,11}, reverberation of the sunlight and absorption of infrared (IR) radiations^{12,13}. While experimental IR transitions of (NaNO₃,nH₂O) (n=1-4) are available in the literature, namely for the internal modes of the nitrate anion^{10,14}, very few theoretical studies are reported on hydrated nitrates. We mention here the harmonic vibrational calculations for the (CaNO₃,10H₂O) cluster¹⁵. Vibrational shifts caused by anharmonicity and solvation have not been investigated so far.

The aim of the paper consists in the proposal of a structural model for the computation of anharmonic vibrational properties for such systems. The quality of the structural model will be evaluated by comparing vibrational theoretical results with their experimental counterparts. The five internal vibrational modes of the anion NO₃⁻ are considered: the symmetric stretching mode called 1, the two asymmetric stretching modes 3A and 3B and the bending modes noted 4A and 4B. They are depicted on the Figure 1.

[Figure 1]

Methods and computational details

Structural optimizations, harmonic and second order perturbative anharmonic wavenumbers¹⁶ were performed with the Gaussian09 program¹⁷. The variational treatment of the anharmonicity is performed with the VCI-P code described in details elsewhere¹⁸ which relies on an improved version of a standard variation-perturbation scheme¹⁹ allowing a massive truncation of the vibrational configuration space with minimum error on the calculated energies. We recall here that, unlike VPT2, the VCI-P treatment is able to treat properly the strongest mode-mode couplings (Fermi or Darling Dennisson resonances) but is in turn much more time consuming than the VPT2 treatment.

Electronic structure calculations were performed at the B3LYP²⁰, PBE0²¹, CAMB3LYP²² and MP2²³ levels. Computations of analytical harmonic wavenumbers at the CCSD(T)²⁴ level were performed with the C4 program²⁵. In all calculations, Pople's valence double (triple) zeta basis set including diffuse and polarization functions, 6-31(1)+G(d,p) and aug-cc-pVTZ Dunning base²⁶ were used. Harmonic calculations for NO₃⁻ anion were also carried out to compare the accuracy of the DFT models with their CCSD(T)/aug-cc-pVTZ counterparts. These DFT models were considered for the following reasons: the B3LYP/6-31+G(d,p) method is known to approach the CCSD(T)/aug-cc-pVTZ results by an average deviation of about 10cm⁻¹ in the case of small organic^{27,28} and radical²⁹ systems. The PBE0 level has been already used to successfully calculate vibrational properties of phosphorus oxides³⁰. CAMB3LYP is used here in reason of the presence of hydrogen bonds within the molecular systems under investigation.

The GSAM⁸ code (Global Search Algorithm of Minima exploration), already tested on atomic and molecular clusters, was used to find the physically reasonable geometrical structures of (NaNO₃,nH₂O)_x (n=1 to 4 ; x=1 to 4) at 298K. This algorithm includes two parts: (i) the

generation of an initial guess that consists of a random generation of several hundreds of structure for each $(\text{NaNO}_3, n\text{H}_2\text{O})_x$ cluster from the geometry of their monomeric units and (ii) their pre-optimization at the PM6 level of theory^{31,32} for which reliable results were reported on hydrated organic systems³³ and molecules exhibiting strong metal/ligand interaction³⁴. The physically reasonable conformers at 298K (Boltzmann probability > 5%) found by the algorithm were then re-optimized with a suitable DFT model chemistry.

Results and discussion

To investigate the $\text{NaNO}_3, n\text{H}_2\text{O}$ clusters with a suitable computational method, several calculations of NO_3^- harmonic frequencies were carried out at different levels of theory and were compared to the more accurate but expensive CCSD(T)/aug-cc-pVTZ method. B3LYP, PBE0, CAMB3LYP and MP2 methods were associated to 6-31+G(d,p), 6-311+G(d,p) and aug-cc-pVTZ basis sets in the present study. Deviations with respect to the CCSD(T)/aug-cc-pVTZ harmonic results are reported in the Table 1.

[Table 1]

Harmonic wavenumbers calculated at the B3LYP level of theory are closer to their CCSD(T) counterparts, unlike the MP2, PBE0 and CAMB3LYP results. For the five NO_3^- modes, the MP2, PBE0 and CAMB3LYP associated with the aug-cc-pVTZ basis set yields an average gap of 52, 57, 53 cm^{-1} respectively while it is only 13 cm^{-1} for B3LYP.

Regarding the basis sets associated to the DFT method, the triple zeta basis set of Pople which is less time consuming and give results close to the aug-cc-pVTZ results was chosen for our final computations on the hydrated systems. It is noteworthy that the double zeta basis set of

Pople, much more computationally cheaper than its triple counterpart, performed a fair good job and was used for qualitative investigations.

The Table 2 qualitatively informs about the strength of the anharmonic couplings occurring in the solute modes for the micro-hydrated nitrates. It is reported the anharmonic wavenumbers of the NO_3^- unit in the $(\text{NaNO}_3, n\text{H}_2\text{O})_1$ molecular systems ($n=1$ to 4) computed with a variational-perturbational scheme (VCI-P) and a second order perturbative treatment (VPT2) from a B3LYP/6-31+G(d,p) quartic force field. The discrepancy between the two results ranges from -9 to $+9 \text{ cm}^{-1}$ whatever the water-to-solute ratio (n). This suggests that the solute-solute and solute-solvent mode couplings are not significantly strong and leads to choose the VPT2 method for the anharmonic treatment of such a systems.

[Table 2]

The choice of the structural model is based on the agreement between the experimental wavenumbers available in literature and the anharmonic values computed from the considered structures. Two experimental sets of frequencies for the nitrate aerosols were reported in the literature^{10,14}. While the first experience was realized in ambient relative humidities corresponding to water-to-solute ratios (n) of 1, 2, 3, 4, the second experience was done for ratios of 0.8, 1.7, 3.3 and 4.3. The standard deviations were estimated to 0.12 and 10 cm^{-1} for the value of the water-to-solute ratio and the experimental wavenumbers, respectively.

The chemical nature of the microhydrated nitrate required a structural representation of this molecular system as an aggregate of several $(\text{NaNO}_3, n\text{H}_2\text{O})$ motifs. The anharmonic wavenumbers of the internal modes related to the nitrate anion were then calculated by considering the supercluster $(\text{NaNO}_3, n\text{H}_2\text{O})_x$ with n and x ranging from 1 to 4.

The fig. 2 depicts the evolution of these wavenumber values in the case of the monohydrated nitrate (WSR=n=1) with respect to the number of motifs within the lowest energy supercluster which varies from $x=1$ to 4 in our study. The wavenumbers were subtracted from their experimental counterparts¹⁰. This shows that the consideration of one structural motif only – $(\text{NaNO}_3, n\text{H}_2\text{O})_{x=1-}$ is not sufficient for a reliable representation of the vibrational property, particularly for the ν_{3A} and ν_{3B} modes for which the disagreement is about 100 cm^{-1} . Going from $x=1$ to $x=4$, the theoretical results fairly converge toward the experimental values and reveal that the structural model $(\text{NaNO}_3, n\text{H}_2\text{O})_3$ as a good compromise between reliability and feasibility.

[Figure 2]

The model chosen to simulate the hydrated nitrate aerosols was applied for WSR=2,3 and 4. About a thousand of structures were first generated with the GSAM procedure for each supercluster. The most probable structures of $(\text{NaNO}_3, n\text{H}_2\text{O})_3$ at 298K obtained at the B3LYP/6-31+G(d,p) level of theory are reported in the Table 3 with their corresponding Boltzmann probability computed from their relative stability based on their zero-point vibrational energy (ZPVE) corrected energies.

[Table 3]

We note a dispersion of NO_3^- (or Na^+) unit when the number of water molecules increases. This can be seen by the evolution of the average of the most sensitive NaNO_3 geometrical parameters w. r. t. the hydration in the supercluster. We observe a global increase of the Na-N

distances, going from 2.77Å for WSR=1 to 2.86Å for WSR=4. Moreover, the number of Na-O bonds lower than 3Å decreases when increases the water-to-solute ratio.

Infrared spectra for $(\text{NaNO}_3, n\text{H}_2\text{O})_3$ ($n=1-4$) were built from the computed VPT2 wavenumbers of each representative structure depicted in Table 3 at the B3LYP/6-311+G(d,p) level of theory. The VPT2 computations were based on a reduced-dimensionality perturbative procedure as detailed in a previous paper³⁵. Here, the quartic force field was spanned over the eighteen dimensions corresponding to all internal modes of the three NO_3^- motifs. It is noteworthy to keep in mind that the third and fourth order force constants between the 6*3 active modes of the solute and the passive modes of the solvent are also estimated in the procedure.

The values of the five modes considered in this study ($\nu_1, \nu_{3A}, \nu_{3B}, \nu_{4A}, \nu_{4B}$) were estimated as follows: for each NO_3^- motif, the Infrared intensity of each VPT2 transitions was weighted by the Boltzmann factor of their corresponding conformer. For each mode of interest, the data was then fitted by a Lorentzian curve for which the full-width at half maximum was set in such way that one peak arises. The maximum values of these peaks, with respect to the water-to-solute ratio, are reported in fig. 3 for three different structural models and compared to their experimental counterpart. The first model (model 1) is built from the lowest energy conformer of $(\text{NaNO}_3, n\text{H}_2\text{O})$ cluster, the second (model 2) corresponds to the lowest energy conformer of the supercluster containing three $(\text{NaNO}_3, n\text{H}_2\text{O})$ motifs and for the third (model 3), all the most probable forms of the supercluster (see Table 3) were considered.

[Figure 3]

Deviations between wavenumbers obtained from the simplest model and the experimental data are very important, reaching 116 cm^{-1} for the mode 3A and for a water-to-solute ratio $n=1$. The average deviation of 45 cm^{-1} cannot be explained neither by the quality of the model chemistry used, nor by the experimental uncertainties. The discrepancies are significantly decreased by using a supercluster model. Deviations in this case range, for the five modes considered, from 0 to 39 cm^{-1} when the most stable structure or all the physically relevant structures were considered. In the framework of this investigation, the average gap between the model 2 and model 3 is lower than 3 cm^{-1} , with a maximum discrepancy of 13 cm^{-1} for the mode 4B of $(\text{NaNO}_3, 2\text{H}_2\text{O})_3$. This gap, however, does not differ significantly in view of the theoretical range of about $\pm 9\text{ cm}^{-1}$ according to whether the VPT2 or VCI-P results are used. This observation is probably specific to the behaviour of semi-rigid modes such as those of the NO_3^- unit.

Conclusion

This study provided theoretical anharmonic calculations for a microhydrated inorganic system: $\text{NaNO}_3, n\text{H}_2\text{O}$ ($n=1-4$) from a B3LYP/6-311+G(d,p) quartic force field. First, the model chemistry was chosen from a set of QM methods for which the quality was assessed relative to the CCSD(T)/aug-cc-pVTZ results on the harmonic frequencies of NO_3^- anion. A representative geometrical model of these forms was then investigated by simulating the molecular cluster as $(\text{NaNO}_3, 1\text{H}_2\text{O})_x$ with $x=1$ to 4. The comparison between the calculated anharmonic frequencies with the B3LYP/6-31+G(d,p) method and their experimental counterparts led to the choice of a supercluster model for this inorganic system. VPT2 anharmonic calculations for water-to-solute ratios (n) ranging from 1 to 4 were performed at the B3LYP/6-311+G(d,p) level of theory from the most probable structures of

(NaNO₃,nH₂O)₃ supercluster determined with our GSAM procedure. The quality of the structural model is illustrated by comparing their computed wavenumbers with those obtained from IR experiences. An average deviation of 15 cm⁻¹ is observed between the two series for the vibrational modes of the solute.

Acknowledgment

The authors thank the Centre Informatique National de l'Enseignement Supérieur (CINES) for computer facilities.

References

- (1) Bégue, D.; Carbonnière, P.; Pouchan, C. *J. Phys. Chem. A* **2001**, *105*, 11379–11382.
- (2) Carbonniere, P.; Begue, D.; Pouchan, C. *J. Phys. Chem. A* **2002**, *106*, 9290–9293.
- (3) Gaigeot, M.-P. *J. Phys. Chem. A* **2008**, *112*, 13507–13517.
- (4) Jin, L.; Simons, J. P.; Gerber, R. B. *Chemical Physics Letters* **2011**, *518*, 49–54.
- (5) Carbonniere, P.; Dargelos, A.; Ciofini, I.; Adamo, C.; Pouchan, C. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4375–4384.
- (6) Yang, S.; Cho, M. *The Journal of Chemical Physics* **2009**, *131*, 135102–135102–8.
- (7) Tribello, G. A.; Cuny, J.; Eshet, H.; Parrinello, M. *The Journal of Chemical Physics* **2011**, *135*, 114109–114109–8.
- (8) Marchal, R.; Carbonnière, P.; Pouchan, C. *The Journal of Chemical Physics* **2009**, *131*, 114105–114105–9.
- (9) Carbonniere, P.; Thicoipe, S.; Very, T.; Assfeld, X. *International Journal of Quantum Chemistry* **2012**, *112*, 2221–2230.
- (10) Gibson, E. R.; Hudson, P. K.; Grassian, V. H. *J. Phys. Chem. A* **2006**, *110*, 11785–11799.
- (11) Korhonen, P.; Kulmala, M.; Hansson, H.-C.; Svenningsson, I. B.; Rusko, N. *Atmospheric Research* **1996**, *41*, 249–266.
- (12) Seinfeld, J. H.; Carmichael, G. R.; Arimoto, R.; Conant, W. C.; Brechtel, F. J.; Bates, T. S.; Cahill, T. A.; Clarke, A. D.; Doherty, S. J.; Flatau, P. J.; Huebert, B. J.; Kim, J.; Markowicz, K. M.; Quinn, P. K.; Russell, L. M.; Russell, P. B.; Shimizu, A.; Shinozuka, Y.; Song, C. H.; Tang, Y.; Uno, I.; Vogelmann, A. M.; Weber, R. J.; Woo, J.-H.; Zhang, X. Y. *Bulletin of the American Meteorological Society* **2004**, *85*, 367–380.
- (13) Clegg, S. L.; Brimblecombe, P.; Wexler, A. S. *J. Phys. Chem. A* **1998**, *102*, 2155–2171.
- (14) Li, X.-H.; Zhao, L.-J.; Dong, J.-L.; Xiao, H.-S.; Zhang, Y.-H. *J. Phys. Chem. B* **2008**, *112*, 5032–5038.
- (15) Hudson, P. K.; Schwarz, J.; Baltrusaitis, J.; Gibson, E. R.; Grassian, V. H. *J. Phys. Chem. A* **2007**, *111*, 544–548.
- (16) Barone, V. *The Journal of Chemical Physic* **2004**, *122*, 014108–014108–10.
- (17) 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.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.;

- Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Krishnan Raghavachari; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M. J.; Knox, E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yasyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.1*; Gaussian Inc., Wallingford CT, 2009.
- (18) Carbonniere, P.; Dargelos, A.; Pouchan, C. *Theoretical Chemistry Accounts* **2010**, *125*, 543
- (19) Pouchan, C.; Zaki, K. *The Journal of Chemical Physics* **1997**, *121*, 342
- (20) Becke, A. D. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.
- (21) Adamo, C.; Barone, V. *The Journal of Chemical Physics* **1999**, *110*, 6158–6170.
- (22) Yanai, T.; Tew, D. P.; Handy, N. C. *Chemical Physics Letters* **2004**, *393*, 51–57.
- (23) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.
- (24) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chemical Physics Letters* **1989**, *157*, 479–483.
- (25) Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P. G.; Auer, A. A.; Bartlett, R. J.; Benedikt, U.; Berger, C.; Bernholdt, D. E.; Bomble, Y. J.; Cheng, L.; Christiansen, O.; Heckert, M.; Heun, O.; Huber, C.; Jagau, T. C.; Jonsson, D.; Juselius, J.; Klein, K.; Lauderdale, W. J.; Matthews, D. A.; Metzroth, T.; O'Neill, D. P.; Price, D. R.; Prochnow, E.; Ruud, K.; Schiffmann, F.; Schwabach, W.; Stopkowitz, S.; Tajti, A.; Vazquez, J.; Wang, F.; Watts, J. D.; the integral packages MOLECULE, PROPS, ABACUS and ECP routines *C4, a quantum chemical program*.
- (26) Dunning, T. H. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.
- (27) Carbonniere, P.; Barone, V. *Chemical Physics Letters* **2004**, *399*, 226–229.
- (28) Carbonniere, P.; Lucca, T.; Pouchan, C.; Rega, N.; Barone, V. *Journal of Computational Chemistry* **2005**, *26*, 384–388.
- (29) Barone, V.; Carbonniere, P.; Pouchan, C. *The Journal of Chemical Physics* **2005**, *122*, 224308–224308–8.
- (30) Carbonnière, P.; Pouchan, C. *Chemical Physics Letters* **2008**, *462*, 169–172.
- (31) Stewart, J. J. P. *Journal of Molecular Modeling* **2007**, *13*, 1173–1213.
- (32) Stewart, J. J. P. *Journal of Molecular Modeling* **2008**, *14*, 499–535.
- (33) Yoshikawa, T.; Motegi, H.; Kakizaki, A.; Takayanagi, T.; Shiga, M.; Tachikawa, M. *Chemical Physics* **2009**, *365*, 60–68.
- (34) Da Costa, L. M.; De Mesquita Carneiro, J. W.; Romeiro, G. A.; Paes, L. W. C. *Journal of Molecular Modeling* **2010**, *17*, 243–249.
- (35) Barone, V.; Biczysko, M.; Bloino, J.; Borkowska-Panek, M.; Carnimeo, I.; Panek, P. *International Journal of Quantum Chemistry* **2012**, *112*, 2185–2200.

TABLE CAPTION

- **Table 1** : Computed harmonic wavenumbers of NO_3^- at the CCSD(T)/aug-cc-pVTZ level of theory. Deviations (in cm^{-1}) with respect to the CCSDT/aug-cc-pVTZ results are reported for different levels of theory.
- **Table 2** : Computed anharmonic wavenumbers of NO_3^- unit in $(\text{NaNO}_3, n\text{H}_2\text{O})_1$ ($n=1$ to 4) from the VCI-P and VPT2 methods supported by a B3LYP/6-31+G(d,p) quartic force field.
- **Table 3** : The most stable structures for $(\text{NaNO}_3, n\text{H}_2\text{O})_3$ at the B3LYP/6-31+G(d,p) level of theory found by the GSAM algorithm with respect to n . The Boltzmann probabilities are reported in parenthesis.

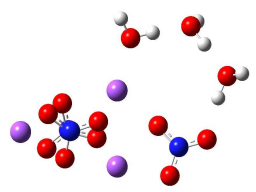
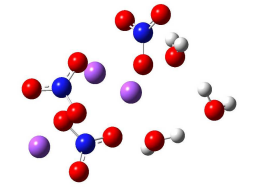
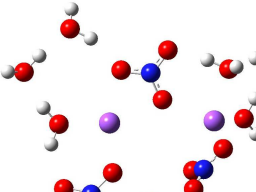
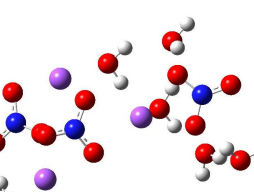
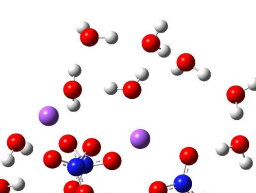
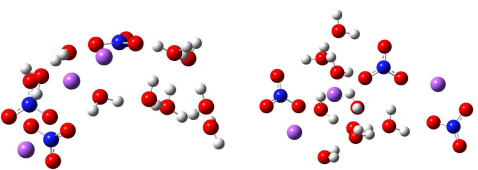
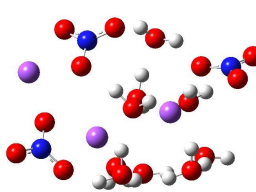
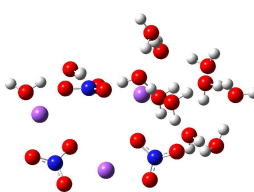
Table 1 : Computed harmonic wavenumbers of NO_3^- at the CCSD(T)/aug-cc-pVTZ level of theory. Deviations (in cm^{-1}) with respect to the CCSDT/aug-cc-pVTZ results are reported for different levels of theory.

Method	ω_1	$\omega_{3A}=\omega_{3B}$	$\omega_{4A}=\omega_{4B}$
CCSD(T)/aug-cc-pVTZ	1046	1372	703
CCSD(T)/6-311G(d)	+7	+67	+11
B3LYP/6-31+G(d,p)	+24	+30	-6
B3LYP/6-311+G(d,p)	+20	+6	+6
B3LYP/aug-cc-pVTZ	+16	-8	+4
MP2/6-31+G(d,p)	+21	+131	-7
MP2/6-311+G(d,p)	+33	+127	+16
MP2/aug-cc-pVTZ	+20	+102	+11
CAMB3LYP/6-31+G(d,p)	+68	+89	+12
CAMB3LYP/6-311+G(d,p)	+63	+64	+24
PBE0/6-311+G(d,p)	+71	+85	+29
PBE0/aug-cc-pVTZ	+64	+69	+24

Table 2 : Computed anharmonic wavenumbers of NO₃⁻ unit in (NaNO₃,nH₂O)₁ (n=1 to 4) from the VCI-P and VPT2 methods supported by a B3LYP/6-31+G(d,p) quartic force field.

n	Method	v ₁	v _{3A}	v _{3B}	v _{4A}	v _{4B}
1	VCI-P	1033	1275	1511	692	714
	VPT2	+4	+1	+9	0	+2
2	VCI-P	1065	1308	1501	709	725
	VPT2	-6	-9	0	0	-3
3	VCI-P	1033	1292	1530	682	716
	VPT2	-3	-4	-8	+1	-1
4	VCI-P	1074	1334	1489	688	716
	VPT2	-4	-6	-5	-3	-2

Table 3 : The most stable structures for $(\text{NaNO}_3, n\text{H}_2\text{O})_3$ at the B3LYP/6-311+G(d,p) level of theory found by the GSAM algorithm with respect to the water-to-solute ratio (n). The Boltzmann probabilities at 298K are reported in parenthesis.

n	The most stable structure	Minor structures
1	 (94%)	 (6%)
2	 (71%)	 (23%)
3	 (76%)	 (14%) (7%)
4	 (74%)	 (23%)

The corresponding Cartesian coordinates are available in the supplementary material.

FIGURE CAPTION

- **Figure 1** : Description of the five internal modes of sodium nitrate considered in this study from Hessian analysis.
- **Figure 2** : Evolution of the computed anharmonic wavenumbers (in cm^{-1}) for the NO_3 unit in $(\text{NaNO}_3, 1\text{H}_2\text{O})_x$ w.r.t the value of x at the B3LYP/6-31+G(p) level of theory. Deviations w.r.t the experimental wavenumbers of $(\text{NaNO}_3, 1\text{H}_2\text{O})$.
- **Figure 3** : Anharmonic wavenumbers at the B3LYP/6-311+G(d,p) level of theory of the NO_3 unit in $(\text{NaNO}_3, n\text{H}_2\text{O})$ cluster w.r.t. the water to solute ratio (n) and for different structural models. Comparison with the experimental data.

Figure 1 : Description of the five internal modes of sodium nitrate considered in this study from Hessian analysis.

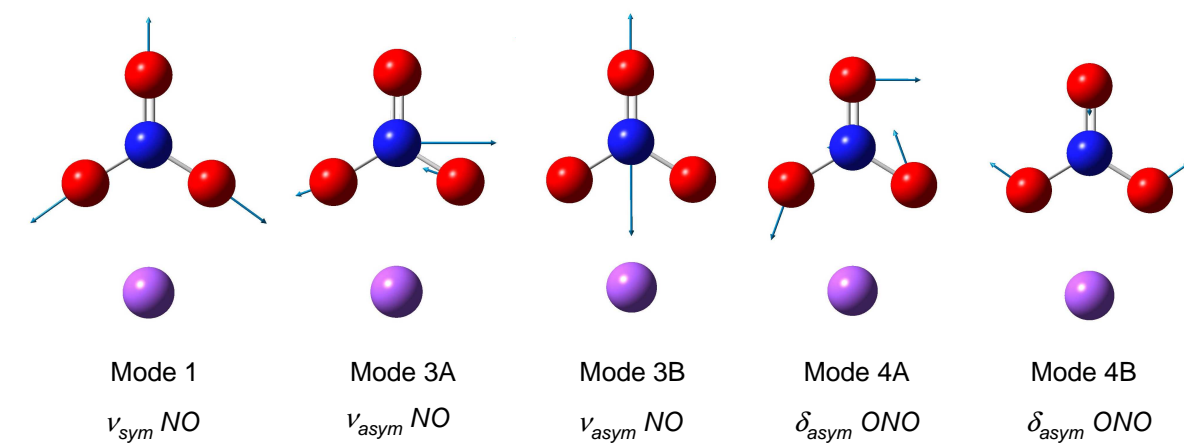


Figure 2 : Evolution of the computed anharmonic wavenumbers (in cm^{-1}) for the NO_3^- unit in $(\text{NaNO}_3, 1\text{H}_2\text{O})_x$ w.r.t the value of x at the B3LYP/6-31+G(p) level of theory. Deviations w.r.t the experimental wavenumbers of $(\text{NaNO}_3, 1\text{H}_2\text{O})$. ($\nu_{1\text{exp}}=1058\text{cm}^{-1}$; $\nu_{3\text{Aexp}}=1367\text{cm}^{-1}$; $\nu_{3\text{Bexp}}=1427\text{cm}^{-1}$; $\nu_{4\text{Aexp}}=717\text{cm}^{-1}$; $\nu_{4\text{Bexp}}=728\text{cm}^{-1}$).

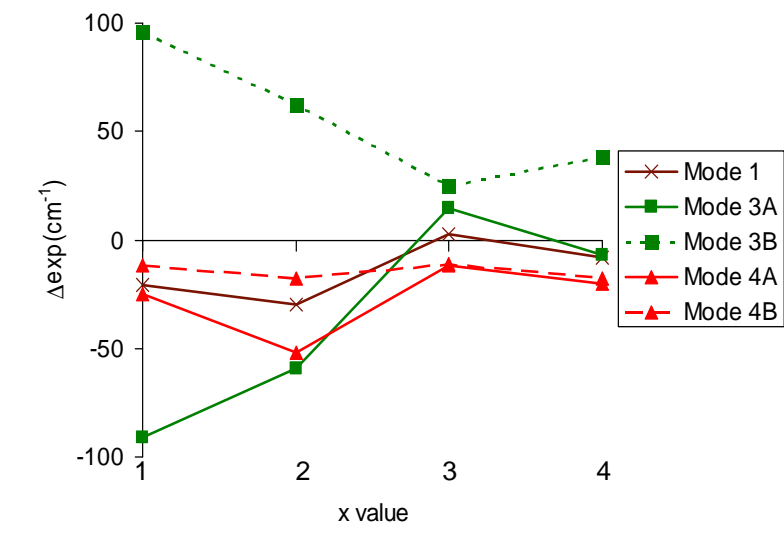
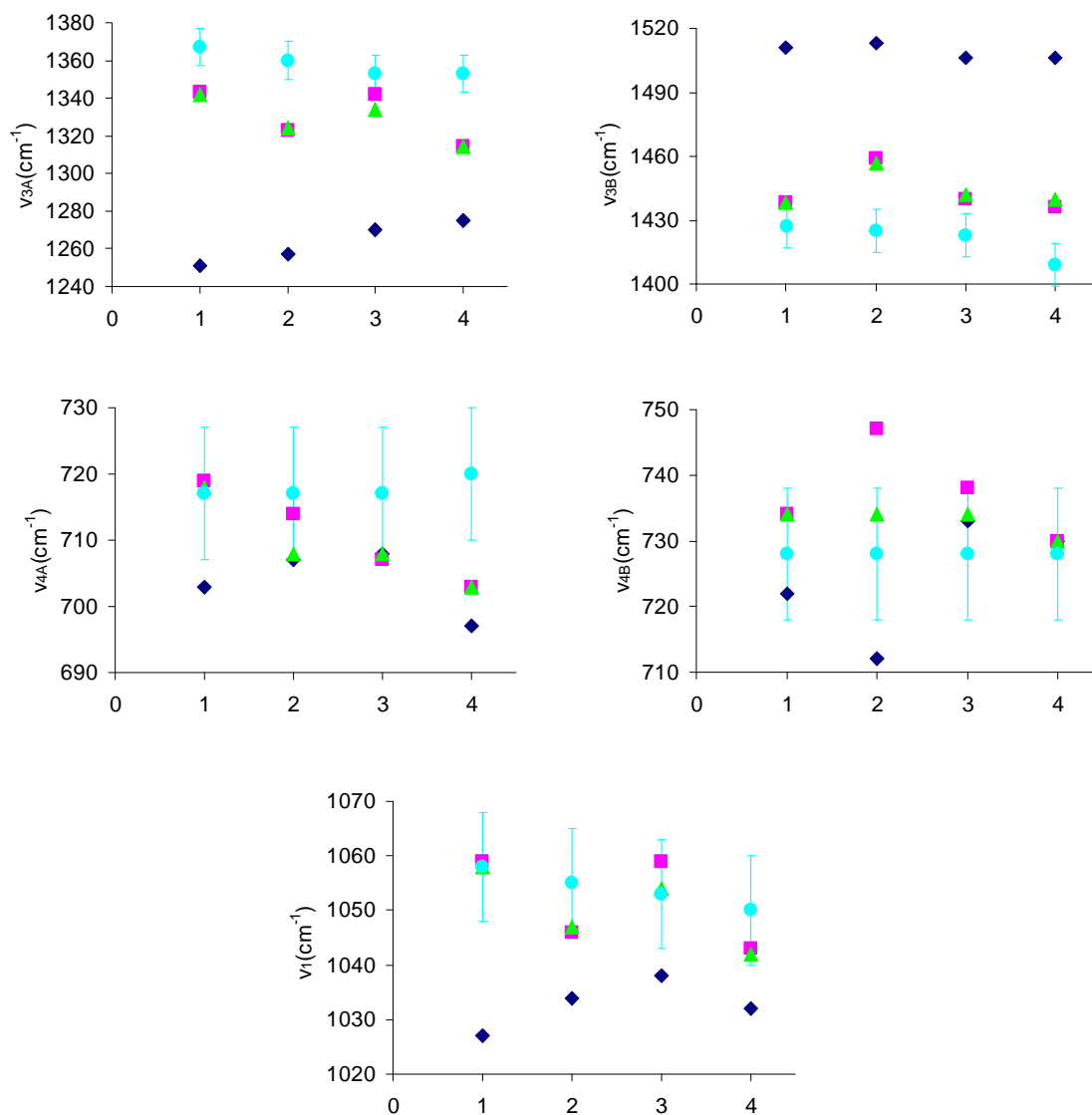


Figure 3 : Anharmonic wavenumbers at the B3LYP/6-311+G(d,p) level of theory of the NO_3^- unit in $(\text{NaNO}_3, n\text{H}_2\text{O})$ cluster w.r.t. the water to solute ratio (n) and for different structural models. Comparison with the experimental data.

- ❖ Model 1: the most stable conformer of $(\text{NaNO}_3, n\text{H}_2\text{O})_1$ cluster (rhombus labels)
- ❖ Model 2: the most stable conformer of $(\text{NaNO}_3, n\text{H}_2\text{O})_3$ cluster (square labels)
- ❖ Model 3: the most probable conformers of $(\text{NaNO}_3, n\text{H}_2\text{O})_3$ cluster (triangle labels)



En conclusion, le modèle le plus complet permet de bien reproduire l'expérience. Cette affirmation ne résulte que des comparaisons des nombres d'onde théoriques et expérimentaux. La validité de notre modèle reposant essentiellement sur l'analyse des nombres d'onde correspondant à chaque transition vibrationnelle, d'autres éléments peuvent permettre de relativiser ou non leurs qualités, dont l'intensité de chaque bande IR. De plus, certains phénomènes de couplage peuvent être mis en exergue.

III.2 Compléments

III.2.1 Intensités des bandes

Expérimentalement, seule l'évolution des intensités des bandes relatives aux modes 3 sont reportées en fonction du taux d'humidité par Gibson[2]. On constate que la bande du mode 3A voit son intensité augmenter avec le taux d'hydratation de plus en plus marqué, tandis que celle correspondant au mode 3B diminue. Par le calcul, on obtient dans la zone spectrale $1200-1600\text{cm}^{-1}$ les spectres présentés sur la figure 3.8 pour $n=1$ à 4.

La bande relative au mode 3A présente une intensité croissante pour $n=1, 2$ et 4. Pour $n=3$, elle diminue, bien que l'ordre de grandeur soit similaire à celui obtenu pour $n=2$. Quant au mode 3B, l'évolution de sa bande est aussi identique à sa contrepartie expérimentale, excepté pour $n=4$, dont l'intensité équivaut à celle de $n=1$. Les modes 3 ne sont pas fortement couplés avec d'autres modes en dehors de ceux considérés dans notre approche anharmonique. Cependant, la somme de quelques faibles couplages avec l'extérieur, notamment dans le cas de $(\text{NaNO}_3, 4\text{H}_2\text{O})_3$, peut engendrer une poussée d'intensité qui conduit à nos résultats. Les évolutions sont par conséquent globalement cohérentes. Un autre paramètre peut entrer en compte dans la validation de nos résultats dans le cadre de notre modèle pré-établi : l'évaluation des incertitudes théoriques qui, au même titre que leurs analogues expérimentales, permettent de donner un peu de flexibilité à ces valeurs.

III.2.2 Incertitudes théoriques

L'incertitude théorique sur un mode a été évaluée en faisant varier la largeur à mi-hauteur choisie pour l'obtention de nos bandes spectrales. Les nombres d'onde présentés dans l'étude correspondent au maximum de la lorentzienne décrivant l'ensemble des valeurs qui représentent une bande particulière. En modifiant cette largeur (en la divisant par 10 ici), plusieurs maxima de forte intensité apparaissent : l'incertitude théorique est considérée comme l'écart entre les maxima situés aux extrémités de la gamme spectrale étudiée et la maximum de la lorentzienne représentative de la globalité. Les valeurs théoriques, associées à leurs incertitudes dans le cadre de la méthode B3LYP/6-311+G(d,p), sont présentées sur la figure 3.9, avec leurs contreparties expérimentales.

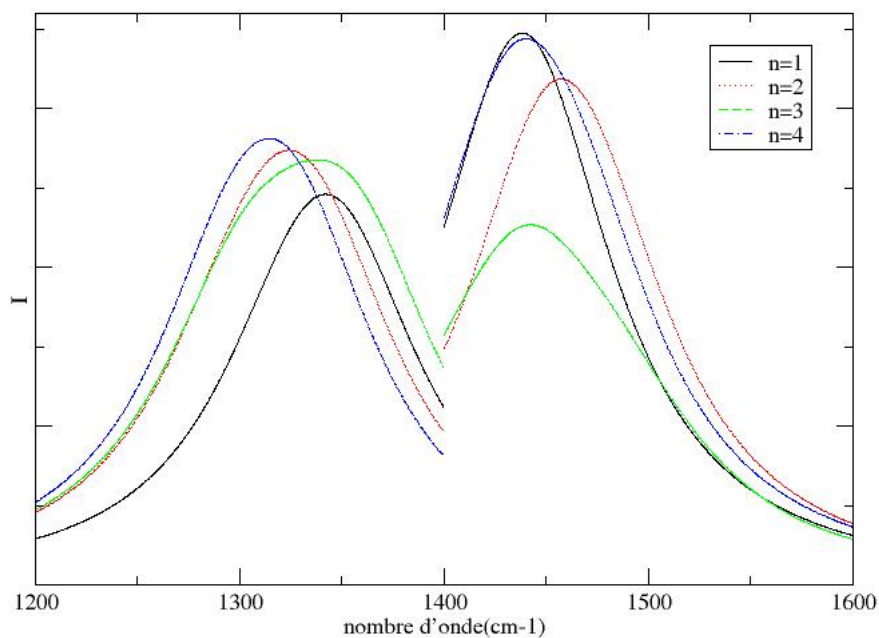


FIGURE 3.8 – Bandes spectrales correspondant aux modes 3A et 3B pour $n=1$ à 4 dans le cadre de la méthode B3LYP/6-311+G(d,p) appliquée au modèle $(\text{NaNO}_3, n\text{H}_2\text{O})_3$

La figure 3.9 montre que la prise en compte des incertitudes calculatoires permet d'obtenir des nombres d'onde théoriques dans la gamme expérimentale. Les écarts entre théorie et expérience ne concerne plus que le mode 3A, pour lequel la différence entre les deux types d'intervalles vaut 6, 7 et 9cm^{-1} pour $n=1, 2$ et 4 respectivement. On peut donc affirmer que les résultats vibrationnels déterminés à partir du modèle géométrique $(\text{NaNO}_3, n\text{H}_2\text{O})_3$ dans le cadre d'une approche anharmonique partielle en B3LYP/6-311+G(d,p) sont quantitativement bons.

IV Conclusion

L'étude théorique de systèmes inorganiques complexes tels que les aérosols de nitrate a nécessité de nombreux calculs préalables afin de déterminer les conditions calculatoires adaptées à la détermination de leurs propriétés structurales puis vibrationnelles. Le modèle géométrique reproduisant le mieux le

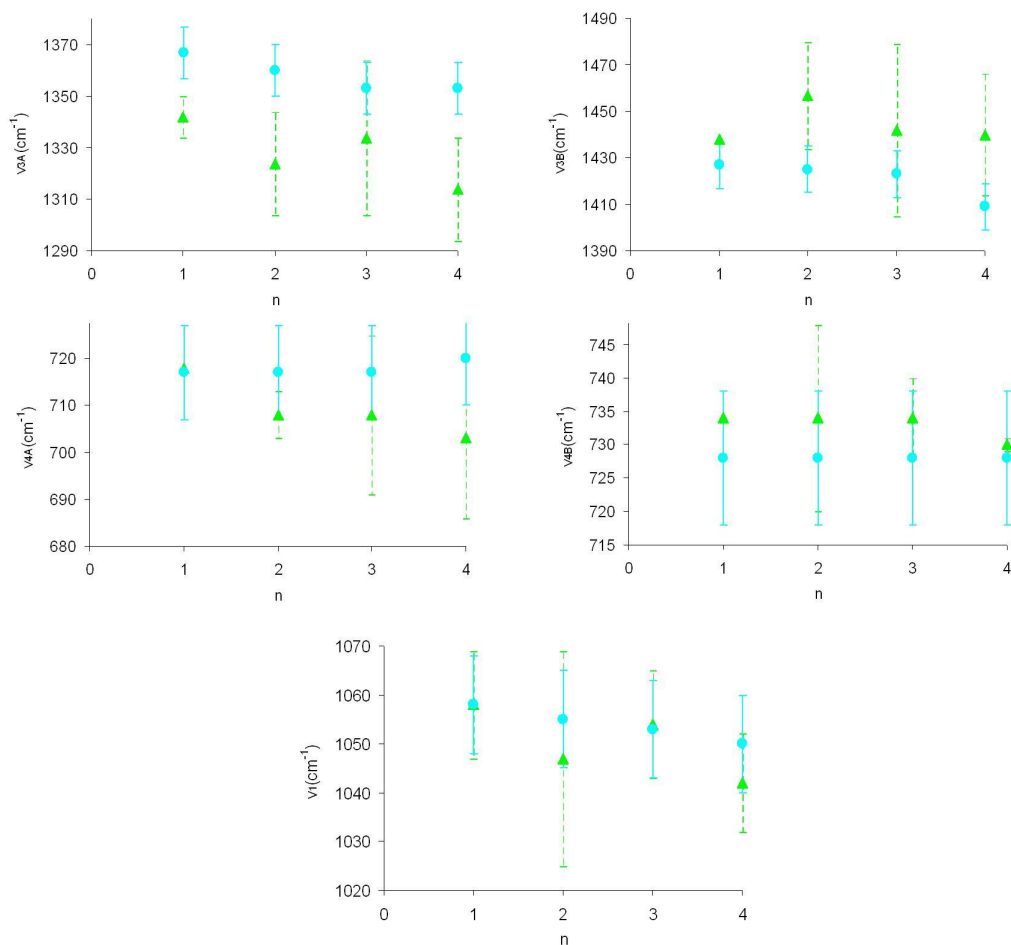


FIGURE 3.9 – Nombres d'onde anharmoniques (en cm^{-1}) obtenus au niveau B3LYP/6-311+G(d,p) pour les 5 modes caractéristiques de l'unité NO_3^- de $\text{NaNO}_3 \cdot n\text{H}_2\text{O}$ en fonction de n (triangle) avec leurs incertitudes (en pointillés) - comparaison avec les données expérimentales (rond)

comportement de NaNO_3 dans un environnement humide régulé a été déterminé : il s'agit d'un modèle du type agrégat d'agrégats de la forme $(\text{NaNO}_3 \cdot n\text{H}_2\text{O})_3$. La prise en compte des conformations les plus stables à 298K (température du laboratoire où ont été réalisées les expériences) est de plus nécessaire. La méthode B3LYP/6-311+G(d,p) a été choisie après l'étude harmonique de l'anion NO_3^- , bien que la base 6-31+G(d,p) fut utilisée pour des tests. Pour l'anharmonicité, la méthode VPT2 implémentée dans

Gaussian09 s'est révélée être celle réalisant le meilleur compromis entre temps de calcul et cohérence des résultats. Une approche partielle est suffisante pour l'étude des 5 modes souhaités, non couplés de façon marquée avec des modes propres au solvant. Ce cadre établi, la comparaison entre les résultats théoriques et leurs contreparties expérimentales conduit à un bon accord, puisque l'écart moyen entre les deux pour les 5 modes est égal à 14cm^{-1} . Si l'on s'attache à considérer les incertitudes issues de la lecture des nombres d'onde théoriques de chaque bande spectrale, les intervalles calculatoires et expérimentaux sont du même ordre de grandeur, ce qui valide notre modèle. L'évolution des intensités des bandes relatives aux modes 3 n'étant pas reproduite, notre modèle est perfectible. Une idée serait d'envisager la présence d'un solvant autour des agrégats d'agrégats via une approche PCM[16, 17, 18, 19] par exemple. Ces calculs nous ont permis de comprendre les interactions se produisant entre les diverses molécules de nitrate à 298K pour différents taux d'humidité. Cette étude met en exergue un composé relativement simple présent dans l'atmosphère et constitue une première étape vers la compréhension des mécanismes d'hydratation.

Bibliographie

- [1] Turšič, J.; Podkrajšek, B.; Grgič, I.; Ctyroky, P.; Berner, A.; Dusek, U.; Hitzenberger, R. *Chemosphere* 2006, 63, 1193 .
- [2] Gibson, E. R.; Hudson, P. K.; Grassian, V. H. *J Phys Chem A* 2006, 110, 11785.
- [3] Hodzic, A.; Vautard, R.; Chazette, P.; Menut, L.; Bessagnet, B. *Atmospheric Chemistry and Physics* 2006, 6, 3257.
- [4] Bauer, S. E.; Koch, D.; Unger, N.; Metzger, S. M.; Shindell, D. T.; Streets, D. G. *Atmospheric Chemistry and Physics* 2007, 7, 5043.
- [5] Korhonen, P.; Kulmala, M.; Hansson, H.-C.; Svenningsson, I.; Rusko, N. *Atmospheric Research* 1996, 41, 249 .
- [6] Hudson, P. K.; Schwarz, J.; Baltrusaitis, J.; Gibson, E. R.; Grassian, V. H. *J Phys Chem A* 2007, 111, 544.
- [7] Seinfeld, J. H.; Carmichael, G. R.; Arimoto, R.; Conant, W. C.; Brechtel, F. J.; Bates, T. S.; Cahill, T. A.; Clarke, A. D.; Doherty, S. J.; Flatau, P. J.; Huebert, B. J.; Kim, J.; Markowicz, K. M.; Quinn, P. K.; Russell, L. M.; Russell, P. B.; Shimizu, A.; Shinozuka, Y.; Song, C. H.; Tang, Y.; Uno, I.; Vogelmann, A. M.; Weber, R. J.; Woo, J.; Zhang, X. Y. *Bulletin of the American Meteorological Society* 2004, 85, 367.
- [8] Clegg, S. L.; Brimblecombe, P.; Wexler, A. S. *J Phys Chem A* 1998, 102, 2155.

-
- [9] Li, X.; Dong, J.; Xiao, H.; Lu, P.; Hu, Y.; Zhang, Y. *Science in China Series B : Chemistry* 2008, 51, 128.
- [10] Li, X.; Zhao, L.; Dong, J.; Xiao, H.; Zhang, Y. *J Phys Chem B* 2008, 112, 5032.
- [11] Zhang, Y.-H.; Choi, M. Y.; Chan, C. K. *J Phys Chem A* 2004, 108, 1712.
- [12] Li, X.-H.; Wang, F.; Lu, P.-D.; Dong, J.-L.; Wang, L.-Y.; Zhang, Y.-H. *J Phys Chem B* 2006, 110, 24993.
- [13] Stanton, J.; Gauss, J.; Harding, M.; Szalay, P.; Auer, A.; Bartlett, R.; Benedikt, U.; Berger, C.; Bernholdt, D.; Bomble, Y.; Cheng, L.; Christiansen, O.; Heckert, M.; Heun, O.; Huber, C.; Jagau, T.; Jonsson, D.; Juselius, J.; Klein, K.; Lauderdale, W.; Matthews, D.; Metzroth, T.; O'Neill, D.; Price, D.; Prochnow, E.; Ruud, K.; Schiffmann, F.; Schwalbach, W.; Stopkowicz, S.; Tajti, A.; Vazquez, J.; Wang, F.; Watts, J.; the integral packages MOLECULE, PROPS, ABACUS and ECP routines C4,a quantum chemical program ;,.
- [14] Carbonniere, P.; Barone, V. *Chem Phys Lett* 2004, 399, 226.
- [15] Carbonniere, P.; Lucca, T.; Pouchan, C.; Rega, N.; Barone, V. *J of Computational Chemistry* 2005, 26, 384.
- [16] Cammi, R.; Tomasi, J. *J of Computational Chemistry* 1995, 16, 1449.
- [17] Tomasi, J.; Persico, M. *Chem Rev* 1994, 94, 2027.
- [18] Miertuš, S.; Scrocco, E.; Tomasi, J. *Chem Phys* 1981, 55, 117.
- [19] Cancès, E.; Mennucci, B.; Tomasi, J. *J Chem Phys* 1997, 107, 3032.

Chapitre 4

Étude de molécules hydratées : cas d'un système organique, structures géométriques des bases d'acides nucléiques

Sommaire

I	Introduction	93
I.1	Présentation générale de l'ADN et de l'ARN	93
I.2	Hydratation de l'ADN	95
I.3	Position du problème	96
II	Étude bibliographique	97
II.1	Cas de l'uracile	97
II.2	Cas de la thymine	97
II.3	Cas de la cytosine	98
II.4	Cas de l'adénine	98
II.5	Cas de la guanine	98
III	Publication	99

IV Conclusion	126
Bibliographie	127

I Introduction

I.1 Présentation générale de l'ADN et de l'ARN

I.1.1 Structure physico-chimique de l'ADN

Une molécule d'ADN se présente sous la forme d'une double hélice enroulée, dont le nombre d'atomes atteint 150 milliards chez l'homme. Cette structure, découverte par Watson et Crick en 1953[1], repose sur l'interaction entre deux brins, dont chacun est constitué d'un seul motif répété plusieurs fois, lui-même étant formé d'une succession de nucléotides. Ce dernier comporte trois éléments de base dont une partie phosphate, qui est un phosphodiester, et une partie sucre (ou ribose) qui donne son nom à l'ADN (pour Acide DésoxyriboNucléique) et à l'ARN (Acide RiboNucléique). La caractéristique principale du nucléotide est la base azotée, qui peut être sous quatre formes différentes pour l'ADN : la guanine(G), la cytosine(C), la thymine(T) et l'adénine(A). Dans le cas de l'ARN, la troisième citée est remplacée par l'uracile(U). Ces molécules sont présentées sur la figure 4.1.

Ainsi, un enchaînement de bases azotées particulier permet de différencier chaque brin entre eux. Ce sont aussi ces bases qui permettent à la molécule d'ADN de s'articuler autour de liaisons hydrogène formées entre celles d'un brin et leurs contreparties issues du second brin : A s'associe à T et G à C dans l'ADN, tandis que l'ARN présente une interaction entre U et A. Tous ces détails sont reportés sur la figure 4.2.

Il faut enfin noter qu'il existe trois types d'ADN : les formes A, Z et B qui est la plus courante.

I.1.2 Fonction biologique de l'ADN

L'ADN est le support de l'information génétique, mais aussi de ses variations : l'ADN permet ainsi l'évolution biologique de chaque espèce. L'information stockée est codée par la succession de bases azotées décrites précédemment. Sa fonction est de fabriquer les protéines dont l'organisme a besoin et dont l'objectif principal est double : assurer l'autonomie de l'organisme (croissance et défense) et la reproduction. Un processus complexe de décodage par des enzymes -non développé ici- permet alors de retranscrire toutes les données contenues dans l'ADN.

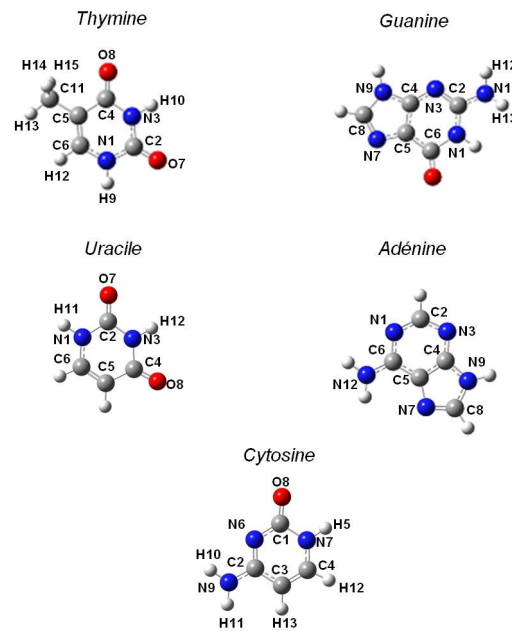


FIGURE 4.1 – Molécules d'ADN (A,T,G,C) et d'ARN (U) ainsi que la numérotation associée aux atomes

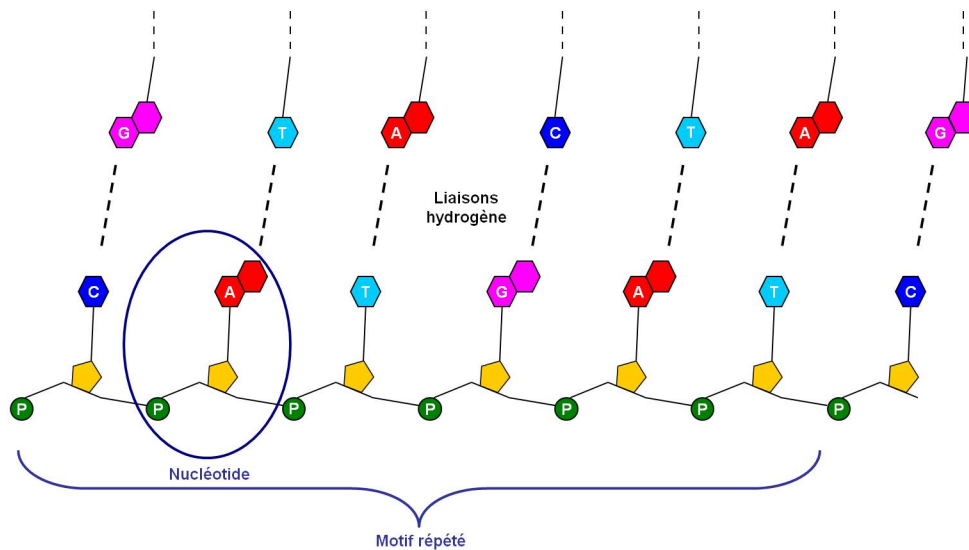


FIGURE 4.2 – Représentation schématique de l'ADN

I.1.3 Relation entre structure et fonction de l'ADN

La fonction de l'ADN se déduit de la lecture du code génétique porté par les paires de bases azotées. Si elle se fait mal, la fonction biologique initiale s'en trouve altérée. Ainsi, la structure physique de l'ADN

peut influencer sur son activité biologique : un ADN trop enroulé ne permet pas l'accès à l'information qu'il porte. La molécule d'ADN doit donc avoir une structure bien particulière pour que sa fonction soit correctement remplie. Quant au couple phosphate/sucre, il apporte la rigidité au système.

I.2 Hydratation de l'ADN

I.2.1 Conséquences du phénomène

Les diverses interactions (ou absences d'interactions) entre l'eau et les nucléotides constituent l'une des bases physico-chimiques du polymorphisme de l'ADN. L'hydratation est de plus nécessaire à la stabilité de la double hélice : les couches d'eau font écran aux charges négatives qu'elles entourent, et réduisent les répulsions entre les groupes phosphates pouvant engendrer un écartement irréversible des brins. D'ailleurs, ce rôle stabilisant est renforcé dans les milieux biologiques par la présence de contre-charges positives dissoutes (type Na^+ et Li^+) qui se lient aux phosphates et compensent leurs charges. Quand l'ADN est dans l'eau, celle-ci vient au contact des empilements de bases et renforce leur cohésion par interactions hydrophobes. Les interactions électrostatiques déstabilisantes entre les brins sont minimisées, à la différence des interactions stabilisantes entre les paires de bases successives.

I.2.2 Explicitation de l'hydratation

Dans un premier temps, des expériences de mesure de densité d'hydratation ont permis d'établir une relation entre le phénomène de solvatation et les différentes formes adoptées par l'ADN[2]. Ensuite, des études aux rayons X ont montré que les deux conformations les plus courantes pour l'ADN, à savoir B et A, sont favorisées dans des conditions d'hydratation différentes. La forme B n'est ainsi permise que si l'activité de l'eau est forte, tandis que la structure A n'est constatée que pour une activité aqueuse faible[3]. D'autres expériences ont permis de compléter ces observations : la spectroscopie IR a relié le nombre de molécules d'eau directement en interaction avec l'ADN et l'activité de l'eau en suivant l'évolution de la bande caractéristique de l'eau liée située à 3400cm^{-1} . Un modèle théorique décrivant les divers sites d'hydratation les plus probables a pu être proposé[4]. L'organisation de l'hydratation dans les sillons¹ de l'ADN a aussi fait l'objet d'études particulières, car elle varie suivant sa forme[5]. La synthèse d'oligonucléotides² a permis de quantifier le nombre moyen de molécules d'eau par nucléotide. Il est de 4 pour la forme B, contre 6 pour la forme A. Or il a déjà été établi que la forme B était prédominante pour de forts taux d'humidité et donc d'hydratation. Ces observations contradictoires implique de considérer une dynamique de l'hydratation : la forme B présente un grand nombre de molécules d'eau, en mouvement

1. zone située entre deux brins, au niveau de l'enroulement.

2. court segment d'acides nucléiques.

et moins organisées, ce qui peut expliquer que la moyennisation du nombre de molécules d'eau soit plus petite que dans le cas de la conformation A.

I.3 Position du problème

Récemment, de nombreuses études par spectroscopie IR ou Raman ont été menées afin de comprendre les interactions au sein de molécules biologiques[6, 7, 8, 9, 10]. En parallèle, des travaux ont été consacrés à l'étude de systèmes organiques microsolvatés par l'eau ou bien un autre solvant[11, 12, 13, 14, 15]. On peut donc envisager que cette technique est adaptée à l'étude de l'hydratation de l'ADN. Expérimentalement, quelques études spectroscopiques ont été réalisées sur l'ADN hydraté. L'évolution des modes bas en fonction du taux d'hydratation a particulièrement été relevé[16, 17]. Le mode de stretching $\nu(P-O)$ entre la partie phosphate et une molécule d'eau a lui aussi fait l'objet d'études particulières[18]. Cependant, leur nombre reste limité en raison de la difficulté d'interprétation des spectres obtenus : les acides nucléiques sont des systèmes complexes qui présentent une basse symétrie. La théorie permet de remédier en partie à ces problèmes d'interprétation, mais nécessite de simplifier les systèmes d'étude, un grand nombre d'atomes augmentant considérablement les temps de calcul. Pour étudier les effets de l'hydratation sur les acides nucléiques, l'approche la plus basique consiste à considérer les systèmes base azotée/eau, en faisant varier le nombre de molécules H_2O pour traduire les modifications du taux d'hydratation. L'objectif ici sera de calculer théoriquement les spectres vibrationnels des bases azotées microhydratées, représentées par X_nH_2O , avec X représentant C, T, U, G ou A et n un entier allant de 0 à 5. Ces calculs se feront à la fois par des approches statiques et dynamiques en raison du comportement observé pour les molécules d'eau autour de l'ADN. Cela nécessitera dans un premier temps de déterminer les structures géométriques représentatives adoptées par ces systèmes. L'algorithme GSAM sera utilisé : les zones privilégiées pour les molécules d'eau seront mises en évidence. On constatera de plus que plusieurs formes coexistent à 298K pour une valeur de n fixe et contribuent donc à l'empreinte vibrationnelle globale. Enfin, un autre problème réside dans le choix de la méthode de calcul utilisée pour répertorier les formes stables de chaque base microhydratée. La méthode B3LYP, déjà validé pour l'étude de systèmes organiques, sera dans un premier temps adoptée. Cependant, la complexité des systèmes étudiés, composés de plusieurs molécules en interaction entre elles via des liaisons hydrogène, suppose que soit envisagée une étude B3LYP-D, qui inclue les effets de dispersion. Les résultats issus de ces deux approches seront discutés pour valider le choix du modèle représentatif de la solution aqueuse. Tous nos résultats seront comparés aux données présentes dans la littérature et détaillées dans la partie suivante.

II Étude bibliographique

La première façon d'appréhender la molécule en solution est de la modéliser associée à un certain nombre de molécules d'eau : c'est la microhydratation. Une forme générale est alors : X, nH_2O avec $X=A, T, G, C$ ou U et n un entier. Le choix de n devient central. De nombreuses publications qui sont répertoriées ci-dessous ont déjà présenté des études sur ces systèmes. D'autres approches tels que le modèle PCM[19, 20, 21, 22] permettent aussi d'intégrer la présence d'un solvant. Ce dernier est représenté par un continuum polarisable caractérisé par une constante diélectrique relative ϵ_r (pour l'eau, ϵ_r vaut 78,39). La partie ci-après présente les différentes études -expérimentales ou théoriques- menées sur les structures des bases d'acides nucléiques anhydres ou bien microhydratées. Les géométries et les principaux sites d'hydratation de chaque base sont présentés et discutés dans ce chapitre.

II.1 Cas de l'uracile

L'uracile est la base la plus étudiée dans la littérature, notamment d'un point de vue théorique puisqu'elle ne comporte que 12 atomes. D'un point de vue expérimental, sa géométrie a été déterminée à l'état solide par des mesures réalisées aux rayons X[23] ou bien par diffraction des électrons[24]. La chimie quantique a permis elle aussi de proposer sa structure géométrique dans plusieurs contextes : gazeux, le plus souvent à partir d'approches B3LYP et MP2[25, 26, 27, 28, 29], et en milieu aqueux, au moyen du modèle PCM associé à la méthode B3LYP[30]. L'étude structurale de l'uracile hydratée n'est pas évidente par l'expérience. En revanche, la théorie est bien adaptée. Ainsi, les structures géométriques de l'uracile avec 1, 2, 3, 4, 5 et même 7 molécules d'eau obtenues par une approche statique DFT ont été répertoriées dans la littérature[27, 28, 31, 32, 33]. La mécanique ou la dynamique moléculaire appliquées respectivement à $U, 11H_2O$ et $U, 49H_2O$ [34, 35] permettent aussi cette description.

II.2 Cas de la thymine

La thymine est d'une forme semblable à l'uracile à un groupement CH_3 près. D'ailleurs, les propriétés de ces deux bases sont souvent étudiées en parallèle[25, 36, 37, 38]. La géométrie de la thymine seule a été déterminée expérimentalement par des techniques telles que celles employées pour l'uracile (rayons X notamment)[39]. La théorie a aussi permis d'établir ses paramètres structuraux au niveau B3LYP[36]. Au même titre que l'uracile, l'étude structurale de la forme hydratée n'est pas évidente alors que la théorie la permet aisément. Ainsi, différentes formes microhydratées telles que $T, 1, 2, 3, 4$ et $5H_2O$ [36, 38, 40, 41, 42] ont été présentées grâce à des méthodes statiques B3LYP, MP2 voire des méthodes de type Monte-Carlo. Des structures possédant 11 molécules d'eau ont aussi été proposées en DFT[35].

II.3 Cas de la cytosine

La cytosine est la base d'ADN la plus petite (13 atomes), ce qui explique qu'elle soit étudiée de façon privilégiée. Sa géométrie ayant déjà été déterminée expérimentalement[43], différents calculs ont été menés au niveau B3LYP[44, 45, 46]. Les paramètres structuraux de la cytosine microhydratée ont été calculés sur de nombreuses formes comprenant de 1 à 5 molécules d'eau le plus souvent[44, 45, 47, 48]. 7, 11, 13 et même 14 H₂O ont aussi été considérées[49, 50]. Les méthodes qui ont servi à ces études sont B3LYP en majorité, MP2 et HF pour les études les plus anciennes. L'influence de la base sur la géométrie de C,1H₂O a été mise en évidence[48]. Enfin, l'hydratation, traitée de façon explicite dans les cas cités précédemment, a été considérée implicitement, via un modèle de type PCM[46].

II.4 Cas de l'adénine

L'adénine est la base complémentaire de la thymine et possède le même nombre d'atomes que cette dernière. La géométrie de la forme anhydre et solide a été déterminée expérimentalement grâce aux rayons X et à la diffraction des neutrons[51, 52]. Théoriquement, les méthodes B3LYP et MP2 ont été employées afin de calculer les différents paramètres géométriques. Plusieurs bases type Pople ou Dunning leur ont été associées[52, 53]. Une étude plus complète menée par Biczysko[54] présente les résultats obtenus à l'aide de différentes méthodes prenant en compte les interactions à longue portée type CAMB3LYP, LC- ω PBE, B3LYP-D, B97-D. L'influence de l'hydratation sur la structure de base de l'adénine a été décrite à plusieurs reprises. L'étude la plus ancienne présente les différentes formes adoptées par les systèmes A,1H₂O et A,2H₂O obtenus au niveau HF/6-31++G**[55]. Les géométries des clusters A,12/13/14/16H₂O ont aussi été déterminées à un niveau calculatoire plus élaboré (B3LYP/6-31G(d))[56]. Mais l'étude la plus aboutie concerne les formes A,nH₂O avec n un entier allant de 1 à 4[57]. De nombreuses géométries ont été envisagées et leur stabilité relative est reportée dans le cadre de la méthode B3LYP à laquelle sont associées les bases 6-31G(d) et 6-311++G(d,p).

II.5 Cas de la guanine

La guanine est la base de l'ADN comportant le plus grand nombre d'atomes (16). ce qui en fait la molécule la moins étudiée jusqu'à présent. De nombreux tautomères découlant de la forme la plus communément rencontrée ont été présentés dans la littérature à des niveaux calculatoires plus ou moins élaborés, tels que HF ou MP2[58, 59, 60, 61]. Peu de données expérimentales sont en revanche disponibles sur la forme anhydre ou bien monohydratée[62, 63]. Les structures adoptées par les formes hydratées telles que G,1H₂O et G,2H₂O sont beaucoup plus étudiées[58, 64, 65, 59, 66], au même niveau de calcul que ceux cités précédemment. L'analyse des paramètres géométriques a même été menée pour des

clusters possédant une hydratation plus importante, allant de 3 à 13 molécules d'eau[67, 64]. Un modèle de traitement implicite du phénomène de solvatation -i.e. PCM- a de plus été considéré pour 7 et 8 molécules d'eau.

III Publication

The use of the GSAM approach for the structural investigation of low lying isomers of molecular clusters from density functional theory-based potential energy surfaces: the structures of microhydrated nucleic acid bases.

Sandrine Thicoipe¹, Philippe Carbonniere^{1*}, Claude Pouchan^{1*}

¹Groupe de Chimie Théorique et Réactivité, IPREM/ECP UMR CNRS 5254, Université de Pau et des Pays de l'Adour, F-64000 Pau, France.

*corresponding authors:

claude.pouchan@univ-pau.fr

philippe.carbonniere@univ-pau.fr

Abstract

This study presents structural properties of microhydrated nucleic acid bases (NAB's) – uracil(U), thymine(T), guanine(G), adenine(A)– investigated by theoretical computations at the B3LYP level of theory. To obtain the different representations of these microhydrated species, the GSAM procedure was applied: the most stable conformers labeled X,nH₂O (X=U,T,G,A and n=1..5) for which the Boltzmann population is higher than 5% at 298K are calculated at the B3LYP and B3LYP-D level of theory. Our calculated geometries are compared to those obtained in the literature. New physically relevant isomers are found with the GSAM algorithm, especially for the tetra and pentahydrated species. The use of DFT-D functional does not strongly modify the relative energies of the isomers for the monohydrated species. On the other hand, when the number of water molecules increases, the results become extremely sensitive to the consideration of dispersion contributions.

Introduction

Water plays an important role in the structure and function of nucleic acids. In living organisms, molecular recognition system, radiation induced DNA damage, stability of the DNA bases tautomers (see ref. ^{1,2,3} and therein) are examples for which the structural investigation of micro-hydrated nucleic acid bases allows a more refined description of these mechanisms and properties.

From a theoretical point of view, the physically reasonable structures of several micro-hydrated forms of adenine^{1,4} (A), thymine⁵ (T), cytosine⁶ (C), guanine^{7,8} (G) and uracil^{2,3,9,10,11,12} (U) were investigated for a number of water molecules ranging from 1 to 7ⁱ, mainly at the B3LYP level of theory^{13,14}. Diffusion Monte Carlo simulations were presented for mono-, di-, and trihydrated uracil¹⁵. Clusters composed of 11, 14 and 16 water molecules were also studied for uracil and thymine¹⁶, cytosine¹⁷ and adenine¹⁷ systems respectively. Furthermore, theoretical investigations were carried out on excited states of microhydrated adenine¹⁸.

The explicit consideration of water molecules surrounding the nucleic acid bases (NAB's) leads to clusters for which the physically reasonable structures cannot be guessed easily. While the approach consisting in generating such structures "by hand" becomes less and less efficient when the number of water molecules increases, the location of the proper nuclear configurations, which corresponds to a particular set of minima within a potential energy surface (PES), can be performed by the use of global search algorithms¹⁸.

We have recently developed an effective algorithm called GSAM (Global search algorithm of Minima) for finding the most stable isomers of clusters. This method is based on the sampling of the most different topological forms before any electronic structure computations allowing

ⁱ More precisely : 4, 3, 5, 2, 7 for A, T, C, G, U, respectively.

the generation of a quite small initial set of promising structures. It has been applied successfully for the structural investigation of atomic clusters such as silicon (Si_n , $n=3,15$)¹⁹, gallium arsenide (Ga_nAs_m , $5 < n+m < 8$)²⁰, tin telluride (Sn_nTe_m , $n=2,8$)²¹; the electrical properties of silicon doped by alkali metal ($\text{Si}_{10}(\text{Li}, \text{Na}, \text{K})_n$, $n=1, 2$)²² and the vibrational signatures of molecular clusters such as micro-hydrated cytosine ($\text{C},n\text{H}_2\text{O}$, $n=1,5$)²³.

This paper presents the capability of the GSAM algorithm to provide the physically reasonable structures of molecular clusters of the microhydrated nucleic acid bases from the monohydrated to the pentahydrated species. We stress that the present aim is to propose a set structures, as complete as possible, of conformers having a non-negligible Boltzmann population. In the next section, the computational details and the computational strategy of the GSAM procedure are detailed. The results are then presented and compared to those obtained in the literature.

Method and computational details

Concerning the investigation of molecular clusters in general, the first step of GSAM algorithm is a stochastic generation of initial structures. The clusters $[\text{A},n\text{H}_2\text{O}]$, $[\text{T},n\text{H}_2\text{O}]$, $[\text{C},n\text{H}_2\text{O}]$, $[\text{G},n\text{H}_2\text{O}]$, $[\text{U},n\text{H}_2\text{O}]$ ($n=1,5$) are considered as an ensemble of two kinds of monomeric units (the nucleic acid base and the water molecule) which are at first separately optimized at the PM6²⁴ level of theory. The generation scheme relies on the Orland procedure²⁵ which consists of growing, monomeric unit by monomeric unit, an ensemble of n -monomeric unit configurations (for instance, five monomeric units for a tetrahydrate of a nucleic acid base). These structures are built by generating a set of random coordinates which correspond to the center of mass of the randomly chosen monomeric units inside a three-dimensional box. The allowed coordinate space for the center of mass of the unit n is defined

by shifting the absolute values of the six extreme abscissa of the n-1 units set by a random value which runs from a value (rmin) to a value (rmax). This restricted three dimensional space serves as a constraint that prevents the sampling of any nonphysical configuration. Additionally, the generated structure is kept within the initial guess if at least one minimal interatomic distance between two monomeric units is within a range corresponding to the sum of the Van Der Waals radii $\pm 0.5 \text{ \AA}$. Furthermore, the generated monomeric unit is re-oriented by rotation around its inertial axes in order to minimize the Coulomb interactions with the n-1 units.

A selection scheme is then applied to detect the relatively similar configurations which most likely will lead to the same stationary point after a full geometry optimization. This consists in calculating the normed quadratic difference²⁶ (D_{ij}) between every couple of structure:

$$D_{ij} = \left(\sum_{\alpha} \frac{|Q_{i\alpha} - Q_{j\alpha}|^2}{\max_{kl} |Q_{k\alpha} - Q_{l\alpha}|^2} \right)^{1/2}$$

where Q is a vector gathering the properties of interest (here all the interatomic distances) for a given structure.

The remaining structures are firstly optimized at the PM6 level of theory. After the discarding of the twin PM6 structures with the same selection scheme, the geometries are refined by using the B3LYP/6-311+G(d,p) and the B3LYP-D^{27,28}/6-311+G(d,p) approaches. Harmonic frequency computations are then performed for isomers having a Boltzmann probability higher than 2 percent at 298K to identify the stable isomers and compute for each of them their zero point vibrational energy (ZPVE) to correct their relative and hydration energies. All the DFT-D computations have been performed using 96 cores of a local computing cluster of Intel Xeon X5660 2.8GHz 12 core processors while the DFT computation were performed on the cluster SGI Altix ICE 8200 (JADE) at the Centre Informatique National de

l'Enseignement Supérieur (CINES). Electronic structure computations, optimizations and harmonic frequencies were performed with the G09 package²⁹.

Results and discussion

An initial guess of a thousand of structures was generated for each microhydrated NAB in which about half of them were discarded after the selection step (from about 80% to 20% for mono-hydrated and penta-hydrated clusters respectively). As illustrated in fig. 1, the process yields from 20 to 70 PM6 structures for 1 to 5 water molecules and from 3 to 50 B3LYP structures. The cartesian coordinates of the stable B3LYP-D isomers having a Boltzmann population higher than 5% at 298K are reported in the supplementary material.

[fig.1]

The labels of the primary binding sites of NAB's (noted A, B, C, D, E) are shown in fig. 2. The corresponding alphabetical order is related to the relative stability of the mono-hydrated isomers computed at the DFT level of theory, as depicted by Kim et al. in their works on microhydration of thymine³⁰, cytosine⁶ and uracil². The tables 1-5 report for [A,nH₂O], [T,nH₂O], [C,nH₂O], [G,nH₂O], [U,nH₂O] (n=1,5) respectively, the B3LYP-D and B3LYP relative and hydration corrected energies of the physically reasonable isomers. The second row of each table provides the name of the structures found in terms of labels of the primary binding sites occupied and in terms of number of water molecules per site. This description is convenient but imprecise, namely when a water molecule is linked between two sites. Such

conformation is mentioned by the label (XY) followed by the number of water molecules hovering between the sites X and Y. Furthermore, the label X'n corresponds to a conformation for which the n water molecules in the site X are oriented out-of-plane of the NAB. In the third and fourth row are reported our theoretical results at the B3LYP-D/6-311+G(d,p) and B3LYP/6-311+G(d,p) level of theory, respectively. The B3LYP-D method has been chosen because of its ability to more properly account for weak intermolecular interactions^{31,32}. If any, the following rows report the most complete theoretical results found in the literature. The structures are ranked according to their relative energy obtained at the B3LYP-D level of theory. On that point, the B3LYP-D structures reported in regular font have a Boltzmann population (%B) higher than 5% at 298K. This corresponds here to isomers for which the relative energy is lower than about 1.5 kcal.mol⁻¹). It is noteworthy that additional structures are reported in italic, either because their Boltzmann population is greater than 5% at the B3LYP level of theory or because they were found in the literature, below 5 kcal.mol⁻¹ with respect to their more stable conformer for most of them. Then, the physically reasonable structures that correspond to the B3LYP-D results are depicted in the figures 3-7.

[fig.2]

Turning back to the fig. 2, the primary binding sites found are in agreement with those previously published in the literature for adenine¹, thymine^{30,33}, cytosine⁶ and uracil^{2,10}. Concerning the guanine, we propose five primary binding sites labeled A, B, C, D, E for which the relative energies of the corresponding monohydrates are 0.0, 2.4, 2.6, 2.8 and 3.2 kcal.mol⁻¹ respectively at the B3LYP/6-311+G(d,p) level which yields the B, C, D sites very close in energy. Note that the picture remains unchanged with the use of the B3LYP-D level

of theory. These results are qualitatively different from those proposed by Hanus et al.³⁴ since the authors found the following ordering: A, C, B, D with relative energies of 0.0, 2.0, 3.0 and 3.4 kcal.mol⁻¹ respectively at the RI-MP2/TZVPP level of theory.

Furthermore, our theoretical results and those available in the literature show that the sites A and B are preferentially occupied for the mono-hydrated species at 298K, excepted in the case of guanine for which only the site A is favored. Both the B3LYP and the B3LYP-D relative energies of the corresponding monohydrates A1 and B1 (see Tables 1-5 for n=1) fit qualitatively and quantitatively the previously reported theoretical results since most of the deviations range from 0.0 and 0.3 kcal.mol⁻¹ and do not overtake 0.5 kcal.mol⁻¹. The comparison is similar for the dihydrated NAB compounds (see Tables 1-5 for n=2) which illustrates that the use of a global search algorithm is in agreement with the pictures previously reported in the literature. In the same line, the consideration of the dispersion contributions has no effect on the ordering of the isomers found. However, in the case of cytosine (see table 3), the subtle changes of the relative energies observed lead to a quite sensitive variations of the Boltzmann probabilities of the forms A1, B1, A2, B2 and A1B1. Unlike the other NAB's, the behaviors of the sites A and B are quite different (see fig 2.) in this case: the water molecule is associated to N₇-H₅ and C₁=O₈ atoms in the site A while the water molecule forms a N₉-H₁₀...O_w-H_w...N₆-C₂ cyclic hydrogen bond in the site B. As said in the ref. 6, H₅ hydrogen is more acidic than H₁₀ and O₈ is more basic than N₆. The effect of these differences is less marked when the B3LYP-D model is taken into consideration.

Combinations of three water molecules on the primary binding sites presented above leads to numbers of 10, 20, 20, 35 and 20 isomers for A, T, C, G and U, respectively. On that point, the differences between our results and the literature are related to the fact that the case of three water molecules linked per binding site was not previously considered for cytosine (see table 3). The GSAM algorithm was also able to locate more than one local minimum

corresponding to a same combination as illustrated for cytosine, guanine and uracil (A3 and A'3: see tables 3-5 and fig. 5-7).

The investigation was also done for the tetra- and pentahydrates of adenine, thymine, cytosine and uracil. For the tetra-hydrated species, the most probable isomers of adenine reported in the literature were found with the GSAM approach as well as all the global minima of thymine, cytosine and uracil. We stress that some physically reasonable isomers found by our work at the B3LYP-D level of theory were not reported in the previous studies. This is the case for the A3B1 form of [T,4H₂O] (%B=7.4) and the A4 form of [U,4H₂O] (%B=36.6). The pentahydrated species referenced as global minima at the B3LYP/DZP++ level of theory in the literature are reported in the tables 2, 3, 5. Their low lying isomers computed by Kim et al.^{2,6,30} involve for all of them a water molecule in the binding site D of the NAB and yield some relative energies ranging from 2 to 6 kcal.mol⁻¹. The structures we found involve water molecules in the binding sites A, B, C and are much more lower in energy (from 0 to 1.6 kcal.mol⁻¹). However a clear distinction of the global minimum is not straightforward since the results are extremely sensitive to the model chemistry used.

The isomers reported in the literature we have not mentioned as physically reasonable according to our B3LYP-D computations (Boltzmann factor below 5%) appear in italic in the tables 1-5. By disabling the Boltzmann factor criterion, most of these isomers were found in the present study. The Boltzmann populations of these isomers are estimated at less than 0.1% at the B3LYP-D level of theory, excepted for the A4' form of [A,4H₂O] (%B=2.2 ; $\Delta E_{\text{B3LYP-D}}=1.7$ kcal.mol⁻¹), the A1(BC)3 form of [C,4H₂O] (%B=4.3; $\Delta E_{\text{B3LYP-D}}=1.8$ kcal.mol⁻¹) and the A2C2 form of [U,4H₂O] (%B=3.2; $\Delta E_{\text{B3LYP-D}}=1.7$ kcal.mol⁻¹). Furthermore the A2C1 form of [T,3H₂O], the A2B1 form and A2C1 form of [U,3H₂O] also merit attention since their corresponding relative energy is calculated at about 1.3 kcal.mol⁻¹ in the literature. However

their corresponding Boltzmann population is low within the set of isomers found by our method (1%, 3%, 2%, respectively).

A more specific comparison of some structures obtained both with the B3LYP and the B3LYP-D methods allows a better understanding of the influence of dispersion on the descriptive forms of each hydrated NAB. For trihydrated adenine, the predominant A3 structure obtained with the B3LYP method is destabilized relatively to the B3LYP-D approach which leads to a new geometry labeled A'3. We note in this case that this destabilization can be essentially explained by an increase of about 0.4Å of the average hydrogen bond length between water molecules. For n=4, the opposite is observed since interactions between solute and solvent decrease by about 0.06Å between the B3LYP A4 form and its B3LYP-D more probable A'4 counterpart. Moreover it should be emphasized in this case that the number of stabilizing interactions is more important when the dispersion effects are taken into consideration. For trihydrated cytosine and guanine, the same observation is noted. Indeed, concerning [C,3H₂O], the B3LYP A3 structure presents an average length of its hydrogen bonds lower by about 0.11Å than those of the A1B2 structure. This last conformer at the B3LYP-D level of theory becomes the most likely form and has an average hydrogen bond length lower by about 0.02Å compared to its B3LYP counterpart. The same explanation can be noted to justify the A3'' structure of the trihydrated guanine. To conclude, we can note that the influence of the dispersion effects imposes some constraints on the hydrogen bonds, which can slightly modify the most likely form and the corresponding energy of the different conformers.

Conclusion

The structural properties of the microhydrated nucleic acid bases, from the mono- to the pentahydration, were investigated with the GSAM procedure with the B3LYP-D/6-311+G(d,p) and B3LYP/6-311+G(d,p). From an initial guess of a thousand of randomly generated structures for each cluster, from 3 to 50 local extrema for each species were located. This produced for each microhydrated species between 1 to 5 stable conformers having a Boltzmann probability higher than 5% at 298K which corresponds to a ZPVE-corrected relative energies ranging from 0 to 1.6 kcal.mol⁻¹. If all the isomers reported in the literature are found for the mono- and dihydrated NAB compounds, the relative energies of the structures obtained by our algorithm, namely in the case of the pentahydrated species, are by 1 to 5 kcal.mol⁻¹ lower than those of the structures previously reported. Moreover, even though the use of DFT-D functional does not strongly modify the relative energies of the isomers for the monohydrated species, the geometries become extremely sensitive to the consideration of dispersion contributions when the number of water molecules increases.

Acknowledgment

The authors gratefully acknowledge the Centre Informatique National de l'Enseignement Supérieur (CINES) for the computational support.

References

1. H.-T. Kim, *J. Mol. Struct. (THEOCHEM)*, 2004, **673**, 121–126.
2. S. Kim and H. F. Schaefer, *J. Chem. Phys.*, 2006, **125**, 144305–144305–9.
3. R. N. Casaes, J. B. Paul, R. P. McLaughlin, R. J. Saykally, and T. van Mourik, *J. Phys. Chem. A*, 2004, **108**, 10989–10996.
4. A. F. Jalbout and L. Adamowicz, *J. Phys. Chem. A*, 2001, **105**, 1033–1038.
5. D. M. Close, C. E. Crespo-Hernández, L. Gorb, and J. Leszczynski, *J. Phys. Chem. A*, 2006, **110**, 7485–7490.
6. S. Kim and H. F. Schaefer, *J. Chem. Phys.*, 2007, **126**, 064301.
7. M. K. Shukla and J. Leszczynski, *J. Phys. Chem. B*, 2005, **109**, 17333–17339.
8. M. K. Shukla and J. Leszczynski, *J. Phys. Chem. B*, 2008, **112**, 5139–5152.
9. M.-P. Gaigeot and M. Ghomi, *J. Phys. Chem. B*, 2001, **105**, 5007–5017.
10. T. Van Mourik, S. L. Price, and D. C. Clary, *J. Phys. Chem. A*, 1999, **103**, 1611–1618.
11. T. Van Mourik, *Phys. Chem. Chem. Phys.*, 2001, **3**, 2886–2892.
12. M.-P. Gaigeot, C. Kadri, and M. Ghomi, *J. Molec. Struct.*, 2001, **565–566**, 469–473.
13. A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
14. P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
15. T. Van Mourik, S. L. Price, and D. C. Clary, *Faraday Discuss.*, 2001, **118**, 95–108.
16. V. I. Danilov, T. van Mourik, and V. I. Poltev, *Chem. Phys. Lett.*, 2006, **429**, 255–260.
17. T. van Mourik, V. I. Danilov, E. Gonzalez, A. Deriabina, and V. I. Poltev, *Chem. Phys. Lett.*, 2007, **445**, 303–308.
18. V. R. Smith, E. Samoylova, H.-H. Ritze, W. Radloff, and T. Schultz, *Phys. Chem. Chem. Phys.*, 2010, **12**, 9632–9636.
19. F. Calvo, *Comput. Materials Science*, 2009, **45**, 8–15.
20. R. Marchal, P. Carbonnière, D. Begue, and C. Pouchan, *Chem. Phys. Lett.*, 2008, **453**, 49–54.
21. R. Marchal, P. Carbonnière, and C. Pouchan, *Comput. and Theo. Chem.*, 2012, **990**, 100–105.
22. P. Karamanis, R. Marchal, P. Carbonnière, and C. Pouchan, *J. Chem. Phys.*, 2011, **135**, 044511–044511–15.
23. P. Carbonniere, S. Thicoipe, T. Very, and X. Assfeld, *Int. J. Quantum Chem.*, 2012, **112**, 2221–2230.
24. J. J. P. Stewart, *J. Molec. Modeling*, 2008, **14**, 499–535.
25. T. Garel and H. Orland, *J. Phys. A: Mathematical and General*, 1990, **23**, L621–L626.
26. G. Maroulis, *J. Chem. Phys.*, 1999, **111**, 583–591.
27. S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799.
28. Q. Wu and W. Yang, *J. Chem. Phys.*, 2002, **116**, 515–524.
29. 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, Krishnan Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. J. Klene, E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yasyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, *Gaussian 09, Revision A.1*, Gaussian Inc., Wallingford CT, 2009.

30. S. Kim, S. E. Wheeler, and H. F. Schaefer, *J. Chem. Phys.*, 2006, **124**, 204310–204310–8.
31. S. N. Steinmann and C. Corminboeuf, *J. Chem. Theory Comput.*, 2011, **7**, 3567–3577.
32. M. Biczysko, P. Panek, and V. Barone, *Chem. Phys. Lett.*, 2009, **475**, 105–110.
33. Y. Nosenko, M. Kunitski, and B. Brutschy, *J. Phys. Chem. A*, 2011, **115**, 9429–9439.
34. M. Hanus, F. Ryjáček, M. Kabelác, T. Kubař, T. V. Bogdan, S. A. Trygubenko, and P. Hobza, *J. Am. Chem. Soc.*, 2003, **125**, 7678–7688.

Table caption

Table 1. B3LYP and B3LYP-D relative (ΔE) and hydration energies (E_{hyd}) (in kcal/mol, ZPVE-corrected values) of the adenine_nH₂O (n=1-5) having a Boltzmann factor higher than 5% at 298 K.

Table 2. B3LYP and B3LYP-D relative (ΔE) and hydration energies (E_{hyd}) (in kcal/mol, ZPVE-corrected values) of the thymine_nH₂O (n=1-5) having a Boltzmann factor higher than 5% at 298 K.

Table 3. B3LYP and B3LYP-D relative (ΔE) and hydration energies (E_{hyd}) (in kcal/mol, ZPVE-corrected values) of the cytosine_nH₂O (n=1-5) having a Boltzmann factor higher than 5% at 298 K.

Table 4. B3LYP and B3LYP-D relative (ΔE) and hydration energies (E_{hyd}) (in kcal/mol, ZPVE-corrected values) of the guanine_nH₂O (n=1-5) having a Boltzmann factor higher than 5% at 298 K.

Table 5. B3LYP and B3LYP-D relative (ΔE) and hydration energies (E_{hyd}) (in kcal/mol, ZPVE-corrected values) of the uracil_nH₂O (n=1-5) having a Boltzmann factor higher than 5% at 298 K.

Table 1. B3LYP and B3LYP-D relative (ΔE) and hydration energies (E_{hyd}) (in kcal/mol, ZPVE-corrected values) of the adenine_nH₂O (n=1-5) having a Boltzmann factor higher than 5% at 298 K.

	Structure	our work B3LYP-D ^a			our work B3LYP ^b			Kim ^c	
		ΔE	$E_{\text{hyd}}^{\text{d}}/n$	%B ^e	ΔE	$E_{\text{hyd}}^{\text{d}}/n$	%B ^e	ΔE	$E_{\text{hyd}}^{\text{d}}/n$
n=1	A1	0	-10.9	72.7	0	-8.6	72.0	0	-10.9
	B1	0.7	-10.2	21.8	0.7	-7.9	22.4	0.2	-10.7
	C1	1.5	-9.4	5.3	1.5	-7.0	5.4	1.2	-9.7
n=2	A2	0	-11.6	92.5	0	-9.5	96.6	0	-12.0
	A1B1	1.5	-10.8	7.2	2.1	-8.4	2.7	1.0	-11.0
	C2	3.4	-10.5	0.0	3.6	-7.7	0.2	3.3	-10.3
	A1C1	3.5	-10.5	0.0	3.5	-7.8	0.3	3.6	-10.2
	B2	5.6	-9.7	0.0	4.4	-7.3	0.1	4.2	-9.9
n=3	A2B1	0	-11.4	94.0	0	-9.2	75.9	0	-11.8
	A2C1	1.7	-10.8	5.4	1.6	-8.6	4.4	1.9	-11.1
	A1B1C1	3.7	-10.1	0.1	4.6	-7.5	0.0	4.4	-10.3
	A1C2	3.8	-10.1	0.1	4.0	-7.8	0.1	3.9	-10.5
	A1B2	4.1	-10.0	0.0	4.2	-7.7	0.0	4.1	-10.4
	A'3	4.4	-9.9	0.1	-	-	-	5.1	-10.1
	B1C2	9.6	-8.2	0.0	8.2	-6.4	0.0	5.0	-10.1
	A3	-	-	-	0.8	-8.9	19.0	1.3	-11.3
	n=4	A3B1	0	-10.8	39.2	0	-8.8	59.1	0
A2B2		0.3	-10.6	21.6	0.9	-8.6	11.9	0.9	-11.1
A2C2		0.3	-10.6	21.6	1.0	-8.6	11.5	0.9	-11.1
A2B1C1		0.1	-10.7	35.4	1.3	-8.5	11.4	1.2	-11.0
A'4		1.7	-10.3	2.2	-	-	-	5.1	-10.0
A3C1		2.3	-10.0	0.5	1.7	-8.4	3.4	1.9	-10.8
A1C3		5.4	-9.5	0.0	6.2	-7.3	0.0	4.9	-10.2
A4		-	-	-	4.8	-7.6	0.0	3.7	-10.4
n=5		A3B1C1	0	-10.4	61.9	2.3	-7.9	1.7	
	A3B2	0.6	-10.2	23.4	0	-8.4	51.6		
	A4B1	1.4	-10.1	6.2	4.9	-7.4	0.0		
	A'3B2	1.5	-10.1	5.2	-	-	-		
	A'3B1C1	2.2	-9.9	1.4	-	-	-		
	A3C2	2.3	-9.8	0.0	0.1	-8.4	45.4		

^aB3LYP-D/6-311+G(d,p) and ^bB3LYP/6-311+G(d,p) and computations.

^cB3LYP/6-311++G(d,p) computations¹. It is not mentioned that results are ZPVE energy-corrected.

^d E_{hyd} : energy of the reaction $A + n\text{H}_2\text{O} \rightarrow A \cdot (\text{H}_2\text{O})_n$.

^eBoltzmann probabilities (%) at 298 K.

Table 2. B3LYP and B3LYP-D relative (ΔE) and hydration energies (E_{hyd}) (in kcal/mol, ZPVE-corrected values) of the thymine_nH₂O (n=1-5) having a Boltzmann factor higher than 5% at 298 K.

	Structure	our work B3LYP-D ^a			our work B3LYP ^b			Close et al ^c	Kim et al ^d	
		ΔE	E_{hyd}^e/n	%B ^f	ΔE	E_{hyd}^e/n	%B ^e	ΔE	ΔE	E_{hyd}^e/n
n=1	A1	0	-10.3	92.2	0	-8.3	88.6	0	0	-8.4
	B1	1.5	-8.8	6.3	1.5	-6.8	7.2	-	1.6	-6.8
	C1	2.5	-7.8	1.3	1.8	-6.4	3.9	2.1	1.9	-6.5
	D1	4.9	-5.4	0.0	3.8	-4.5	0.1	1.7	3.8	-4.6
n=2	A2	0	-10.9	98.8	0	-9.1	97.7	0	0	-9.3
	B2	3.2	-9.3	0.0	2.5	-7.8	1.4	2.8	2.6	-8.0
	A1B1	3.2	-9.3	0.0	2.9	-7.6	0.7	3.4	3.0	-7.8
	C2	3.7	-9.0	0.0	-	-	-	-	3.4	-7.6
	A1C1	4.1	-8.6	0.0	3.7	-7.2	0.0	-	4.0	-7.3
	A1D1	6.2	-7.5	0.0	5.3	-6.4	0.0	-	5.4	-6.6
n=3	A3	0	-10.3	52.0	0	-8.7	77.9	0	-	-
	A2B1	0.2	-10.2	36.8	0.9	-8.4	17.1	1	0	-8.5
	A1B2	0.9	-9.9	10.4	2.0	-8.0	2.7	2.1	1.1	-8.2
	A2C1	1.7	-9.7	1.0	2.5	-7.9	1.1	-	1.3	-8.1
	(BC)3	-	-	-	-	-	-	-	1.9	-7.9
	A2D1	-	-	-	-	-	-	-	2.4	-7.7
	A1C2	-	-	-	-	-	-	-	2.7	-7.6
n=4	A2B2	0	-10.4	89.5	0	-8.5	80.6		0	-7.7
	A3B1	1.5	-10.0	7.4	0.9	-8.3	17.2		-	-
	A4	2.2	-9.8	2.3	4.2	-7.5	0.0		-	-
	A2C2	-	-	-	-	-	-		2.3	-8.1
	A1(BC)3	-	-	-	-	-	-		2.8	-8.0
n=5	A2C2B1	0	-10.0	63.8	0	-8.3	81.4		0	-8.4
	A3B2	0.9	-9.8	13.6	2.3	-7.9	2.4		-	-
	A1B3C1	1.0	-9.8	12.6	2.4	-7.9	1.4		-	-
	A5	1.5	-9.7	5.1	2.9	-7.8	0.0		-	-
	A1B1C3	1.5	-9.7	4.7	3.5	-7.8	0.0		-	-
	A2C3	-	-	-	1.6	-8.0	5.7		-	-
	A2B2D1	-	-	-	-	-	-		3.2	-7.7
	A2C2D1	-	-	-	-	-	-		4.3	-7.5

^aB3LYP-D/6-311+G(d,p) and ^bB3LYP/6-311+G(d,p) computations.

^cB3LYP/6-31++G(d,p) computations⁵.

^dB3LYP/DZP++ computations³⁰.

^e E_{hyd} : energy of the reaction $T + n\text{H}_2\text{O} \rightarrow T(\text{H}_2\text{O})_n$.

^fBoltzmann probabilities (%) at 298 K.

Table 3. B3LYP and B3LYP-D relative (ΔE) and hydration energies (E_{hyd}) (in kcal/mol, ZPVE-corrected values) of the cytosine_nH₂O (n=1-5) having a Boltzmann factor higher than 5% at 298 K.

	Structure	our work B3LYP-D ^a			our work B3LYP ^b			Kim et al ^c	
		ΔE	$E_{\text{hyd}}^{\text{d}}/n$	%B ^e	ΔE	$E_{\text{hyd}}^{\text{d}}/n$	%B ^e	ΔE	$E_{\text{hyd}}^{\text{d}}/n$
n=1	A1	0	-11.3	54.0	0	-9.4	69.6	0	-9.7
	B1	0.1	-11.2	45.9	0.5	-8.9	30.3	0.5	-9.1
	<i>C1</i>	<i>1.0</i>	<i>-11.0</i>	<i>0.0</i>	<i>4.9</i>	<i>-4.5</i>	<i>0.0</i>	<i>5.1</i>	<i>-4.5</i>
n=2	A2	0	-11.5	53.6	0	-9.7	74.8	0	-10.0
	B2	0.5	-11.3	24.5	1.0	-9.2	13.5	1.2	-9.3
	A1B1	0.6	-11.2	20.5	1.3	-9.0	9.0	1.2	-9.3
	<i>(BC)2</i>	<i>1.9</i>	<i>-10.6</i>	<i>1.3</i>	<i>2.0</i>	<i>-8.7</i>	<i>2.5</i>	<i>2.4</i>	<i>-8.8</i>
	<i>A1D1</i>	<i>6.8</i>	<i>-8.1</i>	<i>0.0</i>	<i>5.5</i>	<i>-6.9</i>	<i>0.0</i>	<i>5.7</i>	<i>-7.1</i>
n=3	A2B1	0	-11.4	60.4	0	-9.4	64.6	0	-9.6
	A1B2	0.4	-11.3	31.4	1.5	-8.9	5.1	0.9	-9.3
	(BC)3	1.2	-11.0	7.9	1.5	-8.9	5.4	1.7	-9.1
	<i>A1(BC)2</i>	<i>2.7</i>	<i>-10.6</i>	<i>0.2</i>	-	-	-	<i>2.1</i>	<i>-8.9</i>
	<i>A'3</i>	<i>3.4</i>	<i>-10.2</i>	<i>0.0</i>	<i>1.3</i>	<i>-9.0</i>	<i>7.3</i>	-	-
	<i>A2D1</i>	<i>6.2</i>	<i>-9.3</i>	<i>0.0</i>	<i>4.3</i>	<i>-8.0</i>	<i>0.0</i>	<i>4.5</i>	<i>-8.1</i>
	<i>A1B1D1</i>	<i>6.6</i>	<i>-9.2</i>	<i>0.0</i>	<i>5.6</i>	<i>-7.5</i>	<i>0.0</i>	<i>5.8</i>	<i>-7.7</i>
	<i>A3</i>	-	-	-	<i>0.8</i>	<i>-9.1</i>	<i>15.6</i>	-	-
n=4	A2B2	0	-11.4	93.6	0	-9.4	88.1	0	-9.6
	<i>A1(BC)3</i>	<i>1.8</i>	<i>-10.9</i>	<i>4.3</i>	<i>2.3</i>	<i>-8.8</i>	<i>1.8</i>	<i>2.5</i>	<i>-9.0</i>
	<i>A3B1</i>	<i>3.0</i>	<i>-10.6</i>	<i>0.0</i>	<i>1.4</i>	<i>-9.0</i>	<i>8.1</i>	-	-
	<i>A2B1D1</i>	<i>6.4</i>	<i>-9.8</i>	<i>0.0</i>	<i>4.9</i>	<i>-8.2</i>	<i>0.0</i>	<i>5.2</i>	<i>-8.3</i>
	<i>A2(BC)2</i>	-	-	-	-	-	-	<i>1.6</i>	<i>-9.2</i>
n=5	A2(BC)3	0	-11.0	45.4	0.1	-9.0	39.2	0	-9.1
	A1B2C2	0.6	-10.9	16.2	3.6	-8.3	0.1	-	-
	A2B3	0.9	-10.9	10.7	0	-9.0	43.5	-	-
	A2B'3	0.9	-10.9	10.7	-	-	-	-	-
	(BC)5	1.0	-10.8	8.6	3.1	-8.4	0.2	-	-
	A4B1	1.1	-10.8	6.7	1.0	-8.8	7.4	-	-
	<i>A3B2</i>	<i>2.0</i>	<i>-10.6</i>	<i>0.0</i>	<i>1.4</i>	<i>-8.7</i>	<i>4.1</i>	-	-
	<i>A1(BC)3D1</i>	<i>6.2</i>	<i>-9.8</i>	<i>0.0</i>	<i>5.2</i>	<i>-8.0</i>	<i>0.0</i>	<i>5.2</i>	<i>-8.1</i>
	<i>A2B2D1</i>	<i>7.7</i>	<i>-9.5</i>	<i>0.0</i>	<i>3.5</i>	<i>-8.3</i>	<i>0.1</i>	<i>3.0</i>	<i>-8.5</i>

^aB3LYP-D/6-311+G(d,p) and ^bB3LYP/6-311+G(d,p) computations.

^cB3LYP/DZP++ computations⁶

^d E_{hyd} : energy of the reaction $C + n\text{H}_2\text{O} \rightarrow C(\text{H}_2\text{O})_n$

^eBoltzmann probabilities (%) at 298 K.

Table 4. B3LYP and B3LYP-D relative (ΔE) and hydration energies (E_{hyd}) (in kcal/mol, ZPVE-corrected values) of the guanine_nH₂O (n=1-5) having a Boltzmann factor higher than 5% at 298 K.

	Structure	our work B3LYP-D ^a			our work B3LYP ^b		
		ΔE	E_{hyd}^c/n	%B ^d	ΔE	E_{hyd}^c/n	%B ^d
n=1	A1	0	-11.6	94.4	0	-9.4	93.7
	<i>B1</i>	2.4	-9.2	2.1	2.4	-7.1	1.6
	<i>C1</i>	2.5	-9.2	1.6	2.6	-6.8	1.2
	<i>D1</i>	2.7	-8.9	1.2	2.8	-6.7	1.1
	<i>E1</i>	3.5	-8.1	0.3	3.2	-6.2	0.5
n=2	A1B1	0	-12.6	99.9	0	-10.3	99.7
n=3	A2B1	0	-11.8	57.1	0	-10.0	86.6
	(AB)3	0.7	-11.6	16.7	1.2	-9.6	11.3
	A3'	0.9	-11.5	12.7	-	-	-
	A3''	1.1	-11.5	8.4	3.1	-9.0	0.5
n=4	A1B1E2	0	-11.7	88.0	0.4	-9.2	12.1
	A2B1C1	1.8	-11.2	4.6	0.1	-9.3	36.8
	(AB)3E1	2.0	-11.1	2.3	0.8	-9.1	5.7
	(AB)3'E1	2.0	-11.1	2.3	0.9	-9.1	5.0
	A4	2.2	-11.0	2.1	0.6	-9.2	7.6
	A3B1	3.2	-11.2	0.4	0	-9.4	23.4
	A3'B1	-	-	-	0.7	-9.1	7.0
n=5	A1B1(CD)3	0	-11.5	39.1	-	-	-
	A1B1C1E2	0.2	-11.4	26.5	2.2	-8.9	0.0
	A1B1C1(E2)'	0.2	-11.4	25.4	2.2	-8.9	0.0
	(AB)3E2	0.9	-11.2	8.6	0	-9.3	45.6
	A4B1	1.8	-11.0	1.0	0.8	-9.2	11.2
	A2B1E2	-	-	-	0.2	-9.3	35.1

^aB3LYP-D/6-311+G(d,p) and ^bB3LYP/6-311+G(d,p) computations.

^c E_{hyd} : energy of the reaction $G + n\text{H}_2\text{O} \rightarrow G(\text{H}_2\text{O})_n$.

^dBoltzmann probabilities (%) at 298 K.

Table 5. B3LYP and B3LYP-D relative (ΔE) and hydration energies (E_{hyd}) (in kcal/mol, ZPVE-corrected values) of the uracil_nH₂O (n=1-5) having a Boltzmann factor higher than 5% at 298 K.

	Structure	our work B3LYP-D ^a			our work B3LYP ^b			Van Mourik ^c	Casaes et al. ^d	Kim et al. ^e	
		ΔE	E_{hyd}^f/n	%B ^g	ΔE	E_{hyd}^f/n	%B ^g	ΔE	ΔE	ΔE	E_{hyd}^f/n
n=1	A1	0	-10.0	91.1	0	-8.3	88.3	0	0	0	-8.5
	B1	1.4	-8.6	8.3	1.4	-6.9	8.1	1.5	1.6	1.5	-6.9
	C1	2.0	-6.3	3.0	2.0	-6.3	3.0	1.9	2.0	2.1	-6.3
	D1	3.2	-6.8	0.5	3.1	-5.3	0.5	3.2	3.6	3.0	-5.4
n=2	A2	0	-10.9	98.1	0	-9.1	97.3	0	0	0	-9.3
	B2	2.8	-9.5	0.9	2.4	-7.9	1.7	1.9	-	2.4	-8.1
	A1B1	2.8	-9.5	0.9	2.8	-7.7	0.7	1.1	2.2	2.9	-7.8
	A1D1	4.8	-8.5	0.0	4.7	-6.8	0.0	3.5	-	4.7	-6.9
	D2	6.3	-7.7	0.0	5.5	-6.3	0.0	5.0	-	-	-
	C2	-	-	-	-	-	-	2.2	2.6	3.5	-7.5
	A1C1	-	-	-	-	-	-	2.2	-	4.1	-7.2
n=3	A2B1	0	-10.3	53.1	0.8	-8.5	13.6	0	1.5	0	-8.6
	A3	0.2	-10.3	34.3	0	-8.7	56.2	-	0	-	-
	A'3	1.0	-10.0	5.2	0.5	-8.6	23.1	0.3	-	-	-
	A1B2	1.7	-9.9	3.1	1.9	-8.1	2.3	0.8	-	1.0	-8.3
	A2C1	2.0	-9.8	2.0	2.2	-8.0	1.2	-	-	1.6	-8.1
	A2D1	-	-	-	-	-	-	-	-	1.9	-8.0
	(BC)3	-	-	-	-	-	-	1.4	-	2.1	-7.9
	A1C2	-	-	-	3.5	-7.6	0.1	-	-	3.0	-7.6
	B2D1	10.7	-6.1	0.0	5.7	-6.8	0.0	-	-	4.9	-6.9
n=4	A2B2	0	-10.3	57.9	0	-8.5	49.6	0.8	-	0	-8.8
	A4	0.2	-10.2	36.6	0.4	-8.0	23.3	-	-	-	-
	A2C2	1.7	-9.9	3.2	1.9	-7.6	1.9	-	-	2.6	-8.1
	A3B1	5.2	-9.0	0.0	0.7	-8.0	15.9	-	-	-	-
	A1(BC)3	5.5	-8.9	0.0	5.0	-7.0	0.0	-	-	3.1	-8.0
	A3'B1	-	-	-	1.2	-7.8	7.0	0	-	-	-
	A2(BC)3	0	-10.1	68.1	1.1	-7.9	11.8	-	-	0	-8.3
n=5	A4B1	0.7	-10.0	18.3	1.0	-8.0	12.8	-	-	-	-
	A2C3	0.9	-9.9	8.4	-	-	-	-	-	-	-
	A2B3	9.8	-8.0	0.0	0	-8.3	66.9	-	-	-	-
	A2B2D1	3.7	-9.4	0.0	3.1	-7.6	0.3	-	-	2.1	-7.9
	A2C2D1	-	-	-	-	-	-	-	-	3.7	-7.6
	A1(BC)3D1	-	-	-	-	-	-	-	-	5.0	-7.3

^aB3LYP-D/6-311+G(d,p) and ^bB3LYP/6-311+G(d,p) calculations.

^cMP2/DMA computations¹¹

^dMP2/DZPi computations³

^eB3LYP/DZP++ computations²

^f E_{hyd} : energy of the reaction $U + nH_2O \rightarrow U(H_2O)_n$.

^gBoltzmann probabilities (%) at 298 K.

Figure caption

Figure 1. Number of PM6 and B3LYP/6-311+G(d,p) optimized structures generated with the GSAM algorithm for [A.nH₂O]. [T.nH₂O]. [C.nH₂O]. [G.nH₂O]. [U.nH₂O] (n=1-5).

Figure 2 : Structures of the five nucleic acid bases. A, B, C, D and E indicate the water primary binding sites in the molecules.

Figure 3. Physically reasonable structures of the adenine_nH₂O (n=1-5) obtained at the B3LYP-D/6-311+G(d,p) level of theory with their corresponding Boltzman factor.

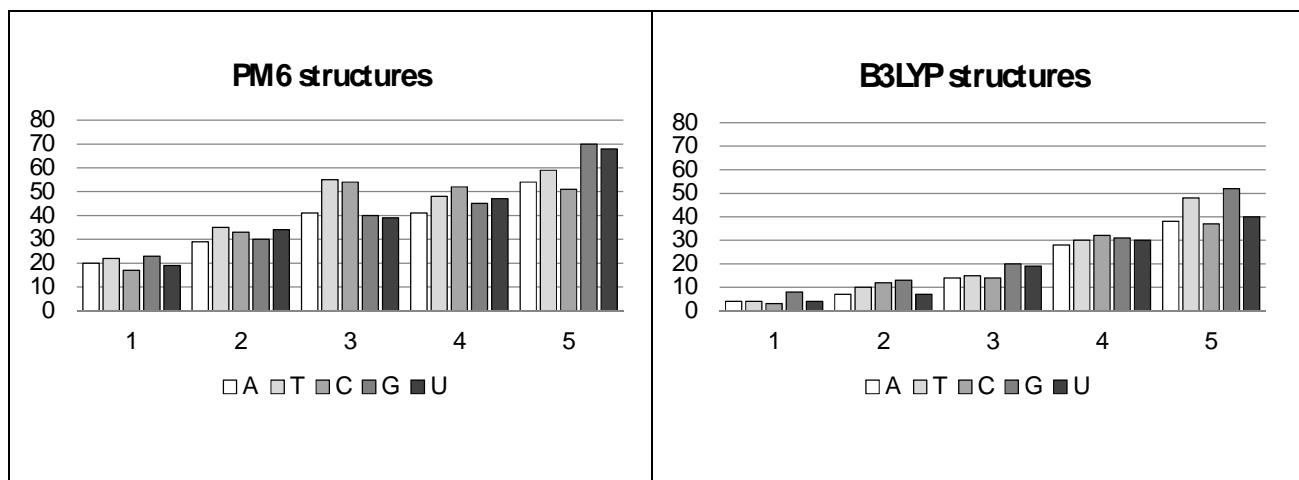
Figure 4. Physically reasonable structures of the thymine_nH₂O (n=1-5) obtained at the B3LYP-D/6-311+G(d,p) level of theory with their corresponding Boltzman factor.

Figure 5. Physically reasonable structures of the cytosine_nH₂O (n=1-5) obtained at the B3LYP-D/6-311+G(d,p) level of theory with their corresponding Boltzman factor.

Figure 6. Physically reasonable structures of the guanine_nH₂O (n=1-5) obtained at the B3LYP-D/6-311+G(d,p) level of theory with their corresponding Boltzman factor.

Figure 7. Physically reasonable structures of the uracil_nH₂O (n=1-5) obtained at the B3LYP-D/6-311+G(d,p) level of theory with their corresponding Boltzman factor.

Figure 1. Number of PM6 and B3LYP/6-311+G(d,p) optimized structures generated with the



GSAM algorithm for [A.nH₂O]. [T.nH₂O]. [C.nH₂O]. [G.nH₂O]. [U.nH₂O] (n=1.5).

Figure 2 : Structures of the five nucleic acid bases. A, B, C, D and E indicate the water primary binding sites in the molecules.

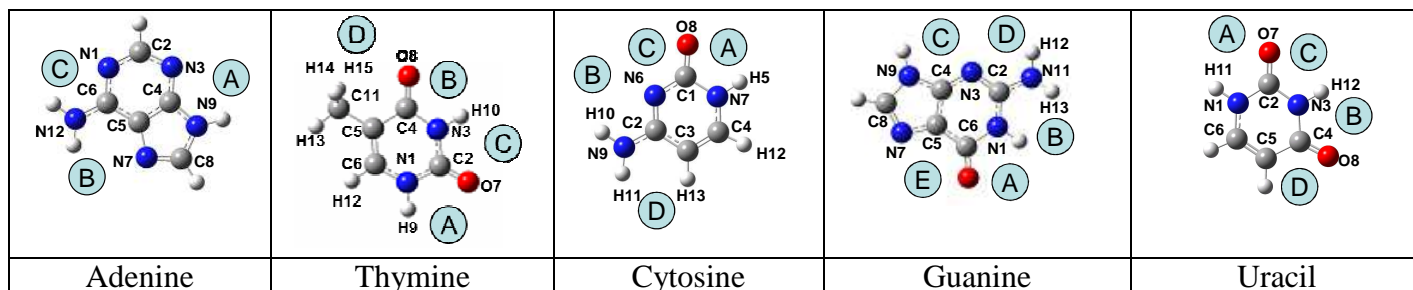


Figure 3. Physically reasonable structures of the adenine_nH₂O (n=1-5) obtained at the B3LYP-D/6-311+G(d,p) level of theory with their corresponding Boltzman factor.

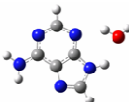



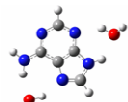
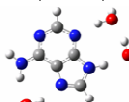


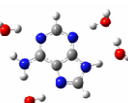
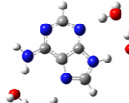

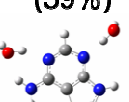
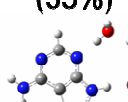


	Structure1	Structure 2	Structure 3	Structure 4
A,1H ₂ O	 (73%)	 (22%)	 (5%)	
A,2H ₂ O	 (93%)	 (7%)		
A,3H ₂ O	 (94%)	 (5%)		
A,4H ₂ O	 (39%)	 (35%)	 (22%)	 (2%)
A,5H ₂ O	 (62%)	 (23%)	 (6%)	 (5%)

Figure 4. Physically reasonable structures of the thymine_nH₂O (n=1-5) obtained at the B3LYP-D/6-311+G(d,p) level of theory with their corresponding Boltzman factor.

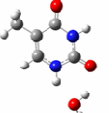
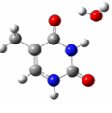
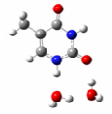
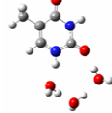
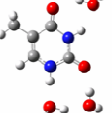
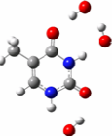
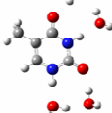
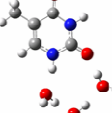
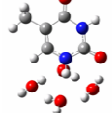

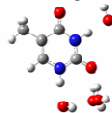
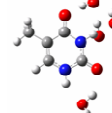
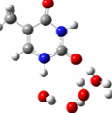
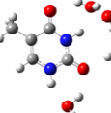
	Structure1	Structure 2	Structure 3	Structure 4	Structure 5
T,1H ₂ O	 (92%)	 (7%)			
T,2H ₂ O	 (99%)				
T,3H ₂ O	 (52%)	 (37%)	 (10%)		
T,4H ₂ O	 (90%)	 (7%)	 (2%)		
T,5H ₂ O	 (67%)	 (14%)	 (13%)	 (5%)	 (5%)

Figure 5. Physically reasonable structures of the cytosine_nH₂O (n=1-5) obtained at the B3LYP-D/6-311+G(d,p) level of theory with their corresponding Boltzman factor.

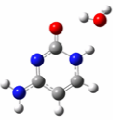

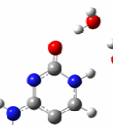

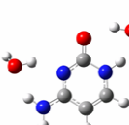
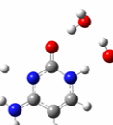
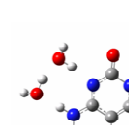
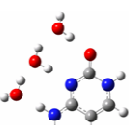

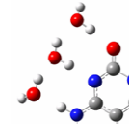

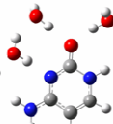

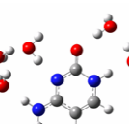

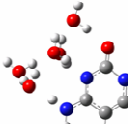
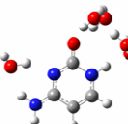
	Structure 1	Structure 2	Structure 3	Structure 4	Structure 5	Structure 6
C,1H ₂ O	 (54%)	 (46%)				
C,2H ₂ O	 (54%)	 (25%)	 (21%)			
C,3H ₂ O	 (60%)	 (31%)	 (8%)			
C,4H ₂ O	 (93%)	 (4%)	 (2%)			
C,5H ₂ O	 (45%)	 (16%)	 (10%)	 (10%)	 (8%)	 (7%)

Figure 6. Physically reasonable structures of the guanine_nH₂O (n=1-5) obtained at the B3LYP-D/6-311+G(d,p) level of theory with their corresponding Boltzman factor.


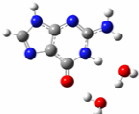

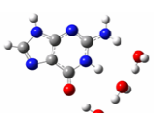
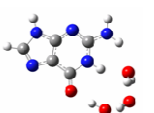
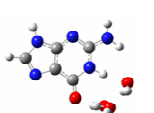

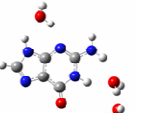
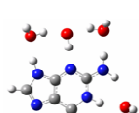
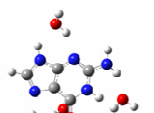
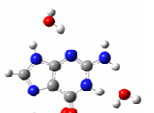
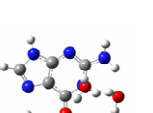

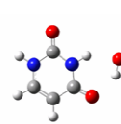
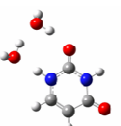

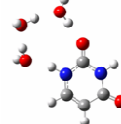
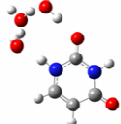

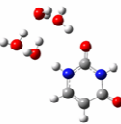

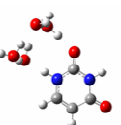
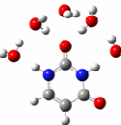
	Structure1	Structure 2	Structure 3	Structure 4
G,1H ₂ O	 (94%)			
G,2H ₂ O	 (100%)			
G,3H ₂ O	 (57%)	 (17%)	 (13%)	 (8%)
G,4H ₂ O	 (90%)	 (5%)		
G,5H ₂ O	 (39%)	 (27%)	 (25%)	 (9%)

Figure 7. Physically reasonable structures of the uracil_nH₂O (n=1-5) obtained at the B3LYP-D/6-311+G(d,p) level of theory with their corresponding Boltzman factor.

	Structure1	Structure 2	Structure 3
U,1H ₂ O	 (91%)	 (8%)	
U,2H ₂ O	 (98%)		
U,3H ₂ O	 (53%)	 (34%)	 (5%)
U,4H ₂ O	 (58%)	 (37%)	
U,5H ₂ O	 (68%)	 (18%)	 (8%)

L'étude des énergies d'hydratation dans le cas des deux méthodes B3LYP et B3LYP-D permet d'aboutir au même constat. Tout d'abord, la valeur d'énergie d'hydratation la plus basse est toujours observé pour $n=2$. L'ajout successif de 2 molécules d'eau permet de stabiliser le système. Le même phénomène est observé pour des taux d'hydratation supérieurs, mais il est moins prononcé puisqu'on note une déstabilisation de l'ordre de 1kcal/mol au maximum entre $n=2$ et $n=5$ pour toutes les bases. Cela paraît logique compte tenu de l'occupation de sites secondaires définis à partir de l'étude de chaque base monohydratée. Le comportement de l'adénine est quantitativement différent. On note certes la même évolution des énergies d'hydratation que pour les autres bases, mais la déstabilisation de l'ensemble se révèle plus rapidement. Ainsi, la différence entre les valeurs de E_{hyd} des formes prépondérantes di- et penta-hydratées est de 0.4 et 1.1kcal/mol respectivement pour la cytosine et l'adénine. La comparaison avec les valeurs E_{hyd} calculées par Kim[57], ou bien Kim et Schaefer[41, 47, 33] montre que les ordres énergétiques des formes les plus probables sont qualitativement équivalents pour la méthode B3LYP. En revanche, la méthode B3LYP-D fournit des résultats différents, qui sont à relier avec les différences observées pour les pourcentages de Boltzmann.

IV Conclusion

Les études structurales menées en parallèle aux niveaux B3LYP et B3LYP-D sur les bases d'acides nucléiques microhydratées nous conduisent à plusieurs conclusions. L'algorithme GSAM s'est révélé être un excellent outil de génération de ces composés. Alors que la littérature présentait des études réalisées *a priori*, cela nous a permis une étude approfondie et générale. Pour $n=1$ et 2, les structures stables décrivant nos systèmes sont identiques : l'utilisation d'une méthode incluant les phénomènes de dispersion de type B3LYP-D, plus coûteuse, n'apparaît pas utile. En revanche, pour des taux d'hydratation plus élevés -3, 4 et 5-, cette méthode a fourni des résultats différents. Ainsi, de nouvelles structures peu ou non favorables en B3LYP le deviennent fortement en B3LYP-D. Ce phénomène est d'autant plus remarquable pour les bases d'ADN bicycliques (A et G). Pour C, T et U, cela est moins prononcé en raison de la présence de sites plus adaptés pour ces interactions (fortes liaisons polarisées du type C=O, associées à des liaisons N-H). En ce qui concerne les énergies d'hydratation par molécule d'eau calculées pour chaque système, le même constat est fait pour les deux méthodes testées. L'ajout d'une molécule d'eau est toujours stabilisant, au moins jusqu'à 5. Mais cette stabilisation est plus prononcée pour $n=2$. On peut supposer que l'encombrement stérique et l'occupation de sites secondaires va défavoriser de plus en plus ces interactions. Globalement, l'utilisation de la DFT-D dépend de la complexité de la molécule hydratée et de ses sites éventuels d'hydratation et du taux considéré. Plus ce dernier est élevé, plus la prise en compte des phénomènes de dispersion semble nécessaire. Cette étude structurale servira de base

à l'étude vibrationnelle des molécules hydratées qui fera l'objet du chapitre 5.

Bibliographie

- [1] Watson, J. *Nature* 1953, 171, 737.
- [2] Feig, M. *Biophysical Journal* 1999, 77, 1769.
- [3] Franklin, R. *Acta Crystallographica* 1953, 6, 673.
- [4] Falk, M.; Hartman, K. A.; Lord, R. C. *J Am Chem Soc* 1963, 85, 387.
- [5] Kopka, M. L.; Fratini, A. V.; Drew, H. R.; Dickerson, R. E. *J Mol Bio* 1983, 163, 129 .
- [6] de Vries, M. S.; Hobza, P. *Annual Review of Physical Chemistry* 2007, 58, 585.
- [7] Roscioli, J.; Pratt, D. *PINAS* 2003, 100, 13572.
- [8] Reha, D.; Valdes, H.; Vondrasek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. *Chem Eur J* 2005, 11, 6803.
- [9] Haber, T.; Seefeld, K.; Engler, G.; Grimme, S.; Kleineremanns, K. *PCCP* 2008, 10, 2844.
- [10] Fricke, H.; Funk, A.; Schrader, T.; Gerhards, M. *J Am Chem Soc* 2008, 130, 4692.
- [11] Clarkson, J.; Baquero, E.; Shubert, V.; Myshakin, E.; Jordan, K.; Zwier, T. *Science* 2005, 307, 1443.
- [12] Becucci, M.; Pietraperzia, G.; Pasquini, M.; Piani, G.; Zoppi, A.; Chelli, R.; Castellucci, E.; Demtroeder, W. *J Chem Phys* 2004, 120, 5601.
- [13] Pasquini, M.; Schiccheri, N.; Piani, G.; Pietraperzia, G.; Becucci, M.; Biczysko, M.; Pavone, M.; Barone, V. *J Phys Chem A* 2007, 111, 12363.
- [14] Biczysko, M.; Piani, G.; Pasquini, M.; Schiccheri, N.; Pietraperzia, G.; Becucci, M.; Pavone, M.; Barone, V. *J Chem Phys* 2007, 127, 144303.
- [15] Schmitt, M.; Böhm, M.; Ratzner, C.; Vu, C.; Kalkman, I.; Meerts, W. L. *J Am Chem Soc* 2005, 127, 10356.
- [16] Lamba, O.; Wang, A.; Thomas Jr, G. *Biopolymers* 1989, 28, 667.
- [17] Wittlin, A.; Genzel, L.; Kremer, F.; Haeseler, S.; Poglitsch, A.; Rupprecht, A. *Phys Rev A* 1986, 34, 493.
- [18] Pohle, W.; Bohl, M.; Boehlig, H. *J Mol Struct* 1991, 242, 333.
- [19] Cammi, R.; Tomasi, J. *J of Computational Chemistry* 1995, 16, 1449.
- [20] Tomasi, J.; Persico, M. *Chem Rev* 1994, 94, 2027.

-
- [21] Miertuš, S. ; Scrocco, E. ; Tomasi, J. *Chem Phys* 1981, 55, 117.
- [22] Cancès, E. ; Mennucci, B. ; Tomasi, J. *J Chem Phys* 1997, 107, 3032.
- [23] Stewart, R. F. ; Jensen, L. H. *Acta Crystallographica* 1967, 23, 1102.
- [24] Ferenczy, G. ; Harsányi, L. ; Rozsondai, B. ; Hargittai, I. *J Mol Struct* 1986, 140, 71 .
- [25] Zhang, R. B. ; Zeegers-Huyskens, T. ; Ceulemans, A. ; Nguyen, M. T. *Chem Phys* 2005, 316, 35.
- [26] Blicharska, B. ; Kupka, T. *J Molec Struct* 2002, 613, 153.
- [27] Palafox, M. ; Iza, N. ; Gil, M. *J Molec Struct : THEOCHEM* 2002, 585, 69.
- [28] Gaigeot, M. ; Kadri, C. ; Ghomi, M. *J Molec Struct* 2001, 565–566, 469.
- [29] van Mourik, T. ; Price, S. L. ; Clary, D. C. *J Phys Chem A* 1999, 103, 1611.
- [30] Brancato, G. ; Rega, N. ; Barone, V. *Chem Phys Lett* 2010, 500, 104 .
- [31] Casaes, R. N. ; Paul, J. B. ; McLaughlin, R. P. ; Saykally, R. J. ; van Mourik, T. *J Phys Chem A* 2004, 108, 10989.
- [32] Gaigeot, M. ; Ghomi, M. *J Phys Chem B* 2001, 105, 5007.
- [33] Kim, S. ; Schaefer, H. F. *J Chem Phys* 2006, 125, 144305.
- [34] Gaigeot, M. ; Sprik, M. *J Phys Chem B* 2003, 107, 10344.
- [35] Danilov, V. I. ; van Mourik, T. ; Poltev, V. I. *Chem Phys Lett* 2006, 429, 255.
- [36] Chandra, A. K. ; Nguyen, M. T. ; Zeegers-Huyskens, T. *J Phys Chem A* 1998, 102, 6010.
- [37] Yarasi, S. ; Billingham, B. E. ; Loppnow, G. R. *J of Raman Spectroscopy* 2007, 38, 1117.
- [38] Graindourze, M. ; Grootaers, T. ; Smets, J. ; Zeegers-Huyskens, T. ; Maes, G. *J Molec Struct* 1991, 243, 37 .
- [39] Gerdil, R. *Acta Crystallographica* 1961, 14, 333.
- [40] Rejnek, J. ; Hanus, M. ; Kabelac, M. ; Ryjacek, F. ; Hobza, P. *PCCP* 2005, 7, 2006.
- [41] Kim, S. ; Wheeler, S. E. ; Schaefer, H. F. ; III *J Chem Phys* 2006, 124, 204310.
- [42] van Mourik, T. ; Benoit, D. M. ; Price, S. L. ; Clary, D. C. *PCCP* 2000, 2, 1281.
- [43] Barker, D. L. ; Marsh, R. E. *Acta Crystallographica* 1964, 17, 1581.
- [44] Chandra, A. K. ; Michalska, D. ; Wysokiński, R. ; Zeegers-Huyskens, T. *J Phys Chem A* 2004, 108, 9593.
- [45] Chandra, A. ; Nguyen, M. ; Zeegers-Huyskens, T. *J Molec Struct* 2000, 519, 1.
- [46] Shishkin, O. V. ; Gorb, L. ; Leszczynski, J. *J Phys Chem B* 2000, 104, 5357.
- [47] Kim, S. ; Schaefer, H. F. *J Chem Phys* 2007, 126, 064301.

-
- [48] Hunter, K. C. ; Wetmore, S. D. *Chem Phys Lett* 2006, 422, 500 .
- [49] Van Mourik, T. ; Danilov, V. I. ; Gonzalez, E. ; Deriabina, A. ; Poltev, V. I. *Chem Phys Lett* 2007, 445, 303.
- [50] Alemán, C. *Chem Phys* 1999, 244, 151.
- [51] Mohamed, T. A. ; Shabaan, I. A. ; Zoghaib, W. M. ; Husband, J. ; Farag, R. S. ; Alajhaz, A. E.-N. M. *J Molec Struct* 2009, 938, 263 .
- [52] Rasheed, T. ; Ahmad, S. *Vibrational Spectroscopy* 2011, 56, 51.
- [53] Zierkiewicz, W. ; Komorowski, L. ; Michalska, D. ; Cerny, J. ; Hobza, P. *J Phys Chem B* 2008, 112, 16734.
- [54] Biczysko, M. ; Panek, P. ; Barone, V. *Chem Phys Lett* 2009, 475, 105.
- [55] Jalbout, A. F. ; Adamowicz, L. *J Phys Chem A* 2001, 105, 1033.
- [56] Sukhanov, O. S. ; Shishkin, O. V. ; Gorb, L. ; Podolyan, Y. ; Leszczynski, J. *J Phys Chem B* 2003, 107, 2846.
- [57] Kim, H. *J Molec Struct : THEOCHEM* 2004, 673, 121.
- [58] Crews, B. ; Abo-Riziq, A. ; Grace, L. ; Callahan, M. ; Kabelac, M. ; Hobza, P. ; de Vries, M. S. *PCCP* 2005, 7, 3015.
- [59] Hanus, M. ; Ryjáček, F. ; Kabelác, M. ; Kubar, T. ; Bogdan, T. V. ; Trygubenko, S. A. ; Hobza, P. *J Am Chem Soc* 2003, 125, 7678.
- [60] Latajka, Z. ; Person, W. ; Morokuma, K. *J Molec Struct : THEOCHEM* 1986, 135, 253.
- [61] Gould, I. R. ; Vincent, M. A. ; Hillier, I. H. *Spectrochimica Acta Part A : Molecular Spectroscopy* 1993, 49, 1727.
- [62] Thewalt, U. ; Bugg, C. E. ; Marsh, R. E. *Acta Crystallographica Section B* 1971, 27, 2358.
- [63] Guille, K. ; Clegg, W. *Acta Crystallographica Section C* 2006, 62, o515.
- [64] Shukla, M. K. ; Leszczynski, J. *J Phys Chem B* 2005, 109, 17333.
- [65] Gorb, L. ; Leszczynski, J. *IJQC* 1997, 65, 759.
- [66] Chandra, A. ; Nguyen, M. T. ; Uchimaru, T. ; Zeegers-Huyskens, T. *J Molec Struct* 2000, 555, 61 .
- [67] Shukla, M. K. ; Leszczynski, J. *J Phys Chem B* 2008, 112, 5139.

Chapitre 5

Étude de molécules hydratées : cas d'un système organique, spectres vibrationnels des bases d'acides nucléiques

Sommaire

I	Choix du modèle de représentation des bases hydratées	131
I.1	Cas de la cytosine	131
I.2	Un cas comparable à la cytosine : étude de la thymine	137
I.3	Validation sur les autres bases dans la zone $1000-1800\text{cm}^{-1}$	141
II	Choix du traitement vibrationnel	146
II.1	Cas de petites molécules azotées	147
II.2	Cas des bases d'acides nucléiques : zone $1000-1800\text{cm}^{-1}$	148
III	Conclusion et Perspectives	149
	Bibliographie	150

L'objectif est de calculer les spectres vibrationnels de chacune des bases présentées ci-dessus. Les résultats obtenus seront comparés aux données expérimentales disponibles dans la littérature. Les spectres

des molécules isolées et microhydratées vont être simulés et comparés aux données issues d'expériences réalisées en matrice d'Argon (ou bien en phase gazeuse) ou en phase aqueuse. Il faut noter que le comportement en matrice d'Ar est quasi-identique à celui observé en phase aqueuse, comme le montre les données expérimentales concernant l'adénine, pour laquelle l'écart moyen sur les fréquences obtenu dans les deux conditions approche les 5cm^{-1} [1]. La détermination théorique des spectres IR de systèmes en solution n'est pas aisée, dans le sens où la qualité des propriétés vibrationnelles obtenues dépendent du modèle de représentation du solvant. Deux types de modèles peuvent être envisagés : soit un modèle implicite, pour lequel le solvant est représenté par un continuum polarisable, soit un modèle explicite pour lequel un nombre connu de molécules d'eau sera considéré autour du soluté. Pour l'étude des bases d'acides nucléiques, un modèle mixte sera utilisé. La première étape consiste donc à déterminer quelle forme microhydratée (représentée par $X, n\text{H}_2\text{O}$) sera adaptée pour la partie explicite nécessaire à l'étude de chaque base d'acide nucléique X en solution.

I Choix du modèle de représentation des bases hydratées

I.1 Cas de la cytosine

I.1.1 Détermination du modèle géométrique

La cytosine est la base d'acide nucléique qui comporte le moins d'atomes (13). Elle est à ce titre étudiée théoriquement de façon privilégiée et constitue donc une molécule test pour cette famille de molécules.

Le premier écueil dans la simulation des bases d'acides nucléiques en phase aqueuse réside dans la détermination d'un modèle géométrique capable de reproduire correctement les effets du solvant sur le soluté. Plusieurs modèles sont envisageables :

- * soit un modèle purement implicite, de type PCM par exemple (noté $C, 0\text{H}_2\text{O} + \text{PCM}$) ;
- * soit un modèle décrivant le phénomène de microhydratation de façon explicite, c'est-à-dire qui considère un nombre entier de molécules d'eau à déterminer (noté $C, n\text{H}_2\text{O}$) ;
- * soit enfin une combinaison des deux modèles précédemment cités (notés $C, n\text{H}_2\text{O} + \text{PCM}$).

La partie explicite du modèle concernant la cytosine en solution aqueuse a été déduite de l'étude des propriétés structurales des systèmes $C, n\text{H}_2\text{O}$ -avec n allant de 1 à 5-, à l'aide du code GSAM. Ainsi, les formes les plus stables au niveau B3LYP issues de cette analyse ont constitué plusieurs possibilités de modèle géométrique. Le modèle PCM a aussi été considéré, d'une part seul et d'autre part associé à la cytosine tétrahydratée.

Le choix du modèle géométrique le plus approprié repose sur l'étude des paramètres géométriques des

différents modèles évoqués et sur un modèle de référence. Ce dernier est issu d'une étude QM/MM, basée sur une molécule de cytosine, considérée au niveau B3LYP/6-31G* et solvatée par 267 molécules d'eau traitées avec un champ de force de type AMBER[2]. Un système comprenant la cytosine et ses 25 plus proches molécules de solvant en a été tiré pour être totalement réoptimisé avec la méthode B3LYP/6-31+G(d,p) : il constituera la représentation la plus proche de la cytosine solvatée. La procédure détaillée de ce travail réalisé par X. Assfeld et T. Véry est décrite dans la publication placée en annexe[3]. La comparaison des paramètres géométriques remarquables de la cytosine -*i.e.* les longueurs C=O, NH et CH- dans le cadre des modèles C,1H₂O, C,2H₂O, C,3H₂O, C,4H₂O, C,0H₂O+PCM et C,4H₂O+PCM est reportée dans le tableau 5.1. Les quatre premiers modèles traitent le solvant de façon explicite avec un nombre de molécules d'eau augmentant progressivement, le modèle suivant correspond au traitement implicite et le dernier correspond à la combinaison des deux modèles.

Paramètres	C,1H ₂ O	C,2H ₂ O	C,3H ₂ O	C,4H ₂ O	C,4H ₂ O+PCM	C+PCM	C,25H ₂ O
C ₍₁₎ O ₍₈₎	1.230	1.233	1.234	1.244	1.253	1.243	1.254
N ₍₇₎ H ₍₅₎	1.021	1.029	1.029	1.032	1.031	1.012	1.030
N ₍₉₎ H ₍₁₀₎	1.007	1.007	1.008	1.018	1.024	1.009	1.034
C ₍₃₎ H ₍₁₃₎	1.080	1.081	1.080	1.082	1.082	1.082	1.079
C ₍₄₎ H ₍₁₂₎	1.083	1.083	1.083	1.085	1.084	1.084	1.083
$\phi(\text{NH}_2)$	168.852	170.22	177.05	173.2	179.3	180.0	172.5

FIGURE 5.1 – Paramètres géométriques de la cytosine obtenus au niveau B3LYP/6-31+G(d,p) selon le modèle de représentation de la solvatation considéré

Les liaisons CH sont très peu sensibles à l'évolution de la complexité du modèle d'hydratation. En revanche, les longueurs des liaisons NH évolue notablement : plus le modèle se complexifie, meilleure est la correspondance avec les valeurs de référence. Les plus faibles écarts sont obtenus pour le système C,4H₂O+PCM. On peut enfin noter que le modèle PCM conduit à un aplatissement de la molécule de soluté : le dièdre $\phi(\text{NH}_2)$ devient proche de 180 degrés.

Globalement, la partie explicite du modèle géométrique adapté est constituée de la forme microhydratée de la cytosine possédant tous ses sites d'hydratation favoris occupés (décrits dans le chapitre précédent) et dont la probabilité de Boltzmann à 298K est suffisamment élevée pour être représentative du taux d'hydratation considéré. Les modèles représentatifs des autres bases d'acides nucléiques en solution aqueuse seront donc ainsi déterminées.

I.1.2 Validation du modèle géométrique : calcul de la signature vibrationnelle

Il est intéressant de confirmer la qualité de notre modèle géométrique de la cytosine en solution aqueuse en analysant les propriétés vibrationnelles obtenues. Pour se faire, les nombres d'onde anharmoniques situés dans la zone $1000-1800\text{cm}^{-1}$ qui constitue la signature vibrationnelle du système ont été déterminés par dynamique moléculaire pour les modèles suivants : C,0H₂O+PCM, C,4H₂O et C,4H₂O+PCM. Nous nous sommes ainsi affranchis du problème éventuel constitué par la forme du potentiel. Les différences (Δ)(en cm^{-1}) entre ces valeurs et les données expérimentales[4] sont reportées dans le tableau 5.2 en fonction du modèle considéré. Les valeurs absolues des moyennes de ces écarts notées Δ_{moy} sont aussi reportées pour chaque modèle.

N°	C,0H ₂ O+PCM	(Δ)	C,4H ₂ O	(Δ)	C,4H ₂ O+PCM	(Δ)	Exp
$\delta(\text{NH}_2)\text{rock}$	1080	-26	1090	-16	1102	-4	1106
$(\delta(\text{C}_{(3)}\text{H},\delta(\text{C}_{(4)}\text{H}))\text{scis}$	1090	-48	1123	-15	1127	-11	1138
$\nu_r \approx \text{C}_{(4)}-\text{N}_{(7)}$	1190	-34	1199	-25	1218	-6	1224
$\nu_r \approx \text{C}_{(1)}-\text{N}_{(6)}$	1237	-52	1246	-43	1259	-30	1289
$(\delta(\text{C}_{(3)}\text{H},\delta(\text{C}_{(4)}\text{H}))\text{rock}$	1320	-43	1326	-37	1336	-27	1363
$\delta(\text{NH})$	1390	-48	1413	-25	1423	-15	1438
$\nu_r \approx \text{C}_{(2)}-\text{N}_{(9)}$	1430	-74	1457	-47	1496	-8	1504
$\nu_r \approx \text{C}_{(2)}-\text{C}_{(3)}$	1503	-25	1507	-21	1514	-14	1528
$\delta(\text{NH}_2)\text{scis}$	1580	-36	1645	29	1628	12	1616
$\nu_r \approx \text{C}_{(3)}=\text{C}_{(4)}$	1607	-54	1609	-52	1637	-24	1661
Δ_{moy}		45		25		15	

FIGURE 5.2 – Fréquences anharmoniques DM entre 100 et 1800cm^{-1} obtenues pour différents modèles représentant l'uracile en solution aqueuse à un niveau B3LYP/6-31+G(d,p) - Calculs des écarts (Δ) entre théorie et expérience

On note que c'est bien le troisième modèle, associant les modèles implicite et explicite du solvant, qui fournit la différence la plus faible (15cm^{-1}), contre 45 et 25cm^{-1} pour chaque modèle considéré séparément. Notre hypothèse est bien validée et peut être appliquée aux bases d'acides nucléiques hydratées.

I.1.3 Méthode de calcul de structure électronique

D'un point de vue géométrique, il a été précédemment remarqué[5, 6] que la méthode B3LYP-D fournissait de meilleurs résultats que sa forme la plus simple B3LYP pour les systèmes présentant des interactions à longue portée tels que les bases d'acides nucléiques microhydratées. Cependant, le calcul des nombres d'onde des modes de vibration par une approche DFT a souvent fourni des résultats en accord avec les valeurs expérimentales[7]. On peut donc envisager une étude vibrationnelle de la cytosine microhydratée à l'aide de ces 2 niveaux calculatoires, tout comme celle de sa forme isolée. La comparaison entre chaque nombre d'onde B3LYP et B3LYP-D, associée aux données expérimentales disponibles dans la littérature, réalisées soit en matrice d'Ar[8, 9], soit en phase aqueuse[4], va être menée. Dans un premier temps, seuls des calculs statiques et dynamiques ont été effectués. Le tableau C présente les écarts B3LYP/B3LYP-D $\Delta\omega$ et $\Delta\nu$ en cm^{-1} entre les fréquences harmoniques et anharmoniques respectivement en fonction du modèle considéré. $\Delta\nu_{\text{VPT2(B3LYP(D)exp)}}$ représente l'écart en valeur absolue calculée entre les valeurs anharmoniques de la méthode calculatoire considérée et ses contreparties expérimentales. On note enfin les valeurs remarquables Δ_{moy} et Δ_{max} , correspondant respectivement aux écarts moyen et maximum observés, après avoir omis les valeurs aberrantes signalées en italique dans le tableau C.

désignation du mode	C ₀ H ₂ O				C ₄ H ₂ O+PCM			
	$\Delta\omega$	$\Delta\nu$	$\Delta\nu$ VPT2(B3LYP/exp)	$\Delta\nu$ DM(B3LYP/exp)	$\Delta\omega$	$\Delta\nu$	$\Delta\nu$ VPT2(B3LYP/exp)	$\Delta\nu$ DM(B3LYP/exp)
$\nu(\text{NH}_2)_{\text{as}}$	3	-6	77	83	10	-7		
$\nu(\text{NH})$	2	4	9	13	52	-44		
$\nu(\text{NH}_2)_{\text{s}}$	10	4	60	56	56	25		
$\nu(\text{CH})_{\text{s}}$	7	8			6	-14	1	32
$\nu(\text{CH})_{\text{as}}$	6	7			5	38		
$\nu(\text{C}=\text{O}), \delta(\text{NH})$	0	4	18	14	32	-25		
$\nu_2 \approx \text{C}_{(17)}=\text{C}_{(4)}, \nu_{17} \approx \text{N}_{(6)}$	1	3	5	8	-1	-63	11	24
$\delta(\text{NH}_2)_{\text{scis}}$	0	5	2	7	-40	-41	39	12
$\nu_2 \approx \text{C}_{(12)}-\text{C}_{(3)}, \nu_{12} \approx \text{N}_{(6)}, \delta(\text{NH})$	2	6	14	20	-1	-17	8	14
$\nu_2 \approx \text{C}_{(27)}\text{N}_{(9)}, \delta(\text{NH}_2)_{\text{scis}}, [\delta(\text{C}_{(3)}\text{H}), \delta(\text{C}_{(4)}\text{H}), \delta(\text{C}_{(4)}\text{H})]_{\text{rock}}$	4	1	3	2	-1	-33	4	8
$\delta(\text{NH}), \nu_2 \approx \text{C}_{(27)}-\text{C}_{(3)}$	2	4	13	14	-4	-37	11	15
$[\delta(\text{C}_{(3)}\text{H}), \delta(\text{C}_{(4)}\text{H})]_{\text{rock}}$	3	3	2	1	1	-29	0	27
$\nu_2 \approx \text{C}_{(17)}-\text{N}_{(6)}, \nu_{17} \approx \text{N}_{(6)}$	3	2			-2	-3	14	40
$[\delta(\text{C}_{(4)}\text{H}), \delta(\text{N}_{(7)}\text{H})]_{\text{scis}}, [\delta(\text{C}_{(3)}\text{H}), \delta(\text{C}_{(4)}\text{H})]_{\text{scis}}$	11	10	43	53	16	-18	11	6
$[\delta(\text{C}_{(3)}\text{H}), \delta(\text{N}_{(7)}\text{H})]_{\text{scis}}$	8	6			1	-32	7	33
$\delta(\text{NH}_2)_{\text{rock}}, \nu_2 \approx \text{C}_{(17)}-\text{N}_{(7)}$	-1	12			0	-104	8	12
$\nu_2 \approx \text{C}_{(27)}-\text{N}_{(3)}$	3	4			-5	35	17	15
$\gamma(\text{CH})_{\text{as}}$	0	0			-3	-8	21	23
$\nu_2 \approx \text{C}_{(17)}-\text{N}_{(7)}, \nu_{17} \approx \text{C}_{(3)}, \delta(\text{NH}_2)_{\text{rock}}$	0	4			-1	-8	21	20
$\gamma(\text{CH})_{\text{s}}$	0	1			-11	-15	32	27
$\nu_2 \approx \text{C}_{(17)}-\text{N}_{(7)}, \nu_{17} \approx \text{C}_{(3)}, \nu_{17} \approx \text{N}_{(6)}, \nu_{17} \approx \text{C}_{(27)}-\text{N}_{(9)}$	3	7			-8	18	14	4
$\gamma(\text{C}=\text{O}), \gamma(\text{C}_{(3)}-\text{H})$	0	0			-20	-2		
$\gamma(\text{C}_{(3)}-\text{H})$	1	2			-36	-128		
$\gamma(\text{NH})$	3	3			-21	-53		
ring deform	1	0			-2	17	36	6
ring deform	-1	7			-2	-39		
ring deform	-2	1			-6	-52		
tor(C ₍₂₇₎ -NH ₂)	-2	-11			-46	-120		
ring deform	1	2			-5	-465	15	450
$\delta(\text{C}_{(27)}-\text{NH}_2)$	-5	1			-110	-381	43	424
ring deform	1	-12			1	-45		
$\gamma(\text{NH}_2)$	25	975			-13	-344		
ring deform	0	3			0	-59		
Amoy			22	25			30	14
Δmax			77	83			87	39
								19
								43
								26
								59

Tableau C : Ecart (en cm^{-1}) entre les valeurs harmoniques ($\Delta\omega$) et anharmoniques ($\Delta\nu$) obtenues pour la cytosine isolée et hydratée aux niveaux VPT2 et DM dans le cadre des méthodes B3LYP et B3LYP-D/6-31+G(d,p)

Comparaison avec les données expérimentales (en italique : valeurs non considérées dans le calcul de Δ moy)

La comparaison des valeurs harmoniques de la cytosine isolée montre que les deux méthodes de calcul de la structure électronique conduisent à des résultats identiques. Cela paraît cohérent compte tenu de l'absence de liaisons hydrogène. La prise en compte de l'hydratation ne modifie que certaines valeurs, relatives aux modes impliqués dans les liaisons hydrogène entre le soluté et le solvant. On peut noter que les valeurs de Δ_{moy} B3LYP et B3LYP-D de la cytosine isolée avec l'approche VPT2 ne permettent pas de conclure sur le choix d'une méthode en particulier, compte tenu de leur écart peu prononcé (3cm^{-1}). Mais la comparaison avec l'expérience pour la forme hydratée nous conduirait à favoriser la méthode B3LYP-D. En effet, les valeurs de Δ_{moy} et Δ_{max} pour le système hydraté (égales à 14 et 39cm^{-1}) dans le cadre de l'approche VPT2 sont inférieures à celles obtenues avec l'autre méthode (30 et 87cm^{-1} respectivement). Cependant, l'étude des valeurs de $\Delta\nu$ permet de relativiser ces observations : établir une conclusion précise serait inopportun. En effet, il apparaît que les valeurs de $\Delta\nu$ sont très dépendantes de la qualité du potentiel. Par exemple, l'écart de 128cm^{-1} observé pour le mode $\gamma(C_{(3)}H)$ est dû à l'absurdité de la valeur anharmonique B3LYP ($\nu=630\text{cm}^{-1}$ pour $\omega=738\text{cm}^{-1}$). Le potentiel considéré à l'ordre 4 n'est pas assez flexible, entraînant des surestimations de nombreux couplages intermodes : on peut citer à titre d'illustration les valeurs des constantes d'ordre 4 B3LYP-D $k_{69,54,54,54}$, $k_{69,54,54,52}$ et $k_{69,52,52,52}$ couplant des modes bas et qui valent respectivement -2127 , -2070 et -2791cm^{-1} . Ces seules constantes, de part leurs valeurs exagérées, conduisent à des valeurs anharmoniques aberrantes.

Puisque la forme du potentiel constitue un obstacle à la détermination de la bonne méthode de calcul de structure électronique, une approche de dynamique moléculaire est envisagée pour s'affranchir de la forme mathématique du potentiel. Seuls les écarts avec l'expérience notés $\Delta\nu_{DM(B3LYP(D)exp)}$ sont reportées sur le tableau C. Certaines valeurs B3LYP-D ne sont pas présentées car elles n'ont pu être attribuées avec certitude. Pour la cytosine isolée, la dynamique moléculaire fournit des résultats légèrement plus éloignés de l'expérience que la méthode VPT2, mais restent qualitativement équivalents. En ce qui concerne les dynamiques de la forme hydratée, la méthode B3LYP fournit des écarts moyen et maximum entre théorie et expérience plus faibles que ceux calculés pour la méthode incluant la dispersion. On peut ainsi conclure que pour un même modèle géométrique, et en s'affranchissant d'une forme mathématique du potentiel, la méthode B3LYP conduit à des propriétés vibrationnelles les plus en adéquation avec l'expérience.

Afin de confirmer ces premières hypothèses, nous reportons dans le paragraphe suivant l'étude d'un autre système présentant le même modèle représentatif de la molécule en solution avec les deux méthodes DFT et DFT-D : il s'agit de la thymine, dont la partie explicite du modèle géométrique est constituée par $T,4H_2O$.

I.2 Un cas comparable à la cytosine : étude de la thymine

Les études des propriétés structurales menées à l'aide du code GSAM menées sur la thymine microhydratée aux niveaux B3LYP et B3LYP-D conduisent à considérer la forme tétrahydratée comme la partie explicite de notre modèle géométrique. En effet, les pourcentages de Boltzmann sont respectivement de 81 et 90%, pour deux sites les plus favorables occupés. Le modèle géométrique est donc constitué de l'association de T₄H₂O avec le modèle PCM. Un calcul des modes de vibration situés dans la zone de signature vibrationnelle 1000-1800cm⁻¹ permet de valider ou non notre modèle. Cette analyse a été menée à un niveau de traitement anharmonique vibrationnel VPT2 puisque cette zone n'est que raisonnablement affectée par la présence de couplages[10]. Les résultats des trois modèles T₀H₂O+PCM, T₄H₂O et T₄H₂O+PCM sont présentés sur le tableau 5.3.

On constate que l'écart moyen Δ_{moy} est divisé par deux avec le modèle combiné par rapport aux données expérimentales[11, 12, 13], ce qui confirme sa meilleure fiabilité. Dans un second temps, les écarts issus de la même analyse vibrationnelle entre 1000 et 1800cm⁻¹ que celle effectuée sur la cytosine via les méthodes B3LYP et B3LYP-D sont présentés : les valeurs en cm⁻¹ de $\Delta\omega$, $\Delta\nu$ VPT2 et $\Delta\nu$ VPT2(B3LYP(D)exp) associées à Δ_{moy} et Δ_{max} sont ainsi reportées sur la figure 5.4 dans les cas isolé et hydraté.

Description du mode	T,0H ₂ O+PCM	(Δ)	T,4H ₂ O	(Δ)	T,4H ₂ O+PCM	(Δ)	Exp
$\nu(\text{C}_{(2)}=\text{O})$	1698	6	1728	-24	1666	38	1704
$\nu(\text{C}_{(4)}=\text{O})$	1648	16	1688	-24	1645	19	1664
$\delta(\text{N}_{(1)}\text{H}),\nu(\text{C}_{(2)}\text{N}_{(3)})$	1425	76	1578	-80	1546	-48	1498
sci(Me), $\nu(\text{N}_{(1)}\text{C}_{(6)})$	1468	-26	1468	-28	1437	5	1442
o sci(Me)	1426	-6	1448	-28	1430	-10	1420
umb(Me)	1400	-16	1414	-30	1381	3	1384
$\delta(\text{C}_{(6)}\text{H}),\delta(\text{N}_{(3)}\text{H}),\nu(\text{C}_{(5)}\text{C}_{(6)})$	1410	-58	1355	-3	1348	4	1352
$\delta(\text{C}_{(6)}\text{H}),\delta(\text{N}_{(1)}\text{H}),$ $\nu(\text{N}_{(1)}\text{C}_{(6)}),$ $\nu(\text{N}_{(3)}\text{C}_{(4)})$	1272	-56	1239	-23	1212	4	1216
$\nu(\text{N}_{(3)}\text{C}_{(4)}),\nu(\text{N}_{(1)}\text{C}_{(6)}),\text{rot}(\text{Me})$	1161	5	1164	2	1159	7	1166
rot(Me),ring def 1	1007	15	1014	8	1007	15	1022
$\nu(\text{N}_{(1)}\text{C}_{(2)}),\text{rot}(\text{Me}),\nu(\text{C}_{(2)}\text{N}_{(3)})$	991	-5	976	10	971	15	986
ring def 1, $\nu(\text{C}_{(5)}-\text{Me})$	792	22	799	15	796	18	814
$\gamma(\text{C}_{(4)}=\text{O})$	758	-6	756	-4	762	-10	752
$\nu(\text{C}_{(4)}\text{C}_{(5)}),\nu(\text{C}_{(5)}-\text{Me})$	730	-90	737	-117	610	10	620
ring def	540	20	551	9	544	16	560
ring def	459	21	468	22	470	10	480
Δ_{moy}		28		25		14	

FIGURE 5.3 – Fréquences anharmoniques VPT2 entre 100 et 1800cm⁻¹ obtenues pour différents modèles représentant la thymine en solution aqueuse à un niveau B3LYP/6-31+G(d,p) - Calculs des écarts (Δ) entre théorie et expérience

Désignation du mode	T,0H ₂ O				
	$\Delta\omega$	$\Delta\nu$	$\Delta\nu$ VPT2 (B3LYPexp)	$\Delta\nu$ VPT2 (B3LYPDexp)	$\Delta\nu$ DM (B3LYPexp)
$\nu(\text{N}_{(1)}\text{-C}_{(2)}), \text{rot}(\text{Me}), \nu(\text{C}_{(2)}\text{N}_{(3)})$	1	1	16	17	9
rot(Me), déformation cycle 1	1	-1	2	3	1
rot(Me)as	4	2	2	0	3
$\nu(\text{N}_{(3)}, \text{C}_{(4)}), \nu(\text{N}_{(1)}\text{C}_{(6)}), \text{rot}(\text{Me})$	4	4	10	14	22
$(\delta(\text{C}_{(6)}\text{H}), \delta(\text{N}_{(1)}\text{H}))\text{as}, \nu(\text{N}_{(1)}\text{C}_{(6)}),$ $\nu(\text{C}_{(2)}\text{N}_{(3)}), \nu(\text{N}_{(3)}\text{C}_{(4)})$	7	4	6	10	13
$\nu(\text{C}_{(5)}\text{-Me}), \nu(\text{N}_{(1)}\text{-C}_{(6)}),$ déformation cycle 1	3	5	20	25	35
$\delta(\text{C}_{(6)}\text{H}), \delta(\text{N}_{(3)}\text{H}), \nu(\text{C}_{(5)}\text{=C}_{(6)})$	7	2	11	13	21
$\delta(\text{N}_{(3)}\text{H}), \delta(\text{C}_{(6)}\text{H})$	-4	0	1	1	10
$\delta(\text{N}_{(1)}\text{H}), \nu(\text{C}_{(2)}\text{N}_{(3)}), \nu(\text{C}_{(4)}\text{C}_{(5)})$	5	0	3	3	13
umb(Me)	5	22	9	13	11
o sci(Me)	13	3	15	12	16
$\delta(\text{N}_{(1)}\text{H}), \text{sci}(\text{Me}), \nu(\text{N}_{(1)}\text{C}_{(6)})$	8	-11	4	15	2
sci(Me), $\nu(\text{N}_{(1)}\text{C}_{(6)}), \delta(\text{N}_{(1)}\text{H})$	5	2	8	10	2
$\nu(\text{C}_{(5)}\text{=C}_{(6)})$	1	1	4	5	28
$\nu(\text{C}_{(4)}\text{=O})$	-1	-13	11	24	27
$\nu(\text{C}_{(2)}\text{=O})$	-1	1	5	4	47
Δ_{moy}			8	11	82
Δ_{max}			20	25	47

FIGURE 5.4 – Écarts en cm^{-1} entre les valeurs harmoniques et anharmoniques obtenues pour la thymine isolée et hydratée dans la zone $1000\text{-}1800\text{cm}^{-1}$ aux niveaux VPT2 et DM dans le cadre des méthodes B3LYP et B3LYP-D/6-31+G(d,p)- Comparaison avec les données expérimentales (suite : page suivante)

T,4H ₂ O+PCM							
Désignation du mode	$\Delta\omega$	$\Delta\nu$	$\Delta\nu$ VPT2 (B3LYPexp)	$\Delta\nu$ VPT2 (B3LYPDexp)	$\Delta\nu$ DM (B3LYPexp)	$\Delta\nu$ DM (B3LYPDexp)	
$\nu(\text{N}_{(1)}-\text{C}_{(2)})\dots$	-5	6	15	21	0		
rot(Me)...	-1	3					
rot(Me)as	5	3					
$\nu(\text{N}_{(3)}, \text{C}_{(4)})\dots$	2	-4	7	3	12	24	
$(\delta(\text{C}_{(6)}\text{H}), \delta(\text{N}_{(1)}\text{H}))_{\text{as}}, \nu(\text{CN})$	3	-14	3	11			
$\nu(\text{C}_{(5)}-\text{Me})\dots$	1	4	22	26	43	47	
$\delta(\text{C}_{(6)}\text{H}), \delta(\text{N}_{(3)}\text{H})\dots$	7	9	4	13	2	13	
$\delta(\text{N}_{(3)}\text{H}), \delta(\text{C}_{(6)}\text{H})$	-79	-70	3	67	26	61	
$\delta(\text{N}_{(1)}\text{H}), \nu(\text{C}_{(2)}\text{N}_{(3)})\dots$	-1	2					
umb(Me)	53	33	10	23	22		
o sci(Me)	31	17	5	22	19		
$\delta(\text{N}_{(1)}\text{H}), \text{sci}(\text{Me})\dots$	-66	-176					
sci(Me), $\nu(\text{N}_{(1)}\text{C}_{(6)})\dots$	87	97	48	49	48		
$\nu(\text{C}_{(5)}=\text{C}_{(6)})$	-15	-7	19	12	29	52	
$\nu(\text{C}_{(4)}=\text{O})$	13	2					
$\nu(\text{C}_{(2)}=\text{O})$	-2	-6	38	32	50	52	
Δ_{moy}			16	25	24	42	
Δ_{max}			48	67	50	52	

Tout comme la cytosine isolée, les écarts harmoniques $\Delta\omega$ et anharmoniques $\Delta\nu$ VPT2 et DM pour T_0H_2O sont semblables et en accord avec les données expérimentales[1, 14, 15], la valeur de Δmoy n'excédant pas 11cm^{-1} . L'étude des valeurs de $\Delta\omega$ après hydratation montre que les vibrations comprenant des atomes impliqués dans les interactions hydrogène ont leurs nombres d'onde harmoniques qui sont sensibles à la considération de la dispersion. La présence conjuguée de deux liaisons C=O et NH explique le plus grand nombre de valeurs $\Delta\omega$ importantes calculées. Le même phénomène est noté pour les valeurs de $\Delta\nu$, ce qui peut être expliqué par la qualité encore plus faible du potentiel dans la reproduction de couplages plus nombreux. Établir une quelconque conclusion à partir de leur seule étude semble donc difficile. La comparaison des valeurs de dynamique moléculaire B3LYP et B3LYP-D avec les données expérimentales conduit à favoriser la première méthode citée, puisque l'écart moyen est inférieur de 18cm^{-1} à celui calculé avec la dispersion. Cela est cohérent avec les écarts observés dans le cadre de la méthode VPT2. Globalement, le cas de la molécule isolée ne nous permet pas de conclure sur le choix d'une méthode. Mais on constate que pour un même modèle géométrique obtenu à partir de deux approches différentes B3LYP ou B3LYP-D, les fréquences vibrationnelles anharmoniques sur les bases d'acides nucléiques hydratés sont mieux représentées au niveau B3LYP, et ce pour les deux méthodes de traitement vibrationnel testées.

I.3 Validation sur les autres bases dans la zone $1000\text{-}1800\text{cm}^{-1}$

L'étude de la cytosine et de la thymine repose sur un même modèle géométrique en B3LYP et en B3LYP-D. Le cas des trois autres bases est différent, puisque le modèle explicite d'hydratation qui se dégage après une étude B3LYP-D est plus complexe que celui observé en B3LYP : le nombre de molécules d'eau considérées est plus important. En appliquant les critères précédemment établis pour son choix, on aboutit aux modèles géométriques suivants, présentés dans le chapitre 4 :

- * l'uracile hydratée peut être représentée par le modèle $U,5H_2O+PCM$, comprenant trois molécules d'eau réparties sur les sites B et C décrits au chapitre 4, à la différence de la géométrie B3LYP, où les molécules sont localisées dans le site B uniquement ;
- * l'adénine hydratée peut être représentée par le modèle $A,3H_2O+PCM$ en B3LYP ou bien $A,5H_2O+PCM$ en considérant la dispersion ;
- * la guanine hydratée peut être représentée par le modèle $G,3H_2O+PCM$ en B3LYP, ou bien $G,4H_2O+PCM$ au niveau B3LYP-D.

Les différents calculs vibrationnels effectués sur ces systèmes et les formes isolées sont identiques à ceux présentés auparavant, excepté les dynamiques moléculaires au niveau B3LYP-D qui n'ont pas été menées. On remarquera que seuls les résultats compris dans la zone de signature vibrationnelle de chaque

système (1000-1800cm⁻¹) sont reportés dans les tableaux suivants.

- * Tableau U : Écart (en cm⁻¹) entre les valeurs harmoniques ($\Delta\omega$) et anharmoniques ($\Delta\nu$) obtenues pour l'uracile isolée et hydratée (*modèles B3LYP et B3LYPD différents : U,5H₂O+PCM) aux niveaux VPT2 et DM dans le cadre des méthodes B3LYP et B3LYP-D/6-31+G(d,p) - Comparaison avec les données expérimentales dans la zone 1000-1800cm⁻¹

- * Tableau A : Écart (en cm⁻¹) entre les valeurs harmoniques ($\Delta\omega$) et anharmoniques ($\Delta\nu$) obtenues pour l'adénine isolée et hydratée (*modèles B3LYP : A,3H₂O+PCM et B3LYPD : A,5H₂O+PCM) aux niveaux VPT2 et DM dans le cadre des méthodes B3LYP et B3LYP-D/6-31+G(d,p) - Comparaison avec les données expérimentales dans la zone 1000-1800cm⁻¹ - La valeur en italique a été écartée du calcul de Δ_{moy}

- * Tableau G : Écart (en cm⁻¹) entre les valeurs harmoniques ($\Delta\omega$) et anharmoniques ($\Delta\nu$) obtenues pour la guanine isolée et hydratée (*modèles B3LYP : G,3H₂O+PCM et B3LYPD : G,4H₂O+PCM) aux niveaux VPT2 et DM dans le cadre des méthodes B3LYP et B3LYP-D/6-31+G(d,p) - Comparaison avec les données expérimentales dans la zone 1000-1800cm⁻¹

Désignation du mode	U, 0H ₂ O						U microhydratée*					
	$\Delta\omega$	$\Delta\nu$	$\Delta\nu$ VPT2 (B3LYPexp)	$\Delta\nu$ VPT2 (B3LYPDexp)	$\Delta\nu$ DM (B3LYPexp)	$\Delta\omega$	$\Delta\nu$	$\Delta\nu$ VPT2 (B3LYPexp)	$\Delta\nu$ VPT2 (B3LYPDexp)	$\Delta\omega$	$\Delta\nu$	
$\nu(\mathbf{C}_{(4)}-\mathbf{C}_{(5)}), \delta(\mathbf{N}_{(3)}\mathbf{H})$	0	1	4	5	12	39	38	24	62			
$\gamma(\mathbf{C}_{(6)}\mathbf{H}, \mathbf{C}_{(5)}\mathbf{H})_{as}$	0	0	23	23	15	-10	7	5	12			
$\nu(\mathbf{N}_{(1)}\mathbf{C}_{(5)}\mathbf{C}_{(6)})$	4	3	13	16	13	-47	-10					
$\nu(\mathbf{N}_{(1)}\mathbf{C}_{(6)}), \delta(\mathbf{C}_{(5)}\mathbf{H})$	7	8	15	23	19	6	4	0	4			
$\nu(\mathbf{C}_{(2)}-\mathbf{N}_{(3)}), \nu(\mathbf{N}_{(3)}-\mathbf{C}_{(4)})$												
$\delta(\mathbf{C}_{(6)}\mathbf{H}, \mathbf{N}_{(1)}\mathbf{H})_{as}$	6	3	2	5	19	-13	-11					
$\delta(\mathbf{C}_{(5)}\mathbf{H}, \mathbf{C}_{(6)}\mathbf{H})_{as}, \delta(\mathbf{N}_{(1)}\mathbf{H})$	11	9	21	30	31	13	26	8	18			
$\nu(\mathbf{C}_{(2)}-\mathbf{N}_{(3)}), \delta(\mathbf{H}),$ $\nu(\mathbf{C}_{(5)}=\mathbf{C}_{(6)})$	1	2	5	7	53	-4	-12	12	0			
$(\delta(\mathbf{N}_{(3)}\mathbf{H}), \delta(\mathbf{C}_{(6)}\mathbf{H}))_s$	1	3	12	15	37	-22	96	0	96			
$(\delta(\mathbf{C}_{(6)}\mathbf{H}), \delta(\mathbf{N}_{(1)}\mathbf{H}),$ $\delta(\mathbf{C}_{(5)}\mathbf{H}))_s, \delta(\mathbf{N}_{(3)}\mathbf{H})$	2	3	11	14	23	-6	49					
$\delta(\mathbf{N}_{(1)}\mathbf{H}), \nu(\mathbf{C}_{(6)}-\mathbf{N}_{(1)})$	3	2	5	3	11	-2	201	92	109			
$\nu(\mathbf{C}_{(5)}=\mathbf{C}_{(6)})$	4	5	1	4	21	3	0	15	15			
$\nu(\mathbf{C}_{(4)}=\mathbf{O})$	-1	-4	51	55	3	-18	-16	40	24			
$\nu(\mathbf{C}_{(2)}=\mathbf{O})$	-1	1	22	21	26	5	17	24	41			
Δ_{moy}			14	17	22			22	38			
Δ_{max}			51	55	53			92	109			

Désignation du mode	A, 0H ₂ O					A microhydratée*				
	$\Delta\omega$	$\Delta\nu$	$\Delta\nu$ VPT2 (B3LYPexp)	$\Delta\nu$ VPT2 (B3LYPDexp)	$\Delta\nu$ DM (B3LYPexp)	$\Delta\omega$	$\Delta\nu$	$\Delta\nu$ VPT2 (B3LYPexp)	$\Delta\nu$ VPT2 (B3LYPDexp)	
$\gamma(\text{C}_{(2)}\text{H})$	1	0	4	4	12	2	-39	42	3	
rock(NH ₂), $\nu(\text{N}_{(1)}\text{C}_{(2)})$	-5	-400	319	81	10	-37	-27			
$(\delta(\text{N}_{(9)}\text{H}), \delta(\text{C}_{(8)}\text{H}))\text{as}$	11	8	5	3	54	35	-4	23	19	
$\nu(\text{C}_{(4)}\text{N}_{(9)}), \delta(\text{C}_{(8)}\text{H})$	1	1	2	3	111	-34	-42	39	3	
rock(NH ₂), $\nu(\text{C}_{(5)}\text{N}_{(7)})$	1	-137	114	23	20	-27	-49			
$(\delta(\text{C}_{(8)}\text{H}), \delta(\text{N}_{(9)}\text{H}))\text{s}$	5	3	7	4	30	-3	-39	23	16	
$(\nu(\text{C}_{(2)}\text{N}_{(3)}), \nu(\text{C}_{(5)}\text{C}_{(6)}))\text{as}$	5	3	17	14	1	-5	-13	8	5	
$\nu(\text{N}_{(1)}\text{C}_{(2)}), \nu(\text{C}_{(5)}\text{N}_{(7)})$	2	31	30	1	13	0	-16	12	28	
$\delta(\text{H})\text{as}, \nu(\text{N}_{(9)}\text{C}_{(8)})$	2	0	9	9	25	-8	-13	15	2	
$(\delta(\text{N}_{(9)}\text{H}), \delta(\text{C}_{(2)}\text{H}))\text{as}$	3	3				-1	37	19	18	
$\nu(\text{C}_{(4)}\text{C}_{(5)}), \delta(\text{C}_{(2)}\text{H})$	1	31	17	14	146	-4	-15			
$\nu(\text{N}_{(1)}\text{C}_{(6)}), \delta(\text{C}_{(2)}\text{H})$	1	1	1	2	148	-5	-42	25	17	
$\nu(\text{N}_{(7)}\text{C}_{(8)})$	2	4	10	6	9	-3	9			
sci(NH ₂), $(\nu(\text{C}_{(4)}\text{C}_{(5)}), \nu(\text{N}_{(1)}\text{C}_{(2)}))\text{s}$	2	-52	62	10	32	-1	-22	26	48	
$(\nu(\text{N}_{(3)}\text{C}_{(4)}), \nu(\text{N}_{(1)}\text{C}_{(6)}))\text{s}$	1	-27	32	5	54	4	-24			
sci(NH ₂), $\nu(\text{C}_{(6)}\text{N}_{(12)})$	-1	28	26	2	66	-20	-82	2	80	
Δ_{moy}			24	12	49			21	22	
Δ_{max}			114	81	148			42	80	

Désignation du mode	G, 0H ₂ O						G microhydratée*			
	$\Delta\omega$	$\Delta\nu$	$\Delta\nu$ VPT2 (B3LYPexp)	$\Delta\nu$ VPT2 (B3LYPDexp)	$\Delta\nu$ DM (B3LYPexp)	$\Delta\omega$	$\Delta\nu$	$\Delta\nu$ VPT2 (B3LYPexp)	$\Delta\nu$ VPT2 (B3LYPDexp)	
$(\delta(\text{NH}))_s, \nu(\text{N}_{(1)}\text{C}_{(6)})$	-4	3	18	15	2	-14	16	5	11	
$(\delta(\text{C}_{(8)}\text{H}), \delta(\text{N}_{(9)}\text{H}))_{\text{as}}$	10	8	29	21	14	4	-34			
$\nu(\text{CN}), (\delta(\text{N}_{(11)}\text{H}))_s$	0	3	1	4	3	-12	41			
$(\delta(\text{N}_{(11)}\text{H}_{(12)}), \delta(\text{N}_{(11)}\text{H}_{(13)}))_s$	0	3	5	2	3	-3	22	16	38	
$\nu(\text{C}_{(4)}\text{N}_{(9)}), (\delta(\text{C}_{(8)}\text{H}), \delta(\text{N}_{(9)}\text{H}))_s$	2	2	10	8	16	-13	-31	34	3	
$\delta(\text{C}_{(8)}\text{H}), \nu(\text{N}_{(7)}\text{C}_{(8)})$	2	0	13	13	6	2	-26	16	42	
$\delta(\text{N}_{(1)}\text{H}), \delta(\text{NH})$	-5	-1				-14	27	21	6	
$(\nu(\text{C}_{(4)}\text{N}_{(9)}), \nu(\text{C}_{(5)}\text{N}_{(7)}))_s$	1	2	7	5	26	-2	-49	33	16	
$\nu(\text{CN}), (\delta(\text{C}_{(8)}\text{H}), \delta(\text{N}_{(9)}\text{H}))_{\text{as}}$	3	-1	13	12	16	-3	17	10	7	
$(\nu(\text{N}_{(1)}\text{C}_{(2)}), \nu(\text{C}_{(4)}\text{C}_{(5)}))_{\text{as}}$	1	4	3	1	38	-5	13	5	8	
$\nu(\text{N}_{(7)}\text{C}_{(8)}), \delta(\text{H})$	0	1	24	25	59	1	6	32	26	
$(\nu(\text{N}_{(1)}\text{C}_{(2)}), \nu(\text{C}_{(4)}\text{C}_{(5)}))_s$	0	1	25	26	55	2	11	19	30	
$\nu(\text{N}_{(3)}\text{C}_{(4)}), \nu(\text{C}_{(4)}\text{N}_{(9)}), \delta(\text{N}_{(9)}\text{H})$	1	1	4	6	42	-4	-4			
$\nu(\text{C}_{(2)}\text{N}_{(3)}), \delta(\text{N}_{(1)}\text{H}), \delta(\text{N}_{(1)}\text{H}_2)$	1	1	7	8	31	-17	-1			
$\delta(\text{NH}_2)$	-3	5	2	3	11	2	3	20	23	
$\nu(\text{C}_{(6)}=\text{O}), \nu(\text{C}_{(5)}\text{C}_{(6)})$	-1	-3	12	15	32	17	24			
Δ_{moy}			12	13	24			19	19	
Δ_{max}			29	26	59			34	42	

L'analyse des valeurs de $\Delta\omega$ conduit aux mêmes conclusions que précédemment : les formes isolées présentent des nombres d'onde harmoniques identiques pour les méthodes B3LYP et B3LYP-D. La prise en compte explicite des molécules de solvant conduit à des écarts sur les modes impliqués dans les liaisons hydrogène, atteignant même 87cm^{-1} pour le mode $sci(Me), \delta(N_{(1)}H)$ de la thymine tétrahydratée. Les différences moyennes Δ_{moy} entre les valeurs anharmoniques et leurs contreparties expérimentales[1, 14, 15][16] correspondantes à l'uracile et à la guanine sont semblables pour les deux niveaux calculatoires utilisés. En revanche, le cas de l'adénine en phase gazeuse est remarquable : malgré la suppression du mode le plus mal traité par le potentiel d'ordre 4 au niveau B3LYP, la méthode B3LYP-D fournit des résultats en meilleur accord avec l'expérience[1, 11] que son analogue. Cela n'est sans doute pas dû qu'à la forme mathématique du potentiel car les résultats dynamiques ne sont pas meilleurs. L'analyse des autres calculs dynamiques concernant l'uracile et la guanine isolées révèle que la DM n'est pas nécessaire pour l'étude de cette zone, puisque la valeur de Δ_{moy} en dynamique est bien supérieure à celle calculée par la méthode VPT2. L'analyse des systèmes hydratés montre que même si le modèle géométrique obtenu en B3LYP-D est plus adapté que celui obtenu en B3LYP, les résultats vibrationnels qui en sont issus ne sont pas meilleurs par rapport aux données expérimentales[17, 18, 4]. La méthode B3LYP est plus favorable que son analogue incluant la dispersion, l'écart entre les moyennes calculées dans le cadre des deux méthodes étant de 6cm^{-1} en faveur de B3LYP. En conclusion, on constate que les mêmes observations sont faites pour les cinq bases d'acides nucléiques étudiées dans la zone $1000\text{-}1800\text{cm}^{-1}$. Pour les formes en phase gazeuse, la dispersion n'améliore pas les valeurs des fréquences vibrationnelles. Quand l'hydratation est prise en compte, la qualité de la méthode B3LYP pour le calcul de la structure électronique suffit largement à compenser la qualité du modèle géométrique issu d'une étude B3LYP-D. Pour la suite, on choisira donc la méthode B3LYP. Reste à déterminer la méthode de traitement vibrationnel la plus adaptée à l'étude de la zone de signature vibrationnelle des bases d'acides nucléiques isolées et surtout hydratées.

II Choix du traitement vibrationnel

Nous allons désormais nous attacher à déterminer quel est la méthode de traitement vibrationnel -statique(VPT2) ou dynamique(MD)- la plus adaptée à l'étude des bases d'acides nucléiques isolées et hydratées dans le cadre d'un calcul de structure électronique au niveau B3LYP/6-31+G(d,p). Un travail préalable a été réalisé sur deux petits systèmes azotés -pyrrole et méthanimine- . L'objectif était de voir si les deux méthodes VPT2 et DM fournissaient des résultats vibrationnels cohérents avec les données expérimentales disponibles dans la littérature sur des systèmes possédant le même type de caractéristiques géométriques que les bases d'acides nucléiques.

II.1 Cas de petites molécules azotées

L'étude d'autres molécules possédant des liaisons NH insérés dans un environnement analogue nous a permis de voir si les méthodes VPT2 et DM permettaient de bien traiter les propriétés vibrationnelles de tels systèmes. Les propriétés vibrationnelles de la méthanimine CH_2NH et du pyrrole $\text{C}_4\text{H}_5\text{N}$ ont ainsi été déterminées par les approches VPT2 et DM et comparées aux données expérimentales disponibles dans la littérature[19, 20, 21, 22]. Seuls certains modes caractéristiques de ces molécules ont été considérées dans cette analyse. Les résultats sont présentés dans les tableaux 5.5.

Mode	méthanimine			Mode	pyrrole		
	VPT2	DM	Exp		VPT2	DM	Exp
$\nu(\text{NH})$	3271	3297	3263	$\nu(\text{NH})$	3517	3568	3531
$\nu(\text{CH}_2)_{\text{as}}$	2962	3020	3024	$\nu(\text{CH})_{\text{s}}$	3153	3167	3149
$\nu(\text{CH}_2)_{\text{s}}$	2877	2883	2914	$(\nu(\text{C}_{(1)}\text{H}, \text{C}_{(2)}\text{H}),$			
$\nu(\text{C}=\text{C})$	1686	1660	1638	$\nu(\text{C}_{(3)}\text{H}, \text{C}_{(4)}\text{H})_{\text{as}}$	3147	3157	3143
$\delta(\text{CH}_2)$	1484	1460	1452	$(\nu(\text{C}_{(2)}\text{H}, \text{C}_{(3)}\text{H}),$			
$(\delta(\text{NH}), \text{rock}(\text{CH}_2))_{\text{s}}$	1329	1343	1344	$\nu(\text{C}_{(1)}\text{H}, \text{C}_{(4)}\text{H})_{\text{as}}$	3131	3150	3128
$(\gamma(\text{NH}), \gamma(\text{CH}))_{\text{as}}$	1131	1137		$(\nu(\text{C}_{(2)}\text{H}, \text{C}_{(4)}\text{H}),$			
$\gamma(\text{CH}_2)$	1076	1077		$\nu(\text{C}_{(1)}\text{H}, \text{C}_{(3)}\text{H})_{\text{as}}$	3121	3123	3119
$(\delta(\text{NH}), \text{rock}(\text{CH}_2))_{\text{as}}$	1053	1050		$\nu(\text{C}_{(1)}=\text{C}_{(2)})$	1545	1537	1518
				$(\delta(\text{C}_{(1)}\text{H}), \delta(\text{C}_{(4)}\text{H}))_{\text{as}}$	1467	1457	1472
				$\delta(\text{NH})$	1411	1410	1424
				$\nu(\text{C}_{(2)}\text{C}_{(3)}),$			
				$(\delta(\text{C}_{(2)}\text{H}), \delta(\text{C}_{(3)}\text{H}))_{\text{as}}$	1398	1375	1401

FIGURE 5.5 – Fréquences anharmoniques VPT2 et DM de la méthanimine et du pyrrole obtenues avec un calcul de structure électronique réalisé à un niveau B3LYP/6-31+G(d,p)

On constate que toutes les valeurs sont cohérentes par rapport à l'expérience : les écarts moyens entre expérience et théorie sont de l'ordre de 34 et 17cm^{-1} pour les approches VPT2 et DM appliquées à la méthanimine, tandis qu'ils valent 8 et 19cm^{-1} pour le pyrrole. Quant aux modes de stretching CH et surtout NH, ils sont bien reproduits par la dynamique moléculaire, la déviation maximale par rapport à l'expérience étant de 37cm^{-1} . L'efficacité des méthodes de traitement vibrationnel ayant été vérifiée, ces dernières peuvent appliquer à de plus gros systèmes tels que les bases d'acides nucléiques.

II.2 Cas des bases d'acides nucléiques : zone 1000-1800cm⁻¹

II.2.1 Comparaison des méthodes VPT2 et DM entre elles

Tous les résultats VPT2 et DM concernant les bases d'acides nucléiques isolées ou bien hydratées sont reportés en annexe. La discussion sera menée sur la zone 1000-1800cm⁻¹. Celle-ci est très peu influencée par les phénomènes de couplages anharmoniques, comme l'ont montré les observations de Handy et al[10]. La méthode VPT2 devrait donc fournir des résultats équivalents à ceux de la dynamique moléculaire. C'est ce que l'on constate pour chaque base isolée : l'écart moyen entre les résultats VPT2 et DM est respectivement de 14, 19, 22, 25cm⁻¹ pour T, A, C et G, U. Pour l'adénine, certaines différences ont été omises pour le calcul de la moyenne : certains modes de vibration sont mal représentés soit par la méthode VPT2 (modes 20 et 23 relatifs à la vibration rock(NH₂)), soit par la DM (modes 22, 29, 30 et 34 relatifs à des stretching CN ou bien CC). Ce même phénomène est observé pour les bases hydratées. On peut citer notamment le mode 54 de la thymine hydratée, pour lequel la valeur DM est inférieure de 96cm⁻¹ à sa contrepartie VPT2, la valeur harmonique étant de 1566cm⁻¹. Le mode 52 de l'uracile hydratée est aussi touché : le même type d'écart de 102cm⁻¹ est observé entre les valeurs VPT2 et DM correspondantes. Pour le système C₄H₂O+PCM, le mode 42 relatif à la vibration δ(NH₂)rock présente un nombre d'onde au niveau VPT2 116cm⁻¹ plus bas que sa valeur harmonique. Globalement, l'écart moyen des fréquences obtenues dans le cadre de ces deux méthodes de traitement vibrationnel varie de 12 (pour la thymine) à 26cm⁻¹(pour l'adénine) sur les systèmes hydratés. On note que les valeurs obtenues en VPT2 et DM sont raisonnablement proches dans la zone 1000-1800cm⁻¹. Les écarts trop importants calculés relèvent soit de la forme mathématique du potentiel VPT2 qui n'est pas optimale, soit de problèmes d'attribution des bandes en DM, notamment dans la zone des vibrations de bending des molécules d'eau. Le cas de l'adénine est différent des autres puisque un certain nombre de couplages semblent être présents déjà sur la forme isolée. Dans un second temps, les valeurs calculées ont été comparées aux données expérimentales disponibles dans la littérature et reportées dans les tableaux situés en annexe.

II.2.2 Comparaison des résultats VPT2 et DM avec les données expérimentales

Les écarts moyens Δ_{moy} entre les valeurs expérimentales et théoriques des formes isolées et aqueuses dans la zone 1000-1800cm⁻¹ sont reportés dans le tableau 5.6.

Dans le cas du système isolé, la méthode VPT2 fournit le meilleur accord entre expérience et théorie, un écart moyen de 14cm⁻¹ étant observé. L'étude des systèmes en solution permet d'aboutir à la même conclusion : seul le cas de la cytosine est différent. La comparaison entre les approches VPT2 et DM montre que la dynamique moléculaire fournit les meilleurs résultats vibrationnels : l'écart moyen atteint 15cm⁻¹, contre 25cm⁻¹ dans le cadre de la méthode statique. Celle-ci n'est pas adaptée sans doute en

Cas	Type de méthode	Cytosine	Thymine	Uracile	Guanine	Adénine
isolé	VPT2	15	7	15	12	23
	DM	30	17	25	23	40
aqueux	VPT2	25	16	24	19	18
	DM	15	25	36	30	29

FIGURE 5.6 – Écarts moyens obtenus (en cm^{-1}) entre les nombres d'onde anharmoniques théoriques (DM ou VPT2) et leurs contreparties expérimentales dans la zone $1000\text{-}1800\text{cm}^{-1}$ pour les différentes bases d'acides nucléiques isolées ou en solution aqueuse à un niveau B3LYP/6-31+G(d,p)

raison de la présence de nombreux couplages impliquant les modes de l'eau. Cela est confirmé notamment par les problèmes d'attribution rencontrés pour les modes 47 et 48 de la cytosine hydratée dont les nombres d'onde se situent entre 1400 et 1600cm^{-1} , pour lesquels 2 pics d'intensité comparable se situaient dans la même zone d'intérêt.

III Conclusion et Perspectives

Les études menées en parallèle sur les bases d'acides nucléiques isolées et hydratées nous ont permis de déterminer les conditions de calcul les plus adaptées pour l'obtention des propriétés vibrationnelles de chacun de ces systèmes. Tout d'abord, le modèle géométrique permettant de représenter chaque base en solution aqueuse est constitué d'un modèle implicite tel que le modèle PCM qui considère le solvant comme un continuum polarisable. Un modèle explicite lui est associé : il prend la forme $X_n\text{H}_2\text{O}$ où X représente la base d'acide nucléique étudiée. La valeur de n est choisie telle que :

- * le pourcentage de Boltzmann de la forme considérée soit élevé pour qu'elle soit représentative du taux d'hydratation correspondant ;
- * ses principaux sites privilégiés pour l'eau soient occupés.

Le modèle géométrique établi, il est nécessaire de déterminer le modèle théorique adapté. La méthode de calcul de structure électronique B3LYP/6-31+G(d,p) permet d'obtenir les résultats les plus en adéquation avec les données expérimentales dans tout le domaine spectral. Si nous nous attachons en particulier à l'étude de la zone $1000\text{-}1800\text{cm}^{-1}$, la méthode de traitement vibrationnel VPT2 fournit les meilleurs résultats, avec en plus un temps de calcul raisonnable pour des systèmes de relativement grande taille (jusqu'à 27 atomes).

L'étude des zones plus extrêmes du spectre pourront être menées ultérieurement afin de déterminer la méthode de traitement vibrationnel la plus adéquate pour chacune d'entre elles. La zone des bas modes semble être la plus évidente à analyser. En effet, elle est marquée par la présence de couplages avec des modes de stretching NH et CH, qui sont mal reproduits par la méthode VPT2. Ce constat est illustré par les cas de la cytosine et de l'adénine. En effet, l'étude de la cytosine révèle les limites des deux approches VPT2 et DM : la méthode VPT2 fournit des valeurs absurdes pour les modes 19 et 20 (correspondant à $\delta(C_{(2)} - NH_2)$ et à une déformation du cycle) et tandis que la dynamique moléculaire les surestime (jusqu'à 100cm^{-1} par rapport à ω_{19} pour le premier). De façon générale, dans cette zone, les valeurs DM sont meilleures que les nombres d'onde issues d'un calcul VPT2. La méthode VPT2 présente toujours l'avantage d'être rapide. Mais pour déterminer les propriétés vibrationnelles dans la zone spectrale en-dessous de 1000cm^{-1} , une mauvaise expression du potentiel qui représente alors mal les couplages anharmoniques peut entraîner des résultats aberrants. La dynamique moléculaire permet de s'affranchir de ce problème. Cependant, une contrainte réside dans le choix du temps de simulation total, qui doit être suffisamment important pour bien reproduire les modes mous compris dans cette zone. Dans notre étude, les résultats dans cette zone peuvent être améliorés en abaissant la température de simulation car le temps de simulation total n'a pas été suffisant dans le cas de la cytosine hydratée notamment (4ps). Quant à la zone des hauts modes de vibration, l'absence de données expérimentales en solution aqueuse est gênant pour pouvoir confronter nos valeurs. De plus, les résultats de dynamique moléculaire ne sont pas faciles à exploiter, la multitude de pics intenses dans la zone entre 2800 et 3500cm^{-1} ne facilitant pas l'attribution. A partir des résultats obtenus, on observe une surestimation de l'anharmonicité pour certains modes. Elle atteint entre 300 et 400cm^{-1} pour les modes de vibration $\nu(N_{(1)}H)$ de la guanine, de l'uracile et de la thymine isolées notamment. Plusieurs stratégies sont envisageables pour pallier au problème de lecture des spectres projetés sur chaque mode normal et pour expliquer de tels écarts. Les valeurs reportées en annexe sont celles issues de ce type de projection, confirmées par une projection des vitesses sur le jeu de coordonnées internes correspondant. Une hypothèse permettant d'expliquer de telles valeurs d'anharmonicité serait que l'énergie cinétique initiale ne se répartisse pas de façon équiprobable sur les modes hauts : une IVR permettrait de le révéler ou pas. Une quantité d'énergie cinétique (en J) égale à $E = \frac{1}{2}h\nu + \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1}$ est alors insufflée au mode que l'on veut étudier. Une simulation de DM permettant de laisser chaque système se relaxer permettra d'établir les répartitions de l'énergie de départ obtenues à la fin de la dynamique.

Bibliographie

- [1] Colarusso, P.; Zhang, K.; Guo, B.; Bernath, P. F. Chem Phys Lett 1997, 269, 39.

-
- [2] Cornell, W. D. ; Cieplak, P. ; Bayly, C. I. ; Gould, I. R. ; Merz, K. M. ; Ferguson, D. M. ; Spellmeyer, D. C. ; Fox, T. ; Caldwell, J. W. ; Kollman, P. A. *J Am Chem Soc* 1995, 117, 5179.
- [3] Carbonniere, P. ; Thicoipe, S. ; Very, T. ; Assfeld, X. *IJQC* 2012, 112, 2221.
- [4] Lord, R. ; Thomas Jr., G. *Spectrochimica Acta Part A : Molecular Spectroscopy* 1967, 23, 2551.
- [5] Steinmann, S. N. ; Corminboeuf, C. *Journal of Chemical Theory and Computation* 2011, 7, 3567.
- [6] Steinmann, S. N. ; Piemontesi, C. ; Delachat, A. ; Corminboeuf, C. *Journal of Chemical Theory and Computation* 2012, 8, 1629.
- [7] Carbonniere, P. ; Barone, V. *Chem Phys Lett* 2004, 399, 226.
- [8] Radchenko, E. ; Sheina, G. ; Smorygo, N. ; Blagoi, Y. *J Molec Struct* 1984, 116, 387.
- [9] Szczesniak, M. ; Szczepaniak, K. ; Kwiatkowski, J. S. ; KuBulat, K. ; Person, W. B. *J Am Chem Soc* 1988, 110, 8319.
- [10] Burcl, R. ; Carter, S. ; Handy, N. C. *Chem Phys Lett* 2003, 373, 357.
- [11] Mathlouthi, M. ; Seuvre, A. ; L. Koenig, J. *Carbohydrate Research* 1984, 134, 23.
- [12] Shang, Z. ; Ting, D. N. ; Wong, Y. T. ; Tan, Y. C. ; Ying, B. ; Mo, Y. *J Molec Struct* 2007, 826, 64.
- [13] Ten, G. N. ; Burova, T. G. ; Baranov, V. I. *J of Applied Spectroscopy* 2005, 72, 104.
- [14] Chandra, A. K. ; Nguyen, M. T. ; Zeegers-Huyskens, T. *J Phys Chem A* 1998, 102, 6010.
- [15] Szczepaniak, K. ; Szczesniak, M. M. ; Person, W. B. *J Phys Chem A* 2000, 104, 3852.
- [16] Gould, I. R. ; Vincent, M. A. ; Hillier, I. H. *Spectrochimica Acta Part A : Molecular Spectroscopy* 1993, 49, 1727.
- [17] Gageot, M. ; Kadri, C. ; Ghomi, M. *J Molec Struct* 2001, 565–566, 469.
- [18] Ten, G. N. ; Bourova, T. G. ; Baranov, V. I. *J of Structural Chemistry* 2005, 46, 998.
- [19] Halonen, L. ; Duxbury, G. *Chem Phys Lett* 1985, 118, 246 .
- [20] Halonen, L. ; Duxbury, G. *J Chem Phys* 1985, 83, 2091.
- [21] Allegrinia, M. ; Johns, J. W. C. ; McKellar, A. R. W. *J Chem Phys* 1979, 70, 2829.
- [22] Mellouki, A. ; Liévin, J. ; Herman, M. *Chem Phys* 2001, 271, 239 .

Chapitre 6

Conclusion

La chimie théorique est un outil performant qui permet de comprendre un certain nombre d'observations expérimentales, notamment les spectres IR ou Raman, facilement simulables par ce biais. L'enjeu fut double dans cette thèse, de par les molécules étudiées (aérosols de nitrate d'une part et bases d'acides nucléiques isolées et hydratées d'autre part), mais aussi par les moyens à mettre en œuvre. Deux modèles doivent être déterminés afin de pouvoir représenter au mieux l'influence sur les propriétés d'un environnement plus ou moins hydraté autour du soluté : un modèle géométrique et un modèle théorique. La recherche du modèle géométrique le plus adapté requiert, pour les deux études, l'utilisation du code GSAM développé au laboratoire. Celui-ci permet de répertorier tous les isomères, notamment les formes les plus stables représentatives de chacun des systèmes étudiés qui correspondent à celles dont le pourcentage de Boltzmann à 298K est supérieur à 5%. Ensuite, il est nécessaire de déterminer le modèle théorique permettant d'obtenir les propriétés vibrationnelles les plus en adéquation avec les données expérimentales disponibles dans la littérature. Ce modèle inclut la méthode de structure électronique, ainsi que la méthode de traitement vibrationnel. Les principales méthodes testées sont issues de la DFT, le choix le plus évident s'étant porté sur la fonctionnelle B3LYP. Compte tenu de la présence de nombreuses interactions de type hydrogène, ont aussi été retenues les méthodes CAMB3LYP et B3LYP-D, qui prennent en compte respectivement les effets dus à la longue portée et à la dispersion. Dans un second temps, les résultats issus des applications de deux méthodes de traitement vibrationnel différentes ont été confrontés. La méthode VPT2 est fortement dépendante de la forme mathématique du potentiel, à la différence de la méthode DM.

La démarche présentée ci-dessus a été appliquée pour cette thèse aux cas des aérosols de nitrate et aux bases d'acides nucléiques. Tout d'abord, en ce qui concerne les aérosols de nitrate, les formes

$(\text{NaNO}_3, n\text{H}_2\text{O})_x$ (avec $n=1$ à 4 et $x=1$ à 4) ont été répertoriées grâce au code GSAM. Il a ainsi été établi que le modèle géométrique le plus adapté pour l'étude de cinq modes de vibration de NaNO_3^- en milieu hydraté est tel que $x=3$. Associé à la méthode de détermination de structure électronique B3LYP/6-311+G(d,p) et en utilisant un traitement vibrationnel du type VPT2, l'écart moyen entre expérience et théorie obtenu est de l'ordre de 14cm^{-1} , contre 40cm^{-1} pour le modèle géométrique le plus simple correspondant à $x=1$. Ensuite, le code GSAM a été appliqué aux bases d'acides nucléiques microhydratées. Les structures les plus stables $\text{X}, n\text{H}_2\text{O}$ avec $\text{X}=\text{A}, \text{T}, \text{C}, \text{G}$ et U et n un entier de 1 à 5 ont été reportées pour les deux niveaux B3LYP et B3LYPD/6-311+G(d,p). Des sites privilégiés identiques pour les molécules d'eau ont été mis en évidence. On constate que pour $n=1$ et 2 , les deux niveaux de calcul de structure électronique fournissent des résultats semblables. En revanche, pour un taux d'hydratation supérieur, la prise en compte de la dispersion modifie l'ordre énergétique, les formes représentatives (*i.e.* dont le pourcentage de Boltzmann à 298K est supérieur à 5%) restant la plupart du temps les mêmes. Cette étude a permis l'établissement d'un modèle géométrique permettant de reproduire le comportement vibrationnel des bases d'acides nucléiques en milieu aqueux, validé sur les cas de la cytosine et de la thymine. Il est constitué de deux modèles de solvation : l'un explicite, issu de l'étude structurale précédente, et l'autre implicite (PCM). La détermination des propriétés vibrationnelles a été limitée à la zone $1000\text{-}1800\text{cm}^{-1}$ car cette dernière représente la signature vibrationnelle du système considéré et comporte le plus de données expérimentales. Après des calculs VPT2 et DM, menés à deux niveaux calculatoires différents (B3LYP et B3LYPD/6-31+G(d,p)), la méthode théorique la plus adaptée à cette région associe le niveau B3LYP/6-31+G(d,p) avec un traitement vibrationnel du type VPT2, les couplages dans cette zone n'étant que peu présents. Les deux autres zones spectrales ne feront pas l'objet ici d'une étude approfondie, surtout celle dont les nombres d'onde sont supérieurs à 2800cm^{-1} , certains problèmes non expliqués pour le moment étant observés.

Pour l'avenir, plusieurs pistes sont à envisager. La zone dont les nombres d'onde sont supérieurs à 3000cm^{-1} doit être analysée dans le détail, notamment par une IVR sur un des modes de stretching dont les nombres d'onde sont largement sous-estimés. La recherche du modèle géométrique concernant l'adénine hydratée avec la méthode B3LYP peut être approfondie. En effet, même si les résultats B3LYP sont meilleurs que leurs contreparties B3LYP-D (l'écart moyen entre les valeurs théoriques et expérimentales étant de 22 et 27cm^{-1} respectivement), le modèle explicite choisi ne correspond pas exactement au critère présenté, puisque seuls les deux sites les plus privilégiés sont occupés par les molécules d'eau, contre trois en théorie. Un taux d'hydratation supérieur à 5 est donc nécessaire afin de déterminer le modèle adéquat représentatif du phénomène d'hydratation. Enfin, les études menées sur ces systèmes hydratés basiques nous ont permis d'établir une démarche type afin de reproduire des propriétés vibrationnelles observées par l'expérience. Elle pourra être ensuite appliquée à d'autres systèmes plus complexes, tels

que des molécules associées avec des glaces interstellaires. On peut noter la glycine, qui est une molécule dont la présence a été révélée dans les météorites¹. En revanche, elle n'a pas été détectée dans le milieu interstellaire. Une des suppositions consiste à envisager cette molécule adsorbée sur des glaces. Une procédure d'analyse théorique pourra être la suivante : l'établissement d'un modèle géométrique traduisant les interactions glace/glycine, puis le choix d'un modèle de calcul de structure électronique adapté et enfin, le calcul des propriétés associées. Elles seront comparées aux valeurs expérimentales mesurées afin de pouvoir confirmer ou infirmer l'hypothèse de départ.

1. Pizzarello, S. ;Huang, Y ;Becker, L. ;Poreda, R.J. ;Nieman, R.A. ;Cooper, G. ;Williams, M. Science, 2001, 293, 2236.

Annexe A

**Nombres d'onde anharmoniques
de la cytosine isolée et en solution
aqueuse déterminés à un niveau
B3LYP/6-31+G(d,p)**

Description du mode	C ₀ H ₂ O					C ₄ H ₂ O+PCM				
	N°	Harm	VPT2	DM	Exp	N°	Harm	VPT2	DM	Exp
	ω_1	135	137	136		ω_{10}	172	111	173	
$\gamma(\text{NH}_2)$	ω_2	164	-2000	300		ω_{21}	429	141	417	
déformation cycle	ω_3	202	158	233		ω_{12}	210	155	214	
$\delta(\text{C}_{(2)}\text{-NH}_2)$	ω_4	358	362	366		ω_{19}	307	43	403	
déformation cycle	ω_5	397	401	371		ω_{20}	415	-15	420	
tor(C ₍₂₎ -NH ₂)	ω_6	525	566	503		ω_{29}	711	596	662	
déformation cycle	ω_7	532	520	507		ω_{27}	563	503	551	
déformation cycle	ω_8	545	529	543		ω_{26}	550	506	536	
déformation cycle	ω_9	578	571	573		ω_{28}	604	582	592	546
$\gamma(\text{NH})$	ω_{10}	624	615	610		ω_{36}	875	846	748 ?823?	
déformation cycle	ω_{11}	724	714	690		ω_{31}	738	630	707	
déformation cycle	ω_{12}	763	759	743		ω_{33}	774	758	748	
déformation cycle	ω_{13}	769	759	765		ω_{34}	793	773	781	787
$\gamma(\text{CH})\text{s}$	ω_{14}	773	759	743		ω_{35}	801	721	789	
déformation cycle	ω_{15}	924	892	903		ω_{39}	976	937	946	966
$\gamma(\text{CH})\text{as}$	ω_{16}	962	942	937		ω_{40}	989	965	971	994
$\nu_r \approx \text{C}_{(1)}\text{-N}^{(\tau)}$	ω_{17}	988	967	973		ω_{41}	1008	987	987	
$\delta(\text{NH}_2)\text{rock}$	ω_{18}	1088	1024	1063		ω_{42}	1135	1019	1102	1106
$(\delta(\text{C}_{(3)}\text{H}, \delta(\text{C}_{(4)}\text{H}))\text{scis}$	ω_{19}	1129	1105	1100		ω_{43}	1148	1098	1127	1138
$\nu_r \approx \text{C}_{(4)}\text{-N}^{(\tau)}$	ω_{20}	1220	1199	1173	1242	ω_{44}	1270	1217	1218	1224
$\nu_r \approx \text{C}_{(1)}\text{-N}^{(e)}$	ω_{21}	1264	1240	1220		ω_{45}	1310	1272	1259	1289

Description du mode	C ₀ H ₂ O					C ₄ H ₂ O+PCM				
	N°	Harm	VPT2	DM	Exp	N°	Harm	VPT2	DM	Exp
($\delta(C_{(3)}H, \delta(C_{(4)}H))$)rock	ω_{22}	1359	1339	1316	1337	ω_{46}	1393	1334	1336	1363
$\delta(NH)$	ω_{23}	1444	1409	1390	1422	ω_{47}	1474	1412	1423	1438
$\nu_r \approx C_{(2)}-N_{(9)}$	ω_{24}	1508	1478	1447	1475	ω_{48}	1530	1467	1496	1504
$\nu_r \approx C_{(2)}-C_{(3)}$	ω_{25}	1570	1525	1503	1539	ω_{49}	1562	1519	1514	1528
$\delta(NH_2)scis$	ω_{26}	1638	1593	1613	1595	ω_{54}	1664	1614	1628	1616
$\nu_r \approx C_{(3)}=C_{(4)}$	ω_{27}	1692	1651	1639	1656	ω_{55}	1681	1587	1637	1661
$\nu(C=O)$	ω_{28}	1775	1738	1697	1720	ω_{56}	1698	1605	1628	
$\nu(CH)as$	ω_{29}	3212	3062	3096		ω_{57}	3231	3119	3134	
$\nu(CH)s$	ω_{30}	3236	3099	3112		ω_{59}	3249	3095	3151	3108
$\nu(NH_2)s$	ω_{31}	3610	3501	3422	3441	ω_{62}	3370	3122	3198	
$\nu(NH)$	ω_{32}	3629	3462	3440	3471	ω_{60}	3256	3053	3033	
$\nu(NH_2)as$	ω_{33}	3751	3642	3549	3565	ω_{65}	3682	3506	3570	

Annexe B

**Nombres d'onde anharmoniques
de l'uracile isolée et en solution
aqueuse déterminés à un niveau
B3LYP/6-31+G(d,p)**

Description du mode	U,0H ₂ O							U,5H ₂ O+PCM						
	N°	Harm	VPT2	DM	Exp(gaz)	Exp(Ar)	N°	Harm	VPT2	DM	Exp			
puckering N3	ω_1	149	154	153			ω_{13}	186	171	175				
puckering N1	ω_2	167	167	166	185	185	ω_{14}	196	182	192				
$\delta(\text{OCNCO})$	ω_3	385	385	330	374	391	ω_{24}	410	404	416	401			
$\gamma(\text{C=CH})$	ω_4	395	390	410	395	411	ω_{26}	428	437	425	423			
déformation cycle	ω_5	520	515	520	512	516	ω_{30}	542	533	529	536			
déformation cycle, $\delta(\text{C=O})$	ω_6	541	532	527		537	ω_{31}	565	556	550	553			
déformation cycle AS, $\delta(\text{C=O})$	ω_7	558	549	547	545	551	ω_{32}	577	568	571	573			
$\gamma(\text{N1H})$	ω_8	563	565	550	588	562	ω_{40}	890	701	708				
$\gamma(\text{N3H})$	ω_9	675	666	623	660	662	ω_{42}	907	758	741				
$\gamma(\text{C}_{(4)}=\text{O})$	ω_{10}	724	714	707	717	718	ω_{35}	726	761	775				
$\gamma(\text{C}_{(2)}=\text{O})$	ω_{11}	743	755	720	757	757	ω_{36}	754	801	787				
ν_r	ω_{12}	771	747	753		759	ω_{38}	795	749	813	782			
$\gamma(\text{C}_{(5)}\text{H}, \text{C}_{(6)}\text{H})\text{s}$	ω_{13}	811	803	797	802	804	ω_{39}	813	817	817				
$\nu(\text{C}_{(4)}-\text{C}_{(6)}), \delta(\text{N}_{(3)}\text{H})$	ω_{14}	968	948	940	952	958	ω_{46}	1021	976	971	1000			
$\gamma(\text{C}_{(6)}\text{H}, \text{C}_{(5)}\text{H})\text{as}$	ω_{15}	969	949	947	972	987	ω_{45}	993	992	974	997			
$\nu(\text{N}_{(1)}\text{C}_{(5)}\text{C}_{(6)})$	ω_{16}	990	977	977	990	980	ω_{44}	992	992	967				
$\nu(\text{N}_{(1)}\text{C}_{(6)}), \delta(\text{C}_{(5)}\text{H})$	ω_{17}	1090	1067	1063	1082	1075	ω_{47}	1116	1091	1083	1091			
$\nu(\text{C}_{(2)}-\text{N}_{(3)}), \nu(\text{N}_{(3)}-\text{C}_{(4)})$														
$\delta(\text{C}_{(6)}\text{H}, \text{N}_{(1)}\text{H})\text{as}$	ω_{18}	1200	1170	1153	1172	1185	ω_{48}	1249	1214	1167				
$\delta(\text{C}_{(5)}\text{H}, \text{C}_{(6)}\text{H})\text{as}, \delta(\text{N}_{(1)}\text{H})$	ω_{19}	1233	1207	1197	1228	1217	ω_{49}	1260	1240	1217	1232			
$\nu(\text{C}_{(2)}-\text{N}_{(3)}), \delta(\text{N}_{(3)}\text{H}), \nu(\text{C}_{(5)}=\text{C}_{(6)}),$														

Description du mode	U,0H ₂ O						U,5H ₂ O+PCM					
	N°	Harm	VPT2	DM	Exp(gaz)	Exp(Ar)	N°	Harm	VPT2	DM	Exp	
$(\delta(\text{C}_{(6)}\text{H}),\delta(\text{N}_{(1)}\text{H}),\delta(\text{C}_{(5)}\text{H}))\text{s}$	ω_{20}	1384	1351	1303	1356	1359	ω_{50}	1409	1375	1384	1387	
$(\delta(\text{N}_{(3)}\text{H}),\delta(\text{C}_{(6)}\text{H}))\text{s}$	ω_{21}	1407	1375	1350	1387	1389	ω_{52}	1495	1413	1308	1413	
$(\delta(\text{C}_{(6)}\text{H}),\delta(\text{N}_{(1)}\text{H}),\delta(\text{C}_{(5)}\text{H}))\text{s},\delta(\text{N}_{(3)}\text{H})$	ω_{22}	1420	1389	1377	1400	1400	ω_{51}	1452	1470	1446		
$\delta(\text{N}_{(1)}\text{H}),\nu(\text{C}_{(6)}-\text{N}_{(1)})$	ω_{23}	1502	1466	1450	1461	1472	ω_{53}	1566	1588	1500	1496	
$\nu(\text{C}_{(5)}=\text{C}_{(6)})$	ω_{24}	1677	1642	1620	1641	1643	ω_{59}	1664	1622	1600	1637	
$\nu(\text{C}_{(4)}=\text{O})$	ω_{25}	1774	1754	1706	1703	1706	ω_{60}	1681	1637	1608	1677	
$\nu(\text{C}_{(2)}=\text{O})$	ω_{26}	1807	1778	1730	1756	1764	ω_{61}	1733	1684	1654	1708	
$(\nu(\text{C}_{(6)}\text{H}),\nu(\text{C}_{(5)}\text{H}))\text{as}$	ω_{27}	3222	3075	3107	3076		ω_{64}	3238	3115	3129		
$(\nu(\text{C}_{(5)}\text{H}),\nu(\text{C}_{(6)}\text{H}))\text{s}$	ω_{28}	3263	3125	3163	3124		ω_{65}	3263	3112	3158		
$\nu(\text{N}_{(3)}\text{H})$	ω_{29}	3608	3439	3197	3436	3435	ω_{62}	3171	2992	2932		
$\nu(\text{N}_{(1)}\text{H})$	ω_{30}	3649	3482	3267	3484	3485	ω_{63}	3194	2974	2988		

Annexe C

**Nombres d'onde anharmoniques
de la thymine isolée et en solution
aqueuse déterminés à un niveau
B3LYP/6-31+G(d,p)**

Description du mode	T ₀ H ₂ O							T ₄ H ₂ O+PCM						
	N°	Harm	VPT2	DM	Exp(gaz)	Exp(Ar)	N°	Harm	VPT2	DM	Exp			
déformations cycle 2 et 3	ω_1	110	112	113			ω_7	124	152	128				
τ (Me)	ω_2	150	143	153		139	ω_{10}	157	112	205				
déformations cycle 1 et 2	ω_3	154	153	154		155	ω_{13}	192	213	323				
δ (C ₍₅₎ -Me), δ (C ₍₄₎ =O)	ω_4	278	278	286		282	ω_{20}	306	273	295				
γ (C ₍₅₎ -Me)	ω_5	296	304	300		307	ω_{21}	316	383	320				
δ (OCNCO), ν (N ₍₃₎ C ₍₄₎)	ω_6	387	385	377		391	ω_{24}	415	438	423				
déformations cycle 1 et 3	ω_7	397	399	390		407	ω_{25}	431	454	423				
déformations cycle 2 et 3	ω_8	461	459	450		455	ω_{28}	481	470	480	480			
déformations cycle 3 et 2	ω_9	546	539	540		540	ω_{29}	563	544	558	560			
γ (N ₍₁₎ H)	ω_{10}	557	558	550		545	ω_{30}	621	610	821	620			
δ (C ₍₄₎ =O), δ (C ₍₂₎ =O)	ω_{11}	604	600	600		601	ω_{33}	749	766	744				
γ (N ₍₃₎ H),déformation cycle 1	ω_{12}	675	667	660		662	ω_{34}	757	745	853				
ν (C ₍₄₎ C ₍₅₎), ν (C ₍₅₎ -Me)	ω_{13}	735	722	717		727	ω_{35}	763	762	616	752			
γ (C ₍₂₎ =O)	ω_{14}	741	752	722		754	ω_{36}	813	796	731	814			
γ (C ₍₄₎ =O)	ω_{15}	763	761	743		763	ω_{37}	883	803	750				
déformation cycle 1, ν (C ₍₅₎ -Me)	ω_{16}	803	787	797		799	ω_{40}	902	882	807				
γ (C ₍₆₎ H)	ω_{17}	907	889	887		889	ω_{41}	942	925	898				
ν (N ₍₁₎ -C ₍₂₎),rot(Me), ν (C ₍₂₎ N ₍₃₎)	ω_{18}	968	943	950		959	ω_{42}	1003	971	986	986			
rot(Me),déformation cycle 1	ω_{19}	1025	1006	1005		1004	ω_{43}	1037	1007	1012				
rot(Me)as	ω_{20}	1069	1048	1043		1046	ω_{44}	1070	1042	1051				
ν (N ₍₃₎ ,C ₍₄₎), ν (N ₍₁₎ C ₍₆₎),rot(Me)	ω_{21}	1157	1129	1117		1139	ω_{45}	1191	1159	1154	1166			

Description du mode	T ₀ H ₂ O						T ₄ H ₂ O+PCM					
	N°	Harm	VPT2	DM	Exp(gaz)	Exp(Ar)	N°	Harm	VPT2	DM	Exp	
$(\delta(C_{(6)}H), \delta(N_{(1)}H), \text{as}, \nu(N_{(1)}C_{(6)}),$ $\nu(C_{(2)}N_{(3)}), \nu(N_{(3)}C_{(4)})$	ω_{22}	1203	1177	1170	1178	1183	ω_{47}	1264	1212	1205	1216	
$\nu(C_{(5)}-Me), \nu(N_{(1)}-C_{(6)}),$ déformation cycle 1	ω_{23}	1232	1200	1185		1220	ω_{46}	1246	1213	1192	1235	
$\delta(C_{(6)}H), \delta(N_{(3)}H), \nu(C_{(5)}=C_{(6)})$	ω_{24}	1375	1346	1336		1357	ω_{48}	1386	1348	1350	1352	
$\delta(N_{(3)}H), \delta(C_{(6)}H)$	ω_{25}	1403	1366	1357		1367	ω_{49}	1422	1381	1410	1384	
$\delta(N_{(1)}H), \nu(C_{(2)}N_{(3)}), \nu(C_{(4)}C_{(5)})$	ω_{26}	1423	1385	1375		1388	ω_{50}	1462	1421	1500		
umb(Me)	ω_{27}	1429	1414	1394	1393	1405	ω_{51}	1468	1430	1398	1420	
o sci(Me)	ω_{28}	1475	1446	1447		1431	ω_{52}	1486	1437	1423	1442	
$\delta(N_{(1)}H), \text{sci(Me)}, \nu(N_{(1)}C_{(6)})$	ω_{29}	1494	1459	1457	1463	1455	ω_{53}	1494	1445	1462		
sci(Me), $\nu(N_{(1)}C_{(6)}), \delta(N_{(1)}H)$	ω_{30}	1503	1464	1470		1472	ω_{54}	1566	1546	1450	1498	
$\nu(C_{(5)}=C_{(6)})$	ω_{31}	1702	1664	1640		1668	ω_{59}	1678	1645	1635	1664	
$\nu(C_{(4)}=O)$	ω_{32}	1759	1722	1684		1711	ω_{60}	1693	1652	1629		
$\nu(C_{(2)}=O)$	ω_{33}	1803	1772	1720	1772	1767	ω_{61}	1730	1666	1654	1704	
$\nu(\text{Me})_s$	ω_{34}	3045	2978	2970		2939	ω_{62}	3044	2935	2949		
$\nu(\text{Me})_{\text{as}}$	ω_{35}	3105	2960	2987	2941	2969	ω_{63}	3100	2948	2974		
$\nu(\text{Me})_{\text{as}}$	ω_{36}	3126	2981	2997	2984	2992	ω_{64}	3132	2996	3013		
$\nu(C_{(6)}H)$	ω_{37}	3214	3075	3110	3076		ω_{65}	3201	2965	3141		
$\nu(N_{(3)}H)$	ω_{38}	3608	3441	3490	3437	3432	ω_{66}	3219	2990	3167		
$\nu(N_{(1)}H)$	ω_{39}	3651	3482	3333	3484	3479	ω_{67}	3231	3078	3013		

Annexe D

**Nombres d'onde anharmoniques
de la guanine isolée et en solution
aqueuse déterminés à un niveau
B3LYP/6-31+G(d,p)**

Description du mode	G ₀ H ₂ O						G ₃ H ₂ O+PCM					
	N°	Harm	VPT2	DM	Exp(Ar)	N°	Harm	VPT2	DM	Exp		
torsion molecule, $\gamma(C_{(2)}NH_2)$	ω_1	139	137	140		ω_7	149	145	143			
torsion molecule, $\gamma(C_{(6)}=O)$	ω_2	160	162	163		ω_9	182	218	183			
butt molecule	ω_3	200	199	200		ω_{10}	210	235	213			
$\delta(\text{molécule}), (\delta(C_{(2)}-NH_2), \delta(C_{(6)}=O))s$	ω_4	318	304	317		ω_{17}	340	344	330	380		
$\gamma(N_{(11)}-H_{(12)})$	ω_5	327	292	310		ω_{16}	310	-17	307			
$(\delta(C_{(2)}-NH_2), \delta(C_{(6)}=O))s$	ω_6	335	328	327		ω_{19}	358	376	350			
$\gamma(C_{(2)}N_{(3)}C_{(4)})$	ω_7	357	353	352		ω_{18}	352	360	347			
$(sqz(C_{(2)}N_{(3)}C_{(4)}), sqz(N_{(1)}C_{(6)}C_{(5)}))s$	ω_8	487	480	483		ω_{22}	498	485	493	500		
$sqz(C_{(2)}N_{(1)}C_{(6)}), \delta(N_{(11)}H_{(13)})$	ω_9	527	513	523	514	ω_{24}	541	542	523			
$(\gamma(N_{(1)}H), \gamma(N_{(9)}H))as$	ω_{10}	534	533	510	520	ω_{25}	552	565	527			
$\gamma(N_{(11)}H_{(13)})$	ω_{11}	545	363	430		ω_{26}	576	611	450	550		
$(\gamma(N_{(1)}H), \gamma(N_{(9)}H))s,$	ω_{12}	596	584	572	589	ω_{31}	698	693	660			
breath molecule	ω_{13}	626	615	618	595	ω_{27}	644	629	633	650		
$\gamma(N_{(7)}C_{(8)}N_{(9)}), (\gamma(C_{(8)}H), \gamma(N_{(9)}H))as$	ω_{14}	659	650	643	645	ω_{28}	657	657	640	642		
$sqz(N_{(1)}C_{(2)}N_{(11)}), \delta(C_{(6)}=O)$	ω_{15}	668	662	654	654	ω_{29}	682	699	673			
$\gamma(N_{(1)}H)$	ω_{16}	694	682	655	709	ω_{37}	856	804	710	865		
$\gamma(N_{(1)}C_{(2)}N_{(3)})$	ω_{17}	730	707	687	721	ω_{32}	729	718	697	700		
$\gamma(N_{(1)}C_{(6)}C_{(5)}), \gamma(C_{(8)}H)$	ω_{18}	765	770	707	778	ω_{33}	771	825	712			
$\gamma(C_{(8)}H), \gamma(C_{(4)}C_{(5)}C_{(6)})$	ω_{19}	820	803	802	855	ω_{36}	845	818	825			
$sqz(C_{(4)}N_{(9)}C_{(8)}), \delta(N_{(11)}H_{(13)})$	ω_{20}	835	822	822	794	ω_{35}	833	815	815			
$sqz(N_{(7)}C_{(8)}N_{(9)})$	ω_{21}	945	931	932	932	ω_{38}	945	939	933	930		

Description du mode	G ₀ H ₂ O					G ₃ H ₂ O+PCM				
	N°	Harm	VPT2	DM	Exp(Ar)	N°	Harm	VPT2	DM	Exp
$(\delta(\mathbf{N}_{(1)}\mathbf{H}), \delta(\mathbf{N}_{(11)}\mathbf{H}_{(12)}), \delta(\mathbf{N}_{(11)}\mathbf{H}_{(13)}))\mathbf{s},$ $\nu(\mathbf{N}_{(1)}\mathbf{C}_{(6)})$	ω_{22}	1047	1028	1012	1010	ω_{40}	1054	1055	1023	1050
$(\delta(\mathbf{C}_{(8)}\mathbf{H}), \delta(\mathbf{N}_{(9)}\mathbf{H}))\mathbf{as}, \nu(\mathbf{C}_{(8)}\mathbf{N}_{(9)})$	ω_{23}	1074	1047	1032	1018	ω_{41}	1091	1039	1063	
$\nu(\mathbf{N}_{(1)}\mathbf{C}_{(6)}), (\delta(\mathbf{N}_{(11)}\mathbf{H}_{(12)}), \delta(\mathbf{N}_{(11)}\mathbf{H}_{(13)}))\mathbf{s},$ $\nu(\mathbf{C}_{(8)}\mathbf{N}_{(9)})$	ω_{24}	1076	1051	1055	1052	ω_{42}	1114	1102	1084	
$(\delta(\mathbf{N}_{(11)}\mathbf{H}_{(12)}), \delta(\mathbf{N}_{(11)}\mathbf{H}_{(13)}))\mathbf{s}, \nu(\mathbf{N}_{(1)}\mathbf{C}_{(6)})$	ω_{25}	1141	1109	1107	1104	ω_{43}	1172	1144	1123	1160
$\nu(\mathbf{C}_{(4)}\mathbf{N}_{(9)}), (\delta(\mathbf{C}_{(8)}\mathbf{H}), \delta(\mathbf{N}_{(9)}\mathbf{H}))\mathbf{s}$	ω_{26}	1176	1150	1136	1140	ω_{44}	1200	1151	1147	1185
$\delta(\mathbf{C}_{(8)}\mathbf{H}), \nu(\mathbf{N}_{(7)}\mathbf{C}_{(8)})$	ω_{27}	1310	1284	1277	1271	ω_{45}	1322	1276	1274	1260
$\delta(\mathbf{N}_{(1)}\mathbf{H}), \delta(\mathbf{N}_{(9)}\mathbf{H}), \delta(\mathbf{N}_{(11)}\mathbf{H}_{(13)})$	ω_{28}	1337	1313	1283		ω_{47}	1377	1371	1300	1350
$(\nu(\mathbf{C}_{(4)}\mathbf{N}_{(9)}), \nu(\mathbf{C}_{(5)}\mathbf{N}_{(7)}))\mathbf{s}, \delta(\mathbf{N}_{(9)}\mathbf{H})$	ω_{29}	1365	1336	1303	1329	ω_{46}	1370	1296	1310	1329
$\nu(\mathbf{C}_{(8)}\mathbf{N}_{(9)}), (\delta(\mathbf{C}_{(8)}\mathbf{H}), \delta(\mathbf{N}_{(9)}\mathbf{H}))\mathbf{as}$	ω_{30}	1390	1348	1345	1361	ω_{48}	1400	1383	1347	
$(\nu(\mathbf{N}_{(1)}\mathbf{C}_{(2)}), \nu(\mathbf{C}_{(4)}\mathbf{C}_{(5)}))\mathbf{as},$ $(\delta(\mathbf{N}_{(1)}\mathbf{H}), \delta(\mathbf{N}_{(11)}\mathbf{H}_{(13)}))\mathbf{s}$	ω_{31}	1441	1408	1367	1405		1428	1395	1365	1390
$\nu(\mathbf{N}_{(7)}\mathbf{C}_{(8)}), \delta(\mathbf{H})$	ω_{32}	1516	1482	1447	1506	ω_{50}	1517	1482	1443	1450
$(\nu(\mathbf{N}_{(1)}\mathbf{C}_{(2)}), \nu(\mathbf{C}_{(4)}\mathbf{C}_{(5)}))\mathbf{s}, \delta(\mathbf{N}_{(11)}\mathbf{H}_{(13)})$	ω_{33}	1559	1522	1492	1547	ω_{51}	1556	1521	1495	1540
$\nu(\mathbf{N}_{(3)}\mathbf{C}_{(4)}), \nu(\mathbf{C}_{(4)}\mathbf{N}_{(9)}), \delta(\mathbf{N}_{(9)}\mathbf{H})$	ω_{34}	1610	1573	1535	1577	ω_{52}	1611	1570	1550	
$\nu(\mathbf{C}_{(2)}\mathbf{N}_{(3)}), \delta(\mathbf{N}_{(1)}\mathbf{H}), \delta(\mathbf{N}_{(1)}\mathbf{H}_2)$ $\delta(\mathbf{NH}_2)$	ω_{35}	1621	1581	1557	1588	ω_{55}	1641	1599	1605	
$\nu(\mathbf{C}_{(6)}=\mathbf{O}), \nu(\mathbf{C}_{(5)}\mathbf{C}_{(6)}), \delta(\mathbf{N}_{(1)}\mathbf{H})$	ω_{36}	1667	1631	1640	1629	ω_{56}	1649	1605	1607	1625
$\nu(\mathbf{C}_{(8)}\mathbf{H})$	ω_{37}	1797	1761	1717	1749	ω_{58}	1691	1665	1623	
$\nu(\mathbf{NH}_2)\mathbf{s}$	ω_{38}	3266	3135	3153		ω_{59}	3275	3118	3153	
	ω_{39}	3585	3439	3480	3428	ω_{64}	3534	3392	3457	

Description du mode	G ₁ 0H ₂ O						G ₃ 3H ₂ O+PCM					
	N°	Harm	VPT2	DM	Exp(Ar)	N°	Harm	VPT2	DM	Exp		
$\nu(\text{N}_{(1)}\text{H})$	ω_{40}	3599	3428	3247	3439	ω_{61}	3348	3177	2970			
$\nu(\text{N}_{(9)}\text{H})$	ω_{41}	3658	3491	3327	3493	ω_{65}	3641	3485	3277			
$\nu(\text{NH}_2)_{\text{as}}$	ω_{42}	3703	3540	3577	3526	ω_{66}	3705	3556	3553			

Annexe E

**Nombres d'onde anharmoniques
de l'adénine isolée et en solution
aqueuse déterminés à un niveau
B3LYP/6-31+G(d,p)**

Description du mode	A ₁ 0H ₂ O						A ₁ 3H ₂ O+PCM					
	N°	Harm	VPT2	DM	Exp(Ar)	Exp(gaz)	N°	Harm	VPT2	DM	Exp	
inv(NH ₂)	ω ₁	26	-79094	167	184	162	ω ₁₇	360	309	364		
γ(NH ₂)	ω ₂	166	226	167	214	208	ω ₁₁	236	-165	213		
oop mode	ω ₃	219	113	214		244ω ₇	167	155	170			
δ(C ₍₆₎ -N ₍₁₂₎)	ω ₄	277	283	282	276	270	ω ₁₆	327	300	303	306	
oop mode (N ₍₁₎ C ₍₂₎ N ₍₃₎)	ω ₅	299	297	214			ω ₂₃	565	546	543	542	
δ(N ₍₉₎ H),twist(NH ₂)	ω ₆	517	636	513	503	506	ω ₁₀	217	206	213		
sqz(N ₍₃₎ C ₍₄₎ N ₍₉₎),rock(NH ₂)	ω ₇	520	590	529			ω ₂₂	539	512	523		
sqz(N ₍₁₎ C ₍₆₎ C ₍₅₎),sqz(C ₍₂₎ N ₍₃₎ C ₍₄₎)	ω ₈	531	529	524			ω ₂₁	537	525	530		
twist(NH ₂)	ω ₉	540	816	513	513	515	ω ₂₆	674	634	647		
(γ(C ₍₂₎ H),γ(N ₍₉₎ H))s	ω ₁₀	575	564	557	566	563	ω ₂₅	646	622	630	626	
ν(C ₍₅₎ C ₍₆₎),rock(NH ₂)	ω ₁₁	618	612	608	610	600	ω ₂₄	621	613	613	578	
(γ(N ₍₉₎ H),γ(C ₍₈₎ H))as	ω ₁₂	666	648	648	655	650	ω ₃₃	887	805	857		
(γ(N ₍₁₎ C ₍₆₎ C ₍₅₎),γ(N ₍₃₎ C ₍₄₎ C ₍₅₎))as	ω ₁₃	685	664	650	678		ω ₃₀	800	787	710		
ν(N ₍₃₎ C ₍₄₎),ν(C ₍₅₎ N ₍₇₎),ν(C ₍₄₎ N ₍₉₎)	ω ₁₄	726	719	717	717		ω ₂₈	725	707	721	728	
(γ(C ₍₆₎ C ₍₅₎ C ₍₄₎),γ(C ₍₈₎ H))as	ω ₁₅	806	787	730	802	801	ω ₂₇	700	655	647		
γ(C ₍₈₎ H)	ω ₁₆	847	847	947	848	847	ω ₃₂	872	818	857		
sqz(N ₍₁₎ C ₍₂₎ N ₍₃₎)	ω ₁₇	899	887	883	887		ω ₃₄	913	892	885		
sqz(N ₍₇₎ C ₍₈₎ N ₍₉₎)	ω ₁₈	942	925	926	927	926	ω ₃₆	964	944	943	914	
γ(C ₍₂₎ H)	ω ₁₉	974	953	946	958	957	ω ₃₇	974	922	953	964	
rock(NH ₂),ν(N ₍₁₎ C ₍₂₎)	ω ₂₀	1010	699	1007	1017	1018	ω ₃₈	1039	10281007			
(δ(N ₍₉₎ H),δ(C ₍₈₎ H))as,ν(N ₍₉₎ C ₍₈₎)	ω ₂₁	1086	1058	1007	1061	1053	ω ₄₀	1182	1139	1127	1162	

Description du mode	A ₁ 0H ₂ O						A ₁ 3H ₂ O+PCM					
	N°	Harm	VPT2	DM	Exp(Ar)	Exp(gaz)	N°	Harm	VPT2	DM	Exp	
$\nu(\text{C}_{(4)}\text{N}_{(9)}), \delta(\text{C}_{(8)}\text{H})$	ω_{22}	1148	1124	1238	1127	1126	ω_{39}	1143	1095	1100	1134	
rock(NH ₂), $\nu(\text{C}_{(5)}\text{N}_{(7)})$	ω_{23}	1245	1120	1209	1229	1234	ω_{41}	1249	1225	1193		
$(\delta(\text{C}_{(8)}\text{H}), \delta(\text{N}_{(9)}\text{H}))\text{s}$	ω_{24}	1271	1247	1210	1240		ω_{42}	1291	1231	1243	1254	
$(\nu(\text{C}_{(2)}\text{N}_{(3)}), \nu(\text{C}_{(5)}\text{C}_{(6)}))\text{as}$	ω_{25}	1340	1297	1291	1290	1280	ω_{43}	1341	1306	1273	1314	
$\nu(\text{N}_{(1)}\text{C}_{(2)}), \nu(\text{C}_{(5)}\text{N}_{(7)})$	ω_{26}	1366	1356	1315	1328	1326	ω_{44}	1376	1344	1316	1332	
$\delta(\text{C}_{(2)}\text{H}), (\delta(\text{N}_{(9)}\text{H}), \delta(\text{C}_{(8)}\text{H}))\text{as},$ $\nu(\text{N}_{(9)}\text{C}_{(8)})$												
ω_{27}		1373	1337	1320	1345	1346	ω_{45}	1392	1353	1328	1368	
$(\delta(\text{N}_{(9)}\text{H}), \delta(\text{C}_{(2)}\text{H}))\text{as}, \nu(\text{C}_{(4)}\text{N}_{(9)})$	ω_{28}	1423	1386	1321	1389		ω_{47}	1489	1435	1412	1454	
$\nu(\text{C}_{(4)}\text{C}_{(5)}), \delta(\text{C}_{(2)}\text{H})$	ω_{29}	1440	1432	1565	1419	1415	ω_{46}	1445	1402	1360		
$\nu(\text{N}_{(1)}\text{C}_{(6)}), \delta(\text{C}_{(2)}\text{H})$	ω_{30}	1509	1467	1326	1474	1468	ω_{48}	1512	1461	1435	1486	
$\nu(\text{N}_{(7)}\text{C}_{(8)})$	ω_{31}	1523	1492	1473	1482		ω_{49}	1529	1473	1450		
sci(NH ₂), $(\nu(\text{C}_{(4)}\text{C}_{(5)}), \nu(\text{N}_{(1)}\text{C}_{(2)}))\text{s}$	ω_{32}	1613	1537	1567	1599		ω_{50}	1615	1568	1530	1542	
$(\delta(\text{N}_{(3)}\text{C}_{(4)}), \nu(\text{N}_{(1)}\text{C}_{(6)}))\text{s}$	ω_{33}	1643	1580	1558	1612		ω_{53}	1637	1571	1533		
sci(NH ₂), $\nu(\text{C}_{(6)}\text{N}_{(12)})$	ω_{34}	1666	1651	1567	1633	1625	ω_{55}	1669	1654	1627	1656	
$\nu(\text{C}_{(2)}\text{H})$	ω_{35}	3192	3050	2853	3057	3061	ω_{56}	3191	3062	3080		
$\nu(\text{C}_{(8)}\text{H})$	ω_{36}	3262	3133	3070			ω_{59}	3278	3137	3173		
$\nu(\text{NH}_2)\text{s}$	ω_{37}	3617	3804	3157	3452	3434	ω_{62}	3452	3349	3400		
$\nu(\text{N}_{(9)}\text{H})$	ω_{38}	3628	3490	3552	3508	3501	ω_{57}	3206	2898	3103		
$\nu(\text{NH}_2)\text{as}$	ω_{39}	3758	4045	3557	3569	3552	ω_{63}	3681	3511	3547		

Annexe F

Publication sur la cytosine

Vibrational Analysis Beyond the Harmonicity from Ab Initio Molecular Dynamics: Case of Cytosine in Its Anhydrous and Aqueous Forms

PHILIPPE CARBONNIERE,¹ SANDRINE THICOIPE,¹
THIBAUT VERY,² XAVIER ASSFELD²

¹Groupe de Chimie Théorique et Réactivité, IPREM/ECP UMR CNRS 5254,
Université de Pau et des Pays de l'Adour, F-64000 Pau, France

²Equipe de Chimie et Biochimie Théoriques, UMR CNRS 7565, Nancy Université, F 54506
Vandoeuvre-lès-Nancy Cedex, France

Received 26 April 2011; accepted 26 May 2011

Published online 23 February 2012 in Wiley Online Library (wileyonlinelibrary.com).

DOI 10.1002/qua.23194

ABSTRACT: This study provides the set of theoretical anharmonic fundamental transitions for cytosine in its anhydrous and hydrated forms from a vibrational analysis issuing from a time-dependent approach. This evaluation is compared to the time-independent VPT2 approach and to the experimental data available in the literature for the two forms. To investigate the vibrational property of a solute in a polar protic solvent, that is, the aqueous form of cytosine in this particular case, a practicable computational strategy to properly take into account the solvent effect is proposed. This is done by considering a combination of polarizable continuum model and of discrete solvent molecules model. The number of solvent molecules was chosen in such a way that the solute-solvent complex presents only one dominant stable conformer at 298 K. The quality of this combined solvent model is illustrated by comparing the structural and anharmonic vibrational properties achieved with those obtained from a purely implicit and purely explicit solvent model. © 2012 Wiley Periodicals, Inc. *Int J Quantum Chem* 112: 2221–2230, 2012

Key words: anharmonic vibrational computations; solvent effect; cytosine; ab initio molecular dynamics

Correspondence to: P. Carbonniere; e-mail: philippe.carbonniere@univ-pau.fr

1. Introduction

The vibrational spectra of nucleic acids are extremely difficult to interpret because of the high complexity and low symmetry of these molecules. Moreover, modifications of the surroundings of these systems can increase in a considerable way the number of structures needed to interpret the spectra.

Cytosine is one of the constituents of nucleic acids, which has been the most studied by infrared (IR) spectroscopy [1]. This method has already revealed that the geometry of the amino group is very sensitive to the interaction with water molecules.

IR studies of cytosine in gas phase have never been carried out. However, experimental IR spectra of cytosine monomer in Ar and N₂ matrixes at very low temperature were realized and most of the fundamental transitions, CH stretching excepted, were reported in the literature [2, 3]. About aqueous cytosine, IR and Raman transitions in the spectral region 1800–400 cm⁻¹ were reported [4, 5] for different values of pH.

Finally, polycrystalline cytosine was studied by IR spectroscopy in its anhydrous and monohydrated forms [4, 6]. A specific study of NH stretching evolutions was also presented as a function of temperature (10–300 K) [1].

From a theoretical point of view, different calculations have been performed on cytosine systems including explicit molecules of water. Cytosine tautomers [7, 8] were first identified at different levels of theory like MP2/TZP [9], B3LYP [10], CCSD/cc-pVTZ [11], and their corresponding calculated harmonic IR spectra with the B3LYP method were reported [10]. This type of method was mainly used with a Pople's double zeta quality basis set. Structural properties were also investigated on protonated [4, 12, 13] and anionic [14] forms at the B3LYP level of theory.

The theoretical structures of hydrated cytosine were first determined considering an implicit aqueous solvent with the polarizable continuum model (PCM) model [15]. On that point, geometries and energies computed at different levels of theory such as M052X [16], wB97XD [17], CAM-B3LYP [18], Time-dependant density functional theory (TD-DFT) model, and MP2 were reported [19, 20]. Structural properties of hydrated cytosine were also investigated with mixed approaches combining PCM and discrete model from one to five water

molecules [21, 22] or purely discrete approaches considering up to 14 water molecules at the B3LYP, MP2, and PM6 levels [23, 24].

Vibrational properties of such species were investigated at the harmonic level from a solvent discrete model containing up to five water molecules with the B3LYP model chemistry [25–28].

To the best of our knowledge, no anharmonic study about hydrated cytosine was reported in the literature. The anharmonic problem is usually undertaken using two different ways. The first way consists of a time-independent/static approach based on a quantum mechanical (QM)/stationary-state picture of the system followed by a perturbational (VPT2) [29], variational (VCI) [30], or variational-perturbational [31, 32] resolution of the vibrational Schrödinger equation. The second way is a time-dependent/dynamic approach in which an anharmonic vibrational spectrum is modeled from a Fourier transform of the time correlation functions of atomic positions or atomic velocities or dipole moment stemmed from the simulation of a molecular dynamics (MD) trajectory. This second approach is neither limited by the size of the molecular system nor by the kind of the vibrational motion we wish to describe since the method do not require an analytical form of the potential electronic surface and is not hampered by a diagonalization procedure.

This study is aiming at providing the first set of theoretical anharmonic fundamental transitions of cytosine from a time-dependent approach in the medium IR region. The results will be compared to the available experimental data and to the VPT2 approach which tends to be the most widely used method for the investigation of large molecular systems in this field. Structural properties of hydrated cytosine are investigated from a discrete solvent containing from 1 to 25 water molecules. In the next section, the computational details and strategy are given. Results are then presented and discussed before the concluding remarks are stated.

2. Computational Details and Methods

Structural optimizations, harmonic frequencies, and anharmonic frequencies issuing from the VPT2 evaluation [33] were performed with the Gaussian09 program [34]. Electronic structure calculations were performed at the DFT level using

the three-parameter hybrid functional of Becke (B3LYP) and a Pople's valence double zeta basis set including diffuse and polarization functions, 6-31+G(d,p). This model chemistry was chosen since it has been previously shown that, for the prediction of harmonic and anharmonic force constants and in the case of small organic [35, 36] and radical [37] systems, it is able to approach the results obtained using the more expensive CCSD(T) [38]/cc-pVTZ [39] level of theory. Geometries of hydrated cytosine clusters [$C_nH_{2n}O$] ($n = 1,5$) were found from the global search algorithm of minima exploration called GSAM code [40]. This algorithm includes two parts: (i) the generation of an initial guess that consists of a random generation of several hundreds of [$C_nH_{2n}O$] molecular clusters from the geometry of their monomeric units and (ii) their preoptimization at the PM6 level of theory followed by a full B3LYP/6-31+G(d,p) optimization of the physically reasonable isomers at a given temperature (Boltzmann probability $> 5\%$ at 300 K).

The VPT2 evaluation of the anharmonic frequencies requires the determination of an analytical potential function determined as follows: from a minimum energy structure of a given system, a quartic force field is built in which the third and fourth derivatives are computed by $6N - 11$ ($N =$ number of atoms) numerical differentiation of analytical second derivatives. The potential provided by the Gaussian code is given by:

$$V(q_1, \dots, q_{3N-6}) = \frac{1}{2} \sum_i \omega_i q_i^2 + \sum_{i,j,k} k_{ijk} q_i q_j q_k + \sum_{i,j,k,l} k_{ijkl} q_i q_j q_k q_l$$

where q_i the i th dimensionless normal coordinate as linear combination of Cartesian coordinates, and ω_i , k_{ijk} , k_{ijkl} the harmonic, cubic, and quartic force constants, respectively, expressed in cm^{-1} . Note that only the anharmonic force constants involving the trimode interactions ($k = l$) are determined by numerical differentiation. From a technical point of view, the quartic force field is obtained using a step size of 0.01 Å for the numerical differentiation of harmonic frequencies, tight convergence criteria for structural optimizations, and fine grids for integral evaluation (that is at least 99 radial and 590 angular points).

Ab initio MD simulations were performed using the atom-centered density matrix (ADMP) formal-

ism [41–43] which uses localized Gaussian basis sets. MD simulations of the selected systems were performed from their optimized geometry in the microcanonical ensemble with a time-step of 0.2 fs over an overall period of 10 ps. Concerning the temperature of the simulation, note that, as mentioned by Tavan and coworkers [44, 45], the classically determined frequencies underestimate anharmonic shifts since the classical amplitudes of the motion are smaller than their quantum counterparts at the same temperature. Thus, the initial averaged kinetic energy was set up to perform simulations averaging the temperature of 600 K to attempt to recover the overall anharmonicity of the vibrational modes. The vibrational analysis from molecular dynamic program (VAMD) code has then been used to perform the spectral analysis from ADMP ab initio trajectories [46].

The assignment of the vibrational bands (see Ref. 46 and therein) is based on the following.

Starting from a set of mass-weighted atomic velocity V_t issuing from a simulation and projected along the $3N - 6$ normal modes $\langle \vec{c}_j \rangle$, the Fourier transform of the projected velocities autocorrelation functions gives rise to a set of $3N - 6$ vibrational densities of states (VDOS) spectra of the generalized modes $\hat{P}_j(\vec{v})$:

$$\hat{P}_j(\vec{v}) = \int dt \langle V_{j0} V_{jt} \rangle e^{i\omega t}$$

As mentioned in the works of Tavan and coworkers [40, 41], the position of the main peak corresponds to the wavenumber of the mode j while the satellite peaks correspond to vibrational states for which the vector displacement is overlapped with the vector $\langle \vec{c}_j \rangle$. This is namely the case for anharmonically coupled states involved in vibrational resonances.

Note moreover that the eigenvectors $\langle \vec{c}_j \rangle$ are obtained by diagonalizing the covariance matrix $[K]$ of mass-weighted velocities issuing from the MD simulation defined as:

$$K_{ij} = \frac{1}{2} \langle (m_i m_j)^{1/2} v_i v_j \rangle$$

It should indeed be noticed that velocity autocorrelation function can be projected on any internal coordinate [47] or related local modes for a more comprehensive assignment of the peaks in terms of functional group involved in the transitions.

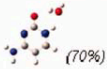
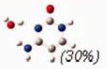
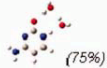
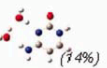
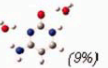
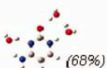
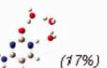
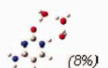
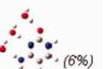
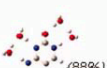
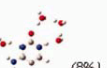

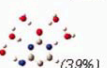

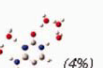
Finally, as discussed in the next section, a preliminary QM/molecular mechanics (MM)

computation of aqueous cytosine was also performed. The system is composed of a cytosine molecule solvated in a cubic box of $21 \times 21 \times 21 \text{ \AA}^3$ with 267 water molecules. All calculations were done using a modified version of the TINKER [48] package coupled with a modified version of Gaussian03 package. The cytosine molecule is treated at B3LYP/6-31G* level of theory while the water molecules are treated with AMBER99 force field [49] with both angles and bond lengths constrained to optimal values defined within the force field using RATTLE algorithm [50]. Periodic boundary conditions were used throughout the calculations. In the first step, the energy of the whole system was minimized to prevent steric clashes inherent to the building of the box. For this step only, the cytosine molecule was described with AMBER99 force field. In the second step, a thermalization phase of 200 fs was carried out using the Beeman integration algorithm [51]. The Berendsen thermostat [52] was used, working in NVT ensemble, with O, i, p time step. From this system, the structure comprising the cytosine molecule and its closest 25 water molecules was then extracted and fully optimized at the B3LYP/6-31+G(d,p) level of theory.

3. Results and Discussion

The model chosen to represent the aqueous cytosine is made by considering a combination of polarizable continuum model and discrete water molecules model. Concerning the explicit part of the solvent model, we chose to consider a quite small number of QM water molecules able to form only one dominating complex structure at 298 K. The structural investigation of $[\text{C},n\text{H}_2\text{O}]$ clusters ($n = 1-5$) was carried out according to the procedure detailed in the previous section, by generating an initial guess of 700 structures for each hydrated cytosine complex to find the most probable isomers at 298 K. These are reported in the Table I with their corresponding Boltzmann probability computed from their relative stability based on their zero-point vibrational energy corrected energies. The results, in perfect agreement with those obtained in Ref. 25, show that at least two forms are favored at this temperature for each cluster, excepted for the $[\text{C},4\text{H}_2\text{O}]$ complex for which the most stable form represents about 90% of the overall population. From these results,

TABLE I
Description of the most stable structures of $\text{C},n\text{H}_2\text{O}$ clusters obtained with the B3LYP/6-31+G(d,p) level of theory, with their corresponding Boltzmann probabilities (%P) at 298 K. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

	Structure 1 (%P)	Structure 2 (%P)	Structure 3 (%P)	Structure 4 (%P)
$\text{C},1\text{H}_2\text{O}$	 (70%)	 (30%)		
$\text{C},2\text{H}_2\text{O}$	 (75%)	 (34%)	 (9%)	
$\text{C},3\text{H}_2\text{O}$	 (68%)	 (17%)	 (8%)	 (6%)
$\text{C},4\text{H}_2\text{O}$	 (88%)	 (8%)		
$\text{C},5\text{H}_2\text{O}$	 (44%)	 (39%)	 (7%)	 (4%)

the tetrahydrated cytosine has been considered to form the discrete part of the model.

Table II illustrates the evolution of the most sensitive geometrical parameters of cytosine with respect to the hydration (see also Fig. 1 for the atom numbering scheme). The values corresponding to anhydrous cytosine, noted $[\text{C},0\text{H}_2\text{O}]$, and to the tetrahydrated cytosine, noted $[\text{C},4\text{H}_2\text{O}]$, are reported. The solvation effects with PCM on the two systems were also considered and lead to the values labeled $[\text{C},0\text{H}_2\text{O}+\text{PCM}]$ and $[\text{C},4\text{H}_2\text{O}+\text{PCM}]$.

The variations of the internal coordinates involving the atoms which take part in the formation of hydrogen bonds cannot be all reproduced with the polarizable continuum model alone. Indeed, the values collected in Table II show that the $\text{N}_{(7)}\text{H}_{(5)}$ and $\text{N}_{(9)}\text{H}_{(10)}$ bond lengths increased by about 0.02 \AA in the complex $[\text{C},4\text{H}_2\text{O}]$. The relevance of the $[\text{C},4\text{H}_2\text{O}+\text{PCM}]$ model is evaluated by comparing the values of the selected geometrical parameters to their counterparts issuing from the $[\text{C},25\text{H}_2\text{O}]$ complex. This reveals some discrepancies which are no greater than 0.003 \AA for the bond lengths. Concerning the dihedral angle between the amino group and the plane of the cytosine cycle, the values differ by less than 7° between the two structures.

TABLE II
Most sensitive geometrical parameters of cytosine w.r.t. the hydration computed at the B3LYP/6-31+G(d,p) level of theory.

Parameters	C, 0H ₂ O	C, 0H ₂ O+PCM	C, 4H ₂ O	C, 4H ₂ O+PCM	C, 25H ₂ O
C ₍₁₎ O ₍₈₎	1,224	1,243	1,244	1,253	1,254
N ₍₇₎ H ₍₅₎	1,011	1,012	1,032	1,031	1,030
N ₍₉₎ H ₍₁₀₎	1,009	1,009	1,028	1,024	1,034
N ₍₉₎ H ₍₁₁₎	1,006	1,008	1,007	1,009	1,010
C ₍₃₎ H ₍₁₃₎	1,082	1,082	1,082	1,082	1,079
C ₍₄₎ H ₍₁₂₎	1,085	1,084	1,085	1,084	1,083
ϕ NH ₂	164.8	180.0	173.2	179.3	172.5

The computed wavenumbers for anhydrous (C,0H₂O) and hydrated (C,4H₂O+PCM) cytosine are compiled in Table III. The experimental data available in the literature, from the IR spectrum of cytosine monomer in Ar matrix [3] and from the IR and Raman spectra of the aqueous cytosine at pH = 7 [5], are also reported and compared to their theoretical counterparts. The harmonic values (Harm) are issued from the Hessian matrix determined at the minimum energy structure. The “static” anharmonic counterpart arises from the evaluation of the wavenumbers using the second-order perturbation theory (VPT2). As reported by Handy and coworkers [53], the VPT2 treatment provides, in absence of strong anharmonic couplings (and in absence of low frequency modes), the correct vibrational wavenumbers, especially for the stretching motions. The “dynamic” anharmonic results arise from a vibrational analysis stemmed from a classical MD trajectory. Note that each MD value corresponds to the maximum am-

plitude of a distribution function which gathers the overall of the values taken by a given vibrational mode during the MD trajectory. On these grounds, the confidence interval related to each wavenumber can be large when the environment around the corresponding mode is strongly modified during the trajectory. It follows that the confidence interval can reach ± 30 cm⁻¹ in these particular cases.

The region ranging between 1800 and 1100 cm⁻¹ is the most abundant in experimental data. The average discrepancy of the VPT2 results with respect to the experimental values is about 12 and 25 cm⁻¹ for [C,0H₂O] and for [C,4H₂O+PCM], respectively. The discrepancies are of the same magnitude between the dynamic results and the observed data in the case of the hydrated cytosine (30 and 15 cm⁻¹ for [C,0H₂O] and [C,4H₂O+PCM], respectively). Here, we point out that some of the MD values cannot be provided univocally since at least two peaks arise with the same order

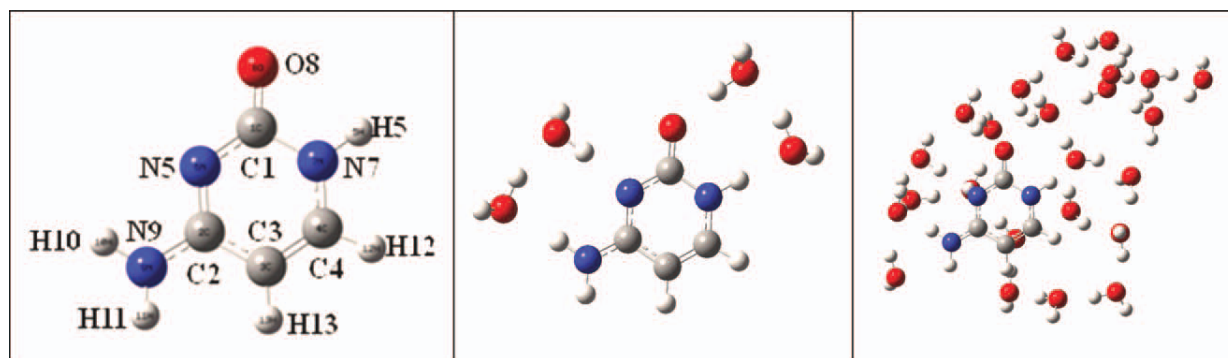


FIGURE 1. The optimized structures of [cytosine,0H₂O], [cytosine,4H₂O], and [cytosine,25H₂O] from B3LYP/6-31+G(d,p) computations with atom numbering scheme. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

TABLE III
Experimental and theoretical wavenumbers of anhydrous and aqueous cytosine between 3600 and 100 cm⁻¹.

Assignment	C,0H ₂ O				C,4H ₂ O+PCM					
	No.	Harm	VPT2	MD	Exp.	No.	Harm	VPT2	MD	Exp.
$\gamma(\text{NH}_2)$	ω_1	135	137	136		ω_{10}	172	111	173	
ring deform	ω_2	164	-2,000	300		ω_{21}	429	141	417	
$\delta(\text{C}_{(2)}-\text{NH}_2)$	ω_3	202	158	233		ω_{12}	210	155	214	
ring deform	ω_4	358	362	366		ω_{19}	307	43	403	
tor($\text{C}_{(2)}-\text{NH}_2$)	ω_5	397	401	371		ω_{20}	415	-15	420	
ring deform	ω_6	525	566	503		ω_{29}	711	596	662	
ring deform	ω_7	532	520	507		ω_{27}	563	503	551	
ring deform	ω_8	545	529	543		ω_{26}	550	506	536	
ring deform	ω_9	578	571	573		ω_{28}	604	582	592	546 ^c
$\gamma(\text{NH})$	ω_{10}	624	615	610		ω_{36}	875	846	748 ? 823 ?	
$\gamma(\text{C}=\text{O}), \gamma(\text{C}_{(3)}-\text{H})$	ω_{11}	724	714	690		ω_{31}	738	630	707	
$\nu_{12}\approx\nu_{(1)}-\nu_{(7)}, \nu_{(2)}-\nu_{(3)}, \nu_{(1)}-\nu_{(6)}, \nu_{(2)}-\nu_{(9)}$	ω_{12}	763	759	743		ω_{33}	774	758	748	
$\gamma(\text{CH})\text{s}$	ω_{13}	769	759	765		ω_{34}	793	773	781	787 ^c
$\nu_{15}\approx\nu_{(1)}-\nu_{(7)}, \nu_{(2)}-\nu_{(3)}, \delta(\text{NH}_2)\text{rock}$	ω_{14}	773	759	743		ω_{35}	801	721	789	
$\nu_{16}\approx\nu_{(1)}-\nu_{(7)}, \nu_{(2)}-\nu_{(3)}, \delta(\text{NH}_2)\text{rock}$	ω_{15}	924	892	903		ω_{39}	976	937	946	966 ^c
$\gamma(\text{CH})\text{as}$	ω_{16}	962	942	937		ω_{40}	989	965	971	994 ^c
$\nu_{17}\approx\nu_{(2)}-\nu_{(3)}$	ω_{17}	988	967	973		ω_{41}	1,008	987	987	
$\delta(\text{NH}_2)\text{rock}, \nu_{18}\approx\nu_{(1)}-\nu_{(7)}$	ω_{18}	1,088	1,024	1,063		ω_{42}	1,135	1,019	1,102	1,106 ^d
$[\delta(\text{C}_{(3)}\text{H}), \delta(\text{N}_{(7)}\text{H})]\text{scis}$	ω_{19}	1,129	1,105	1,100		ω_{43}	1,148	1,098	1,127	1,138 ^d
$[\delta(\text{C}_{(4)}\text{H}), \delta(\text{N}_{(7)}\text{H})]\text{scis}, [\delta(\text{C}_{(3)}\text{H}), \delta(\text{C}_{(4)}\text{H})]\text{scis}$	ω_{20}	1,220	1,199	1,173	1,242 ^a	ω_{44}	1,270	1,217	1,218	1,224 ^d
$\nu_{19}\approx\nu_{(1)}-\nu_{(6)}, \nu_{(2)}-\nu_{(9)}$	ω_{21}	1,264	1,240	1,220		ω_{45}	1,310	1,272	1,259	1,289 ^d
$[\delta(\text{C}_{(3)}\text{H}), \delta(\text{C}_{(4)}\text{H})]\text{rock}$	ω_{22}	1,359	1,339	1,316		ω_{46}	1,393	1,334	1,336	1,363 ^d
$\delta(\text{NH}), \nu_{20}\approx\nu_{(2)}-\nu_{(3)}$	ω_{23}	1,444	1,409	1,390		ω_{47}	1,474	1,412	1,423	1,438 ^d
$\nu_{21}\approx\nu_{(2)}-\nu_{(9)}, \delta(\text{NH}_2)\text{scis}, [\delta(\text{C}_{(3)}\text{H}), \delta(\text{C}_{(4)}\text{H})]\text{rock}$	ω_{24}	1,508	1,478	1,447		ω_{48}	1,530	1,467	1,496	1,504 ^d
$\nu_{22}\approx\nu_{(2)}-\nu_{(9)}, \nu_{(2)}-\nu_{(6)}, \delta(\text{NH})$	ω_{25}	1,570	1,525	1,503	1,595 ^{a}-1,539^b}	ω_{49}	1,562	1,519	1,514	1,528 ^d
$\delta(\text{NH}_2)\text{scis}$	ω_{26}	1,638	1,593	1,613	1,595 ^{b}}	ω_{54}	1,664	1,614	1,628	1,616 ^d
$\nu_{23}\approx\nu_{(3)}=\nu_{(4)}, \nu_{(2)}-\nu_{(6)}$	ω_{27}	1,692	1,651	1,639	1,655 ^{a}-1,656^b}	ω_{55}	1,681	1,587	1,637	1,661 ^d
$\nu(\text{C}=\text{O}), \delta(\text{NH})$	ω_{28}	1,775	1,738	1,697	1,714 ^{a}-1,720^b}	ω_{56}	1,698	1,605	1,628	
$\nu(\text{CH})\text{as}$	ω_{29}	3,212	3,062	3,096		ω_{57}	3,231	3,119	3,134	
$\nu(\text{CH})\text{s}$	ω_{30}	3,236	3,099	3,112		ω_{59}	3,249	3,095	3,151	
$\nu(\text{NH}_2)\text{s}$	ω_{31}	3,610	3,501	3,422	3,438 ^{a}-3,441^b}	ω_{62}	3,370	3,122	3,198	
$\nu(\text{NH})$	ω_{32}	3,629	3,462	3,440	3,468 ^{a}-3,471^b}	ω_{60}	3,256	3,053	3,068	
$\nu(\text{NH}_2)\text{as}$	ω_{33}	3,751	3,642	3,549	3,559 ^{a}-3,565^b}	ω_{65}	3,682	3,506	3,570	

^a Ar matrix taken from Ref. [2].
^b Ar matrix taken from Ref [3].
^c IR wavenumbers taken from Ref [5].
^d Raman wavenumbers taken from Ref [5].

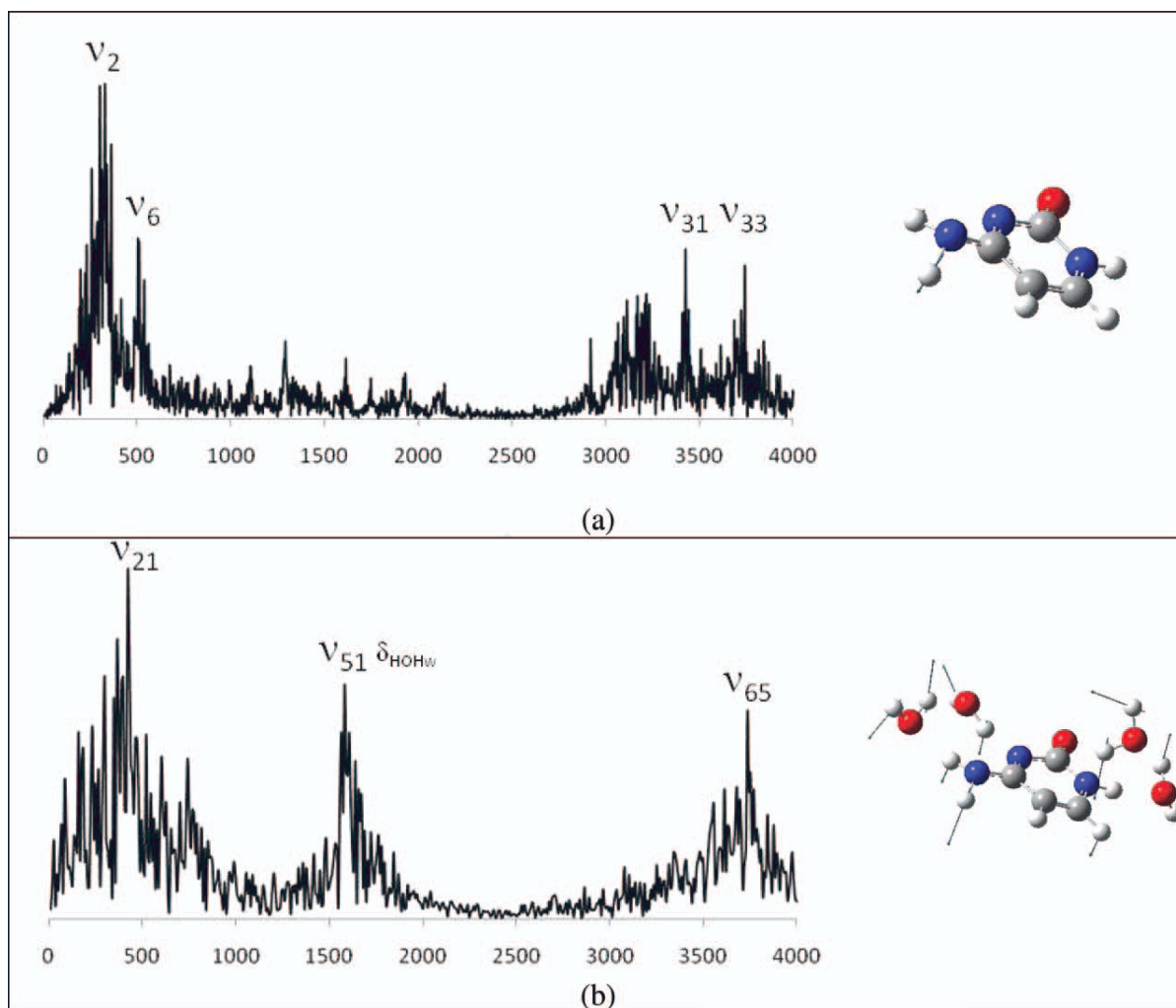


FIGURE 2. Fourier transform of the autocorrelation function of atomic velocities projected on the $\gamma(\text{NH}_2)$ normal mode for (a) [cytosine,0H₂O] and (b) [cytosine,4H₂O+PCM] (wavenumbers in cm⁻¹) and vector displacement of the corresponding motion from Hessian Analysis. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

of magnitude in the spectral region of the main peak, indicating the presence of a strong mode–mode coupling which is not easily attributable. Globally, we note a good agreement between the theoretical values and their experimental counterparts in this range, even though the qualitative variation of the wavenumbers from the anhydrous form to the aqueous form does not always conform to the available experimental trends.

The X–H stretching region comprise the three NH stretching modes noted $\nu(\text{NH}_2)_{\text{as}}$, $\nu(\text{NH}_2)_{\text{s}}$, $\nu(\text{NH})$ and the two CH stretching modes noted $\nu(\text{CH})_{\text{s}}$, $\nu(\text{CH})_{\text{as}}$. The VPT2 evaluation, especially for the $\nu(\text{NH}_2)_{\text{as}}$ and $\nu(\text{NH}_2)_{\text{s}}$ modes, is plagued

by the presence of strong couplings with the lowest modes and particularly with the out of plane motion of amino group $-\gamma(\text{NH}_2)-$ (e.g., $k_{33,2,2} = 3166 \text{ cm}^{-1}$; $k_{31,2,2} = -4760 \text{ cm}^{-1}$ for [C,0H₂O] complex). These anharmonic couplings are misvaluated due to the analytical form and the coordinate system used in such static computations [46]. As shown in Figure 2, which display the Fourier transform of the autocorrelation function of the atomic velocities projected on $\gamma(\text{NH}_2)$ normal mode, these couplings are qualitatively important and should not be neglected for a proper evaluation of the stretching modes for such molecular systems.

TABLE IV

MD evaluations of some fundamental transitions in the range 1800–1100 cm^{-1} from [C,0H₂O + PCM], [C,4H₂O], and [C,4H₂O+PCM] models and deviations w.r.t. the experimental data in aqueous solution.

No.	[C,0H ₂ O+PCM]	(Δ)	[C,4H ₂ O]	(Δ)	[C,4H ₂ O+PCM]	(Δ)	Exp.
$\delta(\text{NH}_2)\text{rock}$	1,080	-26	1,090	-16	1,102	-4	1,106 ^a
$[\delta(\text{C}_{(3)}\text{H}), \delta(\text{C}_{(4)}\text{H})]\text{scis}$	1,090	-48	1,123	-15	1,127	-11	1,138 ^a
$\nu_r \approx \text{C}_{(4)}-\text{N}_{(7)}$	1,190	-34	1,199	-25	1,218	-6	1,224 ^a
$\nu_r \approx \text{C}_{(1)}-\text{N}_{(6)}$	1,237	-52	1,246	-43	1,259	-30	1,289 ^a
$[\delta(\text{C}_{(3)}\text{H}), \delta(\text{C}_{(4)}\text{H})]\text{rock}$	1,320	-43	1,326	-37	1,336	-27	1,363 ^a
$\delta(\text{NH})$	1,390	-48	1,413	-25	1,423	-15	1,438 ^a
$\nu_r \approx \text{C}_{(2)}-\text{N}_{(9)}$	1,430	-74	1,457	-47	1,496	-8	1,504 ^a
$\nu_r \approx \text{C}_{(2)}-\text{C}_{(3)}$	1,503	-25	1,507	-21	1,514	-14	1,528 ^a
$\delta(\text{NH}_2)\text{scis}$	1,580	-36	1,645	29	1,628	12	1,616 ^a
$\nu_r \approx \text{C}_{(3)}=\text{C}_{(4)}$	1,607	-54	1,609	-52	1,637	-24	1,661 ^a
M.D. ^b		-45		-25		-13	

^a Raman wavenumbers taken from Ref. [5].

^b M. D.: mean deviation.

The region below 1100 cm^{-1} is marked by the absence of experimental data. The attention is focused on $-\gamma(\text{NH}_2)-$ and $-\text{tor}(\text{C}_{(2)}-\text{NH}_2)-$ modes of the [C,4H₂O] complex which appear more hindered because of the interaction with water molecules. We point out here that the MD evaluation of the lowest mode frequencies, and particularly the mode ν_2 , can be overestimated, considering the averaged temperature of the simulation.

Another issue is the quality of the solvent model used in this work. Table IV report three sets of MD values gathering the fundamental transitions of aqueous cytosine in the 1800–1100 cm^{-1} range for which an experimental data is available. These sets correspond to the MD evaluations issuing from the purely polarizable continuum model [C,0H₂O+PCM], the discrete model [C,4H₂O], and [C,4H₂O+PCM]. We note that the [C,0H₂O+PCM] model yields a MD evaluation quite far from its experimental counterpart (from -25 to -74 cm^{-1}) for this set of modes which are harmonically fairly coupled with $\delta(\text{NH}_2)\text{rock}$, $\delta(\text{NH}_2)\text{scis}$, and $\delta(\text{NH})$. This deviation is decreased by a factor close to 2 for most of the modes while considering an explicit model of the solvent. The combination of the two models achieves a convergence of about 15 cm^{-1} with the experimental results in aqueous solution.

4. Conclusion

This study provided the set of theoretical anharmonic fundamental transitions for cytosine

in its anhydrous and hydrated forms from a vibrational analysis issuing from a time-dependent approach. This evaluation was compared to the time-independent VPT2 approach and to the experimental data available in the literature for the two forms. In the 1800–1100 cm^{-1} region, a global mean absolute deviation of about 20 cm^{-1} was observed for the two theoretical evaluations with respect to experiment. Nevertheless, the MD evaluation seems more confident below and above this spectral region.

To investigate the vibrational property of a solute in a polar protic solvent, that is, the aqueous form of cytosine in this particular case, a practicable computational strategy to properly take into account the solvent effect was proposed. This was done by considering a combination of polarizable continuum model and discrete solvent molecules model. The number of solvent molecules was chosen in such way that the solute–solvent complex presents only one dominant stable conformer at 298 K. Thus, in the case of the cytosine, the tetrahydrated cytosine has been considered. It has also been shown that the structural properties of this cluster, when considering in addition an implicit solvent model ([C,4H₂O+PCM]), are very close to its counterpart possessing 25 explicit water molecules [C,25H₂O]. This latter form was extracted from a QM/MM simulation containing one molecule of cytosine with 267 water molecules. The quality of this combined solvent model, for the computation of the anharmonic vibrational properties, has been illustrated by comparing the

corresponding MD evaluation and those obtained from the purely implicit solvent model ([C,0H₂O+PCM]) and from the purely explicit solvent model ([C,4H₂O]) with respect to the experiment. The results show that such combined solvent model achieve, in this particular case, a convergence of 13 cm⁻¹ against 25 and 45 cm⁻¹ for the explicit and implicit solvent model, respectively. On these grounds, we will set out to adopt this computational strategy for further vibrational computations beyond the harmonicity.

ACKNOWLEDGMENTS

The authors are particularly grateful to Prof. Claude Pouchan for his valuable scientific involvement in this field, for his long lasting and continuous support, and for giving the seminal idea presented in this work.

References

1. Rozenberg, M.; Shoham, G.; Reva, I.; Fausto, R. *Spectrochim Acta, Part A: Mol Biomol Spectrosc* 2004, 60, 463.
2. Radchenko, E. D.; Sheina, G. G.; Smorygo, N. A.; Blagoi, Y. P. *J Mol Struct* 1984, 116, 387.
3. Szczesniak, M.; Szczepaniak, K.; Kwiatkowski, J. S.; KuBulat, K.; Person, W. B. *J Am Chem Soc* 1988, 110, 8319.
4. Florian, J.; Baumruk, V.; Leszczynski, J. *J Phys Chem* 1996, 100, 5578.
5. Lord, R. C.; Thomas, G. J., Jr. *Spectrochim Acta* 1967, 23A, 2551.
6. Susi, H.; Ard, J. S.; Purcell, J. M. *Spectrochim Acta* 1973, 29A, 725.
7. Ten, G. N.; Zotov, N. B.; Baranov, V. I. *Opt Spectrosc* 2009, 107, 235.
8. Fogarasi, G. *J Phys Chem A* 2002, 106, 1381.
9. Møller, C.; Plesset, M. S. *Phys Rev* 1934, 46, 618.
10. Becke, A. D. *J Chem Phys* 1993, 98, 5648.
11. Purvis, G. D., III; Bartlett, R. J. *J Chem Phys* 1982, 76, 1910.
12. Salpin, J. Y.; Guillaumont, S.; Tortajada, J.; MacAleese, L.; Lemaire, J.; Maitre, P. *ChemPhysChem* 2007, 8, 2235.
13. Bakker, J. M.; Salpin, J. Y.; Maitre, P. *Int J Mass Spectrom* 2009, 283, 214.
14. Smith, D. M. A.; Jalbout, A. F.; Smets, J.; Adamowicz, L. *Chem Phys* 2000, 260, 45.
15. (a) Tomasi, J.; Miertus, S.; Scrocco, E. *Chem Phys* 1981, 55, 117; (b) Tomasi, J.; Persico, M. *Chem Rev* 1994, 94, 2027; (c) Cammi, R.; Tomasi, J. *J Comput Chem* 1995, 16, 1449; (d) Cancés, E.; Mennucci, B.; Tomasi, J. *J Chem Phys* 1997, 107, 3032.
16. Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J Chem Theory Comput* 2006, 2, 364.
17. Chai, J. D.; Head-Gordon, M. *PCCP* 2008, 10, 6615.
18. Yanai, T.; Tew, D. P.; Handy, N. C. *Chem Phys Lett* 2004, 393, 51.
19. Shukia, M. K.; Leszczynski, J. *Mol Phys* 2010, 108, 3131.
20. Shishkin, O. V.; Gorb, L.; Leszczynski, J. *Struct Chem* 2009, 20, 743.
21. Alemán, C. *Chem Phys* 1999, 244, 151.
22. Alemán, C. *Chem Phys Lett* 1999, 302, 461.
23. Van Mourik, T.; Danilov, V. I.; Gonzalez, E.; Deriabina, A.; Poltev, V. I. *Chem Phys Lett* 2007, 445, 303.
24. Shishkin, O. V.; Gorb, L.; Leszczynski, J. *J Phys Chem B* 2000, 104, 5357.
25. Chandra, A. K.; Michalska, D.; Wysokinsky, R.; Zeegers-Huyskens, T. *J Phys Chem A* 2004, 108, 9593.
26. Chandra, A. K.; Nguyen, M. T.; Zeegers-Huyskens, T. *J Mol Struct* 2000, 519, 1.
27. Kim, S.; Schaefer, H. F., III. *J Chem Phys* 2007, 126, 64301.
28. Mazzuca, D.; Marino, T.; Russo, N.; Toscano, M. *J Mol Struct (THEOCHEM)* 2007, 811, 161.
29. Clabo, D. A.; Allen, W. D.; Remington, R. B.; Yamaguchi, Y.; Schaefer, H. F., III. *Chem Phys* 1988, 123, 187.
30. (a) Dunn, K. M.; Boggs, J. E.; Pulay, P. *J Chem Phys* 1986, 85, 5838; (b) Cassam-Chenaï, P.; Lievin, J. *J Comput Chem* 2006, 27, 627; (c) Ribiero, F.; Jung, C.; Leforestier, C. *J Chem Phys* 2005, 123, 54106.
31. Pouchan, C.; Zaki, K. *J Chem Phys* 1997, 107, 342.
32. Carbonnière, P.; Dargelos, A.; Pouchan, C. *Theor Chim Acta* 2010, 125, 543.
33. Barone, V. *J Chem Phys* 2005, 122, 014108.
34. 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.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M. J.; Knox, E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.1*; Gaussian, Inc.: Wallingford, CT, 2009.
35. Carbonnière, P.; Barone, V. *Chem Phys Lett* 2004, 399, 226.
36. Carbonnière, P.; Lucca, T.; Rega, N.; Pouchan, C.; Barone, V. *J Comput Chem* 2005, 26, 384.
37. Barone, V.; Carbonnière, P.; Pouchan, C. *J Chem Phys* 2005, 122, 224308.
38. Ragavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem Phys Lett* 1989, 157, 479.
39. Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J Chem Phys* 1992, 96, 6769.
40. Marchal, R.; Carbonnière, P.; Pouchan, C. *J Chem Phys* 2009, 131, 114105.

CARBONNIERE ET AL.

41. Schlegel, H. B.; Millam, J. M.; Iyengar, S. S.; Voth, G. A.; Daniels, A. D.; Scuseria, G. E.; Frish, M. J. *J Chem Phys* 2001, 114, 9758.
42. Iyengar, S. S.; Schlegel, H. B.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frish, M. J. *J Chem Phys* 2001, 115, 10291.
43. Schlegel, H. B.; Iyengar, S. S.; Li, X.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frish, M. J. *J Chem Phys* 2002, 117, 8694.
44. Schmitz, M.; Tavan, P. *J Chem Phys* 2004, 121, 12233.
45. Schmitz, M.; Tavan, P. *J Chem Phys* 2004, 121, 12247.
46. Carbonnière, P.; Dargelos, A.; Ciofini, I.; Adamo, C.; Pouchan, C. *PCCP* 2009, 11, 4375.
47. Kubo, R.; Toda, M.; Hashitsume, N. *Statistical Physics*, 2nd ed.; Springer: Berlin, 1991, Vol. II.
48. Ren, P.; Ponder, J. W. *J Phys Chem B* 2003, 107, 5933.
49. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J Am Chem Soc* 1995, 117, 5179.
50. Andersen, H. C. *J Chem Phys* 1980, 72, 2384.
51. Beeman, D. *J Comp Phys* 1976, 20, 130.
52. Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J Chem Phys* 1984, 81, 3684.
53. Burcl, R.; Carter, S.; Handy, N. C. *Chem Phys Lett* 2003, 373, 357.