



HAL
open science

Développement de champs de forces polarisables et applications à la spectroscopie vibrationnelle

Florian Thaunay

► **To cite this version:**

Florian Thaunay. Développement de champs de forces polarisables et applications à la spectroscopie vibrationnelle. Chimie organique. Université Paris Saclay (COMUE), 2016. Français. NNT : 2016SACLX037 . tel-01496879

HAL Id: tel-01496879

<https://pastel.hal.science/tel-01496879v1>

Submitted on 28 Mar 2017

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

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

NNT : 2016SACLX037

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
L'ECOLE POLYTECHNIQUE, PALAISEAU

ECOLE DOCTORALE N° 571
Sciences chimiques : molécules, matériaux, instrumentation et biosystèmes

Spécialité de doctorat : Chimie

Par

M. Florian THAUNAY

Développement de champs de forces polarisables
et applications à la spectroscopie vibrationnelle

Thèse présentée et soutenue à Palaiseau, le 02 septembre 2016 :

Composition du Jury :

Présidente du Jury :

Mme Zehnacker-Rentien Anne

Directeur de Recherche au CNRS – Université Paris-Sud

Rapporteurs :

M. Borgis Daniel

M. Bordat Patrice

Directeur de Recherche au CNRS – Ecole Normale Supérieure

Maître de Conférences – Université de Pau et des Pays de l'Adour

Examineur :

M. Calvo Florent

Directeur de Recherche au CNRS – Université Joseph Fourier

Co-directeurs de thèse :

M. Ohanessian Gilles

Mme Clavaguéra Carine

Directeur de Recherche au CNRS – Ecole Polytechnique

Chargé de Recherche au CNRS – Ecole Polytechnique

Dans la nature, tout a toujours une raison.
Si tu comprends cette raison, tu n'as plus
besoin de l'expérience.

Léonard de Vinci

La science est une chose merveilleuse...
tant qu'il ne faut pas en vivre!

Albert Einstein

Remerciements

Ce travail de thèse a été effectué au sein de l'équipe Spectroscopie, Interactions non covalentes et Modélisation (SIM) au laboratoire de Chimie Moléculaire (LCM) de l'Ecole Polytechnique.

Un point important des remerciements, c'est de n'oublier personne. Par chance, le doctorant de chimie théorique est un ermite, qui sort peu de son bureau, interagit avec un nombre généralement faible de personnes, et travaille avec une petite portion d'entre elles. Je n'ai pas fait exception à la règle. En conséquence, chacune des interactions devient importante, pour ne pas dire plus, et un(e) directeur(trice) à l'écoute peu éviter des situations incongrues, de celles où, comme Tom Hanks dans 'Seul au Monde', on se retrouve à peindre un visage sur un ballon de volleyball.

J'ai échappé à ces échanges unilatéraux avec 'Wilson' (de peu, à l'été 2014), et pour ça je tiens à remercier en premier lieu ma co-directrice de thèse **Carine Clavaguéra**, pour son encadrement du quotidien, receptacle de mes réflexions, mes joies (oui, parfois ça compile!), ou frustrations (eh non, ça ne compile pas toujours . . .). Tu as su être là pour tous les problèmes, théoriques, techniques ou autres, et pourtant j'ai vraiment ressenti une grande liberté d'action, qui m'a permis notamment de suivre quantités de formations, ce que je sais ne pas être évident pour bon nombre de doctorants.

Dans un autre registre mais avec la même ferveur, je tiens à remercier également mon second co-directeur de thèse, **Gilles Ohanessian**. Ces longs 'brain storming', à l'heure où j'espérais éviter les bouchons sur la Francilienne, ont pourtant toujours été un vrai plaisir. Quelques années auparavant, la devise de mon Capitaine d'Unité était '*Vous n'allez pas au fond des choses*'. Voici deux hommes qui auraient dû se rencontrer. Merci pour la petite liste du vendredi, ta curiosité insatiable a alimenté la mienne. Je crois pouvoir dire que nous nous sommes bien amusés.

D'autres interactions, et très profitables, il y en a eu. Je remercie chaleureusement **Florent Calvo**, à la fois pour cette très intéressante collaboration autour de l'EVB, et pour avoir bien voulu participer au jury de ma soutenance. Grâce à cette habile transition, j'en profite pour remercier également les autres membres de mon jury, **Anne Zeihnaker-Rentien**, **Daniel Borgis** et **Patrice Bordat**. J'ai été particulièrement heureux qu'un Palois accepte d'être rapporteur de mon travail.

Pendant ces trois années, j'ai également eu des charges d'enseignement. J'ai pu évoluer au milieu d'enseignants passionnés. Beaucoup pourraient être cités, mais je tiens à nommer **Fabien Caillez**, avec qui j'ai beaucoup interagi. Travailler avec toi a été très enrichissant.

J'ai appris des 'anciens', **Ashwani** et **Manjeet** surtout, une pensée pour eux.

Au delà du travail, le quotidien. Merci à **Edith Nicol** d'avoir pris à coeur sa mission de m'emmener au café chaque matin, à mes deux 'collocs', **Yann**, puis **Arnaud**, pour avoir mis de la vie dans ce bureau. Petit regret de ne jamais avoir réussi à vous emmener boire le café, mais il y a toujours un travail incomplet à la fin d'une thèse.

'Mens sana in corpore sano'. **Christophe**, avec toutes tes lubies, me voici avec deux marathons à mon actif. C'est quand tu veux pour remettre ça.

Un petit coucou à tous les autres, ici à l'étage, **Stephan**, **Sophie**, **Ségolène**, **Gilles**, **Houda** ... et au reste du laboratoire.

Faire de la chimie théorique, c'est être condamné à être incompris par son entourage familial. J'embrasse ma famille, vous avez compris l'essentiel. Pensée pour **Charlotte**, courage, tu es au bout !

Je renouvelle tout mon amour à **Capucine**, qui m'a parfois subi, lorsque je dessinais des schémas à des heures tardives pour lui montrer à quel point ce que je fais est compréhensible à un non initié. Ca n'a pas toujours été probant. C'est un plaisir, 3 ans plus tard de te voir expliquer, assez justement, mon sujet de thèse. Cette page de la thèse se tourne, la suivante doit être écrite, la seule certitude, c'est qu'elle le sera avec toi.

Table des matières

Introduction générale	11
Bibliographie	15
I Méthodologie et développements	19
1 La spectroscopie IR théorique	21
1.1 Description du mouvement	21
1.2 Méthodes quantiques pour les calculs d'énergies électroniques	23
1.3 Approche indépendante du temps	25
1.3.1 En mécanique classique	25
1.3.2 En mécanique quantique	25
1.4 Approches dépendantes du temps	30
1.4.1 En mécanique classique	30
1.4.2 En mécanique quantique	31
1.5 Intensités IR calculées	34
1.5.1 Calculs statiques	34
1.5.2 Calculs dynamiques	34
Bibliographie	37
2 Champs de forces	41
2.1 Généralités	41
2.2 Champs de forces additifs de paire (Champs de forces de 1 ^{ère} génération)	42
2.2.1 Interactions de valence	42
2.2.2 Interactions d'atomes non liés	43
2.3 Champs de forces polarisables (Champs de forces de 2 ^{nde} génération)	45
2.3.1 Modèle de multipôles ponctuels	46
2.3.2 Modèle de Drude	47
2.3.3 Modèle des charges fluctuantes	47
2.4 Le champ de forces polarisable AMOEBA	48
2.4.1 Formalisme	48
2.4.2 Extraction des multipôles atomiques	52
2.4.3 Paramètres du champ de forces de l'eau	56
Bibliographie	59

3	Développements théoriques	63
3.1	Recherche de structures de basse énergie	63
3.2	Analyse des modes normaux	66
3.2.1	Implémentation pour l'étude des peptides	68
3.2.2	Implémentation pour l'étude des ions solvatés	70
3.3	Modélisation de la spectroscopie du partage de proton : Empirical Valence Bond	74
3.3.1	Un modèle EVB à 2 états	75
3.3.2	Description d'un partage symétrique : la série des diacides déprotonés $\text{HO}_2\text{C}-(\text{CH}_2)_n-\text{CO}_2^-$	76
	Bibliographie	91
	Annexe A	94
	Annexe B	106
II	Applications	141
4	Dynamique des ions hydratés	143
4.1	Dynamique des ions dans une nanogoutte de 100 H_2O (AMOEBAA03)	144
4.2	Validation du nouveau jeu de paramètres de l'eau AMOEBA14 pour la dynamique des ions solvatés	168
4.2.1	Comportement dynamique des ions	168
4.2.2	Influence des paramètres	169
4.2.3	Temps de résidence	171
	Bibliographie	173
5	Spectroscopie des ions hydratés ($n \geq 36$)	175
5.1	Spectroscopie IR des bandes d'élongation OH. Application aux sulfates	176
5.1.1	Position des massifs liés et apparition des bandes de OH libres en fonction de la taille (AMOEBAA03)	177
5.1.2	Modification de la constante d'anharmonicité du potentiel (AMOEBAA03)	178
5.1.3	Validation du modèle AMOEBA14 de l'eau pour la spectroscopie	179
5.1.4	Position des massifs liés et apparition des bandes de OH libres en fonction de la taille (AMOEBAA14)	182
5.1.5	Modèle à charges fluctuantes	183
5.1.6	Validation du modèle sur les agrégats d'eau pure	186
5.1.7	Attribution des modes normaux de vibration : DMD sur $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$, $n = 36$ et 80	187
5.2	Spectroscopie sur les bandes d'élongation OH. Dépendance à l'ion	193
5.2.1	Spectres expérimentaux d'ions hydratés $\text{M}(\text{H}_2\text{O})_n$, $n = 36$ et 250	193
5.2.2	Spectres DACF d'ions hydratés $\text{M}(\text{H}_2\text{O})_n$, $n = 36$ et 100	194
	Bibliographie	199
	Annexe C	201
	Bibliographie	203

6	Energie et Spectroscopie de l'ion sulfate hydraté dans de petits agrégats d'eau	205
6.0.3	Résultats expérimentaux de l'équipe de Williams	205
6.0.4	Stratégie de calculs	208
6.1	Méthodologie : $[\text{SO}_4(\text{H}_2\text{O})_6]^{2-}$ comme système modèle	209
6.1.1	Problématique : Description d'un spectre multiconfigurationnel par une approche statique	209
6.1.2	Choix d'une méthode pour reproduire la géométrie et la spectroscopie des petits agrégats d'eau incorporant un ion sulfate	210
6.1.3	Détermination des populations relatives d'isomères	211
6.1.4	Spectres statiques multiconfigurationnels	212
6.1.5	Spectres dynamiques AMOEBA	214
6.2	$[\text{SO}_4(\text{H}_2\text{O})_9]^{2-}$	218
6.2.1	Spectres statiques multiconfigurationnels	219
6.2.2	Spectres dynamiques AMOEBA	221
6.3	$[\text{SO}_4(\text{H}_2\text{O})_{12}]^{2-}$	223
6.3.1	Spectres statiques multiconfigurationnels	223
6.3.2	Spectres dynamiques AMOEBA	225
6.4	$[\text{SO}_4(\text{H}_2\text{O})_{13}]^{2-}$	226
6.4.1	Spectres statiques multiconfigurationnels	226
6.4.2	Spectres dynamiques AMOEBA	228
6.5	Tailles intermédiaires : $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$, n = 23-25	229
6.5.1	$[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$	229
6.5.2	$[\text{SO}_4(\text{H}_2\text{O})_{23}]^{2-}$ et $[\text{SO}_4(\text{H}_2\text{O})_{25}]^{2-}$	234
6.6	Conclusion	239
	Bibliographie	241
	Annexe D	243
	Conclusions et perspectives	255

Introduction

La connaissance de la structure des assemblages moléculaires, de leurs interactions et associations statiques et dynamiques, est au coeur de la compréhension du fonctionnement du vivant. La géométrie tridimensionnelle des biomolécules leur confère leurs propriétés biologiques, par la capacité de deux structures fortement complémentaires à se reconnaître et interagir (réactions enzymatiques, peptide-membrane, anticorps-antigène . . .). Dans les systèmes biologiques, l'eau joue un rôle central [1], véritable solvant de la vie. L'hydratation dicterait la dynamique des protéines [2], ou faciliterait le transport de proton [3]. La présence d'ions en solution impacte fortement la solubilité et le comportement d'autres molécules [4]. Les mécanismes de transfert des ions aux interfaces sont d'un intérêt majeur dans des problématiques aussi variées que la catalyse [5] ou le traitement de la pollution de l'eau [6]. Par leurs interactions avec les molécules d'eau, les ions réorganisent le réseau de liaisons hydrogènes [7–9], sous la forme de sphères de solvation. Si l'effet structurant de l'ion au sein de la première sphère est largement admis, la compréhension des effets structuraux à plus longue portée constitue toujours un challenge pour les expérimentateurs, dû au comportement dynamique de l'eau liquide et à la présence de contre-ions en solution [10].

Pour étudier expérimentalement la structure de ces assemblages, une variété de techniques spectroscopiques peuvent être mises en oeuvre, qui ont connu des développements considérables au cours des dernières années et se sont diversifiées : la diffraction des rayons X ou des neutrons, la spectroscopie EXAFS (Extended X-Ray Absorption Fine Structure) ou la résonance magnétique nucléaire (RMN), fournissent des informations structurales telles que les motifs périodiques au sein des biomolécules, ou la coordination et les distances au sein des premières sphères de solvation d'un ion [11]. La spectroscopie vibrationnelle (IR ou Raman) apporte des indices complémentaires sur les interactions spécifiques au sein d'une molécule ou avec son environnement. Le spectre vibrationnel peut aider à l'identification d'isomères, ou permettre d'observer indirectement les modifications du comportement d'une espèce en fonction de son environnement. Depuis une cinquantaine d'années, des bases de données de fréquences IR ont été constituées, particulièrement pour des molécules organiques mais aussi inorganiques, faisant de cette spectroscopie une technique d'analyse de routine.

Les techniques expérimentales se sont longtemps concentrées sur l'étude des propriétés de la phase condensée. Depuis une vingtaine d'années et le développement des sources d'ions dites 'douces' (ESI [12], MALDI [13], . . .), l'analyse de molécules en phase gazeuse par la spectrométrie de masse s'est généralisée, offrant un accès aux propriétés intrinsèques des analytes, en l'absence de solvant. Pour autant, le domaine d'application de la spectrométrie de masse ne se limite pas à l'étude de molécules isolées, mais se révèle un outil de choix pour l'étude de la solvation d'ions gazeux dans des agrégats, permettant, par une sélection en taille très précise, d'observer l'évolution des propriétés de l'ion en fonction du nombre de molécules d'eau. En outre, l'absence de contre-ion élimine l'effet de compétition qui rend parfois délicat l'accès aux propriétés individuelles d'un ion en solution. En augmentant progressivement la taille de l'agrégat, de quelques molécules d'eau à plusieurs centaines, il est possible de caractériser la convergence des propriétés des agrégats vers celles du macroscopique.

Expérimentalement, le principe de la spectroscopie IR repose sur la mesure de l'absorption d'un rayonnement par un échantillon, en fonction de la longueur d'onde du rayonnement émis. Pour l'étude d'espèces en phase gazeuse, une mesure directe de l'absorption est compliquée par la densité trop faible d'ions puisque l'échantillon analysé est un gaz à très faible pression. La détection de l'absorption peut néanmoins être réalisée de façon indirecte, par la mesure du taux de fragmentation induite par l'absorption de photons infrarouge : c'est le principe de la spectroscopie d'action. Sur la gamme du moyen IR ($\sim 500\text{-}2500\text{ cm}^{-1}$), les lasers à électrons libres du Centre Laser Infrarouge d'Orsay (CLIO) et de FELIX, aux Pays-Bas, présentent l'intensité et l'accordabilité requises pour une application à des ions stables. Sur cette gamme de fréquences, l'énergie d'un photon IR est au minimum un ordre de grandeur en dessous de l'énergie de dissociation d'une liaison covalente. L'ion doit donc absorber plusieurs photons pour atteindre son énergie de dissociation. Le principe de la spectroscopie IRMPD (Infrared Multiple PhotoDissociation) a été démontré dans les années 1980 [14] mais les premières expériences de spectroscopie d'ions moléculaires dans l'infrarouge moyen sont plus récentes [15–20], en lien avec la création des lasers à électrons libres. Dans la majeure partie des cas, les spectres IRMPD sont très similaires aux spectres d'absorption à 1 photon [16, 21]. Toutefois, la nature multi-photonique de l'IRMPD peut induire un élargissement et un décalage vers le rouge de certaines bandes par rapport à leur absorption linéaire, du fait de l'anharmonicité [18, 22–24].

La spectroscopie d'action n'est cependant pas limitée aux seuls lasers à électrons libres, puisque des lasers IR accordables de tables, de plus faibles intensités, sont disponibles dans de nombreux laboratoires. La gamme de fréquences disponibles ($\sim 600\text{-}4100\text{ cm}^{-1}$) et leur faible coût ont contribué au développement rapide de l'IRPD ces dernières années. Si on utilise une fragmentation de basse énergie, comme une rupture de liaison hydrogène intermoléculaire, un seul photon est généralement nécessaire, ce qui conduit à des intensités très facilement comparables aux intensités linéaires d'un vrai spectre d'absorption, mêmes si des mécanismes multiphotoniques peuvent également se produire. F. Calvo et coll. modélisent la réponse d'ions gazeux au champ d'un laser pulsé, au cours de simulations de dynamique moléculaire. Lorsque le champ est polarisé circulairement, ils obtiennent une mesure du dichroïsme circulaire, preuve de la chiralité des molécules [25]. Ils évaluent également la probabilité de dissociation d'agrégats chargés et comparent le spectre de dissociation (spectroscopie d'action) au spectre d'absorption linéaire. Le modèle confirme l'effet des anharmonicités dynamiques sur l'élargissement et le déplacement de bandes, et sur les intensités relatives [26, 27].

Au-delà des molécules les plus simples, l'interprétation des spectres vibrationnels d'espèces ioniques en phase gazeuse peut être assez délicate. D'une part, les bases de données accumulées pour des molécules neutres en solution ne sont pas directement utilisables. D'autre part, s'agissant de la solvation d'ions dans des agrégats d'eau, exceptés pour les plus petits, la compréhension de la dynamique de ces systèmes est une étape limitante, avant de pouvoir discuter de la spectroscopie ; certains comportements tels que la préférence de l'ion Cl^- à rejoindre le centre d'un agrégat d'eau ou à rester à sa surface, ont fait encore débat récemment [28, 29]. De plus, très peu d'équipes sont aujourd'hui capables de coupler la production de ces ions solvatés de masse sélectionnée et la spectroscopie IR(M)PD [30–33].

Dans ces problématiques, la modélisation tient un rôle central, fournissant les outils pour l'étude de la dynamique de ces systèmes, ou l'attribution des spectres IR. Dans nos systèmes d'étude (peptides chargés, ions solvatés), établir une relation entre un spectre expérimental et une structure moléculaire implique comme première étape d'identifier les structures stables. La taille et la flexibilité importantes de ces systèmes impose de combiner une description peu coûteuse des atomes (champs de forces [34], modèles semi-rigides [35]), à des méthodes d'exploration globale de la surface d'énergie potentielle, méthodes stochastiques [36], dynamique moléculaire, recuit simulé [37]. . . Le point commun de ces systèmes flexibles et chargés est l'importance des interactions non covalentes dans leur organisation (repliement d'un peptide, distribution des couches de solvation d'un ion). Nous utilisons le champ de forces polarisable AMOEBA (Atomic Multipole Optimized energetics for Biomolecular Applications) [38, 39], développé

pour fournir une description précise des structures et des énergies d’hydratation des ions, des molécules organiques et des biomolécules, au cours de simulations de dynamique moléculaire. Pour les petits ions hydratés $M(\text{H}_2\text{O})_{9-13}$, de longues simulations suffisent généralement à une exploration efficace de l’espace des conformations [40]. Pour les plus grandes tailles, le champ de forces est combiné à la dynamique moléculaire par échange de répliques [41], qui consiste à effectuer plusieurs trajectoires en parallèle à des températures différentes, des répliques, et à les échanger entre elles. Cette méthodologie a également été appliquée avec succès à l’étude de peptides dans notre groupe [42–45]. Pour l’exploration spécifique des surfaces de potentiel des ions solvatés, des outils complémentaires ont été développés afin de mieux tenir compte des réorganisations du réseau de liaisons hydrogènes au sein des agrégats.

La mécanique quantique est ensuite souvent utilisée pour obtenir des énergies fiables de ces structures. La validation d’un niveau de calcul est une tâche parfois délicate. En particulier, pour l’étude des sulfates hydratés, la large gamme de tailles des agrégats ne permet pas de valider un même niveau théorique pour toute l’étude. Si pour les plus petits, des méthodes de haute précision, telles que DLPNO-CCSD(T) [46, 47] ou MP2, sont facilement accessibles, la DFT devient l’outil de référence au-delà. En outre, la précision de la fonctionnelle utilisée est très dépendante du système. Par exemple, la fonctionnelle M11 est jugée performante pour traiter les interactions sulfate-eau [48] et justifie son utilisation pour les petits agrégats, mais décrit beaucoup moins bien les interactions entre molécules d’eau et conduit à lui préférer la fonctionnelle $\omega\text{B97X-D}$ [49, 50] lorsque celles-ci deviennent prépondérantes.

Généralement, les spectres IR sont calculés à la géométrie d’équilibre, dans la double approximation harmonique. Puisque calculés à 0 K, il ne prennent pas en compte les effets de dynamique et de température, tels que l’anharmonicité des modes de vibration, la largeur des bandes, ou le mélange de structures. Ce dernier point peut être évalué *a posteriori* par le calcul de propriétés thermodynamiques incluant les structures les plus stables, dont la dépendance au niveau de calcul doit également être prise en compte. D’autres approches, dites ‘dépendantes du temps’, permettent d’inclure directement les effets d’anharmonicité et de température. Les dynamiques *ab-initio* [51], majoritairement de type Born-Oppenheimer ou Car-Parinello, ont été beaucoup employées [52–54]. Pour la spectroscopie IR [55], Gageot et al. ont calculé la signature de molécules biologiques en phase gazeuse [34, 56–58] ou en solution [59]. Pour les systèmes pour lesquels la dynamique conformationnelle est importante, ou s’agissant d’observer un évènement statistiquement rare, la dynamique *ab-initio* est limitée par les temps de simulations accessibles (quelques dizaines de picosecondes). Pour ces applications, le champ de forces polarisable AMOEBA a été étendu pour le calcul dynamique de spectres IR d’ions en phase gazeuse [60, 61], au travers de la méthode DACF (Dipole Autocorrelation Function). La modélisation de la spectroscopie de l’eau est un problème récurrent pour les champs de forces classiques, s’ils n’y sont pas explicitement consacrés [62]. Plusieurs étapes de calibration d’AMOEBA ont été réalisées (validation de nouveaux paramètres de l’eau [63], augmentation de la flexibilité intramoléculaire ou modèle de charges fluctuantes), afin d’améliorer la description de la signature spectroscopique des gros solvates $M(\text{H}_2\text{O})_n$, $n \leq 100$, en fonction de la température.

Les spectres calculés dynamiquement ne fournissent pas d’accès direct à l’identification des bandes de vibration, qui est généralement traitée *a posteriori* [64, 65]. Nous avons implémenté une méthode d’attribution dynamique des modes normaux de vibration basée sur la DMD (Driven Molecular Dynamics) [66, 67], capable de mimer l’anharmonicité et le couplage entre modes [45], tels qu’observés dans des simulations AMOEBA en fonction de la température. La DMD a également été renforcée par des outils statistiques pour être appliquée à l’étude des massifs d’élongation O-H de molécules d’eau dans des ions solvatés, pour expliquer l’évolution des spectres avec l’augmentation de la taille des agrégats.

Certaines signatures expérimentales, telles que celle associée au partage de proton, ne peuvent pas du tout être décrites par le champ de forces classique, ni par une méthode quantique statique, si le partage est intrinsèquement

dynamique. Un modèle Empirical Valence Bond (EVB) [68, 69] a été récemment implémenté [70], couplé au champ de forces AMOEBA. Il permet de reproduire le comportement dynamique du transfert de proton au sein de petits peptides, ainsi que la signature spectroscopique observée expérimentalement.

Le manuscrit se divise en deux parties. La première regroupe les éléments de méthodologie, dont la description des outils implémentés au cours de cette thèse. Le **chapitre 1** donne un aperçu des méthodes de calculs de spectroscopie infrarouge théorique, plusieurs ayant été utilisées au cours de ces travaux (calculs quantiques statiques harmoniques ou anharmoniques, et spectres dynamiques par fonction d'autocorrélation du dipôle au cours de simulations de dynamique classique). En **chapitre 2**, une description des champs de forces est proposée. Décrire avec précision les ions gazeux nécessite de bien traiter les interactions non covalentes, intra ou intermoléculaires. Les principales méthodes pour traduire les interactions électrostatiques de ces systèmes et notamment la polarisation, sont discutées. Le champ de forces polarisable AMOEBA inclut un calcul explicite de la polarisation, et nous présenterons dans ce manuscrit des exemples de réussite de ce modèle et d'autres cas où des contributions additionnelles (modèle de charges fluctuantes, terme de pénétration de charge) semblent indispensables. Le **chapitre 3** prend la forme d'une boîte à outils. Ceux qui sont consacrés à l'exploration de surface d'énergie potentielle, par la méthode REMD, sont antérieurs à ces travaux mais ont été très utilisés. Les spécificités de la recherche de structures stables d'ions solvatés ont nécessité des outils complémentaires, pour comprendre l'influence des inversions de cycles au sein du réseau de liaisons hydrogènes. Une méthode d'attribution dynamique des modes normaux de vibration a été implémentée dans TINKER. Une première version est consacrée à la description des modes des peptides, incluant leurs couplages et anharmonicités; une autre version permet de discuter la diversité des interactions au sein d'ions solvatés de tailles variables. Enfin, le transfert de proton au cours de simulations de dynamique classique AMOEBA est rendu accessible par l'implémentation d'un modèle Empirical Valence Bond dans TINKER. La méthodologie est discutée au travers de deux exemples, l'aspartate et la série des diacides déprotonés $\text{HO}_2\text{C}-(\text{CH}_2)_n-\text{CO}_2^-$, $n = 2-6$, pour lesquels l'EVB a été en mesure de reproduire la dynamique d'échange du proton et la signature IR expérimentale caractéristique.

La seconde partie est consacrée à des applications à divers ions hydratés mettant en oeuvre les outils précédemment décrits. Le **chapitre 4** traite de la dynamique des agrégats $\text{M}(\text{H}_2\text{O})_{n=100}$ en employant le champ de forces AMOEBA. Pour tous les ions étudiés, de charge comprise entre -II et +III, le champ de forces reproduit les comportements couramment relevés dans la littérature. Les 2 derniers chapitres traitent de spectroscopie IR, avec un intérêt particulier pour l'ion sulfate. En **chapitre 5**, les plus gros agrégats sont traités (36-100 molécules d'eau), principalement au moyen de dynamique moléculaire. Ce chapitre détaille également la calibration du champ de forces pour la spectroscopie. Au cours de la fragmentation des ions dans le spectromètre de masse, l'abondance des fragments peut être suivie au niveau de la source d'ions. Les spectres ESI de Williams révèlent que, pour les sulfates hydratés $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$, certaines tailles montrent une abondance particulière. C'est le cas de $n = 24$ et 36 , et dans une moindre mesure, de $n = 12$ [71]. Williams qualifie ces tailles de 'nombres magiques', attribuant leur abondance relative à une stabilité particulière. Le **chapitre 6** se focalise sur ces gammes de tailles, en traitant dans un premier temps les sulfates de petites tailles ($n = 6-13$), d'un point de vue énergétique et spectroscopique. La mécanique quantique est la méthode de référence pour ces tailles et fournit la majorité des résultats, mais le champ de forces n'est pas restreint à l'exploration de structures, et confirme généralement les prédictions de la chimie quantique sur les structures énergétiquement compétitives ou capables de reproduire un spectre expérimental. Pour les tailles intermédiaires ($n = 23-25, 36$), les deux méthodologies, classique et quantique, fournissent des informations complémentaires.

Bibliographie

- [1] Chaplin, M. *Nat. Rev. Mol. Cell. Biol.* **2006**, 7, 861–866.
- [2] Wood, K.; Frölich, A.; Paciaroni, A.; Moulin, M.; Härtlein, M.; Zaccai, G.; Tobias, D. J.; Weik, M. *J. Am. Chem. Soc.* **2008**, 130, 4586–4587.
- [3] Wraight, C. A. *BBA - Bioenergetics* **2006**, 1757, 886–912.
- [4] Kunz, W.; Henle, J.; Ninham, B. *Curr. Opin. Colloid Interface Sci.* **2004**, 9, 19 - 37.
- [5] Starks, C. M.; Liotta, C. L.; Halpern, M. *Phase Transfer Catalysis*; Chapman & Hall : new York : **1994**.
- [6] Veziriglu, T. N. *Environmental Problems and Solutions : Greenhouse Effects, Acid Rain, Pollution*; Hemisphere : new York : **1990**.
- [7] Conway, B. E. *Ionic hydration in chemistry and biophysics*; Elsevier : Amsterdam : **1981**.
- [8] Richens, D. T. *The Chemistry of Aqua Ions*; John Wiley & Sons, Inc. : New York : **1997**.
- [9] O'Brien, J. T.; Williams, E. R. *J. Am. Chem. Soc.* **2012**, 134, 10228–10236.
- [10] Omta, A. W.; Kropman, M. F.; Woutersen, S.; Bakker, H. J. *Science* **2003**, 301, 347–349.
- [11] Ohtaki, H.; Radnai, T. *Chem. Rev.* **1993**, 93, 1157–1204.
- [12] Cole, R. *Electrospray ionization mass spectrometry : fundamentals, instrumentation and applications*; Wiley, New York : **1997**.
- [13] Karas, M.; Bachmann, D.; Hillenkamp, F. *Anal. Chem.* **1985**, 57, 2935–2939.
- [14] Woodin, R. L.; Bomse, D. S.; Beauchamp, J. L. *J. Am. Chem. Soc.* **1978**, 100, 3248–3250.
- [15] Lemaire, J.; Boissel, P.; Heninger, M.; Mauclaire, G.; Bellec, G.; Mestdagh, H.; Simon, A.; Caer, S. L.; Ortega, J. M.; Glotin, F.; Maitre, P. *Phys. Rev. Lett.* **2002**, 89, 273002.
- [16] Oomens, J.; van Roij, A. J. A.; Meijer, G.; von Helden, G. *Astrophys. J.* **2000**, 542, 404.
- [17] Valle, J. J.; Eyler, J. R.; Oomens, J.; Moore, D. T.; van der Meer, A. F. G.; von Helden, G.; Meijer, G.; Hendrickson, C. L.; Marshall, A. G.; Blakney, G. T. *Rev. Sci. Instrum.* **2005**, 76, 023103.
- [18] Oomens, J.; Sartakov, B. G.; Meijer, G.; von Helden, G. *Int. J. Mass. Spectrom.* **2006**, 254, 1 - 19.
- [19] MacAleese, L.; Maître, P. *Mass. Spectrom. Rev.* **2007**, 26, 583–605.
- [20] Walker, N. R.; Walters, R. S.; Duncan, M. A. *New J. Chem.* **2005**, 29, 1495–1503.
- [21] Macaleese, L. *Spectroscopie Infrarouge d'Intermédiaires Réactionnels Organométalliques*, Thesis, Université Paris-Sud 11, Orsay, France, **2007**.
- [22] Bush, M. F.; O'Brien, J. T.; Prell, J. S.; Saykally, R. J.; ; Williams, E. R. *J. Am. Chem. Soc.* **2007**, 129, 1612–1622.
- [23] MacAleese, L.; Maître, P. *Mass Spectr. Rev.* **2007**, 26, 583–605.
- [24] Vala, M.; Szczepanski, J.; Oomens, J.; Steill, J. D. *J. Am. Chem. Soc.* **2009**, 131, 5784–5791.

- [25] Calvo, F. *Chirality* **2015**, *27*, 253-261.
- [26] Marciante, M.; Calvo, F. *Mol. Sim.* **2014**, *40*, 176 - 184.
- [27] Calvo, F.; Li, Y.; Kiawi, D. M.; Bakker, J. M.; Parneix, P.; Janssens, E. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25956-25967.
- [28] Zhao, Y.; Li, H.; Zeng, X. C. *J. Am. Chem. Soc.* **2013**, *135*, 15549-15558.
- [29] Tielrooij, K. J.; Garcia-Araez, N.; Bonn, M.; Bakker, H. J. *Science* **2010**, *328*, 1006-1009.
- [30] O'Brien, J. T.; Prell, J. S.; Bush, M. F.; Williams, E. R. *J. Am. Chem. Soc.* **2010**, *132*, 8248-8249.
- [31] Zhou, J.; Santambrogio, G.; Brümmer, M.; Moore, D. T.; Woste, L.; Meijer, G.; Neumark, D. M.; Asmis, K. R. *J. Chem. Phys.* **2006**, *125*, 111102.
- [32] Marsh, B. M.; Voss, J. M.; Garand, E. *J. Chem. Phys.* **2015**, *143*, 204201.
- [33] Asmis, K. R.; Neumark, D. M. *Acc. Chem. Res.* **2012**, *45*, 43-52.
- [34] Jaeqx, S.; Oomens, J.; Cimas, A.; Gageot, M.-P.; Rijs, A. M. *Angew. Chem. Int. Ed.* **2014**, *53*, 3663-3666.
- [35] Smeeton, L. C.; Farrell, J. D.; Oakley, M. T.; Wales, D. J.; Johnston, R. L. *J. Chem. Theory Comput.* **2015**, *11*, 2377-2384.
- [36] Metropolis, N. *Los Alamos Science* **1987**, *15*, 125-130.
- [37] Aarts, E. H. L.; van Laarhoven, P. J. M. *Philips J. Res.* **1985**, *40*, 193-226.
- [38] Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933-5947.
- [39] Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. *J. Chem. Theory Comput.* **2013**, 130814070526000.
- [40] Thauay, F.; Clavaguéra, C.; Ohanessian, G. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25935-25945.
- [41] Penev, E. S.; Lampoudi, S.; Shea, J.-E. *Comput. Phys. Commun.* **2009**, *180*, 2013 - 2019.
- [42] Semrouni, D.; Balaj, O. P.; Calvo, F.; Correia, C. F.; Clavaguéra, C.; Ohanessian, G. *J. Am. Soc. Mass. Spectr.* **2010**, *21*, 728-738.
- [43] Joshi, K.; Semrouni, D.; Ohanessian, G.; Clavaguéra, C. *J. Phys. Chem. B* **2012**, *116*, 483-490.
- [44] Semrouni, D.; Ohanessian, G.; Clavaguéra, C. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3450-3462.
- [45] Thauay, F.; Dognon, J.-P.; Ohanessian, G.; Clavaguéra, C. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25968-25977.
- [46] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. *J. Chem. Phys.* **2013**, *139*, 134101.
- [47] Sparta, M.; Neese, F. *Chem. Soc. Rev.* **2014**, *43*, 5032-5041.
- [48] Mardirossian, N.; Lambrecht, D. S.; McCaslin, L.; Xantheas, S. S.; Head-Gordon, M. *J. Chem. Theory Comput.* **2013**, *9*, 1368-1380.
- [49] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *120*, 215-241.
- [50] Leverentz, H. R.; Qi, H. W.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 995-1006.
- [51] Witt, A.; Ivanov, S. D.; Shiga, M.; Forbert, H.; Marx, D. *J. Chem. Phys.* **2009**, *130*, 194510.
- [52] Tangney, P. *J. Chem. Phys.* **2006**, *124*, 044111.
- [53] Tangney, P.; Scandolo, S. *J. Chem. Phys.* **2002**, *116*, 14-24.
- [54] Gageot, M.-P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3336-3359.
- [55] Carbonniere, P.; Thicoipe, S.; Very, T.; Assfeld, X. *Int. J. Quant. Chem.* **2012**, *112*, 2221-2230.
- [56] Gageot, M. P.; Vuilleumier, R.; Sprik, M.; Borgis, D. *J. Chem. Theory Comput.* **2005**, *1*, 772-789.

- [57] Marinica, D. C.; Grégoire, G.; Desfrancois, C.; Schermann, J. P.; Borgis, D.; Gaigeot, M. P. *J. Phys. Chem. A* **2006**, *110*, 8802-8810.
- [58] Cimas, A.; Vaden, T. D.; de Boer, T. S. J. A.; Snoek, L. C.; Gaigeot, M.-P. *J. Chem. Theory Comput.* **2009**, *5*, 1068-1078.
- [59] Gaigeot, M.-P.; Sprik, M. *J. Phys. Chem. B* **2003**, *107*, 10344-10358.
- [60] Semrouni, D. *Energétique et spectroscopie de polypeptides par dynamique moléculaire : champ de force de seconde génération et chimie quantique*, Thesis, Ecole Polytechnique, Palaiseau, France, **2010**.
- [61] Semrouni, D.; Sharma, A.; Dognon, J.-P.; Ohanessian, G.; Clavaguéra, C. *J. Chem. Theory Comput.* **2014**, *10*, 3190-3199.
- [62] Fanourgakis, G. S.; Xantheas, S. S. *J. Chem. Phys.* **2008**, *128*, 074506.
- [63] Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. *J. Phys. Chem. B* **2015**, *119*, 9423-9437.
- [64] Martinez, M.; Gaigeot, M.-P.; Borgis, D.; Vuilleumier, R. *J. Chem. Phys.* **2006**, *125*, 144106.
- [65] Wheeler, R. A.; Dong, H.; Boesch, S. E. *ChemPhysChem* **2003**, *4*, 382-384.
- [66] Bowman, J. M.; Zhang, X.; Brown, A. *J. Chem. Phys.* **2003**, *119*, 646.
- [67] Kaledin, M.; Brown, A.; Kaledin, A. L.; Bowman, J. M. *J. Chem. Phys.* **2004**, *121*, 5646-5653.
- [68] Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218-6226.
- [69] Warshel, A. *Acc. Chem. Res.* **2002**, *35*, 385-395.
- [70] Thauay, F.; Calvo, F.; Ohanessian, G.; Clavaguéra, C. Theory and Applications of the Empirical Valence Bond Approach : From Physical Chemistry to Chemical Biology. In ; John Wiley & Sons, Ltd. : **2016** ; Chapter Empirical valence-bond models based on polarizable force fields for infrared spectroscopy.
- [71] Wong, R. L.; Williams, E. R. *J. Phys. Chem. A* **2003**, *107*, 10976-10983.

Première partie

Méthodologie et développements

Chapitre 1

La spectroscopie IR théorique

La spectroscopie infrarouge est une méthode courante d'analyse et d'identification, qui se base sur la connaissance des mouvements de vibration des molécules : le rayonnement infrarouge excite des modes de vibrations (élongations de liaisons chimiques, déformation d'angles, torsions) qui sont caractéristiques de la molécule étudiée et révélateurs de sa présence. Cette technique est largement répandue dans toute activité économique faisant appel à la chimie analytique. Elle est utilisée dans le contrôle qualité des matières premières, des produits finis et des processus de leurs élaborations par réaction chimique. Elle peut permettre par exemple d'apprécier la qualité d'une fibre polymère ou la pureté d'un médicament. Elle est également utilisée dans le contrôle qualité de l'eau ou des sols, associée au dosage chimique de polluants. Grâce à la mesure de la composition chimique de notre atmosphère, elle aide à surveiller l'évolution du climat sur Terre, notamment par l'évaluation de la concentration des gaz à effet de serre. Au-delà de notre biosphère, elle permet l'identification d'espèces dans l'espace interstellaire et dans les atmosphères planétaires, contribuant à fournir des indices sur l'histoire de la formation de l'Univers et de l'apparition de la Vie. L'interprétation d'un spectre vibrationnel expérimental n'est cependant pas triviale en raison du nombre important de signaux dont la position et l'intensité attendues peuvent être sensiblement affectées par des facteurs propres au système moléculaire d'étude et par l'effet de son environnement (le solvant, par exemple). A l'aide du développement combiné des descriptions mathématiques du phénomène de vibration moléculaire et de la technologie informatique, le chimiste théoricien peut fournir des résultats théoriques à l'expérimentateur. Il peut ainsi apporter une justification théorique à une interprétation et aider à l'identification d'une espèce chimique en prévoyant la totalité d'un spectre ou d'une région spectrale. Les récents progrès des techniques expérimentales permettent d'obtenir des spectres de bonne résolution pour des molécules de grande taille, notamment des molécules d'intérêt biologique, et posent un vrai challenge aux méthodes de calculs, pour rendre compte avec précision de ces systèmes.

1.1 Description du mouvement

La description précise du mouvement des noyaux (une fois moyenné celui des électrons) est le fondement théorique de la spectroscopie IR, apportant de précieux renseignements sur la conformation moléculaire et les champs de forces effectifs entre noyaux. Elle donne des renseignements sur les constantes de force d'équilibre et donc la "force" des liaisons chimiques, et permet de connaître la conformation géométrique du système moléculaire étudié. Dans une approche de mécanique classique, une molécule polyatomique peut être matérialisée par des masses ponctuelles réunies ensemble par des ressorts. Si un tel système reçoit de l'énergie, il se mettra à exécuter un mouvement de vibration impliquant l'ensemble ou une part de ses masses constitutives. Il est possible de décomposer le mouvement

de vibration d'une molécule de N atomes en mouvements élémentaires appelés modes normaux de vibration. Un mode normal de vibration est un mode possédant une fréquence de vibration quantifiée et dans lequel une partie des atomes qui constituent le système à l'équilibre vibrent selon des directions privilégiées en passant simultanément par leur position d'équilibre. Ces directions sont propres au système et dépendent de sa symétrie et de sa géométrie, de la masse de ses atomes et de la force de ses liaisons. Chaque molécule non linéaire possède $3N-6$ modes de vibration. Le nombre d'états vibrationnels d'un système moléculaire dans un état électronique donné est en théorie infini. Selon les principes de la mécanique quantique, chaque mouvement vibrationnel possède un niveau d'énergie (E) quantifié et peut être représenté par une combinaison des $3N-6$ modes de vibration, chaque mode de vibration étant excité ν_i fois (ν_i , entier naturel appelé nombre quantique vibrationnel du mode i). Les niveaux vibrationnels ainsi définis peuvent être exprimés en nombre d'onde ($\bar{\nu}$ en cm^{-1}) par la relation $E = hc\bar{\nu}$, où h est la constante de Planck et c est la célérité de la lumière.

Les transitions vibrationnelles observées par spectroscopie IR ou Raman, par exemple, correspondent alors à l'écart entre deux niveaux d'énergie vibrationnels, l'un correspondant à un état de la molécule initial, peuplé, et l'autre dit final, correspondant à l'état que la molécule atteint après absorption d'un rayonnement IR. Une transition est dite fondamentale si elle correspond au passage du niveau fondamental vers le 1^{er} niveau excité d'un mode. Un spectre de vibration est compliqué par l'apparition d'autres bandes dues à des transitions harmoniques (2ν d'une fondamentale principalement) ou à des vibrations de combinaison (Exemple : $\nu_1 + \nu_2$). L'intensité des harmoniques et bandes de combinaison reste souvent faible devant celle des bandes fondamentales.

Dans la gamme 500 cm^{-1} à 4000 cm^{-1} , les transitions entre ces niveaux correspondent à des fréquences auxquelles vibre une partie des atomes de la molécule, de façon spécifique, alors que les modes les plus mous (souvent en dessous de 500 cm^{-1}) sont attribués à des mouvements collectifs (qui font intervenir une grande partie des atomes de la molécule). En spectroscopie infrarouge expérimentale, on suit la variation du dipôle permanent de la molécule pour identifier ses fréquences propres. Il existe des tables remarquablement étendues pour attribuer les bandes IR de composés organiques et inorganiques. C'est principalement dans la gamme $1000-2000\text{ cm}^{-1}$ et c'est pour cela que celle-ci est appelée gamme des empreintes digitales. Pour autant, attribuer la totalité des bandes, ou prédire celles de composés nouveaux, est impossible en général. Les chimistes ont donc massivement recours à la modélisation, pour prévoir une partie ou la totalité d'un spectre et ainsi aider à l'identification de composés, ou apporter des indices sur une structure géométrique.

Il existe de nombreux modèles mathématiques qui conduisent à l'obtention d'un spectre théorique, et le choix du bon modèle repose en partie sur sa capacité à rendre compte de certaines des propriétés du système. Pour orienter son choix, le chimiste théoricien doit pouvoir répondre à certaines questions telles que :

- La molécule, dans sa structure optimisée, est-elle représentative de son comportement général ?
- Ce dernier varie-t-il au cours du temps, soit parce que la molécule est flexible et que sa géométrie fluctue, soit parce qu'elle est entourée d'autres molécules, et que les interactions qu'elles développent évoluent dans le temps ?
- Les propriétés à décrire sont-elles majoritairement quantiques (système avec un proton mobile, forte délocalisation de charge, ...), ce qui orienterait vers un modèle adapté, peut-être avec des électrons décrits explicitement, ou un modèle plus simple peut-il suffire ?
- Connaissant la taille du système, quel modèle mathématique suis-je capable d'appliquer, parce que le coût calculatoire est acceptable/justifié ? Cette dernière question oriente souvent le choix final.

1.2 Méthodes quantiques pour les calculs d'énergies électroniques

La suite de ce manuscrit décrivant les résultats de nombreux calculs quantiques, les grandes lignes des méthodes utilisées sont rappelées ci-dessous.

L'équation de Schrödinger a été formulée sous sa forme primitive en 1926 par Erwin Schrödinger puis généralisée par Paul Dirac quelques années après, avec la notation bra-ket qu'on lui connaît.

$$\frac{\hat{p}^2}{2m} |\Psi(\hat{r}, t)\rangle + V(\hat{r}, t) |\Psi(\hat{r}, t)\rangle = i\hbar \frac{\partial}{\partial t} |\Psi(\hat{r}, t)\rangle \quad (1.1)$$

Equation de base de la mécanique quantique, elle décrit l'évolution dans le temps du vecteur d'état $|\Psi(\hat{r}, t)\rangle$ d'un système quantique arbitraire. Ce vecteur peut être décomposé sur la base d'états propres de l'hamiltonien $|\Psi(\hat{r}, t)\rangle = \sum_n c_n(t) |\Phi_n(\hat{r})\rangle$ avec $c_n(t) = \langle \Phi_n | \Psi(\hat{r}, t) \rangle$. Si le hamiltonien ne dépend pas explicitement du temps, on résout l'équation aux valeurs propres, ou états stationnaires. La résolution de l'équation de Schrödinger des états stationnaires (parfois nommée équation de Schrödinger indépendante du temps) est couramment utilisée pour l'étude des propriétés électroniques ou vibrationnelles de systèmes moléculaires.

$$\hat{H} |\Phi_n(\hat{r})\rangle = E_n |\Phi_n(\hat{r})\rangle \quad (1.2)$$

Dans l'impossibilité d'atteindre les solutions analytiques de cette équation différentielle au-delà d'un système à un électron, les chimistes théoriciens ont développé, depuis les années 1930, de nombreuses méthodes permettant de calculer les principales observables de systèmes de plus en plus complexes et avec une précision sans cesse croissante. Dans l'approximation de Born-Oppenheimer non relativiste, elles donnent parfois accès à la 'précision chimique' (erreur de l'ordre d'une kcal.mol⁻¹ ou inférieure sur les différences d'énergie électronique pour des molécules contenant quelques dizaines d'atomes).

On peut regrouper les méthodes quantiques de calcul en deux grandes catégories : celles qui ont pour objet le calcul de la fonction d'onde (WFT pour Wave Function Theory) et celles qui utilisent la densité électronique (DFT pour Density Functional Theory).

Dans la première catégorie, la méthode Hartree-Fock, consiste à isoler chaque électron d'un système dans un champ moyen (self consistent field ou SCF), dans lequel chaque électron subit l'interaction des $n - 1$ autres électrons du système dans des positions moyennées. La résolution de l'équation monoélectronique de Hartree-Fock (HF) permet de reproduire le trou de Fermi, qui impose que la probabilité de trouver deux électrons de même spin au même endroit doit être nulle. En revanche, elle ne prend pas en compte le trou de Coulomb, interdisant une probabilité non-nulle de trouver deux électrons au même endroit dans l'espace, à cause d'une répulsion coulombienne alors infinie. L'erreur systématique entre l'énergie réelle et la limite HF de la valeur calculée est appelée énergie de corrélation électronique. Un traitement *a posteriori* des solutions de l'équation de HF permet d'évaluer l'énergie de corrélation. Ce traitement est le plus souvent appliqué par des méthodes de perturbation ou d'interaction de configurations. Dans ce dernier cas, la fonction d'onde est exprimée par une combinaison linéaire de configurations, qui sont des combinaisons linéaires de déterminants de Slater, de l'état fondamental et d'états excités. Dans cette famille, la méthode Coupled-Cluster Single Double and Triple (connectées, traitées par perturbation) excitations (CCSD(T)) [1] fait généralement figure de référence pour les calculs d'énergie, bien que le coût de calcul associé limite son utilisation. Pour traiter des systèmes de plus grande taille, on peut utiliser l'approximation DLPNO-CCSD(T) [2, 3].

Afin de prendre en compte la corrélation électronique, une autre approche, basée sur la théorie des perturbations, a été proposée. Il s'agit de la théorie de perturbation de Møller-Plesset [4]. Le niveau le plus utilisé est le second ordre de perturbation sur l'énergie (MP2) [5], qui fournit une très bonne précision lorsque l'approximation Hartree-Fock

est satisfaisante. A cet ordre de perturbation, les configurations comprenant des excitations électroniques simples et doubles par rapport à la configuration fondamentale sont prises en compte.

La seconde catégorie regroupe les méthodes basées sur la théorie de la fonctionnelle de la densité (DFT) telle que développée dans les deux théorèmes fondateurs de Kohn et al. [6]. Le formalisme repose sur le fait que la densité électronique $\rho(r)$ contient toutes les caractéristiques du système. Notamment, l'énergie est une fonctionnelle de la densité électronique ne dépendant que de 3 variables spatiales, quelque soit la taille du système, ce qui est un moyen élégant de contourner le mur calculatoire auquel se trouvent confrontées les méthodes post-HF. Cette approche permet d'introduire les effets de corrélation électronique dans une approche mono-électronique. La méthode Kohn-Sham (KS) consiste à modéliser un système multi-électronique par un système fictif de type gaz de particules sans interactions, ayant la même densité $\rho(r)$ que le système réel.

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{Ne}[\rho] \quad (1.3)$$

$T[\rho]$ est l'énergie cinétique des électrons, $V_{ee}[\rho]$ est l'énergie de répulsion coulombienne électron-électron et $V_{Ne}[\rho]$ est l'énergie d'attraction coulombienne électron-noyau.

Ni l'expression analytique exacte de $T[\rho]$, ni celle de $V_{ee}[\rho]$ ne sont totalement définies et les termes inconnus sont généralement regroupés dans un terme $E_{XC}[\rho]$, décomposable en deux termes distincts, dont l'un est la fonctionnelle d'échange qui est associée à l'interaction électron-électron de mêmes spins $E_X[\rho]$ et l'autre est la fonctionnelle de corrélation $E_C[\rho]$ qui est associée à l'interaction électron-électron de spins opposés. On trouve différentes manières d'approximer ces termes, avec, par ordre de qualité croissante, des fonctionnelles qui dépendent uniquement de la densité électronique (LDA pour Local Density Approximation), des fonctionnelles qui dépendent de la densité et de son gradient (GGA pour Generalized Gradient Approximation) ou encore du laplacien de la densité (méta GGA), etc ...

La littérature foisonne d'articles et de revues testant la qualité des fonctionnelles [7–18]. Elles sont calibrées sur des jeux de systèmes et propriétés cibles. Cependant, ces informations seules ne suffisent généralement pas à garantir l'efficacité de la méthode sur les systèmes d'une nouvelle étude. Couramment, le chimiste théoricien doit passer par une étape de calibration en testant plusieurs fonctionnelles sur un panel de systèmes et/ou propriétés qu'il espère représentatif de l'étude à mener. Dans ces travaux, tous les résultats DFT sont obtenus à partir de fonctionnelles hybrides, c'est à dire qui contiennent une part d'échange Hartree-Fock [19]. Citons B3LYP (hybride GGA), M06 et M11 (hybrides meta-GGA de la famille des fonctionnelles de la densité du Minnesota [20]). Si ces fonctionnelles n'incluent pas de terme spécifique au calcul des interactions de polarisation ou de dispersion, qui sont importantes dans nos systèmes d'études, leur procédure de paramétrisation (jeux de données expérimentales ou calculs de haute précision) leur garantit certaines performances pour rendre compte de propriétés thermo-chimiques, de cinétique, ou d'interactions non covalentes de systèmes spécifiques. D'autres fonctionnelles sont spécifiquement calibrées pour le calcul des interactions non covalentes, par l'ajout d'un terme de correction empirique. Pour l'étude d'agrégats, nous avons notamment eu recours à la fonctionnelle ω B97X-D [21], qui inclut un terme de correction de dispersion de Grimme [22].

Pour les calculs de fréquences de vibration, l'utilisation de la mécanique quantique aux niveaux DFT et post Hartree-Fock est limitée à des systèmes de quelques dizaines d'atomes. Cette limitation est due, d'une part à la mémoire nécessaire pour effectuer ces calculs et d'autre part à leur durée. Pour les calculs de chimie quantique, nous utilisons, suivant les cas, les logiciels Gaussian [23] ou Turbomole [24].

1.3 Approche indépendante du temps

Le calcul d'un spectre infrarouge statique peut être réalisé dans le cadre de la chimie quantique ou de la mécanique classique. Dans les deux cas, il faut pouvoir décrire la variation de l'énergie potentielle de la molécule lorsqu'elle vibre. Dans la limite de faibles oscillations autour de la position d'équilibre, on peut se placer dans l'hypothèse harmonique, pour laquelle la variation d'énergie potentielle selon une coordonnée d'oscillation peut être assimilée à une fonction quadratique du déplacement. Pour un oscillateur isolé, elle prend la forme :

$$E = \frac{1}{2}kx^2, \quad \text{avec } k = \frac{\partial^2 E}{\partial x^2} \quad (1.4)$$

L'expression peut être généralisée aux oscillations libres au voisinage d'une position d'équilibre pour des systèmes à N degrés de liberté.

Ce modèle, le plus simple qui soit, ne décrit que partiellement le mouvement vibratoire et ne peut rendre compte du déplacement d'un vibreur loin de sa position d'équilibre (cas d'une forte liaison hydrogène par exemple), ni représenter des mouvements de grande amplitude, observables parfois à l'échelle de la molécule.

1.3.1 En mécanique classique

En mécanique classique, on assimile les atomes à des masses ponctuelles liées entre elles par des forces (système 'masses-ressorts'). L'énergie potentielle d'un système à N atomes peut prendre une forme analytique plus ou moins complexe (des exemples seront donnés dans le **chapitre 2**). A une géométrie d'équilibre (un minimum local de la surface d'énergie potentielle), on peut diagonaliser le hessien (la matrice des dérivées secondes de l'énergie). Les 3N valeurs propres obtenues sont les fréquences propres harmoniques du système à cette géométrie d'équilibre. Les 3N vecteurs propres correspondent aux déplacements atomiques associés à ces fréquences. On les nomme modes normaux et correspondent aux combinaisons linéaires de déplacements atomiques pour lesquelles la matrice hessienne est diagonale.

Cette approche n'est utilisée que pour de très grands systèmes [25, 26], l'attrait de la mécanique moléculaire venant surtout de sa capacité à inclure les effets dynamiques, comme nous le verrons dans le **chapitre 2**.

1.3.2 En mécanique quantique

Double approximation harmonique

La première approximation se réfère à la définition de l'oscillateur harmonique précédemment décrit et stipule que dans la limite de faibles oscillations autour de la géométrie d'équilibre, l'énergie peut être assimilée à une fonction quadratique du déplacement. D'un point de vue quantique, le système peut peupler différents niveaux d'énergie vibrationnels quantifiés, qu'il atteint via des excitations de ses modes normaux. Chaque mode normal peut donc être dans son état fondamental, mais également être mono, di, tri . . . excité. La seconde approximation harmonique est basée sur un modèle simplifié de la vibration moléculaire pour lequel on fait l'approximation qu'il n'existe pas de couplages entre les modes, pas plus qu'entre les états d'un mode donné. Cette approche reste la plus utilisée en chimie quantique, à cause du coût souvent élevé que représentent les modèles plus élaborés, qui incluent une part de l'anharmonicité du mouvement. Les plus précis de ces modèles se voient ainsi limités au traitement de systèmes d'une dizaine d'atomes.

Par définition, le modèle harmonique surestime la plupart des fréquences calculées. Il est devenu habituel de proposer une correction *ad hoc* par un facteur d'échelle appliqué sur les fréquences harmoniques. On trouve de plus en plus de benchmarks de ces "scaling factors" [27–30], fonctions des méthodes et bases utilisées, de la gamme de fréquences étudiée, du type de système. Il est possible d'appliquer des corrections différentes sur chaque bande, justifiées par l'anharmonicité différente d'une bande d'élongation CO ou de déformation NH par exemple [31], ou des groupements amides A, I et II [32]. Dans certains cas, les corrections peuvent faire intervenir des modèles plus complexes [33]. Par essence, le facteur d'échelle ne peut que proposer une correction 'moyennée', et ne peut pas prétendre traiter aussi efficacement toutes les contributions d'un spectre sans s'astreindre à une bonne représentation de la surface d'énergie potentielle (SEP) qui les a générées. Le théoricien doit être vigilant à éviter de qualifier d'*ab initio* des calculs trop fortement dépendants de cette correction.

Modèle anharmonique

L'approche générale d'un calcul de spectre statique est la suivante :

(1) Une optimisation de la géométrie, afin d'obtenir une structure pour laquelle l'énergie est minimale (validée par des dérivées secondes de l'énergie par rapport aux coordonnées toutes positives, c'est à dire que toutes les fréquences sont réelles).

(2) Le choix d'une forme analytique pour le potentiel. Au delà du modèle harmonique, le choix le plus commun se porte sur un développement en série de Taylor, c'est à dire une fonction polynomiale dans laquelle les coefficients correspondent aux dérivées successives de l'énergie par rapport aux 3N-6 modes de vibration au point d'énergie minimale. Ce développement est généralement tronqué à l'ordre 4 [34] :

$$\begin{aligned}
 V(q_1, \dots, q_M) = & \frac{1}{2} \sum_{i=1}^M \left(\frac{\partial^2 V}{\partial q_i^2} \right)_0 q_i^2 + \frac{1}{6} \sum_{i=1}^M \sum_{j=1 \leq i}^M \sum_{k=1 \leq j}^M \left(\frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k} \right)_0 q_i q_j q_k \\
 & + \frac{1}{24} \sum_{i=1}^M \sum_{j=1 \leq i}^M \sum_{k=1 \leq j}^M \sum_{l=1 \leq k}^M \left(\frac{\partial^4 V}{\partial q_i \partial q_j \partial q_k \partial q_l} \right)_0 q_i q_j q_k q_l + \dots
 \end{aligned} \tag{1.5}$$

Les dérivées secondes de ce développement correspondent aux constantes de force harmoniques et les dérivées d'ordre supérieur aux constantes anharmoniques. q_i est la $i^{\text{ème}}$ coordonnée normale sans dimension

$$V(q_1, \dots, q_M) = \frac{1}{2} \sum_{i=1}^M \omega_i q_i^2 + \sum_{i,j,k} k_{i,j,k} q_i q_j q_k + \sum_{i,j,k,l} k_{i,j,k,l} q_i q_j q_k q_l + \dots \tag{1.6}$$

(3) Le calcul des différentes constantes de forces (harmoniques, cubiques, quartiques, ...) à la géométrie optimisée. Le calcul et l'ajustement de la SEP peuvent compter pour une grande partie du temps de calcul total, principalement si on cherche à exprimer le potentiel par un polynôme d'ordre élevé. Certains algorithmes optimisent cette recherche en adaptant le degré du développement polynomial (et donc le nombre de calculs de structure électronique) dans une direction de la SEP à la complexité du mouvement. On peut notamment citer AGAPES [35], qui décrit le potentiel autour d'une géométrie d'équilibre par un polynôme dont l'ordre n'est pas figé à l'avance (principe d'une SEP adaptative). L'expansion de la SEP n'est réalisée que jusqu'à un ordre nécessaire, au voisinage de chaque minimum et en ne tenant compte que des couplages multimodes pertinents.

(4) Le traitement à proprement parler du problème vibrationnel se fait par résolution du Hamiltonien vibrationnel, généralement le Hamiltonien de Watson [36]. L'objectif est d'obtenir les énergies des niveaux vibrationnels, et donc les fréquences de transition d'un niveau à l'autre.

$$\left\{ \begin{array}{l} \frac{\hat{H}}{hc} = \left[\frac{1}{2} \sum_i \omega_i P_i^2 + \hat{V}_{harmonic} \right] + \hat{V}_{anharmonic} + \sum_{\alpha} B_{\alpha} P_{\alpha}^2 \\ \frac{\hat{H}}{hc} = \hat{H}_0 + \hat{V}_{anharmonic} + \sum_{\alpha} B_{\alpha} P_{\alpha}^2 \end{array} \right. \quad (1.7)$$

Le Hamiltonien est décomposable en plusieurs contributions. Le terme d'ordre 0 (\hat{H}_0) comprend le terme d'énergie cinétique et la composante harmonique du potentiel. Le terme $\hat{V}_{anharmonic}$ comprend la composante (éventuelle) anharmonique. Ce terme dépend de la fonction analytique du potentiel, comme nous l'avons vu **éq. 1.5**. Enfin, un terme additionnel, dit de Coriolis [37], peut rendre compte du couplage entre vibration et rotation.

La théorie relative à la résolution de cet Hamiltonien est très ouverte, sujette à de nombreuses approximations suivant les modèles utilisés, le traitement total exact étant souvent très compliqué et informatiquement irréalisable. Nous décrivons brièvement les méthodes les plus répandues, en les classant par ordre de précision croissante.

Méthode Self-consistent field

La méthode VSCF (Vibrational Self-Consistent Field) [34] [38], consiste à traiter les interactions d'un mode avec ses multi-excitations (interactions intramodes), dans le champ moyen des autres modes. L'ansatz de l'approximation est basé sur une description de la fonction d'onde sous la forme d'un produit de ses multiexcitations.

$$\Psi_n(Q_1, \dots, Q_N) = \prod_{j=1}^N \Psi_j^{(n)}(Q_j) \quad (1.8)$$

où Q_1, \dots, Q_N sont les coordonnées normales pondérées par la masse.

Elle permet de traiter précisément un grand nombre de degrés de liberté [39–41] mais la corrélation entre modes (interaction intermodes, pouvant donner lieu à des résonances lorsque celles-ci sont fortes) n'étant pas prise en compte, certains mouvements sont mal traduits. Toutefois la méthode VSCF peut fournir une base alternative aux oscillateurs harmoniques (appelée base HO) pour des méthodes plus précises dites post-VSCF ou cc-VSCF (correlation-corrected VSCF). On différenciera dans la suite une base d'états oscillateurs harmoniques (HO) μ ou d'états VSCF Ψ_i^0 .

Méthodes perturbatives

La théorie Moller-Plesset vibrationnelle [38] est une méthode perturbative qui utilise généralement les états VSCF comme Hamiltonien d'ordre 0 (mais peut aussi bien être initiée sur les bases HO). On introduit un opérateur de perturbation qui décrit la différence entre les vraies interactions multimodes et la représentation à champ moyen de VSCF. Cet opérateur peut se décrire comme une séquence de corrections perturbatives de différents ordres (VSCF, VMP2, VMP3, VMP4 ...).

Il y a historiquement deux façons de représenter l'énergie dans la théorie VMP2, en fonction du choix de la partition de l'énergie :

Soit dans la base des HO :

$$E_\nu^{MP2} = E_\nu^{(0)} + \langle \nu | H^{(1)} | \nu \rangle + \langle \nu | H^{(2)} | \nu \rangle + \sum_{k \neq \nu} \frac{\langle \nu | H^{(1)} | \nu \rangle}{E_\nu^{(0)} - E_k^{(0)}} \quad (1.9)$$

soit dans une base d'états VSCF pour laquelle les multi-excitations entre un mode et lui-même sont déjà incluses dans la solution VSCF à l'ordre 0.

$$E_i^{VMP2} = \langle \Psi_i^0 | H | \Psi_i^0 \rangle + \sum_{i \neq j} \frac{(\langle \Psi_i^0 | H | \Psi_j^0 \rangle)^2}{E_i^{(0)} - E_j^{(0)}} \quad (1.10)$$

où E_i^{VMP2} est l'énergie associée à l'état vibrationnel i , Ψ_i^0 représente la configuration de la vibration i qui est vecteur propre du traitement vibrationnel à l'ordre 0, $E_i^{(0)}$ correspond au terme $\langle \Psi_i^0 | \hat{H} | \Psi_i^0 \rangle$ et $\langle \Psi_i^0 | \hat{H} | \Psi_j^0 \rangle$ au terme de la représentation matricielle du Hamiltonien couplant la $i^{\text{ème}}$ et la $j^{\text{ème}}$ configurations. Contrairement à la méthode VSCF, la contribution des couplages intermodes est traitée explicitement et conduit par exemple à une meilleure estimation des élongations à haute fréquence dans les molécules semi-rigides en l'absence de fortes résonances [34]. Toutefois, les méthodes perturbatives surestiment les forts couplages anharmoniques, c'est-à-dire lorsque $E_i^0 \simeq E_j^0$ et/ou quand l'interaction entre les modes i et j , matérialisée par le terme $\langle \Psi_i^0 | \hat{H} | \Psi_j^0 \rangle$, est forte (résonances de Fermi ou de Darling-Dennison). On remarque que pour des niveaux quasiment dégénérés, la méthode diverge, c'est-à dire que la contribution perturbative au second ordre (second terme) tend vers l'infini. La méthode VMP2-m [34], forme canonique modifiée de VMP2, fournit une meilleure stabilité numérique dans le cas de dégénérescences proches. La limitation de la méthode est à la fois dans le degré de perturbation et dans la troncature du nombre de modes couplés permis. La méthode VMP ne peut prétendre représenter l'état exact, la convergence de la série de perturbation ne pouvant être garantie.

La méthode VPT2 (Vibrational perturbation theory second-order) [42] [43] [44] consiste en un partitionnement différent du potentiel anharmonique VMP, qui conduit à une forme simplifiée de l'énergie perturbative à l'ordre 2 :

$$E_i = E_i^{(0)} + \hbar \sum_i^N \omega_i (n_i + \frac{1}{2}) + \hbar \sum_{i \leq j}^N x_{ij} (n_i + \frac{1}{2})(n_j + \frac{1}{2}) \quad (1.11)$$

Dans ce cas encore, les contributions lors de résonances étant fortement surestimées, les alternatives usuelles consistent soient à annuler les termes de résonances (Deperturbed VPT2), soit à les traiter variationnellement (Generalized VPT2) [43]. C'est la méthode utilisée par défaut dans le logiciel GAUSSIAN en utilisant le mot clef *freq=anharmonic*. Certaines études sont réalisées sur la méthode TOSH [45] (Transition optimized shifted Hermite) qui est une nouvelle simplification de VPT2, au moyen d'un hessien partiel, et qui diminue assez fortement l'effort computationnel [46] pour une faible perte de précision dans le cadre de couplages localisés [45, 47].

Méthodes variationnelles

Les méthodes variationnelles permettent de prendre en compte les forts couplages entre les vibrations ou entre vibration et rotation, souvent à l'origine de fortes anharmonicités. Elles reposent sur la diagonalisation complète de la représentation matricielle du Hamiltonien de Watson dans une base HO ou VSCF. La méthode est exacte pour un Hamiltonien donné mais très coûteuse en temps de calcul à cause de la dimension de la matrice à diagonaliser. Elle se révèle généralement impraticable pour des systèmes de taille moyenne. Cette méthode est généralement admise comme la méthode de référence pour les calculs vibrationnels sur de petits systèmes, mais la taille des matrices (le nombre de configurations à prendre en compte pour décrire avec justesse chaque état) limite rapidement la taille des systèmes à étudier [48].

Vibrational Configuration Interaction (VCI) La méthode VCI [38] est une approche qui inclut explicitement la corrélation entre les différents degrés de liberté. Elle consiste à diagonaliser la représentation du Hamiltonien dans une base HO ou VSCF. Elle prend en compte les excitations à partir des modes occupés vers des modes virtuels/vacants et exprime la fonction d'onde comme une combinaison linéaire dans l'espace de tous les produits de Hartree.

$$|VCI\rangle = C_i |\Psi_i\rangle + \sum_{\mu} C_{\mu} \tau_{\mu} |\Psi_i\rangle \quad (1.12)$$

Un Full-VCI inclut la somme de toutes les excitations pour un Hamiltonien donné et une base monomode. La différence d'énergie entre VSCF et Full-VCI est appelée l'énergie de corrélation. Un traitement VCI tronqué (ne prenant qu'une partie de l'espace des excitations) ne fournit qu'une partie de la corrélation électronique, mais pour un coût moindre.

Méthode Coupled-Cluster La méthode VCC (Vibrational Coupled-Cluster), proposée par Christiansen en 2004 [49], prend également en compte la corrélation électronique. Par la forme exponentielle de l'ansatz (la formulation mathématique qui définit les fonctions d'onde vibrationnelles excitées à partir d'un produit de Hartree, généralement un état VSCF, et l'opérateur cluster T), comparée à la forme linéaire en VCI, VCC intègre généralement plus de corrélation et fournit ainsi une précision supérieure pour un Hamiltonien donné (donc pour un nombre de modes couplés donné) et un espace donné.

$$|VCC\rangle = \exp(\hat{T}) |\Phi_i\rangle \quad \text{où } \hat{T} = \sum_{\mu} t_{\mu} \tau_{\mu} \quad (1.13)$$

Par contre, à un même degré d'excitation, l'implémentation de VCC est plus coûteuse que pour VCI, même si de récents progrès ont été faits [50,51]. Pour un espace d'excitations complet, à la fois VCI et VCC peuvent prétendre représenter l'état exact.

Méthodes de variation-perturbation

Pour pallier ce problème de taille des matrices à diagonaliser, des méthodes hybrides, dites de variation-perturbation, combinent les traitements rapides permis par les méthodes perturbatives à la précision des calculs variationnels. Ce principe a été initialement implémenté dans la méthode CIPSI [52], pour des calculs d'énergies électroniques. Concernant le calcul des niveaux vibrationnels [53], une version améliorée est implémentée dans le

code VCI-P [54], qui propose de tronquer massivement l'espace des configurations à inclure dans le processus variationnel en ne sélectionnant que les configurations présentant les interactions les plus fortes avec l'état considéré. Pour chaque état, une petite matrice, contenant les contributions les plus pertinentes, est diagonalisée. Les autres configurations, qui décrivent très faiblement l'état, sont évaluées par une approche perturbative au second ordre. La méthode a notamment été appliquée au radical glycine [55] et au formamide [56].

La construction itérative des fonctions d'ondes vibrationnelles se fait par enrichissement progressif des matrices d'interaction de configuration (IC). Les fonctions d'ondes sont des combinaisons linéaires de fonctions d'onde harmoniques (Φ_i^0), pondérées par l'amplitude du couplage entre ces fonctions ($c_i\nu$). Pour chaque itération, l'énergie de l'état vibrationnel est obtenue par la somme de son énergie issue de la diagonalisation de la matrice IC (ϵ_i^{VCI}) et des contributions faibles calculées par l'approche perturbative.

$$\epsilon_i^{VCI-P} = \epsilon_i^{VCI} + \sum_{weak} \frac{\langle \Psi_i | \hat{H} | \Psi_{weak}^n \rangle}{\epsilon_i^{VCI} - \epsilon_{weak}^{(0)}} \quad (1.14)$$

1.4 Approches dépendantes du temps

Les méthodes statiques les plus coûteuses/élaborées n'échappent pas à un inconvénient majeur dans certains cas, à savoir que toute la description de la vibration n'est fonction que de la seule géométrie optimisée. D'autres approches, dites dépendantes du temps, permettent de simuler un système dynamique et d'explorer une partie de la SEP.

1.4.1 En mécanique classique

En mécanique classique, en assimilant les molécules à des systèmes masses-ressorts, on résout les équations du mouvement par la mécanique Newtonienne, ou par le formalisme Lagrangien, plus adapté.

Si on exprime l'énergie cinétique par :

$$2T = \sum_{i=1}^{3n} (\dot{q}_i)^2 \quad (1.15)$$

et l'énergie potentielle par :

$$2V = \sum_i^{3n} \sum_j^{3n} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_{eq} q_i q_j \quad (1.16)$$

pour lesquelles q_i et \dot{q}_i sont les coordonnées et les vitesses cartésiennes pondérées par les masses, alors le problème vibrationnel harmonique est formulé par 3N équations différentielles de Lagrange :

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_i} + \frac{\partial V}{\partial q_i} \right) = 0 \quad i = 1, 2, \dots, 3n \quad (1.17)$$

ce qui conduit à :

$$\ddot{q}_i + \sum_{j=1}^{3n} k_{ij}^{(q)} q_j = 0 \quad i = 1, 2, \dots, 3n \quad (1.18)$$

Généralement, ce système d'équations est résolu de façon matricielle, par une diagonalisation du Hessien. Les $3N$ valeurs propres obtenues sont les fréquences propres harmoniques du système. Les $3N$ vecteurs propres correspondent aux déplacements atomiques associés à ces énergies. On les nomme modes normaux et correspondent aux combinaisons linéaires de déplacements atomiques pour lesquelles la matrice hessienne est diagonale.

Notons que la propagation des équations du mouvement nécessite un algorithme d'intégration temporelle. Le plus connu est l'algorithme de Verlet, dont nous utilisons une variante, Verlet-Vitesse, pour calculer les positions et les vitesses pour chaque instant t de la trajectoire :

$$\begin{cases} r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \\ v(t + \delta t) = v(t) + \frac{1}{2}[a(t) + a(t + \delta t)]\delta t \end{cases} \quad (1.19)$$

1.4.2 En mécanique quantique

Cette partie s'appuie sur l'article publié par Marx et Hutter [57].

Dynamique de type Ehrenfest : La dynamique *ab initio* de type Ehrenfest est basée sur la résolution de l'équation de Schrödinger dépendante du temps :

$$i\hbar \dot{\psi}(\{r_i\}, \{R_I\}; t) = H\psi(\{r_i\}, \{R_I\}; t) \quad (1.20)$$

où r_i sont les coordonnées électroniques et R_I celles des noyaux.

La première étape consiste à décomposer la fonction d'onde totale en un produit d'une fonction nucléaire et d'une fonction électronique

$$\psi(r_i, R_I; t) \simeq \Psi(\{r_i\}; t)\chi(\{R_I\}; t)\exp\left[\frac{i}{\hbar}\int_{t_0}^t dt' E_e(t')\right] \quad (1.21)$$

E_e est un facteur de phase simplifiant la suite des équations et prenant la forme

$$E_e = \int \int dr dR \Psi^*(\{R_I\}; t)\chi^*(\{R_I\}; t)H_e\Psi(\{R_I\}; t)\chi(\{R_I\}; t) \quad (1.22)$$

A partir des équations (1.19) (1.20) et (1.21), l'équation de Schrödinger peut être transformée en un système de deux équations couplées qui décrivent le comportement des électrons dans le champ moyen des noyaux, et réciproquement :

$$\begin{cases} i\hbar \dot{\Psi} = -\sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 \Psi + \left\{ \int dR \chi^*(\{R_I\}; t) V_{n-e}(\{r_i\}, \{R_I\}) \chi(\{R_I\}; t) \right\} \Psi \\ i\hbar \dot{\chi} = -\sum_i \frac{\hbar^2}{2M_I} \nabla_I^2 \chi + \left\{ \int dr \Psi^*(\{r_i\}; t) H_e(\{r_i\}, \{R_I\}) \Psi(\{r_i\}; t) \right\} \chi \end{cases} \quad (1.23)$$

La résolution de ces équations peut être réalisée de manière itérative et est à la base de la méthode TDSCF (Time Dependant Self-Consistent Field) introduite en 1930 par Paul Dirac [58, 59].

La seconde étape consiste à dériver ces expressions vers la dynamique classique afin de pouvoir considérer les noyaux comme des particules classiques se déplaçant dans le champ moyen créé par les électrons, dont le comportement est défini par la fonction d'onde Ψ .

On obtient finalement deux équations :

$$\left\{ \begin{array}{l} M_I \ddot{R}_I(t) = -\nabla_I \int dr \Psi^* H_e \Psi \\ i\hbar \dot{\Psi} = -\sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 \Psi + V_{n-e}(\{r_{iI}\}, \{R_I(t)\}) \Psi \end{array} \right. \quad (1.24)$$

On a ainsi à la fois une dynamique quantique des électrons et classique des noyaux. La dynamique peut prendre en compte les états électroniques excités. Si l'on contraint la fonction d'onde à l'état fondamental, on se déplace effectivement sur la surface de potentiel de Born Oppenheimer de l'état fondamental. On note que la fonction d'onde électronique étant propagée, il n'est pas nécessaire de la recalculer à chaque pas de dynamique. Toutefois, la dynamique est dictée par les particules les plus rapides (réajustement de la densité électronique) et nécessite un pas d'intégration des équations du mouvement de l'ordre de 10^{-1} à 10^{-2} fs. Ceci rend la méthode très difficile à appliquer à l'étude de systèmes avec de nombreux niveaux de liberté (dynamique en phase condensée par exemple) puisque le potentiel nucléaire doit lui être recalculé à tous ces pas. La méthode a été utilisée dans des problèmes de diffusion de molécules en phase gazeuse [60–62], de collision sur des surfaces (par exemple d'atomes d'oxygène sur une surface de graphite [63]), voire pour l'étude d'espèces en solutions (mélange K-KCl [64]) dans l'approximation de ne traiter explicitement qu'une petite partie des électrons.

Dynamique de type Born-Oppenheimer : Pour de nombreuses applications, il faut avoir accès à de plus longs temps de simulation. Cette condition impose que le pas d'intégration des équations du mouvement ne soit fonction que de la dynamique des noyaux. L'approximation de Born Oppenheimer utilise le rapport important entre les masses des noyaux et des électrons pour déduire que les électrons sont capables de s'adapter instantanément à une variation de géométrie nucléaire. Dans cette hypothèse, on peut découpler le mouvement des noyaux et des électrons, ces derniers étant caractérisés par une fonction d'onde électronique $\Psi_{elec}(\mathbf{r}; \mathbf{R})$ où \mathbf{R} représente la position des noyaux. Les noyaux peuvent ainsi être traités classiquement et propagés au cours d'une dynamique via les équations de Newton. A chaque pas de la dynamique, le problème de structure électronique est résolu par l'équation de Schrödinger indépendante du temps à un jeu de coordonnées nucléaires fixes [65–69].

Le système d'équation associé devient :

$$\left\{ \begin{array}{l} M_I \ddot{R}_I = -\nabla_I \min_{\Psi_0} \{ \langle \Psi_0 | H_e | \Psi_0 \rangle \} \\ E_0 \Psi_0 = H_e \Psi_0 \end{array} \right. \quad (1.25)$$

La différence majeure avec les équations de Ehrenfest est que la fonction d'onde doit être minimisée à chaque géométrie (et non simplement propagée à partir de la fonction d'onde minimum de la géométrie de départ). Les noyaux sont ensuite déplacés dans le champ de potentiel créé par les électrons. Cette optimisation à chaque pas induit un coût important de la méthode, partiellement compensé par la possibilité d'augmenter le pas d'intégration de la dynamique. Les états excités peuvent également être traités.

Dynamique de type Car-Parinello : En 1985, Car et Parinello proposent une solution alternative qui veut combiner les avantages des méthodes de Ehrenfest et Born-Oppenheimer, c'est à dire pouvoir propager la fonction d'onde électronique (et donc conserver une dynamique des électrons), tout en conservant un pas de temps d'intégration assez grand pour l'étude de systèmes complexes. La solution est apportée par l'emploi des fonctions

d'onde électronique comme des variables de la dynamique. Le problème à deux composantes quantique/classique est ramené à un problème à deux composantes classiques, au détriment de l'expression explicite de la dépendance en temps de la partie électronique.

L'approximation de Born-Oppenheimer utilise la séparation par échelle de temps entre les mouvements lents des noyaux et rapides des électrons pour garantir l'adiabaticité de la dynamique. Car et Parinello l'ont réexprimée en termes de séparation par échelle d'énergie des deux parties (noyaux et électrons) d'un même système dynamique. Maintenir l'adiabaticité de la dynamique est primordial pour éviter le transfert d'énergie des noyaux 'chauds' vers les électrons 'froids' et garantir de propager la fonction d'onde au plus proche de la surface de Born-Oppenheimer. En pratique, cela est assuré lorsqu'il n'y a pas de recouvrement des domaines de fréquences de vibrations des électrons et des noyaux.

La façon la plus simple de décrire le formalisme de la méthode de Car-Parinello est de se placer dans une description Lagrangienne de l'équation de Newton. Le Lagrangien \mathcal{L} du système s'écrit, en mécanique classique $\mathcal{L} = T - V$, avec T son énergie cinétique et V son énergie potentielle.

$$\mathcal{L} = \sum_{\alpha} \frac{1}{2} m_{\alpha} \dot{R}_{\alpha}^2 - V(R) \quad (1.26)$$

Avec cette définition, la nouvelle formulation des équations de Newton conduit à

$$m_{\alpha} \ddot{R}_{\alpha} = - \frac{\partial V}{\partial R_{\alpha}} \iff \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{R}_{\alpha}} = \frac{\partial \mathcal{L}}{\partial R_{\alpha}} \quad (1.27)$$

Car et Parinello ont introduit une nouvelle classe de Lagrangiens pour prendre en compte l'énergie du sous-système électronique $\langle \Psi_0 | H_e | \Psi_0 \rangle$ ainsi que les contraintes qui peuvent être imposées au jeu d'orbitales (notamment l'orthonormalité de la matrice des coefficients).

$$L_{CP} = \sum_{\alpha} \frac{1}{2} M_{\alpha} \dot{R}_{\alpha}^2 + \sum_i \frac{1}{2} \mu_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle - \langle \Psi_0 | H_e | \Psi_0 \rangle + \text{contraintes} \quad (1.28)$$

Les équations couplées des mouvement des noyaux et des électrons deviennent :

$$\begin{cases} M_I \dot{R}_I = - \frac{\partial}{\partial R_I} \langle \Psi_0 | H_e | \Psi_0 \rangle + \frac{\partial}{\partial R_I} + \text{contraintes} \\ \mu_i \ddot{\psi}_i = - \frac{\delta}{\delta \psi_i^*} \langle \Psi_0 | H_e | \Psi_0 \rangle + \frac{\delta}{\delta \psi_i^*} + \text{contraintes} \end{cases} \quad (1.29)$$

Les μ_i sont les masses fictives associées aux fonctions d'onde électroniques. Leur choix doit garantir une intégration efficace des équations du mouvement (i.e un pas de temps raisonnable) et une séparation adiabatique des degrés de liberté des noyaux et des électrons. En réalité, tant que μ est différent de 0, il y aura toujours une part de non adiabaticité due au transfert d'énergie des noyaux vers les électrons, et donc une diminution de la température des ions pendant la simulation [70, 71]. Pour minimiser les recouvrements entre les spectres de puissances des noyaux et des électrons, une solution parfois observée est de mimer/remplacer les hydrogènes par des deutériums [72, 73]. Plus généralement, les études de spectroscopies IR par Car-Parinello se limitent au lointain et moyen IR [74–76], tandis que le proche IR est traité par des dynamiques de type Born-Oppenheimer [65–68], mais pas toujours [69, 77–79].

1.5 Intensités IR calculées

Le calcul d'un spectre vibrationnel trouve généralement son intérêt dans l'interprétation d'un spectre IR expérimental, sur des systèmes pour lesquels le nombre important de bandes exclut une identification immédiate. Le plus souvent, il s'agit de spectres d'absorption dont l'intensité des bandes dépend de la variation du moment dipolaire associée à la vibration d'un mode. Bien reproduire un spectre implique de représenter avec précision à la fois les fréquences mais également les intensités de chaque transition.

1.5.1 Calculs statiques

La détermination *ab initio* des intensités vibrationnelles repose sur le calcul des éléments de la matrice de transition du moment dipolaire entre les états initial ψ_i^v et final ψ_f^v :

$$I_{if} = \frac{2\pi^2 N_A}{3hc\epsilon_0} \bar{\nu}_{if} \langle \psi_i^v | \mu(Q) | \psi_f^v \rangle^2 (N_i - N_f) \quad (1.30)$$

où $\mu(Q) = \langle \psi_g^e | \mu(Q) | \psi_g^e \rangle$ est le moment dipolaire en fonction des coordonnées nucléaires Q et ψ_g^e est l'état électronique fondamental. $\bar{\nu}_{if}$ est la fréquence de la transition, en cm^{-1} . Le facteur $(N_i - N_f)$ est la différence du nombre de molécules entre l'état initial et l'état final, en utilisant une distribution de Boltzmann et dépendant de la température.

Dans l'approche la plus répandue de la double approximation harmonique, une fois les fréquences obtenues par diagonalisation du hessien, les intensités IR sont également calculées dans cette approximation, pour laquelle la fonction d'onde vibrationnelle est représentée par un produit d'oscillateurs harmoniques et la fonction moment dipolaire comme une fonction linéaire des coordonnées normales. Pour un mode décrit selon la coordonnée normale Q_s , l'intensité IR est donnée par :

$$I_s = \frac{N_A}{12c^2\epsilon_0} \left(\frac{\partial \mu}{\partial Q_s} \right)_0^2 = 974.892 \mu_s^2 \quad (1.31)$$

où I_s est exprimé en km.mol^{-1} , $\mu_s = (\partial \mu / \partial Q_s)_0$ est exprimé en $\text{e.u}^{-1/2}$ ('u' est l'unité de masse atomique unifiée et 'e' est la charge élémentaire).

Cette forme est complexifiée lorsque sont prises en compte l'anharmonicité mécanique (anharmonicité du potentiel) et électrique (dépendance non linéaire du moment dipolaire aux coordonnées normales) au travers des bandes harmoniques ($k \mu_i$) ou de combinaison ($k \mu_i + l \mu_j$). De telles approches ont été implémentées par Barone dans la suite logicielle Gaussian09, dans la suite des développements sur l'approche perturbative GVPT2 pour les fréquences [80, 81]. Le traitement correct des intensités lors de résonances de Fermi fait l'objet de recherches actuelles (voir [81] pour une liste des méthodes proposées). Une approche variationnelle parallélisée [82] a été proposée par Bégué et al. dans le logiciel P-ANHAR [83].

1.5.2 Calculs dynamiques

L'approche dynamique permet d'avoir accès à des informations supplémentaires, telles que les effets de température sur la position ou l'élargissement des bandes. Des spectres dynamiques peuvent être obtenus par transformée de Fourier de la fonction d'autocorrélation d'une propriété du système.

Théorème de Wiener-Khinchin

Le théorème de Wiener-Khinchin [84] stipule que ‘la densité spectrale de puissance d’un processus stochastique stationnaire f est la transformée de Fourier de la fonction d’autocorrélation correspondante’. La fonction d’autocorrélation à l’instant t pour un processus stationnaire $f(t)$, correspond à la moyenne du produit scalaire entre le vecteur f à l’instant t et à l’instant initial sur un ensemble de systèmes identiques, notée $\langle f(0)|f(t) \rangle$. La fonction d’autocorrélation normée, $R_f(t)$ est exprimée par l’équation :

$$R_f(t) = \frac{\langle f(0)f(t) \rangle}{\langle f(0)f(0) \rangle} \quad (1.32)$$

Dans le cadre de l’hypothèse ergodique pour un processus stationnaire, la moyenne de la fonction $f(t)$ pour un ensemble de systèmes à un instant donné t équivaut à la limite de sa moyenne temporelle pour un système unique sur une durée infinie. Cela revient à définir l’autocorrélation au temps t de la manière suivante :

$$R_f(t) = \int_{-\infty}^{\infty} f(t)f(t+\tau)dt \quad (1.33)$$

Pour un signal de durée finie, celle-ci est divisée en portions de longueur T , qui doit être plus grande que les ‘périodicités’ éventuelles dans le processus $f(t)$. L’autocorrélation est alors reformulée sur l’intervalle fermé $[-T;T]$ et normée par la longueur $2T$ de celui-ci, ce qui donne :

$$R_f(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(t)f(t+\tau)dt \quad (1.34)$$

Dans la mesure où le processus étudié n’est pas strictement stationnaire, l’égalité précédente n’est valable que pour un intervalle de temps de durée infinie. D’après le théorème de Wiener-Khinchin, la densité spectrale à une pulsation ω , $S_f(\omega)$ peut être exprimée

$$S_f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_f(\tau)e^{-j2\pi\omega\tau} d\tau \quad (1.35)$$

comme la transformée de Fourier de $R_f(\tau)$, la fonction d’autocorrélation du signal aléatoire $f(t)$.

Le théorème de Wiener-Khinchin peut être appliqué pour calculer le spectre infrarouge par transformée de Fourier de la fonction d’autocorrélation (ACF) des vitesses (VACF) ou du moment dipolaire (DACF).

Le VACF est facile à mettre en oeuvre car les vitesses sont généralement calculées à chaque pas de la dynamique et les données sont facilement extractibles d’une trajectoire de dynamique moléculaire propagée par les logiciels les plus répandus. Les spectres simulés par VACF incluent les effets thermiques sur le déplacement et l’élargissement des bandes. Cependant, les fluctuations des vitesses atomiques n’ont pas toutes la même influence sur l’évolution du moment dipolaire et ne suffisent pas à caractériser entièrement son évolution. En cela, les intensités VACF n’ont pas de signification physique directe et il est malaisé/dangereux de vouloir les comparer à un spectre d’absorption. Au contraire, si la propriété considérée est le moment dipolaire du système, alors le spectre de puissance obtenu par transformée de Fourier de sa fonction d’autocorrélation (DACF) est bien un spectre de puissance vibrationnel [85, 86]. Le DACF a été implémenté dans TINKER. Le moment dipolaire est stocké dans un fichier à chaque pas de temps. Le programme Fourier de A. Kohlmeier [87] permet le calcul de la transformée de Fourier de la fonction d’autocorrélation d’une suite de valeurs dans le but de simuler un spectre de puissance. En fonction du

fichier d'entrée, il peut être utilisé pour simuler des spectres VACF ou DACF. Le comportement classique des noyaux introduit une approximation sur le calcul de densité spectrale. La formule pour le coefficient d'absorption infrarouge par unité de longueur dans l'approximation du dipôle électrique est donnée en fonction de la densité spectrale I par :

$$\alpha(\omega) = \frac{4\pi^2\omega}{3V\hbar cn(\omega)}(1 - e^{-\beta\hbar\omega})I(\omega) \quad (1.36)$$

avec V le volume de l'échantillon et $n(\omega)$ l'indice de réfraction pour une température $T = \frac{1}{k_B\beta}$. Pour la spectroscopie en phase gazeuse, $n \sim 1$.

Plusieurs facteurs de corrections quantiques peuvent être appliqués, résumés dans [57]. L'utilisation d'un facteur quantique harmonique $Q_{HA} = \frac{\beta\hbar\omega}{1 - e^{-\beta\hbar\omega}}$ est préférée dans deux nombreuses études [88–92] et conduit à un coefficient d'absorption α_{HA} tel que :

$$\alpha_{HA}(\omega) = \frac{4\pi^2\omega^2}{3V\hbar cn(\omega)}I_{CL}(\omega) \quad (1.37)$$

où I_{CL} désigne l'intensité spectrale dans l'approximation classique.

A présent qu'un panel des méthodes couramment utilisées a été décrit, on peut répondre aux questions d'introduction de ce chapitre. Les systèmes étudiés au cours de cette thèse sont tous caractérisés par une grande flexibilité, soit intra-, soit inter-moléculaire, et requièrent au minimum d'être décrits par un mélange de conformations ou de structures. La place des effets dynamiques semble très dépendante des systèmes et de la température d'étude. Dans un peptide à température ambiante, un ion faiblement solvato à 17 K, ou un gros agrégat à plus haute température, la dynamique conformationnelle n'a probablement pas la même importance. Pour certaines propriétés, telles que la signature d'un transfert de proton au sein d'un diacide, reproduire des effets dynamiques semble primordial. Quid de la description explicite des électrons ? La dynamique au sein de nos systèmes d'étude repose sur un traitement précis des interactions non covalentes, dans le domaine de compétences du champ de forces polarisable AMOEBA, qui justifie de plus son utilisation par un faible coût calculatoire et la possibilité de simuler l'évolution des comportements dynamiques sur des temps de simulations suffisamment longs. Nous avons donc, le plus souvent, utilisé une approche de dynamique basée sur le champ de forces AMOEBA. Lorsque la taille des systèmes le permettait, nous avons comparé les spectres DACF avec des superpositions de spectres statiques de plusieurs structures, obtenus par une approche de chimie quantique. Ceci a facilité l'attribution de bandes, mais nous avons aussi développé une approche d'attribution basée sur la dynamique. Enfin, lorsque les limites d'AMOEBA étaient atteintes quant à la représentation des électrons, nous avons cherché à le développer pour maintenir la cohérence de notre approche.

Bibliographie

- [1] Pople, J. A. ; Head-Gordon, M. ; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968-5975.
- [2] Riplinger, C. ; Sandhoefer, B. ; Hansen, A. ; Neese, F. *J. Chem. Phys.* **2013**, *139*, 134101.
- [3] Sparta, M. ; Neese, F. *Chem. Soc. Rev.* **2014**, *43*, 5032-5041.
- [4] Møller, C. ; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618-622.
- [5] Head-Gordon, M. ; Pople, J. A. ; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503-506.
- [6] Kohn, W. ; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133-A1138.
- [7] Zhao, Y. ; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415-432.
- [8] Johnson, E. R. ; Mackie, I. D. ; DiLabio, G. A. *J. Phys. Org. Chem.* **2009**, *22*, 1127-1135.
- [9] Sousa, S. F. ; Fernandes, P. A. ; Ramos, M. J. *J. Phys. Chem. A* **2007**, *111*, 10439-10452.
- [10] Karton, A. ; Gruzman, D. ; Martin, J. M. L. *J. Phys. Chem. A* **2009**, *113*, 8434-8447.
- [11] Amina, E. A. ; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 75-85.
- [12] Santra, B. ; Michaelides, A. ; Fuchs, M. ; Tkatchenko, A. ; Filippi, C. ; Scheffler, M. *J. Chem. Phys.* **2008**, *129*, 194111.
- [13] Thakkar, A. J. ; McCarthy, S. P. *J. Chem. Phys.* **2009**, *131*, 134109.
- [14] Riley, K. E. ; Holt, B. T. O. ; Merz, K. M. *J. Chem. Theory Comput.* **2007**, *3*, 407-433.
- [15] Zhao, Y. ; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289-300.
- [16] Zhao, Y. ; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 5121-5129.
- [17] Zhao, Y. ; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656-5667.
- [18] Riffet, V. ; Frison, G. ; Bouchoux, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18561-18580.
- [19] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- [20] Zhao, Y. ; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157-167.
- [21] Chai, J.-D. ; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 084106.
- [22] Grimme, S. *WIREs Comput. Mol. Sci.* **2011**, *1*, 211-228.
- [23] Frisch, M. J. *et al.* "Gaussian09 Revision E.01", Gaussian Inc. Wallingford CT 2009.
- [24] "TURBOMOLE V6.2 2010, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007 ; available from <http://www.turbomole.com>.", .
- [25] Tama, F. ; Gadea, F. X. ; Marques, O. ; Sanejouand, Y.-H. *Proteins : Struct., Funct., Bioinf.* **2000**, *41*, 1-7.
- [26] Li, G. ; Cui, Q. *Biophysical Journal* **2002**, *83*, 2457-2474.
- [27] Andersson, M. P. ; Uvdal, P. *J. Phys. Chem. A* **2005**, *109*, 2937-2941.

- [28] Jiménez-Hoyos, C. A. ; Janesko, B. G. ; Scuseria, G. E. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6621.
- [29] Alecu, I. M. ; Zheng, J. ; Zhao, Y. ; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 2872-2887.
- [30] Merrick, J. P. ; Moran, D. ; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683-11700.
- [31] Bouteiller, Y. ; Gillet, J.-C. ; Grégoire, G. ; Schermann, J.-P. *J. Phys. Chem. A* **2008**, *112*, 11656-11660.
- [32] Chin, W. ; Mons, M. ; Dognon, J.-P. ; Mirasol, R. ; Chass, G. ; Dimicoli, I. ; Piuzzi, F. ; Butz, P. ; Tardivel, B. ; Compagnon, I. ; von Helden, G. ; Meijer, G. *J. Phys. Chem. A* **2005**, *109*, 5281-5288.
- [33] Nagornova, N. S. ; Guglielmi, M. ; Doemer, M. ; Tavernelli, I. ; Rothlisberger, U. ; Rizzo, T. R. ; Boyarkin, O. V. *Angew. Chem. Int. Ed.* **2011**, *50*, 5383-5386.
- [34] Bounouar, M. ; Scheurer, C. *Chem. Phys.* **2008**, *347*, 194 - 207.
- [35] Richter, F. ; Carbonniere, P. ; Dargelos, A. ; Pouchan, C. *J. Chem. Phys.* **2012**, *136*, 224105.
- [36] Watson, J. K. *Mol. Phys.* **1968**, *15*, 479.
- [37] Carbonniere, P. ; Barone, V. **2004**, *392*, 365 - 371.
- [38] Christiansen, O. ; Luis, J. M. *Int. J. Quant. Chem.* **2005**, *104*, 667-680.
- [39] Adesokan, A. ; Fred, E. ; Brown, E. ; Gerber, R. *Mol. Phys.* **2005**, *103*, 1505 - 1520.
- [40] Pele, L. ; Brauer, B. ; Gerber, R. B. *Theor. Chem. Acc.* **2007**, *117*, 69-72.
- [41] Brauer, B. ; Chaban, G. M. ; Gerber, R. B. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2543-2556.
- [42] Pele, L. ; Gerber, R. B. *J. Phys. Chem. C* **2010**, *114*, 20603-20608.
- [43] Biczysko, M. ; Bloino, J. ; Carnimeo, I. ; Panek, P. ; Barone, V. *J. Mol. Struct.* **2012**, *1009*, 74-82.
- [44] Barone, V. ; Biczysko, M. ; Bloino, J. ; Borkowska-Panek, M. ; Carnimeo, I. ; Panek, P. *Int. J. Quant. Chem.* **2012**, *112*, 2185-2200.
- [45] Hanson-Heine, M. W. D. ; George, M. W. ; Besley, N. A. *J. Chem. Phys.* **2012**, *136*, 224102.
- [46] Besley, N. A. ; Bryan, J. A. *J. Phys. Chem. C* **2008**, *112*, 4308-4314.
- [47] Lin, C. Y. ; Gilbert, A. T. B. ; Gill, P. M. W. *Theor. Chem. Acc.* **2008**, *120*, 23-35.
- [48] Rauhut, G. *J. Chem. Phys.* **2007**, *127*, 184109.
- [49] Christiansen, O. *J. Chem. Phys.* **2004**, *120*, 2140-2148.
- [50] Christiansen, O. *J. Chem. Phys.* **2004**, *120*, 2149-2159.
- [51] Seidler, P. ; Christiansen, O. Vibrational coupled cluster theory. In *Recent Progress in Coupled Cluster Methods : Theory and Applications* ; Springer : **2010**.
- [52] Huron, B. ; Malrieu, J. P. ; Rancurel, P. *J. Chem. Phys.* **1973**, *58*, 5745-5759.
- [53] Pouchan, C. ; Zaki, K. *J. Chem. Phys.* **1997**, *107*, 342-345.
- [54] Carbonnière, P. ; Dargelos, A. ; Pouchan, C. *Theor. Chem. Acc.* **2010**, *125*, 543-554.
- [55] Carbonniere, P. ; Dargelos, A. ; Ciofini, I. ; Adamo, C. ; Pouchan, C. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4375.
- [56] Richter, F. ; Thauay, F. ; Lauvergnat, D. ; Carbonnière, P. *J. Phys. Chem. A* **2015**, *119*, 11719-11728.
- [57] Marx, D. ; Hutter, J. *Ab Initio Molecular Dynamics Basic Theory and Advanced Methods* ; Cambridge University Press : **2009**.
- [58] Deumens, E. ; Diz, A. ; Longo, R. ; Öhrn, Y. *Rev. Mod. Phys.* **1994**, *66*, 917-983.
- [59] Dirac, P. A. M. *Math. Proc. Cambridge* **1930**, *26*, 376-385.
- [60] Delos, J. B. ; Thorson, W. R. ; Knudson, S. K. *Phys. Rev. A* **1972**, *6*, 709-720.

- [61] Meyer, H., D.; Miller, W. H. *J. Chem. Phys.* **1979**, *70*, 3214-3223.
- [62] Saalman, U.; Schmidt, R. *Z. Phys. D Atom. Mol. Cl.* **2014**, *38*, 153-163.
- [63] Isborn, C. M.; Li, X.; Tully, J. C. *J. Chem. Phys.* **2007**, *126*, 134307.
- [64] Selloni, A.; Carnevali, P.; Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1987**, *59*, 823-826.
- [65] Jaqx, S.; Oomens, J.; Cimas, A.; Gageot, M.-P.; Rijs, A. M. *Angew. Chem. Int. Ed.* **2014**, *53*, 3663-3666.
- [66] Cimas, A.; Vaden, T. D.; de Boer, T. S. J. A.; Snoek, L. C.; Gageot, M.-P. *J. Chem. Theory Comput.* **2009**, *5*, 1068-1078.
- [67] Mathias, G.; Marx, D. *Proc. Natl. Acad. Sci.* **2007**, *104*, 6980-6985.
- [68] Marinica, D. C.; Grégoire, G.; Desfrancois, C.; Schermann, J. P.; Borgis, D.; Gageot, M. P. *J. Phys. Chem. A* **2006**, *110*, 8802-8810.
- [69] Beck, J. P.; Cimas, A.; Lisy, J. M.; Gageot, M.-P. *Spectroc. Acta A* **2014**, *119*, 12-17.
- [70] Tangney, P. *J. Chem. Phys.* **2006**, *124*, 044111.
- [71] Tangney, P.; Scandolo, S. *J. Chem. Phys.* **2002**, *116*, 14-24.
- [72] Bhargava, B.; Balasubramanian, S. *Chem. Phys. Lett.* **2006**, *417*, 486 - 491.
- [73] Kirchner, B.; Hutter, J. *Chem. Phys. Lett.* **2002**, *364*, 497 - 502.
- [74] Gageot, M.-P. *J. Phys. Chem. A* **2008**, *112*, 13507-13517.
- [75] Cimas, A.; Maitre, P.; Ohanessian, G.; Gageot, M.-P. *J. Chem. Theory Comput.* **2009**, *5*, 2388-2400.
- [76] Gageot, M.-P.; Sprik, M. *J. Phys. Chem. B* **2003**, *107*, 10344-10358.
- [77] Masia, M.; Forbert, H.; ; Marx, D. *J. Phys. Chem. A* **2007**, *111*, 12181-12191.
- [78] Martinez, M.; Gageot, M.-P.; Borgis, D.; Vuilleumier, R. *J. Chem. Phys.* **2006**, *125*, 144106.
- [79] Gageot, M. P.; Vuilleumier, R.; Sprik, M.; Borgis, D. *J. Chem. Theory Comput.* **2005**, *1*, 772-789.
- [80] Bloino, J.; Barone, V. *J. Chem. Phys.* **2012**, *136*, 124108.
- [81] Bloino, J.; Biczysko, M.; Barone, V. *J. Phys. Chem. A* **2015**, *119*, 11862-11874.
- [82] Bégué, D.; Baraille, I.; Garrain, P. A.; Dargelos, A.; Tassaing, T. *J. Chem. Phys.* **2010**, *133*, 034102.
- [83] Gohaud, N.; Begue, D.; Darrigan, C.; Pouchan, C. *J. Comput. Chem.* **2005**, *26*, 743-754.
- [84] McQuarrie, D. A. *Statistical Mechanics*; HarperCollins Publishers : **1976**.
- [85] Semrouni, D. *Energétique et spectroscopie de polypeptides par dynamique moléculaire : champ de force de seconde génération et chimie quantique*, Thesis, Ecole Polytechnique, Palaiseau, France, **2010**.
- [86] Semrouni, D.; Sharma, A.; Dognon, J.-P.; Ohanessian, G.; Clavaguéra, C. *J. Chem. Theory Comput.* **2014**, *10*, 3190-3199.
- [87] Kohlmeyer, A.; Forbert, H. "Fourier-Version 21.02", **2006**.
- [88] Ramirez, R.; Lopez-Ciudad, T.; Kumar P, P.; Marx, D. *J. Chem. Phys.* **2004**, *121*, 3973-3983.
- [89] Iftimie, R.; Tuckerman, M. E. *J. Chem. Phys.* **2005**, *122*, 214508.
- [90] Lawrence, C. P.; Nakayama, A.; Makri, N.; Skinner, J. L. *J. Chem. Phys.* **2004**, *120*, 6621-6624.
- [91] Schmitz, M.; Tavan, P. *J. Chem. Phys.* **2004**, *121*, 12247-12258.
- [92] Joalland, B.; Rapacioli, M.; Simon, A.; Joblin, C.; Marsden, C. J.; Spiegelman, F. *J. Phys. Chem. A* **2010**, *114*, 5846-5854.

Chapitre 2

Champs de forces

2.1 Généralités

Les systèmes étudiés dans le cadre de ce travail de thèse sont divers, de peptides en phase gazeuse aux ions microsolvatés, leurs tailles variant de quelques dizaines à quelques centaines d'atomes. Ces systèmes sont flexibles, et généralement dynamiques, ce qui signifie que les observables macroscopiques sont une moyenne de leur comportement, qui évolue dans le temps et avec la température. Ainsi, la méthode théorique choisie pour mettre en évidence ces effets ne peut se restreindre à étudier une structure figée du système, mais doit pouvoir rendre compte de ce comportement dynamique sur une échelle de temps caractéristique de l'évènement (repliement d'une chaîne peptidique, transfert de proton, réarrangement de sphères de solvation ...). Les possibilités offertes par les simulations de dynamique sont nombreuses et variées en termes d'échelles de taille, de temps, et de niveaux de théorie. On peut modéliser un système chimique de façon à représenter explicitement les atomes et les électrons (dynamique quantique), choisir de ne prendre en compte les électrons que de façon implicite (modèle 'tout atome' de dynamique classique), ou même ne pas décrire les atomes un à un, mais par agrégats (modèles '*coarse-graining*' ou 'gros grains' [1]). La modélisation de spectres infrarouge ne peut évidemment pas être réalisée à une échelle de précision au-dessus de l'atome. En outre, la dynamique *ab-initio* est drastiquement plus coûteuse que la dynamique moléculaire par champ de forces classiques et son application sur l'échelle de temps est limitée à quelques dizaines de picosecondes pour des systèmes de quelques dizaines d'atomes. Au **chapitre 6**, nous montrerons sur un exemple d'ion sulfate solvato par un agrégat de 25 molécules d'eau que les réponses de certains systèmes ne satisfont pas encore à l'ergodicité au terme de 200 ps de simulation à 200 K, ce qui est bien au-delà des possibilités offertes par la dynamique *ab-initio*. La dynamique moléculaire basée sur un champ de forces représente l'outil adapté à ces problématiques, à condition de contenir les composantes nécessaires pour décrire les interactions parfois complexes des systèmes. Depuis les premiers modèles dans les années 1945-1950 [2-4], les champs de forces se sont perfectionnés et sont devenus des outils essentiels aux études de phénomènes complexes dans de nombreux domaines, dont la science des matériaux ou les simulations de macromolécules biologiques. Le développement de champs de forces est un domaine très actif en chimie théorique 'qui relève autant de l'art que de la science' [5]. Le point commun à tous les champs de forces est qu'ils s'appuient sur la notion de groupement fonctionnel si essentielle en chimie, qui permet de simplifier et généraliser leur utilisation en transposant les formes analytiques et les paramètres définis sur des systèmes modèles à toute une série de systèmes réels plus complexes. Nous revenons rapidement sur les évolutions successives des champs de forces avant de détailler le champ de forces polarisable AMOEBA, utilisé tout au long de cette thèse.

2.2 Champs de forces additifs de paire (Champs de forces de 1^{ère} génération)

Si on décrit l'énergie potentielle V d'un système à N atomes comme une somme de n potentiels V_n à n -corps :

$$V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) = \sum_i^N u^{(1)}(\vec{r}_i) + \sum_{i<j}^N u^{(2)}(\vec{r}_i, \vec{r}_j) + \sum_{i<j<k}^N u^{(3)}(\vec{r}_i, \vec{r}_j, \vec{r}_k) + \dots + \sum_{i<j<\dots<z}^N u^{(n)}(\vec{r}_i, \vec{r}_j, \dots, \vec{r}_z) \quad (2.1)$$

Les vecteurs \vec{r}_i définissent les positions des N atomes, $u^{(1)}$ est le potentiel intramoléculaire (dépendant de champs extérieurs électrique, magnétique, ...) et $u^{(2)}$ est le potentiel d'interaction additif de paire. Les termes suivants sont appelés potentiels à n -corps et mettent en jeu l'interaction simultanée d'au moins 3 atomes. Ils sont généralement beaucoup plus coûteux à calculer et contribuent plus faiblement à l'énergie potentielle que les termes additifs.

Dans les champs de forces de 1^{ère} génération, dits également champs de forces classiques, l'expression du potentiel n'est développée que jusqu'à l'ordre 2 :

$$V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N) = \sum_i^N u^{(1)}(\vec{r}_i) + \sum_{i<j}^N u_{effective}^{(2)}(\vec{r}_i, \vec{r}_j) \quad (2.2)$$

dans laquelle le potentiel d'ordre 2 est dit effectif car il inclut, par sa paramétrisation, une partie des effets à n -corps de façon implicite. Ce terme inclut les interactions de valence (entre atomes liés) et les interactions entre atomes non liés (séparés par plus de deux liaisons).

2.2.1 Interactions de valence

Le potentiel de valence décrit les interactions entre atomes directement liés (interactions 1-2), séparés par deux liaisons (1-3) ou trois (1-4) :

$$V_{valence} = \sum V_{liaison} + \sum V_{angle} + \sum V_{tors} \quad (2.3)$$

L'expression analytique de chaque terme est propre au champ de forces ; les plus courants sont :

- Les termes de liaison et d'angle modélisés par des oscillateurs harmoniques sont utilisés dans les champs de forces de classe I (CHARMM, AMBER, OPL, GROMOS), réservés au traitement de larges systèmes (protéines, acides nucléiques, polymères complexes) :

$$V_{liaison} = \sum_{liaisons} \frac{1}{2} k_r (r - r_0)^2 \quad (2.4)$$

$$V_{angle} = \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \quad (2.5)$$

où $(r - r_0)$ et $(\theta - \theta_0)$ représentent la déviation de la liaison ou de l'angle par rapport à sa position de référence. k_r et k_θ sont les constantes de forces harmoniques correspondantes. Une description plus réaliste des fortes déviations par rapport à la position de référence peut être obtenue par d'autres formes, fonctions de Morse (pour le potentiel d'élongation) ou développement polynômial en série de Taylor. Les champs de forces qui les utilisent sont identifiés de classe II. Citons MMFF94, UFF ou la série MM1, MM2, MM3, et MM4, développés par Allinger et dédiés aux molécules organiques. Ils permettent de mieux calculer les énergies conformationnelles et ou les spectres vibrationnels, et sont plus orientés au traitement de petites et moyennes molécules. Le potentiel intramoléculaire d'AMOEBA dérive de MM3.

- La prise en compte des termes de torsion, angles dièdres et impropres dépend de chaque champ de forces. On peut retenir cette série de Fourier à 3 termes, issue des potentiels MM2 [6, 7], MM3 [8, 9] et MM4 [10, 11] d'Allinger, et utilisée dans de nombreux champs de forces pour rendre compte des interactions 1-4 :

$$V_{dièdre} = \sum_{\phi} [V_1(1 + \cos\phi) + V_2(1 - \cos 2\phi) + V_3(1 + \cos 3\phi)] \quad (2.6)$$

- Dans les champs de forces de classe II, des termes additionnels de couplage entre mouvements (élongation et pliage par exemple) viennent affiner la description du potentiel et contribuent à la flexibilité des molécules, élément primordial au bon traitement des spectres vibrationnels par exemple.

2.2.2 Interactions d'atomes non liés

Il s'agit principalement du terme de van der Waals et de l'électrostatique. Ces interactions existent à la fois au sein d'une molécule, mais également entre molécules, elles peuvent donc être intra- et inter-moléculaires

Le terme de van der Waals (vdW) de répulsion et dispersion contribue à reproduire certaines propriétés telles que la densité, la capacité calorifique, la chaleur de vaporisation, ou encore l'énergie libre d'hydratation.

Il est généralement exprimé via un potentiel 12-6 de Lennard-Jones (LJ) :

$$E_{vdW} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] = 4\epsilon \left(\frac{C_{12}}{r^{12}} - \frac{C_6}{r^6} \right) \quad (2.7)$$

dans lequel σ est la distance pour laquelle $E_{vdW} = 0$, et ϵ est la profondeur du puits. On peut faire apparaître deux coefficients tels que $C_{12} = (C_6)^2$.

La partie attractive en r^{-6} domine à longue distance et rend compte des forces de Keesom (dipôle permanent - dipôle permanent), Debye (dipôle permanent - dipôle induit) et London (dipôle induit - dipôle induit). La partie repulsive, en r^{-12} , dominante à courte distance, est empirique et rend compte de la répulsion de Pauli entre électrons, qui empêche l'interpénétration mutuelle des nuages électroniques.

Le terme électrostatique permet essentiellement de reproduire la réponse électronique du système à l'environnement. Dans son expression la plus simple, il est donné par la loi de Coulomb. Il décrit alors uniquement l'interaction entre les charges ponctuelles portées par deux atomes (A et B) :

$$E_{elec} = \frac{q_A q_B}{4\pi\epsilon_0 r_{AB}} \quad (2.8)$$

Cette forme simplifiée de l'électrostatique ignore le phénomène de polarisation, qui représente la déformation du nuage électronique en réponse au champ électrique induit par les charges des autres atomes. Cette contribution est très coûteuse en temps de calcul, ce qui justifie le choix de la 1^{ère} génération de champs de forces de ne pas inclure de terme explicite. Il y a alors une compensation partielle par les autres termes, généralement par une sous-estimation de la répulsion [12] alliée à une surestimation des charges atomiques, avec le risque que ces compensations ne fonctionnent plus pour un système trop éloigné physiquement ou chimiquement du jeu de molécules ou de géométries de référence.

Les liaisons hydrogènes ont un intérêt particulier dans quantités de systèmes (repliement de molécules biologiques, structure du réseau de molécules d'eau solvatant un ion ...) et doivent être correctement reproduites. Bien qu'elles puissent être vues comme des interactions de vdW un peu particulières, dans lesquelles les forces de Keesom (les interactions dipôle-dipôle) sont généralement un ordre de grandeur plus importantes que dans une interaction vdW classique, certains champs de forces incluent un terme spécifique pour cette liaison, par exemple une expression de type LJ (avec alors des coefficients différents, 12-10 [13]). En outre, il est possible d'inclure dans l'expression la directivité de la liaison avec un facteur $\cos\theta$ où θ est l'angle \widehat{XHY} dans une liaison hydrogène X-H...Y. Dans une version de MM2 [14], les liaisons hydrogènes sont prises en compte par une diminution du rayon de vdW de l'atome d'hydrogène impliqué dans une telle interaction, associée à une augmentation de son puits ϵ , uniquement dans le calcul de ses interactions avec des éléments électrostatifs.

Parmi les nombreux champs de forces de 1^{ère} génération, citons les plus connus ciblés sur la description des systèmes biologiques : AMBER, CHARMM, CVFF, GLYCAM, GROMACS, GROMOS, OPLS. De nombreuses études contribuent à comparer les performances des champs de forces les plus courants [15–18].

- AMBER (Assisted Model Building with Energy Refinement) [19–22] a été développé par le groupe de Kollman. La majorité des paramètres repose sur des données expérimentales (angles, torsions, moments dipolaires, chaleurs expérimentales, énergies libres de vaporisation, solvatation, sublimation ...). La méthode RESP [23] est utilisée pour optimiser les charges atomiques à partir de calculs quantiques. AMBER inclut un terme spécifique pour les liaisons hydrogènes, de type LJ (12-10) [13].
- CHARMM (Chemistry at Harvard Macromolecular Mechanics) [24,25] est développé par le groupe de Karplus à Harvard. Il inclut plusieurs variantes (protéines, lipides, ADN, ARN ...). La forme du potentiel est très proche de celle d'AMBER, à l'exception d'un terme de couplage Urey-Bradley 1-3 et de la paramétrisation des charges atomiques, qui est basée sur les charges de Mulliken.
- GROMOS (GRONingen MOlecular Simulation program package) [26] a été développé à l'Université de Groningen puis à l'ETH à Zurich. Même formalisme qu'AMBER mais le terme de vdW est un LJ un peu particulier puisque les termes C_6 et C_{12} sont calibrés de façon indépendante et ne sont plus liés par la relation $C_{12} = (C_6)^2$.
- OPLS-aa (Optimized Potential for Liquid Simulations-all atoms) [27] est développé par le groupe de Jorgensen. Basé sur le formalisme d'AMBER, il a été initialement développé pour reproduire les densités et les chaleurs de vaporisation des liquides de molécules organiques, puis a été étendu aux peptides. Les simulations OPLS utilisent généralement les modèles TIP4P [28] ou TIP3P [29] pour le modèle de l'eau.

Pour l'essentiel, les expressions analytiques de ces potentiels sont assez semblables. Les auteurs les ont souhaitées les plus simples et générales possibles, dans l'objectif de limiter le nombre de paramètres et de permettre la transférabilité d'un système à un autre. On observe une évolution dans les moyens de paramétrer les champs de forces au cours du temps. Dans les années 70, l'immense majorité des paramètres étaient issus de données expérimentales (cristallographie, thermodynamique et spectroscopie principalement). L'accès aux termes liés était plutôt aisé, mais celui des interactions entre atomes non liés l'était beaucoup moins. Le caractère empirique apporté par les données expérimentales doit être source de prudence, puisqu'il peut facilement être l'objet de compensations entre interactions dans un système spécifique, et nuire à la transférabilité. Avec l'augmentation des puissances de calcul, on a

vu apparaître des approches de mécanique statistique, permettant de reproduire des propriétés thermodynamiques, chaleur de vaporisation ou densité par exemple. Avec l'augmentation de la précision des méthodes quantiques, une plus grande part de la paramétrisation est maintenant issue de calculs de haut niveau théorique. Dans ce cadre, on comprend bien que les systèmes servant à la paramétrisation doivent être les plus restreints possibles (en taille et en nombre). Un attribut essentiel des champs de forces est donc la transférabilité des paramètres, l'objectif étant, à partir d'un jeu limité de données, d'étendre les propriétés du champ de forces à des systèmes de taille plus importante. A cet égard, un regard critique doit être porté sur les interactions non liées, qui apportent généralement une contribution plus importante dans les systèmes à taille réelle étudiés que dans les modèles de petite taille qui ont permis de les définir.

Pour décrire des systèmes contenant des macromolécules ou des ions, il est important de pouvoir rendre compte de la déformation de la densité électronique sous l'effet de charges. C'est dans ce but qu'a vu le jour une nouvelle génération de champs de forces, avec pour objectif la reproduction plus précise des structures moléculaires, l'obtention des conformations d'équilibre, ou la description des vibrations moléculaires (spectres infrarouge).

2.3 Champs de forces polarisables (Champs de forces de 2^{nde} génération)

En 1930, London postule que les interactions entre molécules peuvent être décomposées en 4 contributions : l'électrostatique, l'induction, la dispersion et l'échange. A grandes distances, les nuages électroniques ne se recouvrent pas, et il est possible d'exprimer ces termes à l'aide d'un développement multipolaire [30]. Pour les molécules chargées, les termes prépondérants sont l'électrostatique, qui résulte de l'interaction entre les moments multipolaires permanents de chaque molécule, et l'induction, qui est définie par l'interaction d'un moment multipolaire permanent d'une molécule avec le moment multipolaire induit de l'autre molécule. Dans les systèmes neutres ou apolaires, la dispersion régit le comportement à longue distance. En première approximation du développement multipolaire, le terme dominant est celui de l'interaction entre moments multipolaires induits, soit la polarisation mutuelle des distributions de charge. Enfin, l'échange rend compte de la répulsion provoquée par les recouvrements de densité électronique à courte distance.

Il existe deux approches en mécanique quantique pour calculer ces interactions : la théorie supermoléculaire, pour laquelle $E_{int} = E_{AB} - E_A - E_B$, dans le cas d'un dimère, et la théorie de la perturbation, qui décrit l'énergie d'interaction comme une série infinie de termes correctifs, appelés énergies de polarisation :

$$E_{int} = \sum_{i=0}^{\infty} E_{pol}^i \quad (2.9)$$

pour laquelle l'ordre 0 de l'énergie de polarisation est un terme purement électrostatique, les premier et second ordres admettent les énergies d'induction et de dispersion, etc ...

Les champs de forces ont été en majorité développés pour l'étude de phase condensée, dans laquelle l'effet polarisant du solvant joue un rôle important. En première approximation, il peut être vu comme un continuum diélectrique [31, 32], mais ne rend alors pas compte des interactions spécifiques entre soluté et solvant [33]. Des études montrent les difficultés des champs de forces à charges ponctuelles à rendre compte des énergies libres de solvation de petites molécules polaires [34]. L'importance de la polarisation dans la reconnaissance entre protéine et ligands a été identifiée [35, 36]. De nombreux modèles de polarisation explicite [37–42] ont été développés. Trois types de modèles sont très utilisés dans les applications sur des systèmes d'intérêt biologique : les modèles basés sur des multipôles ponctuels, le modèle de l'oscillateur de Drude et le modèle de charges fluctuantes.

2.3.1 Modèle de multipôles ponctuels

Ce modèle considère une molécule comme une distribution de charges localisées sur des centres (atomiques ou non, i.e. paires libres par exemple). Ainsi, la distribution de charge totale M_i sur un site i est la somme des charges permanentes et induites :

$$M_i = M_i^0 + M_i^{ind} \quad (2.10)$$

Le groupe de Claverie et coll. l'ont approximée par un ensemble de charges et dipôles ponctuels et polarisables [43] et modélisé la topologie électrostatique au voisinage d'une molécule [44]. Un modèle de dipôles ponctuels a été proposé par Silberstein [45] puis appliqué par Applequist [46]. Ce modèle reproduit bien les polarisabilités moyennes, mais la polarisation devient infinie lorsque les dipôles induits interagissent à une distance $R \leq (4\alpha_1\alpha_2)^{1/6}$, où α_1 et α_2 sont les polarisabilités atomiques. Pour corriger ce phénomène de 'catastrophe de polarisation', Tholé propose d'utiliser un facteur d'amortissement [47]. Des variantes utilisent la calibration de ce facteur d'amortissement comme un moyen de rendre les polarisabilités atomiques indépendantes de l'environnement [48].

Le développement multipolaire peut être étendu aux charges, dipôles, quadrupôles et/ou moments d'ordres supérieurs (champ de forces NEMO [49]). Dans le cas d'une distribution multipolaire M_i développée jusqu'aux quadrupôles :

$$M_i = [q_i, \mu_{ix}, \mu_{iy}, \mu_{iz}, Q_{ixx}, Q_{ixy}, Q_{ixz}, \dots, Q_{izz}] \quad (2.11)$$

où q_i est la charge ponctuelle localisée sur l'atome i , μ est le dipole, et Q est le quadrupôle. Seules 5 composantes des quadrupôles sont indépendantes, pour des raisons de symétrie.

L'interaction électrostatique entre les distributions de charges sur les sites i et j est donnée par :

$$E_{elec} = \frac{1}{2} \sum_i \sum_{j \neq i} M_i^t T_{ij} M_j \quad (2.12)$$

avec la distribution de charge induite :

$$M_i^{ind} = \alpha_i^{-1} \sum_{j \neq i} T_{ij} (M_j^0 + M_j^{ind}) \quad (2.13)$$

α_i^{-1} étant la polarisabilité de l'atome i et T_{ij} la matrice des interactions multipôle-multipôle. L'équation (1.13) est résolue de manière itérative jusqu'à un seuil de convergence afin d'obtenir les dipôles induits. On peut alors écrire l'énergie électrostatique finale :

$$E_{elec} = \frac{1}{2} \sum_i \sum_{j \neq i} (M_i^0)^t T_{ij} M_j^0 + \frac{1}{2} \sum_i \sum_{j \neq i} (M_i^{ind})^t T_{ij} M_j^0 \quad (2.14)$$

où le premier terme correspond à l'énergie électrostatique issue des moments 'permanents' et le second terme à l'énergie de polarisation générée par la présence des moments induits.

Ce modèle de polarisation est notamment utilisé par les champs de forces SIBFA [50,51], AMOEBA ou NEMO.

2.3.2 Modèle de Drude

L'oscillateur de Drude [52], du nom de Paul Drude et développé en 1902, est parfois appelé 'shell model', ou 'charge on spring model'. A l'origine, il a été imaginé pour l'étude de matériaux ioniques à l'état solide. Dans ce modèle, le phénomène de polarisation repose sur la présence d'une particule additionnelle, chargée et sans masse, attachée au centre de chaque centre atomique (à l'exception généralement des hydrogènes, pour des questions d'économies de calcul) par un ressort harmonique mimant les redistributions électrostatiques. La position de la particule est ajustée de façon autocohérente, en fonction des variations du champ électrostatique dues à l'environnement, pour n'importe quelle configuration du système. Elle prend ainsi en compte à la fois le champ électrique permanent dû aux charges fixes et la contribution issue des dipôles induits. L'avantage de l'oscillateur de Drude est qu'il préserve le modèle simple d'interaction particule-particule des interactions coulombiennes, et donc que son implémentation dans les codes de simulation biomoléculaire est relativement peu coûteuse. L'énergie électrostatique est la somme des interactions de paires entre les charges atomiques et les charges partielles des particules de Drude, dans une formulation qui permet d'éviter le phénomène de 'catastrophe de polarisation'. Il a été implémenté dans une version polarisable de CHARMM [53, 54].

2.3.3 Modèle des charges fluctuantes

Le modèle des charges fluctuantes est basé sur le modèle d'équilibration de charges [55]. Il utilise des charges variables, localisées sur les atomes et couplées à leur environnement. Les charges sont calculées par minimisation de l'énergie à une structure géométrique donnée, à chaque pas de temps. La relation de Coulomb est remplacée par une forme basée sur les valeurs des électronégativités et des duretés des atomes, qui sont équilibrées pendant la simulation. L'avantage est d'obtenir la polarisabilité sans introduire de nouvelles interactions. En contre-partie, le modèle réclame une diminution significative du pas d'intégration de la simulation, et donc une augmentation du coût calculatoire. Pour des macromolécules, la charge peut se déplacer à faible coût énergétique le long des liaisons covalentes et tend à surestimer la réponse de polarisation. Différents modèles ('atom-atom charge transfer' (AACT) [56], 'bond-charge increment' (BCI) [57, 58], 'atom-bond electronegativity equalization method' (ABEEM) [59]) proposent des solutions à ce problème de sur-polarisation.

Dans le formalisme dit des charges fluctuantes, le potentiel chimique est équilibré via une redistribution de la densité de charge. L'énergie d'un système apparaît ainsi comme dépendant de la charge et s'exprime pour un système de M molécules contenant N_i atomes par molécule sous la forme :

$$\begin{aligned}
 E(R, Q) = & \sum_{i=1}^M \sum_{\alpha=1}^N \chi_{i\alpha}^2 + \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \sum_{\alpha=1}^{N_i} \sum_{\beta=1}^{N_j} J_{i\alpha i\beta} Q_{i\alpha} Q_{j\beta} \\
 & + \frac{1}{2} \sum_{i=1}^{MN'} \sum_{j=1}^{MN'} \frac{Q_i Q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{j=1}^M \lambda_j \left(\sum_{i=1}^N Q_{ij} - Q_j^{Total} \right)
 \end{aligned} \tag{2.15}$$

où λ_i est le multiplicateur de Lagrange qui assure la conservation de la charge totale, Q_i est la charge partielle d'un site atomique et χ_i l'électronégativité atomique. J_i est la dureté atomique, proportionnelle à la dérivée seconde de l'énergie totale d'un système chimique par rapport au nombre d'électrons dans un environnement nucléaire fixe, et mesure de la résistance à la déformation et au transfert de charge. Ces paramètres sont optimisés afin de reproduire les moments dipolaires moléculaires et la réponse par polarisation moléculaire associée. Ce modèle a été utilisé pour étendre CHARMM en CHARMM-FQ [60, 61] couplé au modèle polarisable dépendant de la charge de l'eau TIP4P-QDP [62]. Il a également été utilisé pour évaluer le transfert de charge entre l'ion Gd^{3+} et sa sphère de solvation en post-traitement de dynamique AMOEBA [63]. L'introduction d'une fluctuation à la fois des charges

et des dipôles a été présentée comme un moyen de rendre compte des effets à *n corps* à la fois de polarisation et de transfert de charge [57]. Le champ de forces hybride QM/MM COSMOS-NMR (Computer Simulation of Molecular Structure) [64] utilise des calculs de mécanique quantique pour calculer les fluctuations de charges à chaque pas, dans le formalisme de la théorie de la polarisation de liaison [65].

2.4 Le champ de forces polarisable AMOEBA

La suite logicielle TINKER [66], développée par Ponder, comprend un package de dynamique moléculaire, principalement destiné à l'étude des biomolécules. TINKER est compatible avec plusieurs jeux de paramètres usuels, tels qu'Amber (ff94, ff96, ff98, ff99, ff99SB), CHARMM (19, 22, 22/CMAP), Allinger MM (MM2-1991 et MM3-2000), OPLS (OPLS-UA, OPLS-AA), Merck Molecular Force Field (MMFF), le modèle polarisable de Liam Dang, mais surtout avec le champ de forces polarisable AMOEBA (2004, 2009, 2013). En 2015, les auteurs ont publié la nouvelle version, TINKER 7, qui pour la première fois utilise le formalisme Fortran95 et l'utilisation de modules à la place des commons (répertoires de partage des variables). En pratique, à part cette convention d'écriture, l'architecture a peu varié, si ce n'est dans l'utilisation de l'allocation dynamique de mémoire, qui permet par exemple de traiter des systèmes plus grands que précédemment. La flexibilité apportée par l'allocation dynamique est un avantage conséquent pour l'optimisation de nouvelles fonctionnalités dans le code. Le code prévoit également une part de parallélisation (OpenMP), mais pour l'instant uniquement pour les calculs d'interactions en conditions périodiques. Une parallélisation plus complète, basée sur MPI, ainsi que pour les systèmes basés sur GPU, est attendue dans une prochaine version.

Le champ de forces polarisable AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) est développé par Ponder et coll. [67,68]. Un de leurs objectifs est d'atteindre une précision chimique dans les propriétés thermodynamiques de protéines en solution, en s'appuyant sur une évaluation des énergies conformationnelles et d'interaction comparable aux modèles de mécanique quantique. Nous l'avons choisi pour ses capacités à reproduire diverses propriétés physico-chimiques en phase gazeuse ou condensée. Notamment, grâce à la transférabilité de ses paramètres, AMOEBA a su reproduire les énergies libres de solvatation de cations [69–72] tels que K^+ , Na^+ , Mg^{2+} , Ca^{2+} et Zn^{2+} . Sa totale flexibilité intramoléculaire, aliée à son traitement fin de l'électrostatique, est un atout pour le calcul de spectres vibrationnels. AMOEBA a été comparé à d'autres champs de forces (OPLSAA, AMBER, CHARMM) par Jensen et coll. [73,74] pour reproduire les énergies conformationnelles d'acides aminés calculées en chimie quantique. Il a été montré que les champs de forces qui utilisent uniquement des charges ponctuelles pour le calcul de l'énergie électrostatique ne peuvent reproduire que la moitié des conformations étudiées, alors qu'AMOEBa atteint un taux de 80 %, grâce à la distribution multipolaire et la prise en compte des effets de polarisation.

2.4.1 Formalisme

Les interactions liées Les termes entre atomes liés de manière covalente dérivent directement du potentiel MM3 d'Allinger [8,9] :

$$E_{liés} = E_b + E_\theta + E_{b\theta} + E_{oop} + E_{tors}(+E_{ub}) \quad (2.16)$$

où E_b est la composante d'élongation, E_θ celle de déformation angulaire, $E_{b\theta}$ un terme croisé couplant les deux précédents, E_{oop} se réfère au pliage hors du plan (*oop* pour 'out of plane'), E_{tors} à la torsion et E_{ub} la composante d'Urey-Bradley (interaction entre atomes séparés par deux liaisons). Cette dernière tend à disparaître des champs de forces modernes. Elle est implémentée dans TINKER mais, en pratique, le seul terme proposé dans AMOEBA est celui de la molécule d'eau.

Le potentiel d'élongation prend la forme d'un développement en série de Taylor à l'ordre 4 du potentiel de Morse :

$$E_b = K_b(b - b_0)^2[1 - cbnd(b - b_0) + qbnd(b - b_0)^2] \quad (2.17)$$

$$E_b = K_b(b - b_0)^2[1 - 2.55(b - b_0) + (7/12)2.55^2(b - b_0)^2] \quad (2.18)$$

avec K_b la constante d'élongation, $cbnd$ et $qbnd$ respectivement des facteurs de proportionnalité pour les termes cubique et quartique (indépendants du système étudié), b la longueur de liaison et b_0 sa valeur de référence.

Ce potentiel mime le comportement anharmonique d'un potentiel d'élongation avec pour seul paramètre de calibration la constante K_b . La forme analytique de ce potentiel se veut universelle, indépendante de la liaison décrite. D'après les auteurs de MM3, la constante d'ordre 3 suffit dans la grande majorité des cas à apporter l'effet d'anharmonicité, le terme d'ordre 4 ne se justifiant alors que comme un garde-fou, permettant au potentiel de tendre vers $+\infty$ à longue distance pour éviter des situations problématiques d'états dissociatifs lorsque la structure rencontrée est loin des distances standards (structure créée mais pas encore optimisée par exemple), ce qui posait problème dans MM2 et avait été géré de différentes façons suivant les versions.

La déformation angulaire est décrite par un potentiel d'ordre 6 :

$$E_\theta = K_\theta(\theta - \theta_0)^2[1 - 0.014(\theta - \theta_0) + 5.6 \times 10^{-5}(\theta - \theta_0)^2 - 7 \times 10^{-7}(\theta - \theta_0)^3 + 2.2 \times 10^{-8}(\theta - \theta_0)^4] \quad (2.19)$$

où θ est la valeur de l'angle et θ_0 sa valeur de référence.

Le couplage entre élongation et déformation angulaire prend la forme :

$$E_{b\theta} = K_{b\theta}[(b - b_0) + (b' - b'_0)](\theta - \theta_0) \quad (2.20)$$

avec $K_{b\theta}$ la constante de force, $(\theta - \theta_0)$ la déformation angulaire, et $(b - b_0)$ et $(b' - b'_0)$ les élongations des deux liaisons formant l'angle considéré.

Pour les atomes trivalents sp^2 , le terme de déformation hors du plan E_{oop} est contrôlé par une fonction de Wilson-Decius-Cross [75] :

$$E_{oop} = K_\chi \chi^2 \quad (2.21)$$

où K_χ est la constante de force et χ est l'angle dièdre formé par les 4 atomes considérés.

L'énergie de torsion est décrite par un développement en série de Fourier à l'ordre 6 d'une fonction périodique :

$$E_{tors} = \sum_{n=1}^6 K_{n\omega} [1 + \cos(n\omega \pm \gamma)] \quad (2.22)$$

En pratique, seuls les 3 premiers termes sont utilisés dans AMOEBA. L'importance des deux premiers termes pour reproduire les barrières rotationnelles des alcanes a été décrite [76, 77] et Allinger propose une procédure pour calculer le 3^{ème} terme [8]. A chacun est associée une phase $\gamma = 0$ ou 180° .

Les autres termes de couplage présents dans MM3 (élongation-pliage, torsion-élongation, torsion-pliage, pliage-pliage) n'ont pas été conservés dans AMOEBA.

Les interactions non liées AMOEBA se limite aux termes d'énergie électrostatique, de polarisation, de dispersion et de répulsion d'échange (les deux derniers au travers du terme de vdW) :

$$E_{non-liés} = E_{elec} + E_{pol} + E_{vdw} \quad (2.23)$$

Les phénomènes de transfert de charge ne sont pas explicitement pris en compte dans le champ de forces mais pourraient être inclus dans une prochaine version AMOEBA 2.0 [78].

Les termes de répulsion d'échange et de dispersion sont décrits dans un formalisme 'buffered 14-7' proposé par Thomas A. Halgren [79], à la place du LJ 12-6 standard. Le potentiel s'est montré plus efficace que le LJ pour reproduire les énergies d'interaction des gaz rares. En effet, par comparaison avec des calculs quantiques, Halgren a montré que la partie répulsive en r^{-12} du LJ conduit à un potentiel répulsif trop mou, bien amélioré par la forme en r^{-14} du buffered 14-7.

$$E_{vdw}(ij) = \epsilon_{ij} \left(\frac{1.07}{\rho_{ij} + 0.07} \right)^7 \left(\frac{1.12}{\rho_{ij}^7 + 0.12} - 2 \right) \quad (2.24)$$

Le potentiel est fonction de la distance de séparation R_{ij} entre les atomes i et j au travers du terme $\rho_{ij} = R_{ij}/R_{ij}^0$, avec R_{ij}^0 la distance pour laquelle l'énergie est minimale, distance qui devient $R_{ij}^0 = \frac{(R_{ii}^0)^3 + (R_{jj}^0)^3}{(R_{ii}^0)^2 + (R_{jj}^0)^2}$ lors d'une combinaison sur une paire d'atomes différents. Dans ce cas, la profondeur du puits est donnée par $\epsilon_{ij} = \frac{4\epsilon_{ii}\epsilon_{ij}}{(\sqrt{\epsilon_{ii}} + \sqrt{\epsilon_{ij}})^2}$.

L'énergie électrostatique est calculée avec des multipôles dérivés jusqu'aux quadripôles (eq 2.11).

La polarisation électronique décrit la redistribution de la densité électronique due à un champ extérieur :

$$\mu_{i,\alpha}^{ind} = \alpha_i E_{i,\alpha} (\alpha \in \{x, y, z\}) \quad (2.25)$$

α_i est la polarisabilité atomique et $E_{i,\beta}$ la composante selon l'axe β du champ électrique sur l'atome i . Le dipôle induit $\mu_{i,\beta}^{ind}$ sur le site i est :

$$\mu_{i,\beta}^{ind} = \alpha_i \left(\sum_j T_{\beta}^{ij} M_j + \sum_{j'} T_{\beta\gamma}^{ij'} \mu_{j',\gamma}^{ind} \right) \quad \text{avec } \beta, \gamma = 1, 2, 3 \quad (2.26)$$

T est la matrice d'interaction et M_j contient les composantes des moments multipolaires permanents. Le premier terme entre parenthèses correspond au champ de polarisation dû aux moments permanents, et le second à la polarisation produite par les dipôles induits des autres atomes. Le terme de polarisation doit donc être calculé de manière auto-cohérente, jusqu'à convergence de tous les dipôles induits. L'approche itérative d'induction développée

initialement par Tholé est adoptée afin d'éviter la divergence de l'énergie de polarisation. Dans ce cadre, la fonction d'amortissement prend la forme :

$$\rho = \frac{3a}{4\pi} \exp(-au^3) \quad (2.27)$$

avec $u = r_{ij}/(\alpha_i\alpha_j)^{1/6}$, r_{ij} étant la distance réelle entre les sites i et j , α_i et α_j leurs polarisabilités atomiques, et a un facteur sans dimension contrôlant la force de l'amortissement.

Lorsque deux atomes sont suffisamment proches, leurs densités de charges peuvent se recouvrir, avec pour conséquence que l'écrantage de la charge d'un noyau par sa propre densité électronique diminue. Cette interaction, appelée 'pénétration électrostatique' ou 'pénétration de charge', est attractive à très courte portée. Différentes stratégies ont été proposées pour tenir compte de cet effet purement quantique [80]. Le modèle de champ de forces GEM ('Gaussian Electrostatic Model') [81, 82] modélise la densité électronique à partir de multipôles distribués sur des fonctions gaussiennes. Le modèle permet de rendre compte de la polarisation et du transfert de charge, et de calculer des effets à courte distance comme la répulsion d'échange. Le modèle X-Pol (the Explicit Polarization Theory), introduit par Gao [83, 84], est basé sur la fragmentation du système en sous-unités moléculaires (molécules, ions, ligands, résidus amino-acides ...). La fonction d'onde totale est décrite comme un produit de Hartree des fonctions d'onde des fragments, et son optimisation est limitée à l'intérieur de chaque fragment, polarisé par le reste du système. S'il n'y a pas explicitement de transfert de charges entre les fragments, un transfert effectif peut être observé par la polarisation de l'atome servant de liaison entre deux fragments. La méthode a pu être employée dans des simulations de MD sur des protéines solvatées dans des boîtes d'eau en conditions périodiques (15000 atomes) [85]. Le modèle polarisable SIBFA (Sum of Interactions Between Fragments Ab Initio Computed), développé par Gresh et coll. [50] a été doté de termes additionnels mimant le recouvrement dans le développement multipolaire [51], et dont le formalisme pourrait être étendu à AMOEBA [78, 86].

Notions de types et classes Dans le formalisme d'AMOEBA, chaque atome est défini par un type et une classe. Les interactions liées (élongation, déformation angulaire, torsion ...) et de van der Waals utilisent des paramètres définis en fonction des classes des atomes. Celles-ci sont assez génériques. Les jeux de paramètres usuels AMOEBA_{bio09} ou AMOEBA_{pro13} en proposent 42 pour décrire tous les C, N, O, H, S des composés organiques. Une distinction plus fine est réalisée au niveau des interactions électrostatiques (via la définition des multipôles et de la polarisation), dont on pressent qu'elles sont plus sensibles à l'environnement chimique. Ces interactions sont définies par les types d'atomes. Avec notre expérience des paramètres AMOEBA, il est devenu routinier, avant d'entreprendre une étude sur un nouveau système, d'extraire les multipôles associés (et donc de générer de nouveaux types d'atomes). La modification de paramètres d'une classe est moins fréquente mais répond à certains besoins : ajustement de paramètres de torsion sur une surface *ab initio* [87], ou des constantes intramoléculaires de l'eau afin d'en corriger les fréquences d'élongation OH [88].

2.4.2 Extraction des multipôles atomiques

Définition des multipôles atomiques Dans AMOEBA, les multipôles sont localisés sur les centres atomiques, et définis dans des repères locaux pour chaque site (atome ou paire libre), construits à partir des atomes voisins (**Fig 2.1**). Pendant que la molécule vibre, tourne, ou diffuse au cours de la simulation, les multipôles atomiques demeurent constants dans le cadre local défini.

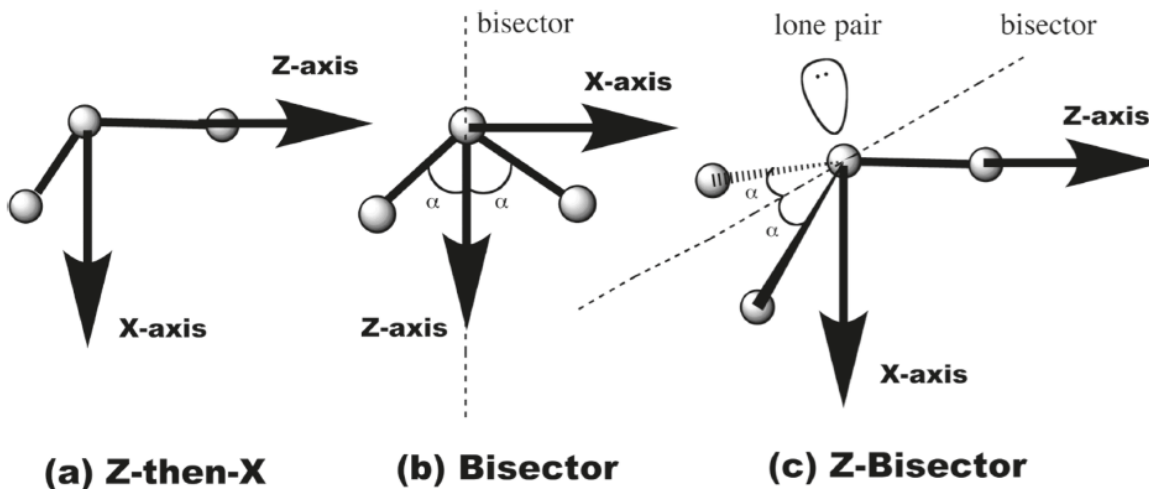


FIGURE 2.1 – 3 définitions du repère local pour les sites de multipôles atomiques : (a) La convention *Z-then-X* est utilisée pour la majorité des sites. En ajoutant l’axe orthogonal Y, on peut traiter les centres chiraux. (b) La convention *Bisector* est utilisée lorsque l’axe Z peut être défini par la bissectrice de l’angle formé avec les deux voisins. Typiquement utilisé dans le cas de (pseudo)symétrie c_{2v} , comme l’eau ou un carboxylate. (c) La convention *Z-then-bisector* est utilisée par exemple pour les amines NH_2 . L’axe Z est alors défini par la liaison N-R, puis la bissectrice entre les deux liaisons N-H définit le plan ZX. Dans chaque cas, l’axe Y est obtenu par construction du trièdre direct $(\vec{i}; \vec{j}; \vec{k})$. Figure 1 issue de [89]

Analyse de la distribution des multipôles Dans AMOEBA, les multipôles sont extraits par la méthode ‘Distributed multipôles Analysis’ (DMA) [90], implémentée dans le logiciel GDMA [91] de Stone. L’analyse consiste en une répartition de la densité de charge sur les sites atomiques, afin de déterminer les moments dipolaires. La procédure est simple, rapide et *a priori* ‘exacte’ dans le sens où les moments multipolaires distribués sur les atomes reproduisent effectivement la distribution de charge. Cependant, lorsque la procédure fait appel à des bases d’orbitales de grandes tailles, et notamment à des fonctions diffuses, la distribution des multipôles conduit fréquemment à des contributions non physiques, qui peuvent conduire à de vrais problèmes, lors d’interactions à courte distance par exemple. Dans la seconde version de GDMA, pendant que les fonctions compactes continuent d’être traitées de façon exacte via un algorithme de Gauss-Hermite, les contributions des fonctions diffuses sont intégrées à la densité électronique par une quadrature basée sur une grille (algorithme de Lebedev). On notera que la troncature de DMA au niveau des quadripôles, et l’absence d’information sur le potentiel électrostatique dans la procédure, peuvent conduire à des erreurs significatives, parfois jusqu’à 10 % par rapport au potentiel électrostatique (ESP) [92]. De meilleurs résultats peuvent être attendus par un ajustement explicite des moments dipolaires atomiques sur l’ESP [93], donc toutes les extractions de multipôles présentées dans ce manuscrit sont réalisées avec la première version de GDMA.

Pour une molécule donnée, si les multipôles M_i^{DMA} sont extraits d’un conformère arbitraire présentant certaines interactions spécifiques, on voudrait extraire un jeu de multipôles M_i indépendants de la conformation :

$$M_i^{DMA} = M_i + \mu_i \quad (2.28)$$

où μ_i est le dipôle induit par la polarisation intramoléculaire de M_i . La contribution de la polarisation doit être retirée des multipôles M_i^{DMA} afin d'éviter un double comptage lorsque le modèle de polarisation du champ de forces est appliqué. Cette correction a déjà été appliquée par des procédures empiriques [94, 95]. Ren et Ponder proposent une méthode basée sur un fit du potentiel électrostatique qui doit garantir l'obtention de multipôles 'permanents' [67]. Les multipôles extraits dans DMA sont optimisés sur une surface de potentiel électrostatique *ab initio*, MP2/aug-cc-pVTZ lorsque la taille du système le permet. L'inclusion de fonctions diffuses est important pour bien décrire les interactions intermoléculaires, ou reproduire les énergies libres d'hydratation [96]. En général, les charges partielles sur les atomes sont gardées fixes durant la phase d'optimisation des dipôles et quadrupôles. La méthode n'est cependant pas sans biais et Wu [97] voit dans la procédure d'optimisation sur l'ESP une source probable d'erreurs de l'électrostatique d'AMOEBa. Sur le n-octane, il observe de larges déviations des quadrupôles par rapport à ceux obtenus par DMA et propose de limiter le critère de convergence de l'ajustement sur l'ESP à $0.5 \text{ kcal.mol}^{-1}.\text{electron}^{-2}$ au lieu de 0.1, ou à défaut d'éviter les zones avec de fortes interactions de vdW. De façon assez contre-intuitive, cela revient à proposer que, si on s'intéresse à reproduire le comportement d'un système avec une forte liaison hydrogène, l'extraction de multipôles devrait être faite sur une géométrie tout autre, plutôt dans l'esprit d'éviter les situations délicates, que de guider une bonne reproduction de l'interaction.

Au cours de nos recherches, nous avons pu observer l'influence de la géométrie d'extraction des multipôles sur la capacité d'AMOEBa à reproduire des ordres énergétiques de conformères, ou les spectres IR associés. La définition d'un jeu de multipôles n'est souvent qu'une étape dans le processus de paramétrisation d'un modèle, et il peut être délicat de juger de la responsabilité des multipôles dans la réussite ou l'échec du modèle à reproduire la propriété visée. Cependant, l'exemple suivant donne quelques pistes sur les difficultés rencontrées et l'importance du choix du niveau théorique, ou de la géométrie d'extraction, dans cette étape parfois délicate.

Energies relatives et spectroscopie des conformères du dipeptide Ace-Phe-Ala-NH₂ :

Cette validation d'un jeu de multipôles est un avant-propos à l'étude que nous avons réalisée sur le dipeptide protégé Ace-Phe-Ala-NH₂ [98], présentée en **chapitre 3** pour ses aspects méthodologiques sur l'analyse dynamique de modes normaux de vibration. Dans une première étape, il s'agissait d'explorer l'espace conformationnel du dipeptide et d'en déterminer les structures de basse énergie. Cette étape a été réalisée par couplage entre la méthode *Replica Exchange Molecular Dynamics* (REMD) [99], qui sera décrite également au **chapitre 3**, et le champ de forces AMOEBA. Notons simplement qu'à l'issue de cette étape, un nombre important de conformations avaient été engendrées et réoptimisées avec AMOEBA. 10 structures de basse énergie (notées Fa_a à Fa_j dans la **Fig 2.2**) ont ensuite été sélectionnées pour optimisation par des méthodes de chimie quantique. Trois fonctionnelles ont été testées pour l'optimisation de géométrie, M06, B97-D et B3LYP, associées à la base 6-311+G(d,p). Aux géométries M06, des calculs d'énergie ont été réalisés avec la même fonctionnelle et la base de Dunning cc-pVTZ. Les géométries obtenues avec les 3 fonctionnelles sont très similaires, de même que l'ordre de stabilité des conformations (**Tab2.1**). La grande flexibilité du peptide permet une variété d'interactions intramoléculaires non covalentes dont la nomenclature est également donnée en **Fig 2.2**. Ces interactions correspondent souvent à des liaisons hydrogènes entre une liaison N-H d'amine ou d'amide et un oxygène voisin dans un motif C₅ ou C₇. De plus, des interactions de type C₈ ou C₁₀ sont possibles entre groupes plus distants. Deux sortes d'interactions C₇B existent. Dans FA_d, l'interaction intervient dans un cycle régulier C₇ tandis que le cycle est tordu dans Fa_f. Il a été montré précédemment dans différentes structures avec des distances C₇ différentes que la force de la liaison hydrogène n'était pas le seul facteur

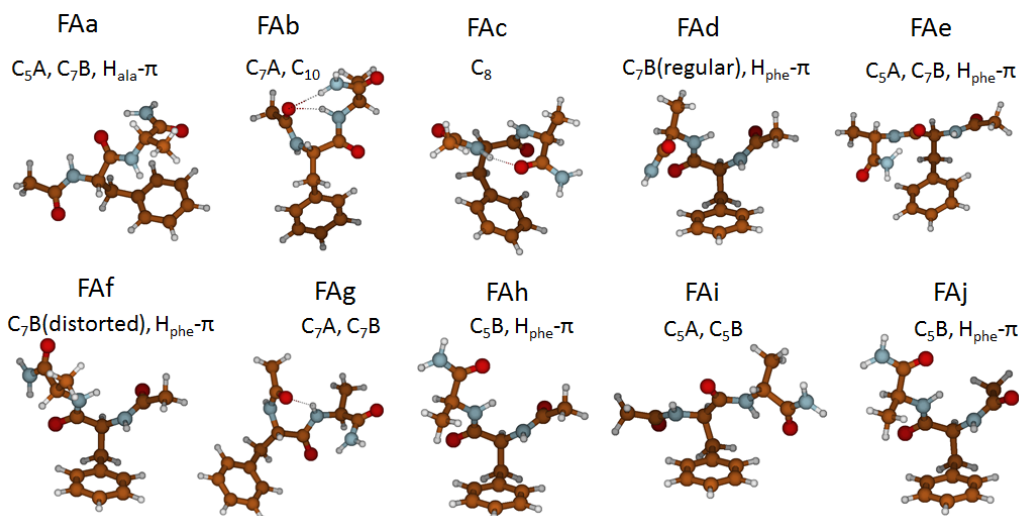


FIGURE 2.2 – Structures de basse énergie de FA au niveau $B3LYP/6-311+G(d,p)$, et leurs principales interactions non covalentes

déterminant de la stabilité de la structure [100]. Un autre effet stabilisant rencontré est l'interaction π qui peut exister entre un hydrogène de l'amide et le cycle aromatique. Pour tous les niveaux de théorie utilisés, FA_a est la structure la plus favorable énergétiquement, stabilisée par des interactions C_5 , C_7B et $H_{ala}-\pi$. Ce résultat est en accord avec les structures obtenues par Mons et al. [101].

La validation d'un jeu de multipôles est une tâche qui peut être longue, en fonction de la propriété ciblée. Pour le dipeptide, la procédure a été basée sur une approche itérative. Un jeu de multipôles préexistants avait été engendré à partir des deux fragments, l'alanine protégée $Ace-Ala-NH_2$ et la phenylalanine protégée $Ace-Phe-NH_2$, à partir de la densité électronique calculée au niveau $MP2/cc-pVTZ$. En dépit du bon niveau théorique proposé, des doutes ont été émis sur sa fiabilité, dûs à l'absence des interactions spécifiques du dipeptide dans les fragments. Des structures présentant ces interactions spécifiques ont donc été engendrées par REMD. A partir de la structure la plus stable, d'autres jeux de multipôles ont été extraits sur le dipeptide total. Cependant, comme il est impossible de maintenir le même niveau de théorie pour extraire les multipôles du dipeptide dans son ensemble, deux niveaux de théorie ont été choisis : $MP2/cc-pVDZ$ et $M06/6-311++G(d,p)$. La fonctionnelle M06 a été choisie pour ses performances dans la reproduction des interactions non covalentes [102]. Plusieurs cycles ont été nécessaires (extraction de multipôles au niveau M06, suivie d'une exploration REMD), pour converger vers la détermination de FA_a comme la structure la plus stable en AMOEBA (i.e que la structure est la plus stable engendrée pour deux générations REMD successives, avec des jeux de multipôles différents). Pour cette structure uniquement, une extraction de multipôles a été réalisée en utilisant la fonction d'onde MP2. De plus, dans un cycle précédent, une extraction de multipôles avait été réalisée au niveau M06 sur la structure FA_f (alors la plus stable proposée par la REMD).

Deux points méritent d'être discutés. Le premier concerne l'influence très importante du jeu de multipôles sur l'ordre énergétique des structures. Sur ce point, la procédure REMD peut être mise hors de cause, la plupart des structures finalement sélectionnées pour l'étude avaient été engendrées dès la première itération, mais n'avaient simplement pas passé la sélection sur le critère énergétique AMOEBA. Les fortes variations de l'ordre de stabilité des structures s'expliquent par la force relative des interactions intramoléculaires non covalentes dérivant des multipôles. Le second point concerne la sensibilité des spectres DACF aux jeux de multipôles ; à leur niveau de théorie, mais plus encore à la conformation d'extraction. Certains massifs sont plus touchés que d'autres, tels que la gamme des elongations N-H ($3300-3600\text{ cm}^{-1}$), les modes d'elongation C=O ($1600-1800\text{ cm}^{-1}$) ou les bandes amides ($1300-$

TABLE 2.1 – Energies relatives en kcal.mol⁻¹ pour les 10 conformations de basse énergie de FA, en utilisant différentes fonctionnelles et bases. Tableau publié dans le *Supporting Information* de [98]

	^a B3LYP	^a B97-D	^a M06	^b M06
FAa	0.00	0.00	0.00	0.00
FAb	1.02	2.11	2.67	2.74
FAc	1.26	1.32	5.38	5.47
FAd	1.27	1.54	2.79	2.67
FAe	1.91	1.54	1.65	1.48
FAf	1.95	1.75	3.35	3.32
FAg	2.15	3.61	2.97	2.68
FAh	2.36	4.00	4.03	3.84
FAi	5.68	5.59	3.31	2.93
FAj	9.84	2.35	6.70	6.60

^a : 6-311+G(d,p), ^b : M06/cc-pVTZ//M06/6-311+G(d,p)

1500 cm⁻¹). Pour ces modes, les déplacements de fréquences apparaissent du fait de leur réponse fine aux liaisons hydrogènes intramoléculaires.

Par exemple, dans la gamme des élongations N-H (**Fig 2.3, partie de droite**), 3 bandes sont présentes dans le spectre calculé avec les multipôles obtenus des fragments, et seulement 2 avec les multipôles extraits sur le dipeptide complet. De plus, le déplacement vers le rouge de cette troisième bande peut être attribué à l'extraction à partir des fragments séparés, qui ne tient pas compte des liaisons hydrogènes entre les deux résidus. Des effets similaires peuvent être observés sur les modes d'élongations C=O. L'effet dérivant du choix de la conformation du peptide, FA_a ou FA_f, pour l'extraction des multipôles au même niveau de théorie (courbes rouge et bleue), est de nouveau surtout marqué dans la gamme des élongations N-H, dû à la mémoire des liaisons hydrogènes à la conformation de référence. Lorsqu'on s'intéresse au niveau de théorie (DFT/M06 vs MP2) pour une même géométrie d'extraction, les différences sont beaucoup moins marquées. Sur la gamme des élongations N-H, l'écart moyen entre fréquences expérimentales [101] et calculées (T = 50 K) est inférieur à 6 cm⁻¹ en utilisant chacun des deux jeux de multipôles. Au niveau des intensités, le massif des élongations C-H est affecté par l'utilisation des multipôles extraits des fragments du fait du manque de l'interaction NH- π dans la structure de référence, qui peu induire une plus faible polarisation des carbones et hydrogènes du cycle aromatique. Les intensités de la gamme des élongations N-H sont perturbées pour des raisons similaires.

Cet exemple semble indiquer qu'inclure les interactions non covalentes dans l'extraction des multipôles, si possible couplées à un haut niveau de chimie théorique, est capital. La question de l'extension aux plus grands peptides voire aux protéines est posée. Dans l'idéal, il serait souhaitable que les briques élémentaires de la protéine soient les dipeptides qui présentent les interactions qui seront retrouvées dans la protéine, bien qu'il semble difficile de toujours avoir cette possibilité. Dans le cas de FA, à la fois les multipôles M06 et MP2 peuvent être utilisés pour réaliser des spectres DACF. Pour être en accord avec le cadre général d'AMOEBA, le reste de l'étude (qui avait pour sujet l'attribution dynamique de modes normaux) utilisait les multipôles dérivés au niveau MP2/cc-pVDZ. A partir de la procédure DMA, un jeu de multipôles a été assigné sur chaque site atomique et les multipôles ont été moyennés sur les atomes équivalents, exception faite des NH_{ala} et NH_{phe}, afin de reproduire finement la différence d'environnement entre les structures FA_a et FA_f.

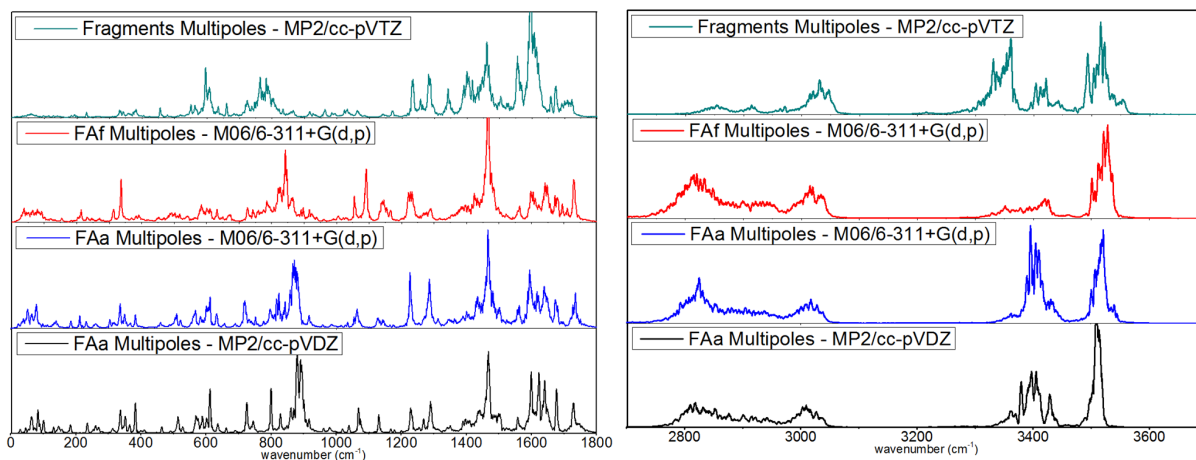


FIGURE 2.3 – Spectres d’absorption DACF-AMOEBAs de la conformation FA_a , avec 4 jeux de multipôles, dans deux gammes spectrales. De haut en bas : multipôles extraits des fragments séparés alanine protégée et phénylalanine protégée au niveau MP2/cc-pVTZ (vert), multipôles extraits de la structure FA_f au niveau M06/6-311++G(d,p) (rouge), multipôles extraits de la structure FA_a au niveau M06/6-311++G(d,p) (bleu) et multipôles extraits de FA_a au niveau MP2/cc-pVDZ (noir). Les spectres sont enregistrés pendant 300 ps à 200 K.

2.4.3 Paramètres du champ de forces de l’eau

En 2003, Ren et Ponder ont publié la version 03 de leur champ de forces polarisable AMOEBA [67, 68]. Dans l’article de 2003 sont notamment proposés des paramètres pour l’eau et des benchmarks sur les capacités du champ de forces à reproduire certaines propriétés en phase gazeuse ou liquide pour l’eau pure. Pendant 10 ans, de nombreux articles sont venus en compléments, avec notamment des paramètres additionnels pour toute une série d’ions [69, 70, 103, 104]. Ces dernières années, les auteurs ont cherché à automatiser les procédures d’optimisation de paramètres [105], notamment via la procédure ForceBalance [106] (Fig 2.4), qui vise à fitter un jeu de paramètres sur un large ensemble de données de références (calculs *ab-initio* et données expérimentales). Cette méthode a été appliquée au modèle de l’eau dans le champ de forces AMOEBA et est transférable à d’autres situations.

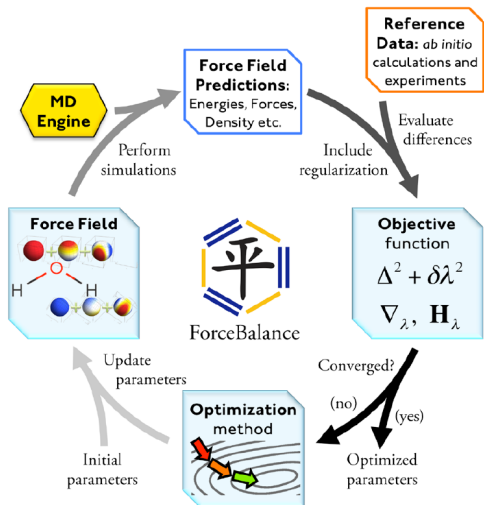


FIGURE 2.4 – Procédure ForceBalance : une méthode d’optimisation des paramètres pour le champ de forces. Un jeu de paramètres initiaux définit un champ de forces et permet de lancer des simulations. Une fonction *objectif* est décrite comme une somme des écarts aux moindres carrés entre les données calculées pendant la simulation et un jeu de référence. Le jeu de paramètres est amélioré jusqu’à convergence de cette fonction. Figure 1 de l’article [106].

Dans le même article, les auteurs présentent le champ de forces iAMOEBAs (‘inexpensive AMOEBA’), qui, outre cette méthode de paramétrisation, réalise une approximation sur le calcul de la polarisation. Dans AMOEBA, le moment dipolaire induit résulte de deux contributions, l’interaction avec les multipôles permanents des autres sites (ce qu’on nomme la polarisation directe) et les interactions avec les moments dipolaires induits des autres sites. Ce

deuxième terme demande une procédure itérative pour converger et représente un coût calculatoire significatif (1^{ère} source de temps de calcul). Dans iAMOEBA, seuls les termes de polarisation directe sont calculés. Il est très difficile de savoir quelle part de la précision est perdue avec cette approximation, puisque les deux modifications (le calcul simplifié de la polarisation et la procédure ForceBalance) sont présentées simultanément. Un indice est apporté par le **Fig 2.5** sur l'exemple du calcul de la densité de l'eau. iAMOEBA (courbe vert clair) semble très intéressant partout où le champ de forces sera fortement guidé par des données expérimentales ou *ab-initio* suffisantes. Si on observe ce que devient AMOEBA (courbe bleue) lorsqu'on lui retire la contribution mutuelle de la polarisation (courbe vert foncé), on peut craindre pour le comportement d'iAMOEBA sur des propriétés non calibrées. Si l'intérêt d'iAMOEBA semble pour l'instant restreint à l'étude de systèmes pour lesquels une approche complète du calcul de la polarisation n'est pas envisageable, la procédure ForceBalance semble apporter de sérieuses facilités au processus de paramétrisation.

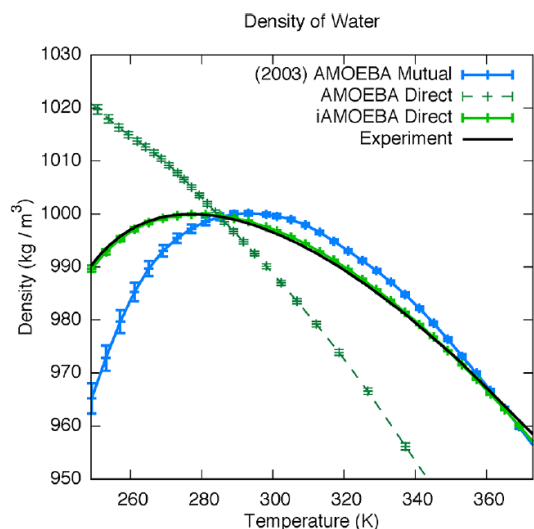


FIGURE 2.5 – Evolution de la densité de l'eau en fonction de la température, avec différents modèles. Les paramètres dans iAMOEBA sont optimisés pour reproduire spécifiquement cette donnée, incluse dans la fonction objectif de la procédure ForceBalance. Pour AMOEBA, on voit l'effet catastrophique de lui retirer la deuxième contribution à la polarisation (différence entre AMOEBA mutual et AMOEBA direct). Figure 2 de l'article [106]

Basé sur la procédure ForceBalance, un nouveau jeu de paramètres pour l'eau a été publié [88] sous le nom d'AMOEBA14 (**Tab 4.1**). La calibration repose sur des données expérimentales en phase condensée (densité, enthalpie de vaporisation, compressibilité isotherme, capacité calorifique isobare, coefficient d'expansion thermique et constante diélectrique) et sur des propriétés dérivées de calculs *ab-initio* (informations de structures, énergies et gradients) sur des agrégats d'eau $(\text{H}_2\text{O})_n$, $n=1-22$. On notera que le jeu de données n'inclut aucune donnée spectroscopique. C'était déjà le cas pour AMOEBA03, pour lequel les fréquences des élongations OH de l'eau étaient inversées. Des corrections des termes intramoléculaires de liaison K_b , d'angle K_θ , et d'Urey-Bradley K_{ub} et r_{ub}^{H-H} , avaient été proposées par notre équipe [107] pour calibrer les fréquences de la molécule isolée pour un calcul statique ou dynamique. Ces corrections ont été incluses dans la procédure d'optimisation de la version AMOEBA14, et restent quasiment nos valeurs proposées à l'issue de la procédure ForceBalance, même si on peut regretter que le jeu utilisé soit celui préconisé pour les calculs des fréquences statiques.

La différence majeure entre les deux versions concerne donc les interactions non liées. Pour les paramètres de vdW, une augmentation du diamètre et du puits de l'oxygène, accompagnée d'une diminution pour l'hydrogène, rapproche le modèle d'une situation à un seul site de répulsion-dispersion, centré sur l'oxygène. Les polarisabilités atomiques ont augmenté de 8 et 10 % entre les versions, (ce que Ponder qualifie de déviation non significative [88]), et les facteurs d'amortissement proposés par Tholé ont été conservés. Les plus grandes variations sont observées pour l'électrostatique. Les charges de l'oxygène et l'hydrogène diminuent de 18 % en valeur absolue, les dipôles varient tout autant, avec -57 % sur la composante z sur l'oxygène, alors que les composantes x et y du dipôle sur l'hydrogène augmentent respectivement de 162 et 367 %. Avec ces nouveaux multipôles, Ponder note que le moment dipolaire du dimère de l'eau à la géométrie d'équilibre est moins bien reproduit qu'avec AMOEBA03, mais meilleur

TABLE 2.2 – Comparatif des paramètres pour l’eau dans AMOEBA03 et AMOEBA14. La colonne *prior width* indique l’importance de chaque paramètre dans les fonctions “objectifs” de la procédure ForceBalance. Valeurs publiées dans [88].

parameter	units	AMOEBA03	AMOEBA14	prior width
O monopole	e	-0.51966	-0.42616	0.4
O dipole Z	e bohr	0.14279	0.06251	0.1
O quadrupole XX	e bohr ²	0.37928	0.17576	0.2
O quadrupole YY	e bohr ²	-0.41809	-0.23160	0.2
O quadrupole ZZ	e bohr ²	0.03881	0.05584	0.2
H monopole	e	0.25983	0.21308	0.4
H dipole X	e bohr	-0.03859	-0.10117	0.1
H dipole Z	e bohr	-0.05818	-0.27171	0.1
H quadrupole XX	e bohr ²	-0.03673	0.12283	0.2
H quadrupole YY	e bohr ²	-0.10739	0.08950	0.2
H quadrupole ZZ	e bohr ²	0.14412	-0.21233	0.2
O polarisability	Å	0.837	0.920	0.1
H polarisability	Å	0.496	0.539	0.1
damping factor	Å	0.39	0.39	none
O vdw diameter	Å	3.405	3.5791	0.3
O vdw epsilon	kcal/mol	0.11	0.1512	0.1
H vdw diameter	Å	2.655	2.1176	0.3
H vdw epsilon	kcal/mol	0.0135	0.0105	0.1
H vdw reduction factor	none	0.91	0.8028	0.1
O-H bond length	Å	0.9572	0.9565	0.1
bond force constant	kcal/mol/Å ²	556.85*	556.82	50
H-O-H angle value	degree	108.5	107.91	5
angle force constant	kcal/mol/rad ²	48.7*	48.98	40
H-H Urey-Bradley length	Å	1.5326*	1.5467	none
Urey-Bradley force const.	kcal/mol/Å ²	-7.6*	-8.62	25

* valeurs corrigées, issues de [107]

pour une série de géométries [88]. Enfin, d’importants changements sur les quadrupôles atomiques conduisent à une amélioration de la composante xx du moment quadrupolaire moléculaire (les composantes yy et zz restant à peu près inchangées).

AMOEBA14 ayant été publié 1 an après le début de cette thèse, certains résultats avaient donc été obtenus avec la version ‘historique’ AMOEBA03. De plus, à ce jour, seuls les paramètres de l’eau ont été fournis, sans ajustement des paramètres pour les ions (dont la calibration est nécessairement liée à un champ de forces de l’eau). L’utilisation de cette nouvelle version a nécessité une longue série de tests, sur ses capacités à reproduire la dynamique des ions hydratés, ou leur spectroscopie.

Dans le **chapitre 4** traitant de la dynamique des ions hydratés, nous présenterons essentiellement des calculs utilisant la version AMOEBA03, puis quelques indications des modifications engendrées par l’utilisation de AMOEBA14. Dans le **chapitre 5** traitant la spectroscopie des ions solvatés dans des agrégats de grande taille ($n_{H_2O}=36-100$), les premiers résultats utilisent également la version 03, dont nous jugeons les capacités et limites. Nous y incluons une proposition de modification du modèle de potentiel d’élongation OH, qui permet à bas coût d’améliorer les spectres obtenus. La nouvelle version du champ de forces AMOEBA14 apporte de vraies améliorations du modèle en terme de spectroscopie, et a permis de discuter des systèmes très variés, en taille ou en nature de l’ion, et l’essentiel de nos résultats s’appuient sur cette version. Plusieurs problèmes persistent néanmoins, et nous proposons plusieurs modifications de paramètres, ainsi qu’un modèle de charges fluctuantes, qui améliorent assez significativement les spectres calculés. A la fin du **chapitre 5**, le champ de forces AMOEBA14 est validé et est utilisé dans les études spectroscopiques du **chapitre 6**.

Bibliographie

- [1] Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press : **2008**.
- [2] Hill, T. L. *J. Chem. Phys.* **1946**, *14*, 465-465.
- [3] Westheimer, F. H.; Mayer, J. E. *J. Chem. Phys.* **1946**, *14*, 733-738.
- [4] Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208-1209.
- [5] Halgren, T. A. *Curr. Opin. Struct. Biol.* **1995**, *5*, 205 - 210.
- [6] Frierson, M. R.; Imam, M. R.; Zalkow, V. B.; Allinger, N. L. *J. Org. Chem.* **1988**, *53*, 5248-5258.
- [7] Allinger, N. L.; Zhou, X.; Bergsma, J. *J. Mol. Struct. : THEOCHEM* **1994**, *312*, 69 - 83.
- [8] Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551-8566.
- [9] Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8566-8575.
- [10] Langley, C. H.; Allinger, N. L. *J. Phys. Chem. A* **2003**, *107*, 5208-5216.
- [11] Lii, J. H.; Allinger, N. L. *J. Phys. Chem. A* **2008**, *112*, 11903-11913.
- [12] Dehez, F.; Angyan, J.; Gutiérrez, I. S.; Luque, F. J.; Schulten, K.; Chipot, C. *J. Chem. Theory Comput.* **2007**, *3*, 1914-1926.
- [13] Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.
- [14] Allinger, N. L.; Kok, R. A.; Imam, M. R. *J. Comput. Chem.* **1988**, *9*, 591-595.
- [15] Gundertofte, K.; Liljefors, T.; Norrby, P. O.; Pettersson, I. *J. Comput. Chem.* **1996**, *17*, 429-449.
- [16] Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *PLoS ONE* **2012**, *7*, 1-6.
- [17] Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. In *Protein Simulations*, Vol. 66; Academic Press : **2003**.
- [18] Paton, R. S.; Goodman, J. M. *J. Chem. Inf. Model.* **2009**, *49*, 944-955.
- [19] 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*, 5179-5197.
- [20] P. Kollman, R. Dixon, W. C. T. F. C. C. A. P. *The development/application of a minimalist organic/biochemical molecular mechanic force field using a combination of ab Initio calculations and experimental data*; volume 3 Springer : **1997**.
- [21] Ross, W. S.; Hardin, C. C. *J. Am. Chem. Soc.* **1994**, *116*, 6070-6080.
- [22] Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845 - 862.
- [23] Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049-1074.
- [24] A. D. MacKerell, J. *et al.* *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

- [25] Foloppe, N.; MacKerell, Jr., A. D. *J. Comput. Chem.* **2000**, *21*, 86–104.
- [26] Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennel, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
- [27] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- [28] Abascal, J. L. F.; Vega, C. *J. Chem. Phys.* **2005**, *123*, 234505.
- [29] Mark, P.; Nilsson, L. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.
- [30] Jackson, J. D. *Classical electrodynamics*; Wiley, NY : **1996**, Second edition.
- [31] Gilson, M. K.; Honig, B. *Proteins : Struct., Funct., Bioinf.* **1988**, *4*, 7–18.
- [32] Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297–1304.
- [33] J. R. Pliego, J.; Riveros, J. M. *J. Phys. Chem. A* **2001**, *105*, 7241–7247.
- [34] Klimovich, P.V., M. D. *J. Comput-Aided. Mol. Des.* **2010**, *24*, 307–316.
- [35] Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. *Proc. Natl. Acad. Sci.* **2008**, *105*, 6290–6295.
- [36] Jiao, D.; Zhang, J.; Duke, R. E.; Li, G.; Schnieders, M. J.; Ren, P. *J. Comput. Chem.* **2009**, *30*, 1701–1711.
- [37] Masella, M.; Flament, J.-P. *J. Chem. Phys.* **1999**, *111*, 5081–5090.
- [38] Burnham, C. J.; Li, J.; Xantheas, S. S.; Leslie, M. *J. Chem. Phys.* **1999**, *110*, 4566–4581.
- [39] Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. *PLoS Comput Biol* **2010**, *6*, 1–17.
- [40] Harder, E.; Kim, B.; Friesner, R. A.; Berne, B. J. *J. Chem. Theory Comput.* **2005**, *1*, 169–180.
- [41] Archambault, F.; Chipot, C.; S., I.; Luque, F. J.; Schulten, K.; Dehez, F. *J. Chem. Theory Comput.* **2009**, *5*, 3022–3031.
- [42] Xie, W.; Pu, J.; A. D. MacKerell, J.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1878–1889.
- [43] Claverie, P.; Rein, R. *Int. J. Quant. Chem.* **1969**, *3*, 537–551.
- [44] Vigné-Maeder, F.; Claverie, P. *J. Chem. Phys.* **1988**, *88*, 4934–4948.
- [45] Silverstein, L. *Philos. Mag. Series 6* **1917**, *33*, 215 - 521.
- [46] Applequist, J.; Carl, J. R.; Fung, K.-K. *J. Am. Chem. Soc.* **1972**, *94*, 2952–2960.
- [47] Thole, B. *Chem. Phys.* **1981**, *59*, 341 - 350.
- [48] van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399–2407.
- [49] Hermida-Ramón, J. M.; Brdarski, S.; Karlström, G.; Berg, U. *J. Comput. Chem.* **2003**, *24*, 161–176.
- [50] Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *J. Chem. Theory Comput.* **2007**, *3*, 1960–1986.
- [51] Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. *J. Phys. Chem. A* **2003**, *107*, 10353–10359.
- [52] Rick, S. W.; Stuart, S. J. Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. In *Reviews in Computational Chemistry*; John Wiley & Sons, Inc. : **2003**.
- [53] Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; A. D. MacKerell, J. *J. Chem. Theory Comput.* **2005**, *1*, 153–168.
- [54] Vorobyov, I. V.; Anisimov, V. M.; Jr., A. D. M. *J. Phys. Chem. B* **2005**, *109*, 18988–18999.
- [55] Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141–6156.
- [56] Chelli, R.; Procacci, P.; Righini, R.; Califano, S. *J. Chem. Phys.* **1999**, *111*, 8569–8575.
- [57] Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, *103*, 4730–4737.

- [58] Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741-754.
- [59] Yang, Z.-Z.; Wang, C.-S. *J. Phys. Chem. A* **1997**, *101*, 6315-6321.
- [60] Patel, S.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1 - 15.
- [61] Patel, S.; Mackerell, A.D., J.; Brooks, C. *J. Comput. Chem.* **2004**, *25*, 1504 - 1514.
- [62] Bauer, B. A.; Warren, G. L.; Patel, S. *J. Chem. Theory Comput.* **2009**, *5*, 359-373.
- [63] Clavaguéra, C.; Calvo, F.; Dognon, J.-P. *J. Chem. Phys.* **2006**, *124*, 074505.
- [64] Möllhoff, M.; Sternberg, U. *Mol. Mod. Annu.* **2001**, *7*, 90-102.
- [65] Sternberg, U.; Koch, F. T.; Möllhoff, M. *J. Comput. Chem.* **1994**, *15*, 524-531.
- [66] Ponder, J. "TINKER - Software Tools for Molecular Design (version 7)", <http://dasher.wustl.edu/tinker>, **2015**.
- [67] Ren, P.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497-1506.
- [68] Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933-5947.
- [69] Grossfield, A.; Ren, P.; Ponder, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 15671-15682.
- [70] Grossfield, A. *J. Chem. Phys.* **2005**, *122*, 14671-14682.
- [71] Piquemal, J.-P.; Perera, L.; Cisneros, G. A.; Ren, P.; Pedersen, L. G.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*, 054511.
- [72] Wu, J. C.; Piquemal, J.-P.; Chaudret, R.; Reinhardt, P.; Ren, P. *J. Chem. Theory Comput.* **2010**, *6*, 2059-2070.
- [73] Rasmussen, T. D.; Ren, P.; Ponder, J. W.; Jensen, F. *Int. J. Quant. Chem.* **2007**, *107*, 1390-1395.
- [74] Kaminsky, J.; Jensen, F. *J. Chem. Theory Comput.* **2007**, *3*, 1774-1788.
- [75] Wilson, E. D.; J.C. Cross, P. . In *Molecular Vibrations : The Theory of Infrared and Raman Vibrational Spectra*; New York : McGraw-Hill : **1955**.
- [76] Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127-8134.
- [77] Bartell, L. S. *J. Am. Chem. Soc.* **1977**, *99*, 3279-3282.
- [78] Ponder, J. W. "personal communication", **2016**.
- [79] Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827-7843.
- [80] Alecu, I. M.; Zheng, J.; Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 2872-2887.
- [81] Elking, D. M.; Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A.; Pedersen, L. G. *J. Chem. Theory Comput.* **2010**, *6*, 190-202.
- [82] Chaudret, R.; Gresh, N.; Narth, C.; Lagardère, L.; Darden, T. A.; Cisneros, G. A.; Piquemal, J.-P. *J. Phys. Chem. A* **2014**, *118*, 7598-7612.
- [83] Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657-663.
- [84] Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890-1900.
- [85] Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459-467.
- [86] Narth, C.; Lagardère, L.; Polack, E.; Gresh, N.; Wang, Q.; Bell, D. R.; Rackers, J. A.; Ponder, J. W.; Ren, P. Y.; Piquemal, J.-P. *J. Comput. Chem.* **2016**, *37*, 494-506.
- [87] Semrouni, D.; Ohanessian, G.; Clavaguéra, C. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3450-3462.

- [88] Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. *J. Phys. Chem. B* **2015**, *119*, 9423-9437.
- [89] Ren, P.; Wu, C.; Ponder, J. W. *J. Chem. Theory Comput.* **2011**, *7*, 3143-3161.
- [90] Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233 - 239.
- [91] Stone, A. J. "Distributed Multipole Analysis for Gaussian Wavefunctions, version 2.2.03", **2005-2007**.
- [92] Ángyán, J. G.; Chipot, C. *Int. J. Quant. Chem.* **1994**, *52*, 17-37.
- [93] Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. *J. Phys. Chem. A* **1999**, *103*, 9872-9882.
- [94] Engkvist, O.; Astrand, P.-O.; Karlstöm, G. *J. Phys. Chem.* **1996**, *100*, 6950-6957.
- [95] Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048-1057.
- [96] Shi, Y.; Wu, C.; Ponder, J. W.; Ren, P. *J. Comput. Chem.* **2011**, *32*, 967-977.
- [97] Wu, J. C. *Development of Accurate and Efficient Models for Biological Molecules*, Thesis, The University of Texas at Austin, **2011**.
- [98] Thaumay, F.; Dognon, J.-P.; Ohanessian, G.; Clavaguéra, C. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25968-25977.
- [99] Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607-2609.
- [100] Chin, W. *Conformational landscape of small model peptides in the gas phase : UV/IR spectroscopy and theoretical approach*, Thesis, Paris Sud University, France, **2005**.
- [101] Chin, W.; PiuZZi, F.; Dognon, J.-P.; Dimicoli, I.; Mons, M. *J. Chem. Phys.* **2005**, *123*, 084301.
- [102] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *120*, 215-241.
- [103] Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549-2564.
- [104] Marjolin, A.; Gourlaouen, C.; Clavaguéra, C.; Ren, P. Y.; Piquemal, J.-P.; Dognon, J.-P. *J. Mol. Mod.* **2014**, *20*, 2471-2478.
- [105] Wu, J. C.; Chattree, G.; Ren, P. *Theor. Chem. Acc.* **2012**, *131*, 1138.
- [106] Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. *J. Phys. Chem. B* **2013**, *117*, 9956-9972.
- [107] Semrouni, D. *Energétique et spectroscopie de polypeptides par dynamique moléculaire : champ de force de seconde génération et chimie quantique*, Thesis, Ecole Polytechnique, Palaiseau, France, **2010**.

Chapitre 3

Développements théoriques

Dans notre équipe, une activité principale est la modélisation de spectres IR d'ions gazeux, en relation avec des expériences IR(M)PD. Nos approches combinent généralement les outils de la chimie quantique et de la mécanique classique, en vue de comprendre la dynamique des systèmes et de proposer des attributions aux signatures expérimentales. Depuis quelques années, des outils ont été développés autour de la dynamique moléculaire et du champ de forces polarisable AMOEBA, pour des applications aussi diverses que l'étude de la dynamique des ions hydratés [1], l'exploration de surface d'énergie potentielle de peptides [2, 3] ou d'agrégats [4], ou la spectroscopie IR dynamique [3, 5–7]. Les systèmes étudiés au cours de cette thèse (peptides chargés, ions solvatés) ont nécessité l'emploi de toute la méthodologie pré-existante et justifié la création de nouveaux outils. Nous commencerons par décrire ceux qui sont consacrés à l'exploration de surface d'énergie potentielle. La méthode REMD a été implémentée au sein de TINKER avant le début de ces travaux, pour explorer l'espace des conformations de peptides gazeux. Les spécificités de la recherche de structures stables d'ions solvatés ont nécessité des outils complémentaires, pour comprendre l'influence des inversions de cycles au sein du réseau de liaisons hydrogènes entre molécules d'eau. La spectroscopie IR dynamique, par transformée de Fourier de la fonction d'autocorrélation du moment dipolaire (DACF, décrite en **chapitre 1**), est implémentée dans TINKER. Elle ne donne pas directement accès à l'identification des bandes IR. Une méthode d'attribution dynamique des modes normaux de vibration a été implémentée dans TINKER et couplée au DACF. Une première version est consacrée à la description des modes des peptides, incluant leurs couplages et anharmonicités; une autre version permet de discuter la diversité des interactions au sein d'ions solvatés de tailles variables. Certaines signatures expérimentales, telles que celle associée au partage de proton, ne peuvent pas être décrites par le champ de forces classique. Un modèle Empirical Valence Bond (EVB) [8, 9] a été récemment développé et implémenté [10], couplé au champ de forces AMOEBA. La méthodologie est discutée au travers de deux exemples, l'aspartate et la série des diacides déprotonés $\text{HO}_2\text{C}-(\text{CH}_2)_n-\text{CO}_2^-$, $n = 2-6$, pour lesquels l'EVB a été en mesure de reproduire la dynamique d'échange du proton et la signature IR expérimentale caractéristique.

3.1 Recherche de structures de basse énergie

La première étape pour comprendre ou reproduire par le calcul des spectres expérimentaux est d'identifier les structures de basse énergie. Pour des systèmes de grande taille, l'espace conformationnel est complexe et son exploration nécessite des méthodes spécialisées. Les méthodes analytiques de recherches de points stationnaires sont nombreuses, basées sur l'énergie (simplex [11, 12], sequential univariate [13]), le gradient du potentiel (plus grande pente [14], gradients conjugués [15]) ou son hessien (Newton-Raphson [16], Quasi-Newton [17, 18]) mais sont facilement piégées dans des minima locaux. Elles ne permettent pas de trouver le minimum absolu d'une surface

de potentiel si la structure de départ de la recherche appartient à un puits de potentiel local éloigné du minimum global.

Il faut alors avoir recours à des méthodes qui, mieux que d'explorer finement un minimum local, sont capables de se déplacer sur la surface de potentiel et d'explorer les bassins éloignés. Nous ne détaillerons pas toutes les méthodes d'exploration disponibles, mais citons simplement quelques unes des plus connues : la méthode Monte-Carlo [19–21], éventuellement couplée à l'échange de répliques [22–24], la méthode du recuit simulé [25], les algorithmes inspirés de la génétique des populations [26,27], et les dynamiques moléculaires quantique ou classique. Au laboratoire, nous utilisons la dynamique moléculaire classique avec AMOEBA, couplée à de l'échange de répliques (REMD ou Replica Exchange Molecular Dynamics [28]).

Le REMD (ou parallel tempering [29]) permet d'effectuer plusieurs trajectoires en parallèle (des répliques) à des températures différentes, et de les échanger entre elles. Lorsqu'une réplique est échangée vers une température de simulation plus basse, elle explore mieux le fond des bassins énergétiques. Lorsqu'elle est échangée vers une température plus élevée, elle est plus apte à franchir des barrières de potentiel et quitter un bassin. La méthode permet ainsi de surmonter la tendance des méthodes d'exploration déterministes à rester piégées dans des bassins énergétiques locaux. L'implémentation du REMD avec TINKER a été réalisée par Penev et coll. via le logiciel TiReX [30] puis modifiée au laboratoire [5]. Puisque TINKER effectue les portions de trajectoires de dynamique moléculaire, on peut réaliser des simulations REMD avec le champ de forces polarisable AMOEBA. La parallélisation de TiReX est basée sur MPI (Message Passing Interface). Chaque portion de trajectoire d'une réplique est un processus indépendant entre deux tentatives d'échange successives. Chacune de ces portions est attribuée à un coeur de calcul. En pratique, nous utilisons de 12 à 16 répliques évoluant à des températures entre 100 et 350-800 K, en fonction du système d'étude. La distribution des températures se fait par un critère géométrique qui autorise les répliques à échantillonner souvent le fond des bassins. Les portions de trajectoires sont propagées avec un pas de temps de 1 fs. Les tentatives d'échanges ont lieu tous les 10 000 pas de dynamique, soit tous les 10 ps. Une trajectoire complète dure 5 ns et est sauvegardée tous les 5 ps. A l'issue, c'est donc entre 12 000 et 16 000 structures qui sont conservées pour être réoptimisées par le champ de forces AMOEBA, puis étiquetées par leur énergie. Afin de s'assurer une exploration efficace, des générations successives de REMD peuvent être réalisées à partir des structures de basse énergie issues des générations REMD précédentes. La méthode a été souvent appliquée à l'étude de peptides [2, 3, 7, 31, 32] et trouve un nouvel intérêt dans la recherche de structures stables de petits sulfates hydratés $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$.

En effet, si pour la gamme $n = 3-7$ molécules d'eau, il peut être assez aisé d'imaginer des structures stables, fondées sur l'intuition chimique (voir refs 12-21 de [4], **annexe D**), la tâche se complique rapidement lorsque la taille augmente. On remarque d'ailleurs que, au delà de $n = 7$, une approche fréquente consiste à (dés)habiller une structure connue pour une taille voisine [33–35]. Cette approche est difficilement justifiable par d'autres raisons que la difficulté de générer les structures d'intérêt. Ce point est soulevé par Wong et al. [36] puis Johnston et al. [37], qui réalisent des explorations de surface de potentiel de sulfates hydratés par des méthodes Monte-Carlo. Wong et al., sur la gamme $n = 3-17$, complètent la simulation Monte-Carlo par des simulations de dynamique moléculaire avec le champ de forces MMFFs à 400 ou 1000 K. Johnston et al. explorent toute la gamme $n = 3-50$, couplant le modèle TIP4P pour l'eau (modèle à 4 sites, 3 charges sur chacun des atomes plus un centre de Lennard-Jones, les coordonnées intramoléculaires sont rigides) [38], à une exploration par 'basin-hopping' [39]. Seules 3 classes de mouvements sont permises : la translation d'une molécule d'eau, sa rotation autour de son oxygène, ou du sulfate autour de son soufre, et une inversion de toutes les liaisons hydrogènes au sein d'un cycle. Le point commun des deux approches est d'utiliser la dynamique uniquement comme un moyen de générer un grand jeu de structures qui seront réoptimisées quantiquement, et non comme un outil de préselection des structures stables (la rigidité des modèles et les températures de simulation n'autorisant pas la dynamique à explorer efficacement des bassins de potentiel). Nous avons montré pour les tailles $n = 12$ et 13 [4] que les structures précédemment publiées étaient

loin d'être les plus stables et que des outils d'exploration performants doivent être utilisés, même à ces tailles. Pour $[SO_4(H_2O)_9]^{2-}$, nous montrerons au **chapitre 6** sans ambiguïté qu'il existe plusieurs structures bien plus compétitives que celle proposée par Johnston et al. [39].

Appliquée à l'étude de sulfates hydratés, notre méthodologie d'exploration s'est initialement basée sur des trajectoires de dynamique AMOEBA de 4 ns à 200 K, en profitant de la flexibilité du champ de forces polarisable. Le pas d'intégration est fixé à 1 fs et la trajectoire est sauvegardée toutes les 10 ps. Certaines structures jugées intéressantes ont initié d'autres trajectoires afin de compléter l'exploration de la surface d'énergie potentielle. Les plus petits agrégats autorisent la construction de structures complémentaires, fondées sur le sens chimique. Parallèlement, des simulations REMD avec TiReX ont été réalisées dans les conditions précédemment énoncées. Afin d'éviter une éventuelle évaporation de molécules d'eau aux hautes températures de simulation, l'agrégat est entouré d'un potentiel de Lennard-Jones, sphérique de rayon 12 Å, assez grand pour ne pas être ressenti par les molécules d'eau en surface de l'agrégat, mais capable de les y rediriger en cas d'évaporation. L'ensemble de ces outils peut aisément engendrer de grandes quantités de structures dont un échantillon doit être sélectionné, en vue d'optimisations par des méthodes quantiques ou comme points de départ de simulations DACF/AMOEBA. La sélection est réalisée en partie sur la base des énergies AMOEBA, et nécessite d'évaluer la confiance à attribuer à ces énergies, et sur des critères plus subjectifs, tels qu'une haute compacité, une certaine symétrie, ou la présence d'agencements spécifiques à mêmes de reproduire certaines bandes caractéristiques.

Dans ce large éventail de structures de départ, la plupart possèdent des réseaux cycliques de liaisons hydrogènes, c'est à dire un enchaînement de liaisons hydrogènes dont on peut faire varier le sens sans influencer qualitativement le reste de la structure, et que l'on qualifiera de cycles indépendants. En **Fig 3.1**, un exemple est donné pour une structure de $[SO_4(H_2O)_9]^{2-}$.

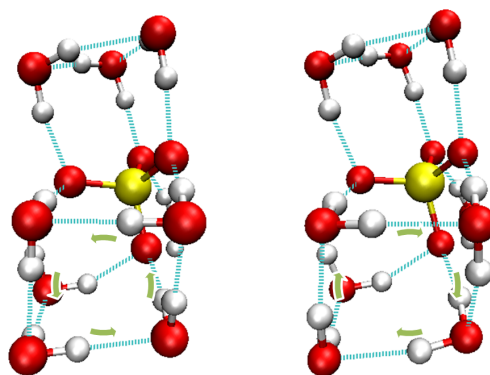


FIGURE 3.1 – Exemple d'un cycle de liaisons hydrogènes (cycle à 4 molécules d'eau), pour une structure de $[SO_4(H_2O)_9]^{2-}$. Les molécules peuvent donner une liaison hydrogène dans le sens horaire (à droite) ou anti-horaire (à gauche), sans nécessiter de réorganisation du reste de la structure.

La question de l'équivalence (en termes d'énergie et de signature spectrale) de ces structures est posée. Nous discuterons dans le **chapitre 6**, sur les structures $[SO_4(H_2O)_6]^{2-}$ et $[SO_4(H_2O)_9]^{2-}$, l'impact de ces inversions sur la stabilité des structures engendrées. Afin de comptabiliser et engendrer toutes les structures issues de ces inversions de cycles, nous avons créé un outil en langage PYTHON, basé sur le package NetworkX [40], également utilisé par Johnston [37]. NetworkX est un outil d'analyse de données qui réalise des opérations sur des graphes. Dans notre application, les oxygènes sont les noeuds d'un réseau, avec une orientation dictée par les liaisons covalentes OH. Notre outil détermine, pour un réseau de liaisons hydrogènes, tous les cycles fermés indépendants de liaisons hydrogènes (i.e tous les cycles qui ne peuvent pas être formés intégralement par combinaison d'autres cycles). L'application répertorie également toutes les combinaisons d'au plus 4 inversions simultanées et génère les nouvelles

structures. Le choix de se limiter à 4 cycles inversés simultanément n'est dicté que par le nombre de combinaisons, qui croît de façon exponentielle avec la taille de l'agrégat (**Fig 3.2**). Pour $n = 6$ et 9, les énergies relatives des structures inversées ont été calculées et seront discutées en **chapitre 6**. Au-delà de $n = 13$, il devient illusoire de vouloir accéder à toutes les structures par un calcul quantique de haut niveau théorique mais, sous certaines conditions, ces informations peuvent contribuer au calcul d'une population de Boltzmann.

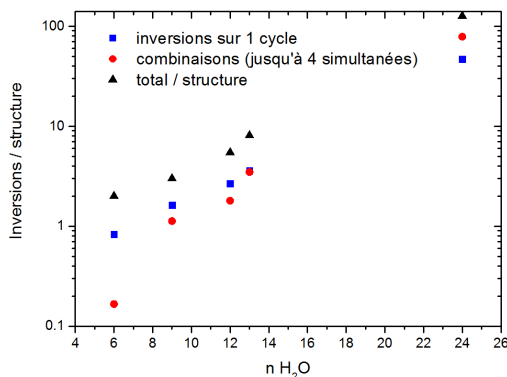


FIGURE 3.2 – Nombre de structures générées par inversion d'un à 4 cycles de liaisons hydrogènes. Les données sont moyennées sur une sélection de structures : $n = 6$ (6), $n = 9$ (8), $n = 12$ (15), $n = 13$ (23), $n = 24$ (11). Pour $n = 9$, 6 des structures engendrées par inversions de cycles sont des doublons, pour des questions de forte symétrie de l'agrégat, et ont donc été retirées du total. Le nombre d'inversions augmentant exponentiellement avec la taille de l'agrégat, une échelle logarithmique est utilisée.

3.2 Analyse des modes normaux

Les spectres IR calculés dynamiquement par DACF (**chapitre 2**) permettent d'accéder aux effets d'anharmonicité et de température. En contrepartie, la fonction d'autocorrélation du dipôle ne fournit pas les informations nécessaires à l'attribution des bandes. Au-delà de l'approche par diagonalisation du Hessien, qui ne peut prétendre capturer ces effets, plusieurs approches existent, fiables mais souvent limitées à l'étude de petits systèmes (voir refs 13, 27-48 de [7]). Wheeler et al. ont proposé la méthode 'principal mode analysis' (PMA) [41, 42] pour laquelle les modes normaux et leurs fréquences associées sont calculés à partir d'une trajectoire de dynamique moléculaire classique en évaluant les vecteurs et valeurs propres de la matrice de covariance des déplacements des coordonnées moléculaires autour de leur position d'équilibre [42–44]. L'approche est simple mais n'est exacte que sous certaines conditions, dont l'harmonicité du potentiel, un équilibre thermique complet au sein de la molécule, et un hamiltonien moléculaire indépendant du temps. Cependant, aucune de ces conditions n'est généralement respectée au cours d'une simulation de dynamique moléculaire. L'analyse instantanée de mode normaux [45, 46], est basée sur une diagonalisation du hessien à chaque pas de dynamique. La description des modes normaux évolue avec le temps, ce qui complique leur identification [47]. De plus, la méthode peut être très coûteuse pour des systèmes importants. La procédure de Kohanoff [48] permet d'extraire les valeurs et vecteurs propres d'une trajectoire de dynamique sans équilibre thermique. Une première estimation globale des fréquences est suivie par la détermination des vecteurs propres selon un ajustement aux moindres carrés de la trajectoire incluant des contraintes d'orthogonalité. Puis la trajectoire est propagée sur chacun des modes normaux. De cette façon, chacune des trajectoires projetées décrit principalement une fréquence, qui est ensuite réestimée, généralement en utilisant une transformée de Fourier de la fonction d'autocorrélation des vitesses (VACF). Vuilleumier et al. [49] ont proposé une méthode basée sur un

critère de localisation de la transformée de Fourier des fonctions d'autocorrélation de la vitesse des modes effectifs. Un choix de la fonction de localisation conduit la méthode à être équivalente à une approche de type PMA, basée sur la diagonalisation de la matrice de covariance; un autre critère de localisation conduit à une méthode offrant une forte analogie avec l'analyse usuelle des modes normaux des structures à l'équilibre, mais où le hessien moyenné en température remplace le hessien. Le hessien lui-même n'est jamais calculé. Bowman et al. ont proposé une nouvelle approche d'analyse des modes normaux de vibration, ne reposant pas sur une diagonalisation du Hessien. La méthode Driven Molecular Dynamics (DMD) [50, 51] repose sur l'ajout d'une force d'entraînement $U(t)$, de forme sinusoïdale, à la dynamique. Dans ce cas, le Hamiltonien de la dynamique moléculaire entraînée devient, pour un système de N atomes :

$$H(\mathbf{p}, \mathbf{q}, t) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{q}) + U(t) \quad (3.1)$$

avec \mathbf{q} et \mathbf{p} respectivement les $3N$ coordonnées atomiques et leurs moments conjugués, et V le potentiel moléculaire.

Le terme d'entraînement est calculé à chaque pas de la dynamique, fonction des distances entre noyaux :

$$U(t) = \sum_{i,j} \lambda_{ij} r_{ij} \sin(\omega_n t) \quad (3.2)$$

où λ_{ij} sont les constantes de couplage (i.e les paramètres d'entraînement), r_{ij} les distances internucléaires et ω_n la pulsation de la force d'entraînement, constante pendant la dynamique. Notons qu'en pratique, tous les λ_{ij} sont choisis égaux (on notera alors directement λ).

Lorsque cette force sinusoïdale est propagée pendant la dynamique, elle est ressentie par les vibreurs du système. Ceux dont la fréquence propre est très proche de ω_n absorbent rapidement de l'énergie, augmentant ainsi l'amplitude de leurs oscillations. Ce phénomène de résonance, qui peut être localisé ou plus global, est assimilable à un mode normal de vibration. Les autres vibreurs, de fréquence propre non résonante avec ω_n , peuvent absorber temporairement puis redistribuer de l'énergie. L'effet est illustré en **Fig 3.3** sur le N-méthyl-acétamide. Une pulsation $\omega_n = 3518 \text{ cm}^{-1}$ a été choisie comme périodicité de la force d'entraînement. Cette fréquence est très proche de la fréquence propre du vibreur N-H, qui résonne et absorbe de l'énergie graduellement durant les 5 ps de simulation. A l'inverse, le vibreur C=O ne subit qu'une faible perturbation.

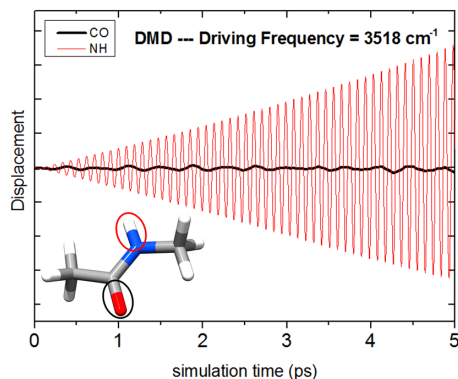


FIGURE 3.3 – Suivi des élongations C=O et N-H du NMA pendant une simulation DMD de 5 ps, avec $\omega_n = 3518 \text{ cm}^{-1}$.

Bowman et al. ont montré sur H_5O_2^+ [50], que dans la limite des faibles oscillations, les modes normaux obtenus

par DMD sont quasi-identiques à ceux obtenus dans une approche statique. Sur des mouvements collectifs de molécules plus complexes, tels que le mode amide II du N-méthyl-acétamide (**Fig 3.4**), nous notons [7] que toutes les contributions du mode sont reproduites, mais que certaines intensités peuvent varier, en raison des mécanismes d'absorption d'énergie et d'inévitables petites variations de géométries entre les structures optimisée et de départ de l'enregistrement du mode normal par DMD.

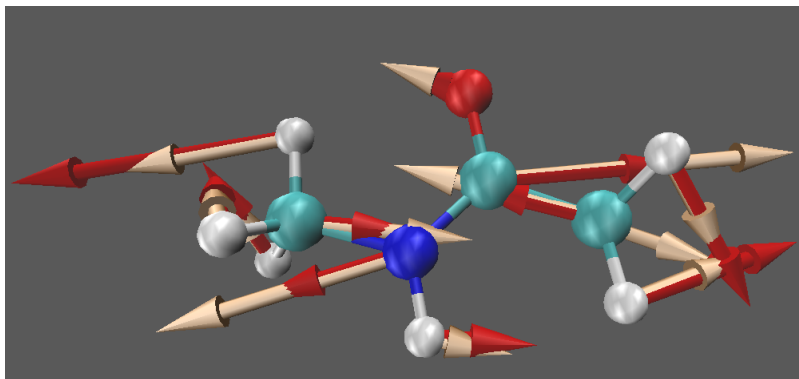


FIGURE 3.4 – Mode amide II du N-méthyl-acétamide (1551 cm^{-1}) : en blanc, modes normaux à partir de calculs statiques et en rouge, à partir de simulation DMD (10 000 pas, 0,1 fs par pas, $\lambda = 10^{-6}$ Hartree. Bohr $^{-1}$, filtre = 5 cm^{-1}).

Mais l'avantage de l'approche réside dans une augmentation des paramètres d'entraînement (λ et longueur de simulation), qui autorise une plus grande amplitude des oscillations et l'apparition de couplages anharmoniques. Nous avons ainsi proposé, pour l'étude du dipeptide Ace-Ala-Phe-NH₂ en phase gazeuse [7], plusieurs jeux de conditions d'entraînement, $1 \times 10^{-6} \leq \lambda \leq 2 \times 10^{-5}$ Hartree. Bohr $^{-1}$, associée à des simulations de 5 000 à 10 000 pas de 0.1 fs, pour capturer soit les modes isolés dans l'approximation des faibles oscillations, soit des mouvements couplés tels qu'observés à plus hautes températures.

Bowman propose d'utiliser la DMD comme moyen de calculer des spectres vibrationnels non statiques, tout en conservant l'information sur les modes normaux de vibration [50]. L'inconvénient majeur est l'absence de signification des intensités calculées, fonctions de l'amplitude des oscillations et non de la variation du moment dipolaire. Toutefois, l'outil DMD a l'avantage de pouvoir être facilement couplé au DACF, du fait qu'ils reposent tous les deux sur des simulations avec le champ de forces AMOEBA. Les spectres DACF englobent les effets de dynamique et de température, leurs intensités sont de vraies intensités IR, directement comparables aux spectres expérimentaux IRPD. Il suffit alors d'extraire de ces spectres les fréquences des bandes dont on souhaite une interprétation et de réaliser pour chacune une courte simulation DMD avec le ω_n correspondant.

3.2.1 Implémentation pour l'étude des peptides

Notre première implémentation de la DMD dans TINKER6 visait à attribuer les modes de vibration de petits peptides (~ 40 atomes) sur la gamme $300\text{-}3600 \text{ cm}^{-1}$. L'implémentation est détaillée dans l'article [7] en **annexe A** mais peut être résumée succinctement :

Calcul d'un spectre DACF en température : La dynamique est propagée en utilisant l'algorithme Verlet-vitesse, avec un pas de simulation de 0.1 fs, qui permet de bien échantillonner les hautes fréquences. Chaque trajectoire débute par une phase d'équilibration de 10 à 50 ps, dans l'ensemble canonique, afin d'atteindre la température recherchée. Cette température est maintenue via le thermostat Nosé-Hoover. A l'issue, les positions

et vitesses sont enregistrées pour initier une seconde simulation de 200 ps réalisée dans l'ensemble microcanonique, pendant laquelle le moment dipolaire est collecté à chaque pas. A la fin, un spectre est calculé par la procédure DACF. L'ensemble de la procédure est réalisé 3 fois, pour lesquelles les géométries initiales sont identiques mais les vitesses atomiques et les énergies cinétiques sont distribuées de façon aléatoire. Le spectre final est une moyenne des 3 simulations. Les fréquences ω_n à scanner en DMD sont choisies à partir de ce spectre DACF.

Analyse de modes dynamiques par DMD : Une simulation DMD est réalisée pour chaque ω_n d'intérêt. Les amplitudes d'oscillation de chacune des 3N-6 coordonnées internes (CI) sont enregistrées à chaque pas de la simulation et leur fréquence est calculée à l'issue. Les CI qui obéissent au critère $|\omega - \omega_n| \leq \epsilon$, sont identifiées comme résonantes. Le choix du filtre en fréquence ϵ , entre 0.5 et 5 cm^{-1} , dépend de la taille du système et de la gamme de fréquences étudiée, et sera discuté ultérieurement. Parmi les CI résonantes, celle dont l'écart $|\omega - \omega_n|$ est le plus faible sert de référence pour définir le mode normal. Ses amplitudes permettent de définir deux pas dans la simulation (**Fig 3.5**). Pour chaque atome impliqué dans la définition d'une CI résonante, le déplacement associé au mode normal est donné par l'écart de position entre ces deux pas. Pour tous les autres, le déplacement associé est nul. Le déplacement est converti en coordonnées cartésiennes et un fichier de visualisation du mode normal est écrit dans un format lu par Molden ou VMD.

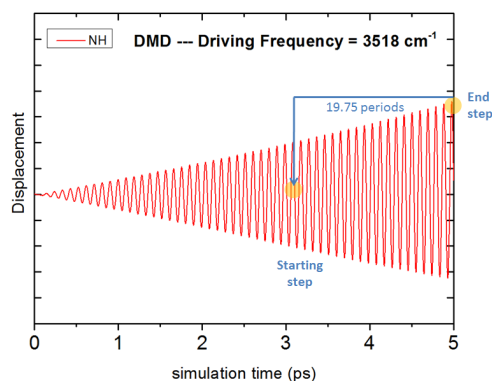


FIGURE 3.5 – Choix de deux pas de simulation pour décrire le mode normal de vibration. La CI de l'élongation N-H est la référence du mode normal. Le pas de fin correspond à la dernière grande amplitude de la CI. Le pas de départ est déduit en retranchant 19.75 périodes. A ce pas (et toutes les $\frac{1}{4}$ périodes), la CI est à sa valeur d'équilibre.

Détails techniques : Les géométries initiales peuvent être optimisées avec TINKER jusqu'à un RMS sur le gradient inférieur à 10^{-5} kcal.mol^{-1} , afin d'éviter une période plus ou moins longue pendant laquelle l'énergie est distribuée sur plusieurs CI, avant que n'émergent les CI résonantes. C'est à cette condition que l'émergence de résonance peut être clairement observée après seulement quelques milliers de pas de simulation. Il est généralement recommandé d'utiliser un pas d'intégration maximum de 1/10 de la période du signal observé (par exemple 0.8 fs à 4000 cm^{-1}). Nos outils d'analyse de modes sont basés sur un calcul très précis des fréquences d'oscillation des coordonnées internes, et requièrent d'augmenter le critère de précision : un pas d'intégration de 0.1 fs dans la gamme 1000-4000 cm^{-1} et 0.2 fs en dessous. Pour la même raison, la précision numérique sur les coordonnées fournies dans TINKER a dû être augmentée de 10^{-6} Å pour les coordonnées cartésiennes, 10^{-5} Å pour les longueurs des liaisons et 10^{-4} ° pour les angles, à 10^{-8} Å pour les coordonnées cartésiennes, 10^{-7} Å pour les longueurs de liaisons et 10^{-6} ° pour les angles, via le mot clef DIGITS. Le paramètre λ contrôle l'amplitude de la force d'entraînement et doit être choisi avec précaution. Il doit permettre une rapide excitation des modes et reproduire une anharmonicité proche de celle du spectre DACF à analyser. Il existe ici un biais entre les méthodes, le mode d'absorption de l'énergie par DMD étant par essence sélectif en fréquence, critère que l'on renforce d'autant plus que l'on cherche à reproduire des modes à basse température. Toutefois, pour un mode isolé, il peut être utile de corrélérer le déplacement en fréquence

d'une bande en fonction de la température (en DACF) ou de λ (en DMD). En **Fig 3.6**, la fréquence du mode local d'élongation N-H du NMA est suivie en fonction des deux paramètres. Les conditions de simulation sont données sur la figure. On observe que l'on peut choisir λ de façon à reproduire le déplacement de la bande DACF, jusqu'à mimer la température de 300 K. Au-delà, on atteint les limites de l'absorption différente des deux modèles. A 350 K, une grande quantité d'énergie est absorbée par l'ensemble des vibreurs de la molécule et conduit à des variations de géométrie que ne reproduit pas l'absorption plus spécifique de la DMD. Pour l'attribution de spectres DACF à 1 K, on cherche à rester dans la limite des oscillations harmoniques. Nous avons utilisé $\lambda = 10^{-6}$ Hartree.Bohr⁻¹ dans la gamme 1000-4000 cm⁻¹ et 5.10^{-7} Hartree.Bohr⁻¹ en dessous, car les modes sont en moyenne plus mous. A 200 K, la température autorise de nombreux couplages entre modes et des comportements anharmoniques. Pour représenter ces modes, nous avons ajusté λ à 2.10^{-5} Hartree.Bohr⁻¹ dans la gamme 1000-4000 cm⁻¹ et 10^{-5} Hartree.Bohr⁻¹ en dessous.

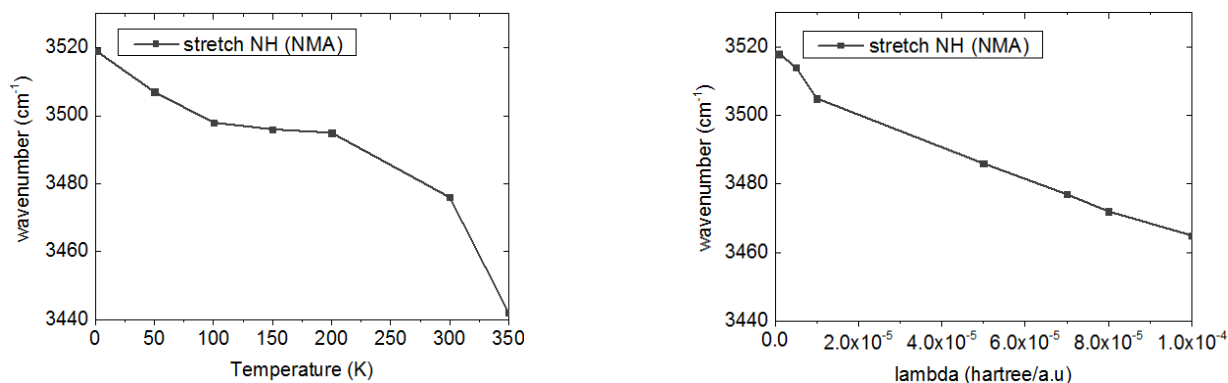


FIGURE 3.6 – Déplacement du maximum du mode d'élongation N-H du NMA en fonction de la température ou de l'énergie absorbée. A gauche : 100 ps de simulation DACF dans la gamme 0-350 K. A droite : chaque point correspond à une simulation DMD de 20 ps dans la gamme 3400-3550 cm⁻¹ pour différentes valeurs de λ . A 200 K, l'intensité maximale de la bande d'élongation N-H est observée pour une fréquence proche de 3495 cm⁻¹. Le déplacement de la bande peut être reproduit avec une simulation DMD de 20 ps avec $\lambda = 3.10^{-5}$ Hartree.Bohr⁻¹.

Pour l'étude du dipeptide Ace-Ala-Phe-NH₂ [7], une recherche de structures par REMD a validé les structures précédemment publiées [52, 53]. Le DACF a permis de reproduire les signatures expérimentales dans la gamme des empreintes digitales et des élongations N-H. Enfin, la DMD a fourni des attributions de modes normaux pour ces deux gammes, et a mis en lumière l'apparition de couplages anharmoniques avec la température.

3.2.2 Implémentation pour l'étude des ions solvatés

Encouragés par les capacités de l'outil DMD à attribuer des modes normaux dynamiques, nous avons cherché à l'employer dans un autre de nos domaines d'intérêt : la spectroscopie des ions hydratés, et plus particulièrement l'attribution des massifs d'élongation OH de sulfates hydratés $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$, pour $n = 24-80$ molécules. Les enjeux de la compréhension des interactions au sein de ces agrégats seront détaillés au **chapitre 5**, nous discuterons ici uniquement de l'aspect méthodologique. A ces tailles, de (très) nombreuses configurations sont compétitives et les spectres expérimentaux sont le reflet de ce mélange. Si pour les plus petits ($n = 24-36$), une exploration efficace de la surface d'énergie potentielle associée à de nombreux spectres statiques, peut encore être une partie de la solution, la démarche devient impraticable au-delà. Les spectres dynamiques à température finie deviennent la méthode de référence, capable de reproduire, avec une assez bonne précision, à la fois les massifs des élongations des OH liés et l'apparition des bandes caractéristiques de groupements de OH libres dans les agrégats de plus grande taille (**chapitres 5 et 6**). Pour un ion hydraté par 2 à 4 sphères de solvation et autant d'interactions spécifiques, la réponse des vibreurs OH de l'eau se distribue sous un massif large de $\sim 600-700$ cm⁻¹, et fluctue au gré de la

dynamique de réarrangement de l'agrégat.

Pour identifier ces contributions, la couplage DACF-DMD fait face à de nouveaux enjeux dont le plus important est de conserver la précision de réponse malgré la densité de vibreurs dans la gamme observée. Pour $n = 80$, il s'agit de 160 vibreurs OH, impliqués dans des modes locaux ou de combinaison, et répartis sur la gamme 3100-3750 cm^{-1} . A cela s'ajoute la flexibilité de réorganisation des molécules d'eau au travers de leur interactions par liaison hydrogène, qui peut facilement induire une réponse collective si on utilise les conditions de simulation précédemment définies. En effet, la redistribution d'énergie entre les CI est d'autant plus facile que leurs fréquences propres sont proches, et des phénomènes de compétition sont fréquemment observés. En **Fig 3.7**, à gauche, on observe les amplitudes d'oscillation de deux vibreurs OH_1 et OH_2 d'une même molécule d'eau sur une simulation DMD de 0.2 ps. $\omega_n = 3600 \text{ cm}^{-1}$ est propre à exciter le mode impliquant la combinaison antisymétrique des deux vibreurs de cette molécule de seconde sphère d'un agrégat $[\text{SO}_4(\text{H}_2\text{O})_{80}]^{2-}$. Dans la figure de droite, la simulation est prolongée jusqu'à 1 ps. On suit toujours les oscillations du vibreur OH_2 mais à présent également celles d'un vibreur O_AH_B d'une autre molécule d'eau. On observe que, par le jeu de réorganisations du réseau de liaisons hydrogènes, l'environnement autour du vibreur OH_2 a changé sur la gamme 0.4-0.6 ps, et que celui-ci ne résonne plus. A l'inverse, le vibreur O_AH_B absorbe plus fortement sur cette gamme temporelle et pourrait être défini comme CI résonante si la durée de la simulation DMD était de 0.6 ps, bien que les profils des deux vibreurs indiquent clairement que OH_2 est un meilleur candidat.

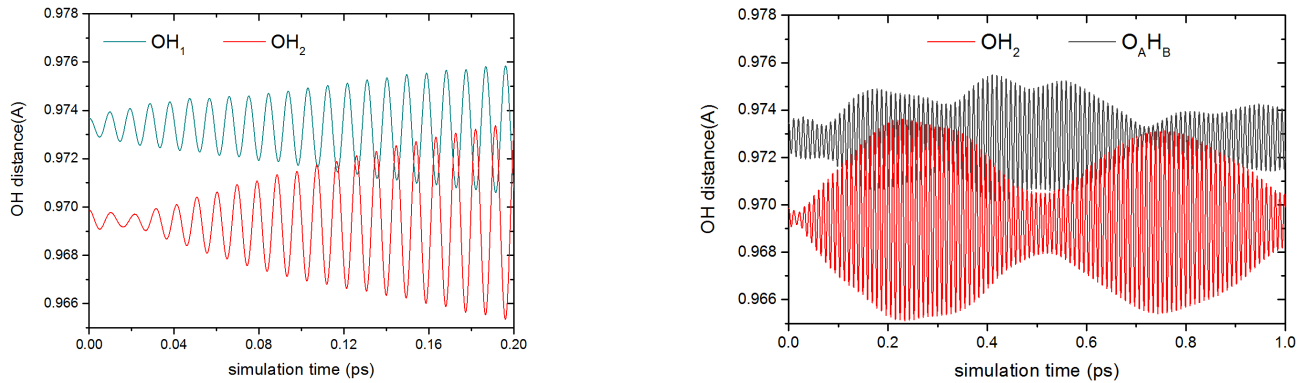


FIGURE 3.7 – Suivi des élongations OH de molécules d'eau dans un agrégat $[\text{SO}_4(\text{H}_2\text{O})_{80}]^{2-}$, pendant des simulations de DMD à $\omega_n = 3600 \text{ cm}^{-1}$. A gauche, les deux composantes de la combinaison antisymétrique au sein d'une molécule, capturée pendant les 1^{ères} 0.2 ps de simulation. A droite, la simulation est prolongée à 1 ps et permet d'observer la compétition avec un vibreur d'une autre molécule, via des réorganisations du réseau de liaisons hydrogènes.

Il devient essentiel de se placer dans des conditions qui limitent la redistribution d'énergie, mais sans toutefois se restreindre aux très faibles oscillations harmoniques. Plusieurs combinaisons de λ et de durée de simulation ont été testées, la sensibilité aux conditions d'entraînement augmentant avec la taille des agrégats et la facilité de réorganisation du réseau des liaisons hydrogènes. Le point délicat du choix des conditions est que, spécialement pour les agrégats de grande taille, l'environnement autour des molécules d'eau des sphères extérieures (3^{ème} et 4^{ème}) est très flexible, ce qui induit qu'elles peuvent répondre à une gamme assez large de fréquences d'excitation, pour peu que les conditions soient assez souples. A l'inverse, la 1^{ère} sphère est beaucoup plus rigide et notamment les vibreurs liés au sulfate ont besoin d'absorber un minimum d'énergie pour présenter des oscillations visibles. Il faut ainsi permettre une excitation suffisante du système pour capturer au moins une résonance par vibreur sur la gamme scannée, sans bruyier tout le signal par des réorganisations dans les sphères extérieures. Dans la version actuelle du modèle, une durée de simulation de 0.2 ps, soit une vingtaine de périodes, alliée à $2.10^{-7} \leq \lambda \leq 5.10^{-7} \text{ Hartree.Bohr}^{-1}$, nous apparait comme un bon compromis. Dans une future implémentation orientée vers

le traitement des agrégats de plus grande taille ($n \geq 80$), on pourra revenir à une définition de λ_{ij} dépendant du vibreur excité, avec un traitement différent en fonction des sphères de solvation.

A la différence des peptides, pour lesquels le spectre présente des bandes résolues, les élongations OH des agrégats se distribuent sous un large massif (par exemple pour $[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$ en **Fig 3.9**) et il n'y a plus lieu d'identifier des fréquences spécifiques. A présent, la gamme complète des fréquences DACF est analysée. Pour chaque fréquence entière de la gamme, une simulation DMD est lancée, de façon systématique, à l'issue de laquelle les fréquences d'oscillations de chaque CI sont relevées. On applique un filtre en fréquences ϵ tel que $0.5 \text{ cm}^{-1} \leq \epsilon \leq 1 \text{ cm}^{-1}$, proche de la valeur du pas de scan en fréquence. Seules sont conservées les fréquences d'oscillations des CI qui obéissent au critère $|\omega - \omega_n| \leq \epsilon$. La position de chacune des coordonnées internes dans la structure est relevée. Chaque vibreur est identifié comme appartenant à une molécule de 1^{ère}, 2^{ème}, 3^{ème} ou 4^{ème} sphère de solvation. Est également stocké le type de molécule vers laquelle l'hydrogène pointe (vers le sulfate, vers 1 molécule de 1^{ère} sphère, de 2^{ème} sphère ou une autre).

La nomenclature des différents vibrateurs est illustrée en **Fig 3.8**, sur la structure notée W24-A. Pour cette structure, et une autre, W24-B (les deux ayant été extraites d'une trajectoire AMOEBA à 200 K), les attributions DMD sont réparties en fonction des sphères de solvation (1^{ère} sphère sur la figure de gauche, et 2^{nde} sphère sur celle de droite), et superposées au spectre DACF à 200 K (**Fig 3.9**). Dans cet exemple, les deux structures ont chacune 12 molécules en 1^{ère} et en 2^{nde} sphères. Chacun des 48 vibrateurs OH est associé au minimum à une fréquence de résonance. La barre verticale associée est d'autant plus haute que $|\omega - \omega_n|$ est faible. En 2^{nde} sphère, l'étiquette *Other* désigne un vibreur d'une molécule qui pointe vers une molécule qui n'est pas en 1^{ère} sphère (dans cet exemple particulier, il s'agit évidemment d'une molécule de 2^{nde} sphère). Une simulation DMD de 0.2 fs correspond à une photo du système à l'échelle de la dynamique, et ne peut prétendre décrire tout le spectre DACF. Toutefois, un petit jeu de simulations DMD à partir de structures de la dynamique DACF (telles que les deux présentées ici), fournit de précieuses informations sur la localisation des types de modes en fonction de la taille de l'agrégat. Par exemple, les fréquences les plus rouges du spectre DACF de $[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$ sont à attribuer aux molécules de 1^{ère} sphère, au travers de leurs vibrateurs dirigés vers la 2^{nde} sphère. Ce résultat est confirmé par les calculs DFT et n'est pas accessible par une simple diagonalisation du hessien basée sur le potentiel AMOEBA (ce point sera discuté au **chapitre 6**).

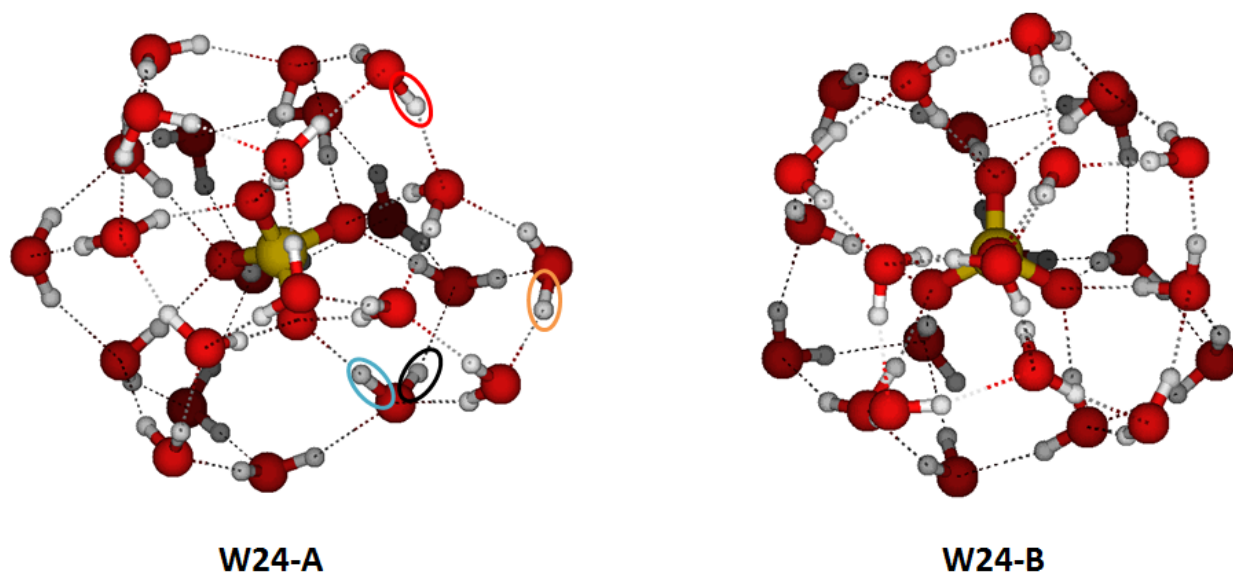
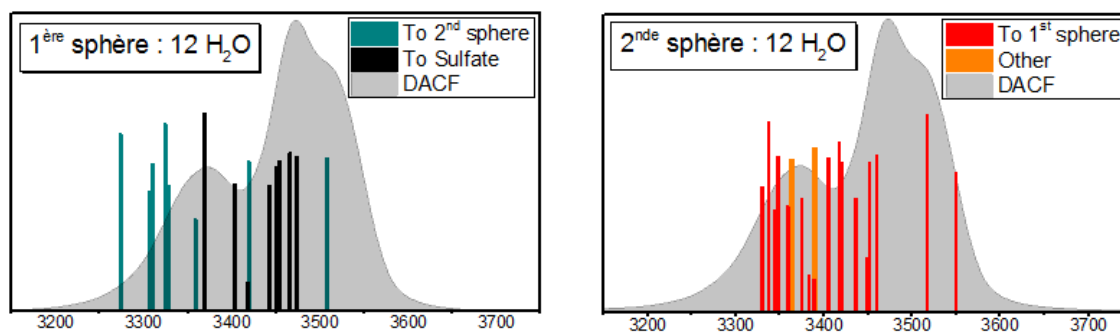


FIGURE 3.8 – Deux agrégats $[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$, dont les fréquences sont attribuées en DMD. Sur W24-A sont identifiés différents vibrateurs. En noir : vibrateur de 1^{ère} sphère et hydrogène dirigé vers le sulfate; en bleu : vibrateur de 1^{ère} sphère et hydrogène dirigé vers une molécule de 2^{nde} sphère; en rouge : vibrateur de 2^{nde} sphère et hydrogène dirigé vers une molécule de 1^{ère} sphère; en orange : vibrateur de 2^{nde} sphère et hydrogène dirigé vers une molécule autre que de 1^{ère} sphère (ici de 2^{nde}).

Structure W24-A : $\text{Lambda} = 5.10^{-7}$ Hartree Bohr⁻¹, 2 000 steps, 1 cm⁻¹ filter



Structure W24-B : $\text{Lambda} = 5.10^{-7}$ Hartree Bohr⁻¹, 2 000 steps, 1 cm⁻¹ filter

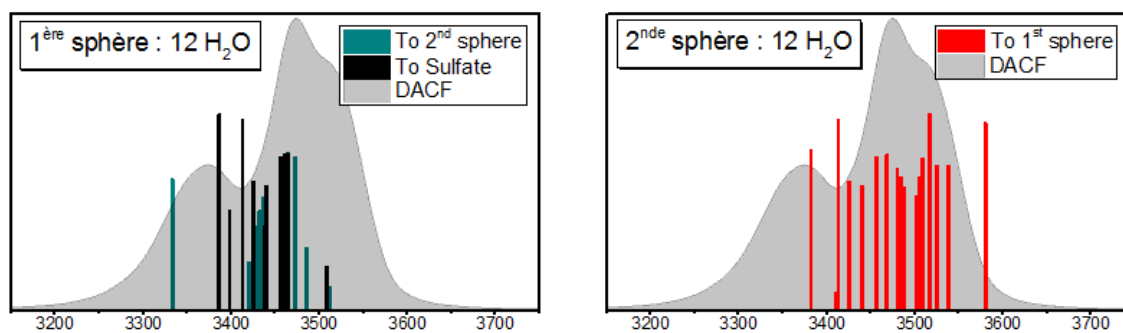


FIGURE 3.9 – Attributions DMD des deux structures W24-A et W24-B comparées à une moyenne de spectres DACF (enveloppe grise) à 200 K.

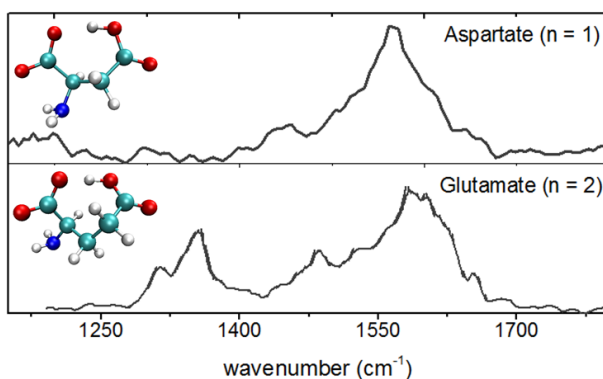


FIGURE 3.10 – Spectres IRMPD de bases conjuguées d’acides-amino $\text{CO}_2\text{-CH}_2(\text{NH}_2)\text{-(CH}_2)_n\text{-COOH}$, enregistrés à CLIO.

3.3 Modélisation de la spectroscopie du partage de proton : Empirical Valence Bond

Dans la suite de cette partie, on considère la contribution électrostatique comme l’ensemble des interactions provenant de la distribution multipolaire et des effets de polarisation pris en compte comme un terme à n -corps.

La dynamique moléculaire basée sur le champ de forces polarisable AMOEBA fournit des outils très efficaces, au travers du DACF et de la DMD, pour l’étude de la spectroscopie IR de systèmes aussi variés que des peptides gazeux ou des ions solvatés. La description sophistiquée de l’électrostatique, la prise en compte de l’anharmonicité due aux effets de température, et la possibilité d’enregistrer un spectre dynamique, représentatif d’un environnement évolutif, lui permettent de proposer une description réaliste de systèmes parfois très complexes. Pourtant, par définition, le champ de forces classique est confronté à la limite intrinsèque qu’aucune liaison covalente ne saurait être rompue pendant la dynamique. Outre les limites évidentes en termes d’étude de la réactivité chimique, certaines signatures spectrales expérimentales ne peuvent être reproduites. Il s’agit notamment de la signature caractéristique du transfert de proton, mécanisme d’une importance majeure dans quantités de processus chimiques ou biochimiques [54]. Un exemple de ces signatures caractéristiques a été proposé par Oomens [55] sur une série de spectres IRMPD d’acides-amino déprotonés. Dans la région des empreintes digitales, il observe la différence de forme entre Trp, Tyr, Phe, Cys et Ser déprotonés d’une part, et Asp et Glu d’autre part. Pour les premiers, deux bandes autour de 1330 et 1630 cm^{-1} peuvent être attribuées aux modes d’élongation symétrique et antisymétrique COO^- , alors que pour les deux derniers, un massif large est observé sur la gamme 1300-1700 cm^{-1} et attribué aux élongations C=O et OCO en présence d’un proton fortement délocalisé entre les groupements carboxylates. Utilisant la même technique IRMPD, nous avons ré-enregistré les spectres à température ambiante de l’aspartate et du glutamate dans la région 1100-1800 cm^{-1} (**Fig. 3.10**), au Centre Laser Infrarouge d’Orsay (CLIO). Une attention particulière a été portée sur les conditions d’irradiation et de fragmentation (pour les détails techniques, voir [55, 56]), et conduit à des spectres plus résolus, mais en accord qualitatif avec ceux d’Oomens et coll. [55]. Pour l’aspartate, on observe une large bande sur la gamme 1470-1680 cm^{-1} , avec un maximum d’intensité à 1570 cm^{-1} . Du côté rouge de cette bande, un coude apparaît vers 1450 cm^{-1} . Le spectre du glutamate est assez similaire, avec le massif principal décalé de 25 cm^{-1} dans le bleu par rapport à l’aspartate. Un second massif, entre 1275 et 1400 cm^{-1} , peut être lié à l’interaction de groupements CH_2 avec les groupements acides, interaction qui n’est pas possible pour l’aspartate, car la chaîne carbonée est plus courte.

Notre objectif est de pouvoir reproduire, au cours de simulations de dynamique moléculaire, la mobilité d’un proton qui s’échange entre deux sites accepteurs. Une telle propriété peut être calculée au moyen de dynamique de type Born-Oppenheimer (BOMD) ou Car-Parinello (CPMD) [57–59], même si les coûts calculatoires importants

peuvent empêcher d'obtenir la statistique nécessaire à un résultat significatif. Cette contrainte motive l'utilisation de méthodes semi-empiriques, telles que AM1 ou PM3, récemment reparamétrées pour l'étude du transfert de proton dans l'eau [60, 61]. Lorsque le choix est porté sur l'utilisation de champs de forces classiques, il s'agit de champs de forces réactifs [62], dont la procédure de paramétrisation peut être assez complexe. La première description d'un transfert de proton entre deux anions OH^- , par une fonction de plusieurs états, a été proposée par Goldblum [63], mais a été largement diffusée grâce au modèle Empirical Valence Bond (EVB) de Warshel [8, 9]. L'approche a été étendue par plusieurs groupes, notamment pour l'étude du transport du proton dans l'eau par Voth et al. [64] ou Borgis et al. [65].

Nous discutons l'implémentation d'un modèle EVB dans le programme TINKER7, réalisée en collaboration avec F. Calvo (Grenoble), afin de coupler deux états diabatiques du système, chacun décrit par le champ de forces polarisable AMOEBA. Basée sur une description de calcul classique, cette approche pourrait, pour un coût très raisonnable, fournir de précieuses informations sur la dynamique d'échange. En outre, grâce au couplage avec le DACF, elle pourrait reproduire la signature caractéristique du transfert de proton au sein d'amino-acides ou de diacides déprotonés.

Nous avons choisi l'aspartate comme premier système d'application, la présence d'un transfert de proton ayant été démontrée par Oomens par spectroscopie IRMPD [55]. Le modèle est détaillé en **annexe B**, le cas de l'aspartate ayant fait l'objet d'un chapitre dans le livre "Theory and Applications of the Empirical Valence Bond Approach : From Physical Chemistry to Chemical Biology", sous presse, mais est résumée succinctement en section 1.3.1.

L'implémentation d'un modèle EVB est généralement liée à une longue phase de paramétrisation. Un de nos objectifs est de profiter de la description précise des interactions non covalentes par le champ de forces polarisable pour proposer un modèle de couplage relativement simple. De plus, toute la dynamique loin du transfert de proton étant gérée par AMOEBA, nous espérons une transférabilité partielle du modèle entre des systèmes similaires. Cette transférabilité sera étudiée en section 1.3.2 sur la série des diacides déprotonés $\text{HO}_2\text{C}-(\text{CH}_2)_n-\text{CO}_2^-$, $n = 1-6$.

3.3.1 Un modèle EVB à 2 états

L'énergie de l'état fondamental du système est décrite comme une combinaison linéaire de deux états diabatiques de valence ψ_i , dans lesquels le proton est lié de façon covalente à l'un ou l'autre des deux oxygènes impliqués dans l'échange.

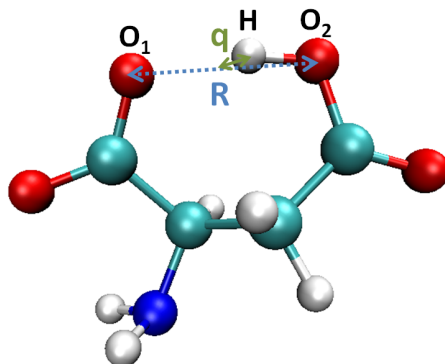
$$\Psi = g_1\psi_1 + g_2\psi_2 \quad (3.3)$$

Les potentiels V_i des états diabatiques sont décrits par AMOEBA. Le Hamiltonien, dans sa description matricielle 2×2 , fait apparaître un terme de couplage V_{12} autorisant le partage du proton entre les deux états. La plus petite valeur propre du Hamiltonien est donnée par :

$$\begin{aligned} V_{\text{EVB}} &= \frac{1}{2} \left[V_1 + V_2 - \sqrt{(V_1 - V_2)^2 + 4V_{12}^2} \right] \\ &= \frac{1}{2}(V_1 + V_2 - D), \\ \text{avec } D &= [(V_1 - V_2)^2 + 4V_{12}^2]^{1/2} \end{aligned} \quad (3.4)$$

et les vecteurs propres (g_1, g_2) prennent la forme :

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} [1 + (V_{\text{EVB}} - V_1)^2/V_{12}^2]^{-1/2} \\ (1 - g_1^2)^{1/2} \end{pmatrix} \quad (3.5)$$

FIGURE 3.11 – Définition des distances R et q sur la structure coudee de l'aspartate

La trajectoire sur la surface EVB est générée en propageant les équations du mouvement à partir du théorème de Hellmann-Feynman :

$$\frac{\partial V_{\text{EVB}}}{\partial x} = g_1^2 \frac{\partial V_1}{\partial x} + g_2^2 \frac{\partial V_2}{\partial x} + 2g_1g_2 \frac{\partial V_{12}}{\partial x}. \quad (3.6)$$

Le moment dipolaire $\vec{\mu}$ est obtenu par combinaison linéaire de ses valeurs sur les états diabatiques :

$$\vec{\mu}_{\text{EVB}} = g_1^2 \vec{\mu}_1 + g_2^2 \vec{\mu}_2. \quad (3.7)$$

Couplage des états diabatiques Le terme de couplage hors-diagonal V_{12} permet l'échange du proton par le mélange des deux états diabatiques. L'expression a été choisie volontairement simple afin d'en limiter la paramétrisation. Elle dépend de la distance R entre les deux sites [64–66] et de la distance q du proton au centre de la ligne séparant les sites [65, 66] (**Fig. 3.11**).

$$q = \left\| \vec{r}_{\text{H}} - \frac{1}{2}(\vec{R}_{\text{O}_1} + \vec{R}_{\text{O}_2}) \right\|.$$

La forme choisie pour le potentiel de couplage est inspirée de l'expression proposée par Vuilleumier et Borgis [65] pour l'eau protonée, mais lisse la décroissance de l'effet du couplage lorsque $q > 0.4 \text{ \AA}$.

$$V_{12} = A \frac{\exp(-\alpha R)}{1 + \gamma q^2}, \quad (3.8)$$

où A , α et γ sont 3 paramètres à ajuster.

La simplicité de l'expression sert deux objectifs : d'abord, on mise sur la précision d'AMOEBa pour décrire la plus grande partie du mouvement du proton, le rôle du potentiel de couplage se limitant à connecter les deux états de valence ; ensuite, la procédure de paramétrisation doit être la plus simple possible afin de faciliter la transférabilité.

3.3.2 Description d'un partage symétrique : la série des diacides déprotonés $\text{HO}_2\text{C}-(\text{CH}_2)_n-\text{CO}_2^-$

Plusieurs spectres IRMPD ont été enregistrés par E. Nicol à CLIO sur la série des diacides déprotonés $\text{HO}_2\text{C}-(\text{CH}_2)_n-\text{CO}_2^-$, pour $n = 1-4$ (**Fig 3.12**). Les échantillons sont injectés en électrospray dans la cellule de mesure du FT-ICR, à des concentrations en solution de 5.10^{-5} M dans un mélange $\text{H}_2\text{O}/\text{acétonitrile}$ 50/50, à l'exception de l'acide glutarique ($n = 3$), pour lequel le solvant est l'acétonitrile pur. Les spectres des acides adipique ($n = 4$) et glutarique ($n = 3$) présentent chacun une signature claire d'un proton partagé. La bande caractéristique, moins large que celle observée sur le spectre réalisé dans les mêmes conditions sur l'aspartate, est en accord avec un partage plus fort, facilité par la symétrie des acides. L'analogie avec les plus petits acides est plus difficile, l'acide succinique ($n =$

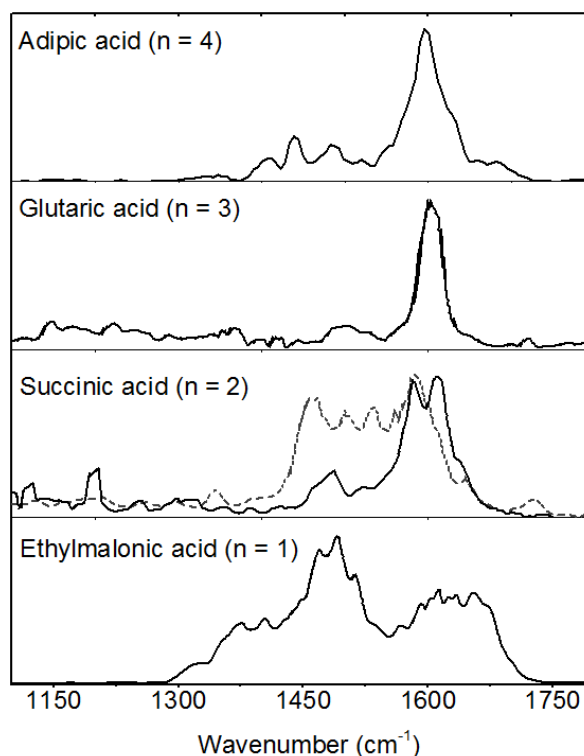


FIGURE 3.12 – Spectres IRMPD des diacides déprotonés $\text{HO}_2\text{C}-(\text{CH}_2)_n-\text{CO}_2^-$, $n = 1-4$, enregistrés à CLIO à température ambiante.

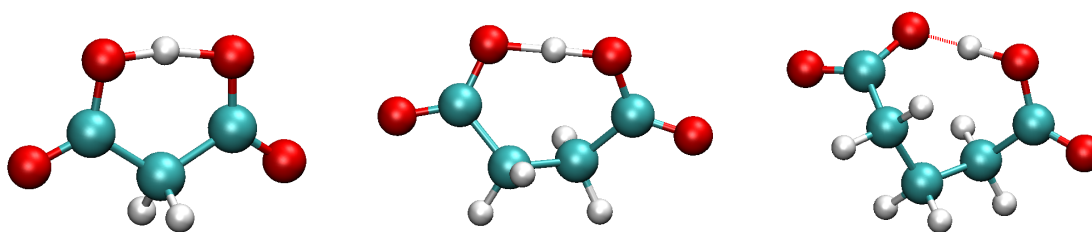


FIGURE 3.13 – Structures coudées des diacides déprotonés $\text{HO}_2\text{C}-(\text{CH}_2)_n-\text{CO}_2^-$, (de gauche à droite, $n = 1-3$), optimisées au niveau MP2/cc-pVTZ.

2) présentant une signature moins nette, et très dépendante des conditions d’irradiation (2 spectres sont présentés parmi les 4 enregistrés, et témoignent du manque de reproductibilité). Le spectre de l’acide éthylmalonique ($n = 1$) prend la forme d’un double massif étendu sur la gamme $1300-1750 \text{ cm}^{-1}$, qui pourrait témoigner d’un partage gêné par les contraintes géométriques (pseudo-cycle à 6 atomes qui défavorise la linéarité de l’angle \widehat{OHO}) (**Fig 3.13**).

Dans une approche EVB-AMOEBa, la symétrie des diacides face à l’échange de proton est bien accueillie car elle simplifie la paramétrisation d’AMOEBa. En effet, un même jeu de paramètres (et notamment de multipôles) permet de calculer les potentiels V_1 et V_2 des deux états diabatiques. De plus, la haute symétrie des petits diacides ($n = 1-3$) laisse prévoir une surface de potentiel symétrique face à l’échange de proton, et donc potentiellement plus simple à modéliser que celle de l’acide aspartique. Malgré cela, les étapes de calcul des surfaces d’énergie potentielle (SEP) et d’ajustement du potentiel de couplage restent coûteuses si elles doivent être appliquées à chaque système étudié. Pour les plus grandes tailles ($n = 3, 4$ et peut-être 2), les spectres sont assez similaires pour espérer que l’environnement autour du proton ne soit que faiblement influencé par la longueur de la chaîne carbonée. Dans ce cadre, la paramétrisation de la surface EVB pourrait être réalisée pour une seule taille, puis appliquée au reste de la série, dans les limites de la transférabilité du modèle.

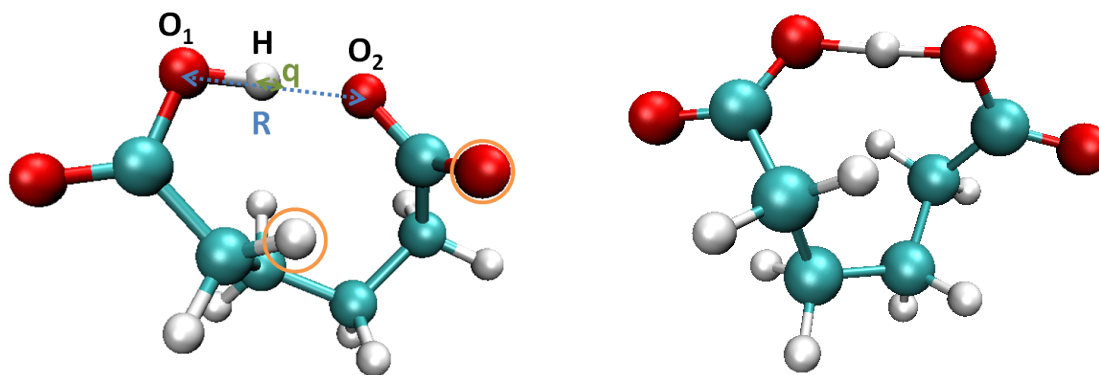


FIGURE 3.14 – Conformères de l’acide adipique déprotoné $\text{HO}_2\text{C}-(\text{CH}_2)_4-\text{CO}_2^-$ optimisés au niveau MP2/cc-pVDZ. A ce niveau de calcul, les deux conformères, l’un avec une liaison hydrogène $\text{CH}_2\text{—OCO}$ (conformère A, à gauche), et l’autre non (conformère B, à droite), sont isoénergétiques.

1.3.2.1 L’acide adipique $\text{HO}_2\text{C}-(\text{CH}_2)_4-\text{CO}_2^-$ comme système de référence

Dans cette série, l’acide adipique a été choisi comme système modèle, son spectre IRMPD bien résolu présentant la signature caractéristique d’un proton partagé (**Fig 3.15**). En outre, la longueur de sa chaîne carbonée doit lui apporter toute la flexibilité nécessaire à la bonne orientation des groupements carboxylates, facilitant le transfert du proton. A la différence des plus petits diacides, sa surface d’énergie potentielle peut ne pas être totalement symétrique face à l’échange de proton, dû à la présence de torsions, et la possibilité de former des liaisons hydrogènes secondaires de type $\text{HCH}\dots\text{OCO}$ (**Fig 3.14**, à gauche). De ce fait, un modèle paramétré sur ce système serait également plus facilement transférable aux tailles supérieures.

Calculs de chimie quantique Dans l’exemple précédent sur l’aspartate, les méthodes quantiques statiques ont échoué à reproduire la signature du partage de proton, ce qui a justifié d’inclure les effets dynamiques de transfert via le modèle EVB. La capacité de ces méthodes est de nouveau évaluée sur l’acide adipique, qui a été optimisé à différents niveaux de théorie, en vue de calculer les spectres statiques harmoniques. Quatre niveaux de théorie ont été choisis : DFT/M06/6-31++G(d,p), MP2/6-31++G(d,p), MP2/cc-pVDZ et MP2-cc-pVTZ. Au niveau M06, le minimum global (conformère A) présente une liaison hydrogène de type $\text{HCH}\dots\text{OCO}$ (**Fig 3.14**, à gauche). Aux niveaux MP2/6-31++G(d,p) et MP2/cc-pVDZ, une autre géométrie (conformère B) est isoénergétique (respectivement plus stable de 0.2 et 5.10^{-3} $\text{kJ}\cdot\text{mol}^{-1}$), pour laquelle cette interaction n’existe pas, et qui permet au proton d’être plus symétriquement partagé (**Fig 3.14**, à droite). Cette géométrie représente un minimum local en M06/6-31++G(d,p) ($+4.0$ $\text{kJ}\cdot\text{mol}^{-1}$). Au niveau MP2/cc-pVTZ, elle est plus favorable de 2.2 $\text{kJ}\cdot\text{mol}^{-1}$. Les géométries autour du proton sont données en **Tab 3.1**. Pour le conformère A, l’augmentation du niveau de théorie favorise un rapprochement des oxygènes O_1 et O_2 , un allongement de la liaison covalente O_1H et une diminution de la longueur de la liaison hydrogène. Pour le conformère B, les géométries MP2/cc-pVDZ et MP2/cc-pVTZ sont équivalentes, les deux niveaux de calcul étant peu différents et le proton déjà très fortement partagé au niveau MP2/cc-pVDZ.

Les spectres statiques dans l’approximation harmonique sont calculés à la géométrie d’équilibre obtenue avec chaque méthode (**Fig 3.15**). Au niveau MP2/cc-pVDZ, les deux conformères sont isoénergétiques donc les deux spectres sont proposés. Le premier constat est qu’aucun spectre statique ne reproduit le spectre expérimental. La description de la vibration est très dépendante de la géométrie de calcul. Pour le conformère A, les bandes les plus bleues (de haut en bas, respectivement à 1825, 1771 et 1823 cm^{-1}) sont attribuées à une combinaison de l’élongation $\text{O}=\text{COH}$ et du pliage COH . La seconde bande (respectivement à 1712, 1654 et 1728 cm^{-1}) correspond à

TABLE 3.1 – Coordonnées internes impliquées dans le transfert de proton aux géométries optimisées avec différents niveaux de théorie, impliquant DFT/M06 ou MP2. Au niveau MP2, deux géométries sont compétitives.

	conformère A				conformère B		
	M06/b1	MP2/b1	MP2/b2	MP2/b3	MP2/b1	MP2/b2	MP2/b3
R_{OO}	2.617	2.578	2.565	2.546	2.432	2.408	2.403
$R_{O...H}$	1.609	1.564	1.550	1.531	1.325	1.241	1.239
r_{OH}	1.011	1.018	1.021	1.021	1.109	1.167	1.164

b1 : 6-31++G(d,p), b2 : cc-pVDZ, b3 : cc-pVTZ

la combinaison antisymétrique O=C=O. A la géométrie du conformère B, le proton est tellement partagé qu'apparaît une double bande, à 1787 et 1793 cm^{-1} en MP2/cc-pVDZ, 1749 et 1758 cm^{-1} en MP2/cc-pVTZ, de type O=COH, pour chacun des groupements carboxyliques. De même, les deux bandes, respectivement à 1552 et 1633 cm^{-1} , et 1527 et 1614 cm^{-1} , sont attribuées au pliage COH, avec l'angle défini à partir de l'un ou l'autre des carboxylates. Toutefois, aucune de ces descriptions ne permet de reproduire le spectre, mais confirme que reproduire une géométrie d'équilibre ne suffit pas, et qu'il est nécessaire d'inclure des effets dynamiques.

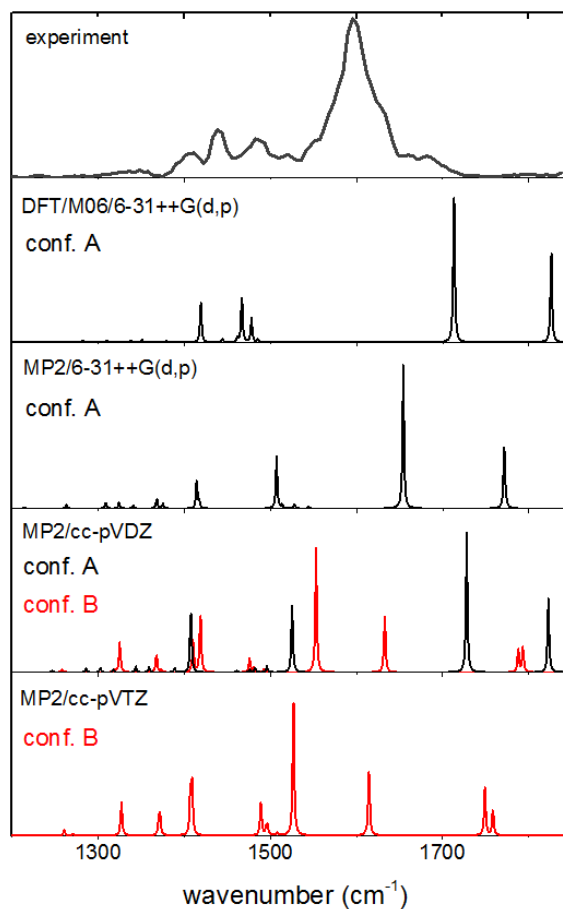


FIGURE 3.15 – Spectre IRMPD de l'acide adipique enregistré à CLIO et prédit par des calculs statiques de chimie quantique, dans l'approximation harmonique, en utilisant différentes méthodes, DFT/M06/6-31++G(d,p), MP2/6-31++G(d,p), MP2/cc-pVDZ, MP2/cc-pVTZ, et deux conformères stables A et B. Aucun facteur d'échelle n'a été appliqué sur les fréquences.

Spectres IR AMOEBA à température finie Un moyen d'inclure les effets dynamiques d'anharmonicité, souvent à l'origine d'un élargissement et d'un déplacement de bandes, est de réaliser une simulation de dynamique

moléculaire. Le champ de forces polarisable AMOEBA a souvent montré son efficacité dans de telles descriptions [3, 5–7], même s’il est par construction incapable de décrire le partage de proton de l’aspartate dans une approche classique. Tous les paramètres du champ de forces sont issus de la version AMOEBAbio-09 [67] à l’exception des multipôles, qui doivent être recalculés. Le conformère A est choisi comme géométrie d’extraction pour 3 raisons. D’abord, le conformère B n’a été identifié que lorsque de très courtes distances entre les oxygènes O_1 et O_2 étaient permises ($\sim 2.4 \text{ \AA}$). Dans une approche de dynamique, les fluctuations de géométrie vont probablement engendrer des structures pour lesquelles le caractère stabilisant de la liaison hydrogène HCH...OCO, présente dans le conformère A, va être important. Ensuite, concernant la procédure d’extraction des multipôles par DMA, puis leur ajustement sur le potentiel électrostatique, Wu [68] a relevé les incertitudes de la procédure dans le cas de très fortes liaisons hydrogènes et de zones de fortes interactions de van der Waals (vdW). La géométrie du conformère B est une illustration parfaite de la situation à éviter. Enfin, une géométrie optimisée d’une liaison hydrogène AMOEBA est un équilibre entre, d’une part, les interactions électrostatiques, attractives, et d’autre part, le vdW, répulsif. De façon peut-être un peu contre-intuitive, plus le proton est partagé dans la géométrie d’extraction, plus les charges extraites sont homogènes entre les deux groupements carboxylique/carboxylate, ce qui conduit à un affaiblissement de la liaison hydrogène lorsqu’elle est calculée par un potentiel AMOEBA. Un jeu de multipôles a donc été extrait par la procédure DMA appliquée à la fonction d’onde MP2/aug-cc-pVTZ sur la géométrie du conformère A, optimisée au niveau MP2/cc-pVDZ. A l’issue, les dipôles et quadrupôles ont été ajustés sur le potentiel électrostatique au même niveau de calcul, en conservant les charges fixes. Dans la procédure d’optimisation, la densité électronique est évaluée autour de chaque atome avec un rayon fixe $r = 0.65 \text{ \AA}$ pour tous les atomes à l’exception des hydrogènes, où la valeur recommandée de $r_H = 0.325 \text{ \AA}$ a été choisie.

Un spectre DACF a été enregistré au cours d’une dynamique AMOEBA à 300 K (**Fig 3.16**). Dans le formalisme AMOEBA, et à la différence des calculs quantiques, les deux interactions de l’hydrogène sont décrites de façon fondamentalement non équivalentes, puisque l’une est traitée par des termes intramoléculaires, et l’autre par des interactions non liées. La géométrie optimisée AMOEBA est caractérisée par $R^{AMOEBA} = 2.73 \text{ \AA}$ et $r_{OH}^{AMOEBA} = 0.99 \text{ \AA}$, indiquant une liaison hydrogène nettement plus faible que celle calculée par les calculs quantiques. La position des bandes est d’ailleurs en excellent accord avec les fréquences statiques M06/6-31++G(d,p), la méthode qui décrit le plus faiblement la liaison hydrogène dans notre sélection. Ces spectres soulignent l’importance d’une description équivalente des deux sites d’appartenance du proton, et du transfert entre ces sites, et justifient le recours au modèle EVB.

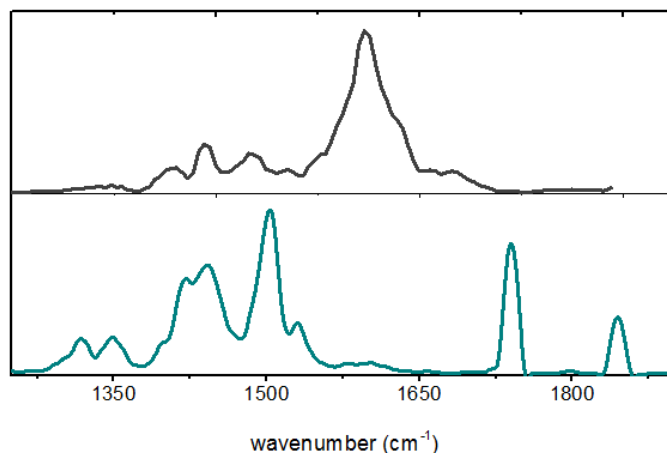


FIGURE 3.16 – Spectres DACF-AMOEBA de l’acide adipique à 300 K (bleu) et exp. enregistré à CLIO (gris).

Représentation de la surface *ab initio* La première étape de paramétrisation du modèle EVB consiste à construire une surface d'énergie potentielle *ab initio*. Nous avons choisi le niveau MP2/cc-pVTZ, et une gamme de distances R et q compatible avec le transfert de proton.

Il n'est pas immédiat de déterminer si la rotation autour d'une torsion, ou la présence d'une liaison hydrogène secondaire de type HCH...OCO, va influencer de façon notable sur le transfert de proton. Dans ces conditions, déterminer l'état de transition, i.e la configuration d'énergie maximale le long de la coordonnée de transfert de proton, peut se révéler délicat. Sans cette information, deux voies semblent éligibles pour la construction de la SEP. Le point de départ commun est l'optimisation de 6 géométries au niveau MP2/cc-pVTZ en contraignant R , tous les 0.1 Å sur la gamme $2.4 \text{ \AA} \leq R \leq 2.9 \text{ \AA}$.

- Le scan rigide consiste, pour chacune des 6 distances R , à geler l'ensemble de la géométrie et à déplacer pas à pas le proton le long de la coordonnée d'échange. Le scan est rapide, mais l'erreur sur les énergies augmente à mesure que l'on s'éloigne de la géométrie optimisée. Cette approche ne se justifierait que par une vitesse de transfert du proton trop grande pour permettre une relaxation de la molécule. Les simulations EVB sur l'aspartate ont montré qu'un échange peut se produire sur plusieurs femtosecondes, soit quelques dizaines de pas de simulation, ce qui semble suffisant pour certaines réorganisations.
- La seconde méthode consiste en une optimisation sous contrainte pour chaque point de la SEP. Les coordonnées internes concernées sont la distance R , la distance r_{O_2H} , et l'angle $\widehat{O_2O_1H}$, que l'on fixe à 2° , la totale linéarité n'étant pas permise par l'algorithme d'optimisation sous contrainte de GAUSSIAN. Le reste de la géométrie est relaxé au niveau MP2/cc-pVTZ. A ce niveau de calcul, la géométrie d'équilibre est caractérisée par $R^{MP2} = 2.56 \text{ \AA}$ et la distance la plus courte entre le proton et un oxygène est $r_{OH}^{MP2} = 1.12 \text{ \AA}$. En se déplaçant le long de la coordonnée R , les minima locaux MP2/cc-pVTZ augmentent en moyenne de $0.75 \text{ kcal.mol}^{-1}$ par pas de $dR = 0.1 \text{ \AA}$.

Nous avons opté pour l'optimisation sous contrainte, qui pour la gamme de distance $2.5 \text{ \AA} \leq R \leq 2.9 \text{ \AA}$, conduit à une surface de potentiel symétrique face à l'échange proton. Pour toutes ces distances R , l'optimisation MP2/cc-pVTZ conduit toujours à des géométries avec une interaction stabilisante HCH...OCO, qui forment un pseudo cycle à 8 centres (**Fig 3.14**, à gauche), permis par la rotation du groupement carboxylate (type conformère A). Pour $R = 2.4 \text{ \AA}$, la compétition entre les géométries de type A et B rend difficile l'évaluation de la SEP MP2 de la géométrie du conformère A par cette méthode. Pour cette taille, nous aurions pu lui préférer un scan rigide, toutefois, il a pu être évalué que, en conservant l'interaction HOH...OCO, le minimum local à $R = 2.4 \text{ \AA}$ est $\sim 2 \text{ kcal.mol}^{-1}$ moins stable que le minimum global (à $R = 2.56 \text{ \AA}$), du même ordre de grandeur que le minimum local de $R = 2.8 \text{ \AA}$. Nous montrerons dans la suite que ces distances ne sont pas ou peu échantillonnées au cours de la dynamique EVB-AMOEBA.

Procédure de paramétrisation pour l'ajustement du potentiel EVB sur le potentiel *ab initio*

Le potentiel EVB doit pouvoir reproduire, d'une part, la géométrie du minimum global et les énergies relatives des minima locaux, et d'autre part, les barrières de potentiel pour chaque distance R . Pour ce dernier point, il est attendu que la symétrie des diacides favorise les situations de fort partage du proton ($q \sim 0$), bien plus que dans le cas de l'aspartate, ce qui conduit à réévaluer le rôle du potentiel de couplage.

Aux géométries des minima, le proton est localisé sur un site, et la description de l'énergie EVB est très majoritairement décrite par l'énergie AMOEBA de l'état diabatique concerné. AMOEBA échoue à reproduire la géométrie optimisée MP2, de même que les énergies relatives des minima locaux. La géométrie optimisée AMOEBA est caractérisée par $R^{AMOEBA} = 2.73 \text{ \AA}$ et $r_{OH}^{AMOEBA} = 0.99 \text{ \AA}$, indiquant une liaison hydrogène nettement plus faible qu'en MP2/cc-pVTZ. En se déplaçant le long de la coordonnée R , on observe une faible augmentation des énergies des minima locaux AMOEBA, de l'ordre de $0.3\text{-}0.4 \text{ kcal.mol}^{-1}$ par pas de $dR = 0.1 \text{ \AA}$, en augmentant R , et à l'inverse, une forte augmentation pour les plus courtes distances R . Dans le formalisme d'AMOEBA, la longueur

d'une liaison hydrogène est déterminée par les interactions électrostatiques et le terme de vdW. L'équilibre des deux termes permet de décrire correctement les longueurs de liaisons hydrogènes pour des interactions standard, mais le champ de forces est dans l'incapacité de décrire des liaisons hydrogènes beaucoup plus courtes, la composante répulsive du terme de vdW augmentant en $1/r^{14}$. Pour améliorer la description des liaisons hydrogènes fortes, il faudrait implémenter un terme de pénétration de charge, stabilisant à courte distance, afin de mimer le recouvrement des densités électroniques. Ce terme a été implémenté dans SIBFA comme un développement sur la distribution multipolaire [69–72]. Selon Ponder [73], la pénétration de charge se révèle être un terme manquant important dans AMOEBA, découlant de l'utilisation des multipôles ponctuels. Sa prise en compte améliore nettement l'accord entre les énergies électrostatiques AMOEBA et SAPT. Il pourrait apparaître dans une prochaine version AMOEBA 2.0. En attendant, nous proposons de traiter le déséquilibre des interactions non liées pour le cas spécifique du partage de proton, en reparamétrant le potentiel de vdW. En effet, l'électrostatique fixe la directionnalité de la liaison hydrogène et inclut des effets à *n-corps* qui le rendent délicat à manipuler. À l'inverse, les contributions de vdW sont additives et il est généralement admis que ce terme a en charge d'inclure toutes les interactions qui ne sont pas traduites explicitement.

On rappelle le formalisme 'buffered 14-7' proposé par Thomas A. Halgren [74], et utilisé dans AMOEBA pour rendre compte des effets de répulsion d'échange et de dispersion :

$$E_{vdW}(ij) = \epsilon_{ij} \left(\frac{1.07}{\rho_{ij} + 0.07} \right)^7 \left(\frac{1.12}{\rho_{ij}^7 + 0.12} - 2 \right) \quad (3.9)$$

Le potentiel est fonction de la distance de séparation R_{ij} entre les atomes i et j au travers du terme $\rho_{ij} = R_{ij}/R_{ij}^0$, avec R_{ij}^0 la distance pour laquelle l'énergie est minimale, distance qui devient $R_{ij}^0 = \frac{(R_{ii}^0)^3 + (R_{jj}^0)^3}{(R_{ii}^0)^2 + (R_{jj}^0)^2}$ lors d'une combinaison sur une paire d'atomes différents. Dans ce cas, la profondeur du puits est donnée par $\epsilon_{ij} = \frac{4\epsilon_{ii}\epsilon_{ij}}{(\sqrt{\epsilon_{ii}} + \sqrt{\epsilon_{ij}})^2}$.

La paramétrisation vise à reproduire la géométrie d'équilibre MP2/cc-pVDZ autour de la liaison hydrogène (i.e $R = 2.56$ Å, $R_{O\dots H} = 1.55$ Å et $r_{OH}=1.02$ Å). Les 4 paramètres ajustables (si on conserve le formalisme 'buffered 14-7') sont les puits ϵ_{HH} et ϵ_{OO} , et les rayons R_{HH}^0 et R_{OO}^0 . Cherchant à rééquilibrer spécifiquement l'interaction $O\dots H$, l'approche la moins perturbative serait de n'appliquer les corrections qu'au calcul de cette interaction spécifiquement. Toutefois, les outils de décomposition de l'énergie de TINKER montrent que les autres interactions de vdW de ces atomes sont très faibles en comparaison. Les combinaisons d'ajustement de paramètres pour reproduire la géométrie MP2 sont nombreuses, nous en avons choisi deux, l'une impliquant la diminution conjointe des deux puits ϵ_{HH} et ϵ_{OO} de -44 %, l'autre celle des rayons de référence R_{HH}^0 et R_{OO}^0 de -9 % (**Tab 3.2**). On note que, si l'environnement immédiat de la liaison hydrogène du conformère A est reproduit, la géométrie AMOEBA (**Fig 3.17**, à droite) ne fait pas apparaître de liaison hydrogène HCH...OCO, les contraintes dues aux termes de vdW, à l'électrostatique entre les oxygènes O_1 et O_3 , et à la torsion, n'étant pas compensées par la stabilisation offerte par cette liaison hydrogène faible. Toutefois, cette paramétrisation permet, au travers de l'une ou l'autre des corrections de vdW, de reproduire les énergies relatives des minima locaux en fonction de R avec une précision de l'ordre de la kcal.mol⁻¹ (**Fig 3.18**).

AMOEBA ayant été paramétré pour reproduire les énergies relatives des minima, il reste à bien traduire celles des états de transition, rôle attribué au potentiel de couplage. Malgré la forme choisie (**eq. 1.8**), V_{12} a un effet assez local, autour de $q=0$. Plus les distances R sont grandes, plus les énergies à $q \sim 0$ sont surestimées en AMOEBA et demandent une forte compensation du potentiel de couplage pour ajuster la surface EVB sur l'énergie MP2. Dans sa forme actuelle, V_{12} n'a pas la flexibilité nécessaire pour reproduire finement les barrières à chaque distance R . Nous lui introduisons une dépendance en R au travers de son paramètre A , donc l'expression devient : $A(R) = 2000 \times R^2 - 10240 \times R + 13345$, permettant de reproduire les barrières de potentiel sur la gamme $2.5 \text{ \AA} \leq R$

TABLE 3.2 – Coordonnées internes impliquées dans le transfert de proton à la géométrie optimisée MP2/cc-pVDZ, reproduite par optimisation AMOEBA en faisant varier les paramètres de vdW des atomes O et H impliqués dans la liaison hydrogène.

	MP2	- 9 % R_{ii}^0	- 44 % ϵ_{ii}
R_{OO}	2.565	2.551	2.556
$R_{O\dots H}$	1.550	1.555	1.559
r_{OH}	1.021	0.996	0.998

i = O ou H impliqués dans la liaison hydrogène

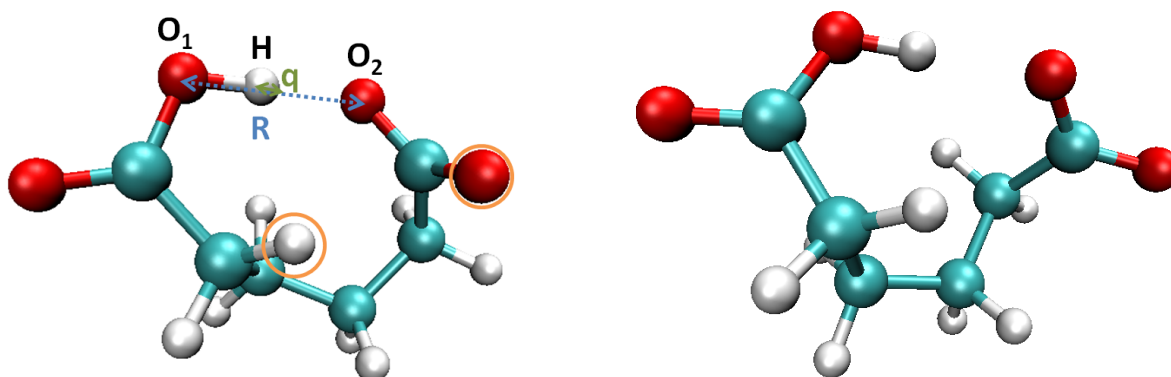


FIGURE 3.17 – Structure coudée de l’acide adipique déprotoné $\text{HO}_2\text{C}-(\text{CH}_2)_4-\text{CO}_2^-$. A gauche, géométrie optimisée en MP2/cc-pVDZ, et à droite, avec le champ de forces AMOEBA.

$\leq 2.9 \text{ \AA}$ (Fig 3.18). Les deux autres paramètres restent identiques à ceux utilisés pour l’aspartate, $\alpha = 0.745 \text{ \AA}^{-1}$ et $\gamma = 5.35 \text{ \AA}^{-2}$. Les deux propositions de correction du terme de vdW apparaissent équivalentes et permettent de reproduire la surface MP2 avec un excellent accord.

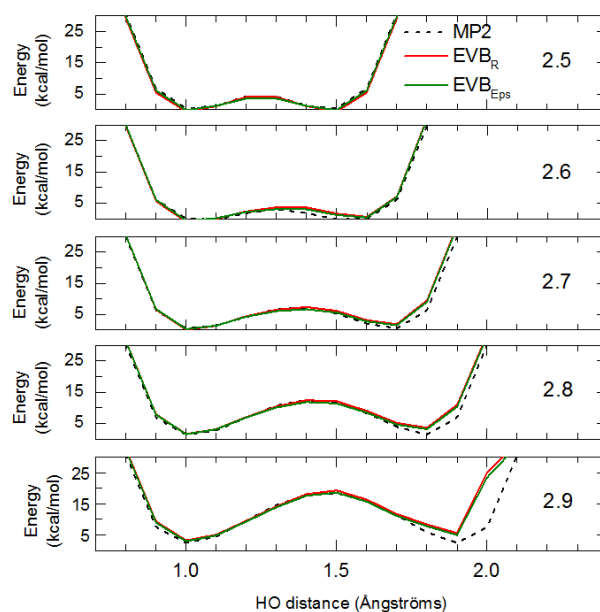


FIGURE 3.18 – Surfaces d’énergie potentielle EVB de l’acide adipique comparées à la SEP MP2/cc-pVTZ. Pour la SEP EVB_R , les rayons de vdW des O et H impliqués dans la liaison hydrogène ont été diminués de -9 %, pour la SEP EVB_{Eps} , leurs puits de vdW ont été diminués de -44 %.

Suivi de la dynamique d'échange de proton par simulations EVB : Les simulations EVB sont réalisées dans l'ensemble NVT. Chaque propriété présentée (spectres, distributions de distances) est une moyenne sur 10 trajectoires de 50 ps. Sans description explicite du potentiel EVB pour $R = 2.4 \text{ \AA}$, nous nous assurons que cette distance est peu échantillonnée au cours des simulations EVB (**Fig 3.19**, à gauche). Elle n'est jamais atteinte à $T = 50 \text{ K}$, et très rarement à 300 K . On peut suivre également cette distance directement pendant une simulation EVB-AMOEBa à 100 K (**Fig 3.20**, en bleu). Par ailleurs, la plupart des échanges de proton ont lieu pour $2.45 \text{ \AA} \leq R \leq 2.50 \text{ \AA}$, ce qui est un peu en dessous de la distance d'équilibre MP2 mais facilement justifiable par les écarts entre la surface EVB calculée (basée sur les géométries relaxées MP2) et la surface réellement explorée pendant la simulation EVB (en absence de liaison hydrogène HOH...OCO). De manière générale, la température, dans la gamme 50-300 K, conduit à un élargissement faible de la distribution des distances R échantillonnées au cours des dynamiques EVB. Le proton, lui, est d'autant plus facilement partagé que la température est basse (**Fig 3.19**, à droite), en accord avec la distribution de R , les barrières étant plus difficiles à passer lorsque R s'éloigne de sa valeur de référence. On observe une très grande symétrie du système par rapport à l'échange de proton, et des barrières faciles à passer, puisque à 100 K , le proton est échangé d'un site à l'autre plus d'une fois toutes les 500 fs (**Fig 3.20**, en haut).

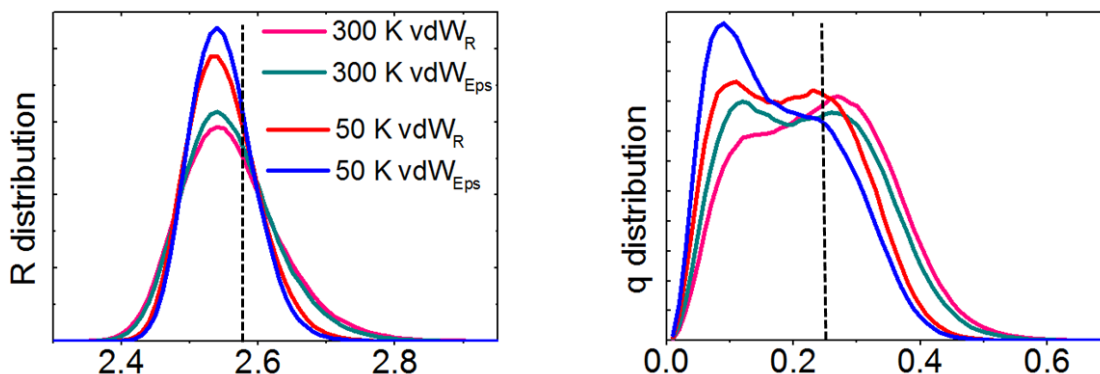


FIGURE 3.19 – Distribution des distances R et q , moyennées sur 10 simulations de 50 ps chacune, pour 2 températures 50 K et 300 K, et pour 2 corrections apportées au terme de vdW. En trait pointillé noir sont positionnées les valeurs de R et q correspondant à la géométrie d'équilibre MP2.

Spectres EVB-AMOEBa En accord avec les faibles variations de géométrie, les spectres dynamiques EVB évoluent peu avec la température (**Fig 3.21**). Le massif s'élargit vers le bleu avec l'augmentation de la température, qui crée des situations de plus faible partage. Le maximum principal est idéalement positionné et est peu affecté par la température. Toutefois, les fluctuations de géométries apportées par la température, élargissent nettement les spectres calculés à 200 et 300 K vers le bleu. A 50 K, la largeur de la bande expérimentale est très fidèlement reproduite. Le reste du spectre est en bon accord, avec un massif de 3 bandes sur la gamme $1400-1500 \text{ cm}^{-1}$ et une bande de très faible intensité entre 1300 et 1350 cm^{-1} .

1.3.2.2 Transférabilité vers les petits acides $\text{HO}_2\text{C}-(\text{CH}_2)_n-\text{CO}_2^-$, ($n = 2-3$)

L'acide glutarique Le spectre enregistré à CLIO sur l'acide glutarique $\text{HO}_2\text{C}-(\text{CH}_2)_3-\text{CO}_2^-$ (**Fig 3.23**, en haut)) présente une bande intense à 1600 cm^{-1} , à la même position que celle de l'acide adipique, bien que plus fine, signature claire d'un transfert de proton. Il laisse deviner une contribution supplémentaire vers 1500 cm^{-1} , là où l'acide adipique faisait apparaître 3 bandes peu intenses entre 1400 et 1500 cm^{-1} .

On souhaite tester la transférabilité du modèle EVB à l'acide glutarique, en évitant les étapes de calcul de la SEP et d'ajustement du potentiel de couplage. Deux approches sont envisagées :

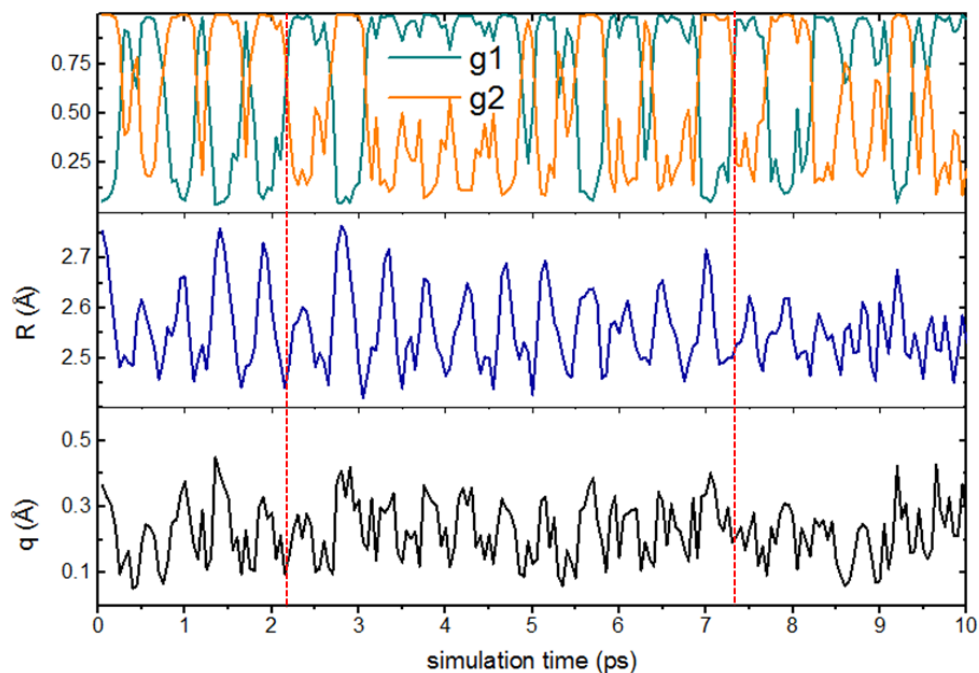


FIGURE 3.20 – Simulation EVB-AMOEBa de 10 ps à 100 K, et évolution des vecteurs propres EVB g_1 (vert) et g_2 (orange), et des distances R (bleu) et q (noir) (en Å). Deux repères (traits pontillés rouges) signalent des exemples d'échanges de proton.

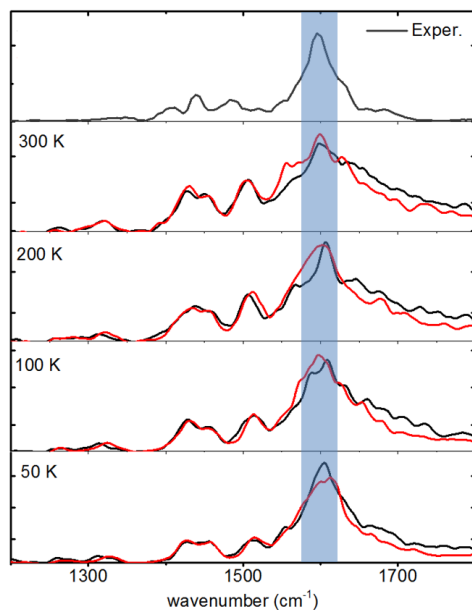


FIGURE 3.21 – Spectres EVB-AMOEBa en fonction de la température entre 50 et 300 K, pour les deux jeux de correction vdW_R (noir) et vdW_{Eps} (rouge), et comparés au spectre IRMPD.

- La similarité des signatures des acides adipique et glutarique laisse supposer un comportement proche et leurs différences peuvent être dues aux géométries d'équilibre. L'optimisation de l'acide glutarique au niveau MP2/cc-pVTZ (Fig 3.13, à droite) conduit à une géométrie pour laquelle le proton est fortement partagé, $R = 2.428 \text{ \AA}$, $r_{OH} = 1.125 \text{ \AA}$ et $r_{O...H} = 1.305 \text{ \AA}$, l'angle $\widehat{O_1O_2H}$ étant de 0.16° . A cette géométrie, un jeu de multipôles est extrait au niveau MP2/aug-cc-pVTZ, suivi d'un ajustement des dipôles et quadrupôles sur le potentiel électrostatique, au même niveau de théorie. A l'issue, les paramètres de vdW sont ajustés pour

TABLE 3.3 – Coordonnées internes impliquées dans le transfert de proton à la géométrie optimisée MP2/cc-pVTZ, reproduite par optimisation AMOEBA en faisant varier les paramètres de vdW des atomes O et H impliqués dans la liaison hydrogène.

geom C ₅ O ₄ H ₇	MP2	- 25.5 % R _{ii} ⁰	- 82 % ε _{ii}
R _{OO}	2.428	2.426	2.426
R _{O...H}	1.305	1.424	1.421
r _{OH}	1.125	1.003	1.005
geom C ₆ O ₄ H ₇	MP2	- 20.5 % R _{ii} ⁰	- 76.5 % ε _{ii}
R _{OO}	2.565	2.568	2.564
R _{O...H}	1.550	1.581	1.574
r _{OH}	1.021	0.992	0.994

i = O ou H impliqués dans la liaison hydrogène

reproduire la géométrie optimisée MP2/cc-pVTZ (**Tab 3.3**). L’ajustement est plus important que pour l’acide adipique, impliquant une diminution de -25.5 % des rayons de référence R_{HH}⁰ et R_{OO}⁰, ou -82 % des puits ε_{HH} et ε_{OO}. A ce prix, la distance R est reproduite, mais pas les deux longueurs r_{OH} et $r_{O...H}$, cette fois à cause d’un manque de flexibilité du potentiel d’élongation. Les simulations EVB à 300 K (**Fig 3.22**, courbes noires) échantillonnent les distances R autour de la géométrie d’équilibre MP2. A ces courtes distances, le proton est évidemment très fortement partagé. De telles distances sortent des limites de transférabilité du modèle, qui impliquent d’échantillonner des géométries assez semblables à celles de l’acide adipique, et pour lesquelles le champ de forces AMOEBA ne joue pas un rôle secondaire face au potentiel de couplage.

- Pour reproduire le comportement de transfert du proton observé sur l’acide adipique, il faut se déplacer sur une SEP similaire. Une optimisation sous contrainte de l’acide glutarique est réalisée au niveau MP2/cc-pVTZ, en fixant la géométrie locale autour du proton aux valeurs optimisées de l’acide adipique, soit $R = 2.565$ Å, $r_{OH} = 1.021$ Å et $\widehat{O_2O_1H}$ à 2°. Un jeu de multipôles est extrait à cette géométrie (mêmes conditions que précédemment), puis les corrections de vdW sont appliquées afin de reproduire cette géométrie locale en AMOEBA. La correction à appliquer, -20.5 % sur les rayons de référence R_{HH}⁰ et R_{OO}⁰, ou -76.5 % sur les puits ε_{HH} et ε_{OO} (**Tab 3.3**), est logiquement moins importante que pour obtenir la géométrie optimisée de l’acide glutarique, mais plus que dans le cas de l’acide adipique, la géométrie étant plus contrainte. Notamment, un hydrogène aliphatique est distant de 2.51 Å de l’oxygène du carboxylate, très en dessous de la distance d’équilibre de la contribution de vdW (3.42 Å avant correction sur R_{OO}⁰ et 2.96 Å avec). Un cas similaire apparaît par symétrie pour l’oxygène OH.

Des simulations EVB entre 100 et 300 K ont été réalisées, à partir des deux jeux de corrections du vdW. Les distributions de distances (**Fig 3.22**) font apparaître que la distance d’équilibre R est plus courte que pour l’acide adipique, et que les fluctuations de géométrie sont moins importantes. A 300 K, R est comprise entre 2.4-2.6 Å, maximale pour $R = 2.48$ Å, géométrie locale intermédiaire entre les minima globaux des acides adipique et glutarique. L’explication est apportée par les corrections de vdW. En effet, pour ce système, il n’y a plus d’équivalence entre les deux corrections de vdW, du fait de la proximité des CH₂. En appliquant les corrections sur les puits (vdW_{Eps}), le proton partagé devient assez insensible aux interactions extérieures et ne s’écarte pas de la zone de partage symétrique. A l’inverse, avec la correction sur les distances, le proton peut être momentanément repoussé par un CH₂ et forcé de se déporter sur un site. Cette gêne occasionne également un élargissement des distances R , faible à 300 K, mais très net à 100 K, avec deux situations, un partage à courte distance ($R \sim 2.48$ Å) lorsqu’aucun CH₂ ne vient repousser le proton, et à plus longue distance ($R \sim 2.55$ Å) en cas d’interaction. L’allongement de la distance R la rapproche de la distance observée pour l’acide adipique. Celle-ci implique une augmentation des distances OH et O...H, et contribue à déplacer la bande vers le rouge. Ce sont donc les interactions

extérieures (torsions, liaisons hydrogènes secondaires) qui fixent en partie la distribution des distances R , et ces interactions sont moins destabilisantes pour l'acide glutarique.

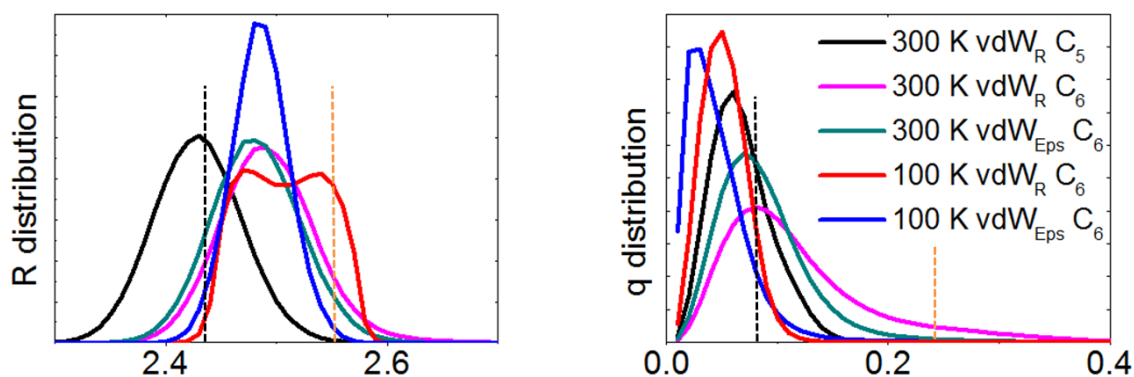


FIGURE 3.22 – Distribution des distances R et q , moyennées sur 10 simulations de 50 ps chacune. En trait pointillé sont positionnées les valeurs de R et q correspondant aux géométries d'équilibre MP2 de l'acide glutarique (en noir) et de l'acide adipique (en orange). Les notations C₅ et C₆ désignent les géométries MP2 de référence (acides glutarique ou adipique) qui ont servies de cibles à l'ajustement des paramètres de vdW.

Sur les spectres (**Fig 3.23**), la plus grande mobilité du proton avec le jeu de corrections vdW_R est répercutée en termes de largeur de bande et de déplacement vers le rouge, beaucoup moins bien reproduite par les corrections vdW_{Eps}. Le meilleur accord est obtenu à 300 K, où la bande calculée est déplacée de 15 cm⁻¹ dans le bleu.

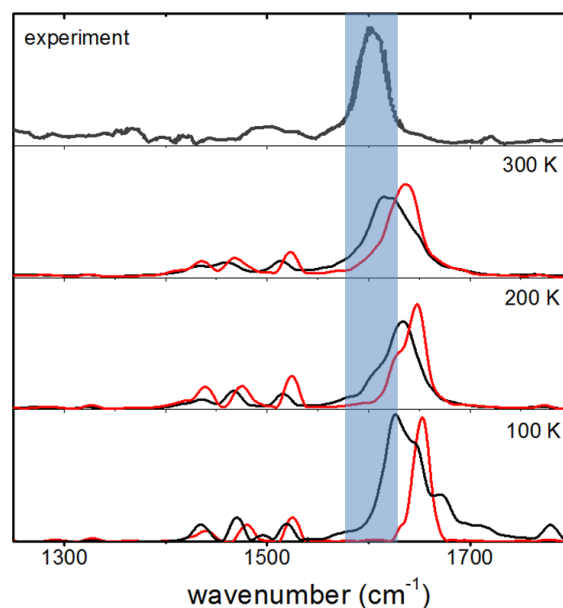


FIGURE 3.23 – Spectres EVB-AMOEBA en fonction de la température entre 100 et 300 K, pour les deux jeux de correction vdW_R (noir) et vdW_{Eps} (rouge), enregistrés et comparés au spectre IRMPD.

L'acide succinique (n = 2) La transférabilité vers l'acide glutarique a été permise par la présence des CH₂, qui gênant le proton à l'équilibre, favorisaient une augmentation des distances R (avec toutefois un effet moindre que pour l'acide adipique). Pour HO₂C-(CH₂)₂-CO₂⁻, la molécule forme un pseudo-cycle à 7 côtés, imperturbable (**Fig 3.13, au centre**). Lorsque les mêmes conditions que précédemment lui sont appliquées (optimisation contrainte, extraction de multipôles et corrections de vdW), les dynamiques EVB conduisent à une situation de très fort partage.

Les distances d'équilibres échantillonnées, autour de 2.4-2.45 Å sont hors de la zone d'ajustement du modèle. Pour reproduire la signature des petits acides, la procédure complète devrait être envisagée.

1.3.2.3 Conclusion et perspectives : transférabilité vers $\text{HO}_2\text{C}-(\text{CH}_2)_6-\text{CO}_2^-$

A l'inverse, il semble possible de reproduire la signature de plus gros acides, tels que $\text{HO}_2\text{C}-(\text{CH}_2)_6-\text{CO}_2^-$, puisque la structure présente des contraintes similaires à l'acide adipique (Fig 3.24). La même procédure est appliquée, en se limitant aux corrections vdW_R , puisqu'il a été observé que les corrections vdW_{Eps} pouvaient rendre le proton insensible à une partie de son environnement. Au cours des simulations EVB, la distribution des distances R échantillonnées (Fig 3.25) est très similaire à celle de l'acide adipique (entre 2.4 et 2.7 Å à 50 K, entre 2.4 et 2.8 Å à 300 K). Le proton alterne des situations de fort partage, privilégiées à basse température, et des périodes pendant lesquelles il est localisé sur un site, ce qui semble en accord avec l'intuition chimique que les fluctuations au sein des longues chaînes peuvent augmenter la fréquence de rupture de la liaison hydrogène. Les spectres EVB-AMOEBA en température révèlent tous ces comportements. À 50 K, la bande caractéristique autour de 1600 cm^{-1} témoigne du transfert de proton, dans des conditions très similaires à celles de l'acide adipique. L'apparition d'une bande vers 1800 cm^{-1} est attribuée à l'élongation antisymétrique COO^- en cas de liaison hydrogène plus faible. La signature du transfert de proton s'atténue à mesure que la température et les distances R échantillonnées augmentent.

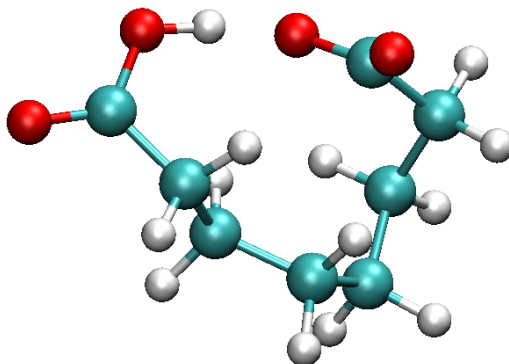


FIGURE 3.24 – Géométrie optimisée de $\text{HO}_2\text{C}-(\text{CH}_2)_6-\text{CO}_2^-$ au niveau MP2/cc-pVTZ

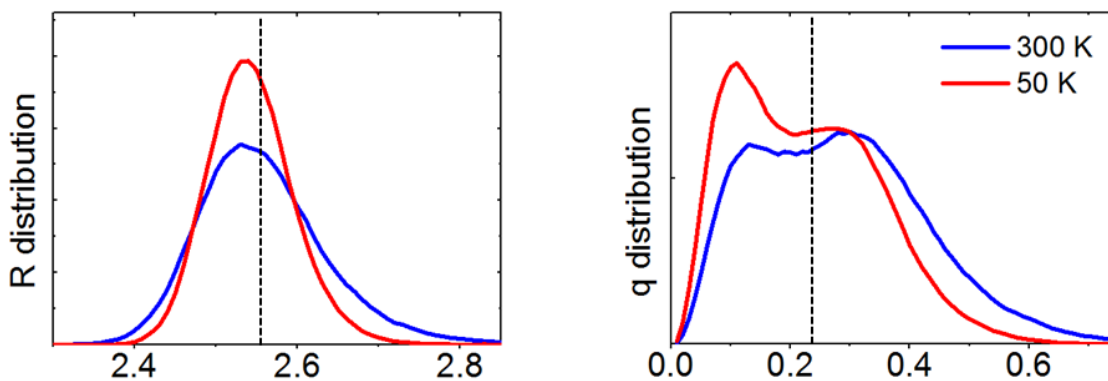


FIGURE 3.25 – Distribution des distances R et q , moyennées sur 10 simulations de 50 ps chacune. En trait pointillé noir sont positionnées les valeurs de R et q correspondant aux géométries d'équilibre MP2 de l'acide glutarique (en noir) et de l'acide adipique (en orange).

Le modèle EVB-AMOEBA se révèle capable de traduire le partage au sein d'amino-acides ou de diacides déprotonés. Couplé à la dynamique moléculaire, de longues simulations sont possibles, permettant d'avoir accès à toute la statistique de la dynamique d'échange. Grâce au DACF, le modèle peut reproduire la signature IR caractéristique de

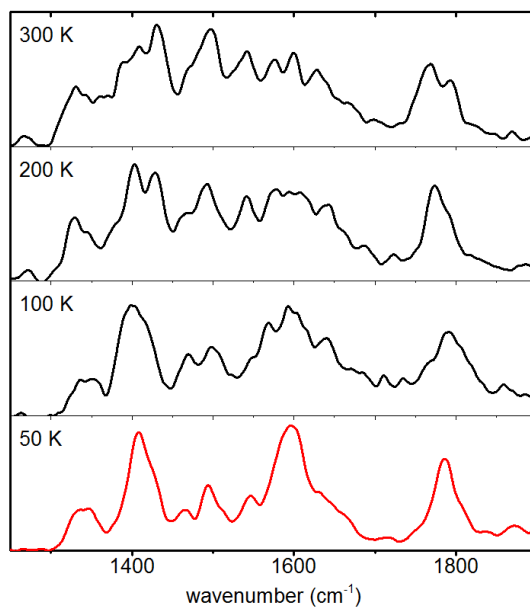


FIGURE 3.26 – Spectres EVB-AMOEBAs en fonction de la température entre 50 et 300 K.

l'échange de proton. La procédure de paramétrisation a un coût raisonnable, mais conduit à chercher une meilleure transférabilité entre systèmes. Dans cette optique, l'implémentation d'un terme de 'pénétration de charge' dans AMOEBA pourrait grandement alléger la procédure de paramétrisation [73].

Bibliographie

- [1] Clavaguéra, C. ; Calvo, F. ; Dognon, J.-P. *J. Chem. Phys.* **2006**, *124*, 074505.
- [2] Semrouni, D. ; Ohanessian, G. ; Clavaguéra, C. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3450-3462.
- [3] Semrouni, D. ; Balaj, O. P. ; Calvo, F. ; Correia, C. F. ; Clavaguéra, C. ; Ohanessian, G. *J. Am. Soc. Mass. Spectr.* **2010**, *21*, 728–738.
- [4] Thaunay, F. ; Clavaguéra, C. ; Ohanessian, G. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25935-25945.
- [5] Semrouni, D. *Energétique et spectroscopie de polypeptides par dynamique moléculaire : champ de force de seconde génération et chimie quantique*, Thesis, Ecole Polytechnique, Palaiseau, France, **2010**.
- [6] Semrouni, D. ; Sharma, A. ; Dognon, J.-P. ; Ohanessian, G. ; Clavaguéra, C. *J. Chem. Theory Comput.* **2014**, *10*, 3190-3199.
- [7] Thaunay, F. ; Dognon, J.-P. ; Ohanessian, G. ; Clavaguéra, C. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25968–25977.
- [8] Warshel, A. ; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218-6226.
- [9] Warshel, A. *Acc. Chem. Res.* **2002**, *35*, 385–395.
- [10] Thaunay, F. ; Calvo, F. ; Ohanessian, G. ; Clavaguéra, C. Theory and Applications of the Empirical Valence Bond Approach : From Physical Chemistry to Chemical Biology. In ; John Wiley & Sons, Ltd. : **2016** ; Chapter Empirical valence-bond models based on polarizable force fields for infrared spectroscopy.
- [11] Nelder, J. A. ; Mead, R. *Comput. J.* **1965**, *7*, 308-313.
- [12] Lagarias, J. C. ; Reeds, J. A. ; Wright, M. H. ; Wright, P. E. *SIAM J. Optimiz.* **1998**, *9*, 112-147.
- [13] Hansen, P. ; Shi-Hui, L. ; Jaumard, B. *Int. J. Comput. Math.* **1989**, *28*, 183 - 193.
- [14] Fletcher, R. ; Powell, M. J. D. *Comput. J.* **1963**, *6*, 163-168.
- [15] Payne, M. C. ; Teter, M. P. ; Allan, D. C. ; Arias, T. A. ; Joannopoulos, J. D. *Rev. Mod. Phys.* **1992**, *64*, 1045–1097.
- [16] Hilderbrandt, R. L. *Comput. & Chem.* **1977**, *1*, 179 - 186.
- [17] Shanno, D. F. *Math. Comp.* **1970**, *24*, 647–656.
- [18] J. E. Dennis, J. ; Moré, J. J. *SIAM Rev.* **1977**, *19*, 46-89.
- [19] Metropolis, N. ; Ulam, S. *J Am. Stat. Assoc.* **1949**, *44*, 335-341.
- [20] Metropolis, N. *Los Alamos Science* **1987**, *15*, 125-130.
- [21] Metropolis, N. ; Rosenbluth, A. W. ; Rosenbluth, M. N. ; Teller, A. H. ; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087-1092.
- [22] Hukushima, K. ; Nemoto, K. *J. Phys. Soc. Jpn* **1996**, *65*, 1604-1608.
- [23] Athènes, M. ; Calvo, F. *ChemPhysChem* **2008**, *9*, 2332–2339.

- [24] Kofke, D. A. *J. Chem. Phys.* **2002**, *117*, 6911-6914.
- [25] Aarts, E. H. L.; van Laarhoven, P. J. M. *Philips J. Res.* **1985**, *40*, 193-226.
- [26] Holland, J. H. *Adaptation in Natural and Artificial Systems : An Introductory Analysis with Applications to Biology, Control, and Artif. Intell.*; Bradford Book : **1992**.
- [27] Mitchell, M. *Artif. Intell.* **1998**, *100*, 325 - 330.
- [28] Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607-2609.
- [29] Earl, D. J.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910-3916.
- [30] Penev, E. S.; Lampoudi, S.; Shea, J.-E. *Comput. Phys. Commun.* **2009**, *180*, 2013 - 2019.
- [31] Joshi, K.; Semrouni, D.; Ohanessian, G.; Clavaguéra, C. *J. Phys. Chem. B* **2012**, *116*, 483-490.
- [32] Albrieux, F.; Calvo, F.; Chirot, F.; Vorobyev, A.; Tsybin, Y. O.; Lepère, V.; Antoine, R.; Lemoine, J.; Dugourd, P. *J. Phys. Chem. A* **2010**, *114*, 6888-6896.
- [33] Gao, B.; Liu, Z.-F. *J. Chem. Phys.* **2004**, *121*, 8299-8306.
- [34] Zhan, C.-G.; Zheng, F.; Dixon, D. A. *J. Chem. Phys.* **2003**, *119*, 781-793.
- [35] Zhou, J.; Santambrogio, G.; Brümmer, M.; Moore, D. T.; Woste, L.; Meijer, G.; Neumark, D. M.; Asmis, K. R. *J. Chem. Phys.* **2006**, *125*, 111102.
- [36] Wong, R. L.; Williams, E. R. *J. Phys. Chem. A* **2003**, *107*, 10976-10983.
- [37] Smeeton, L. C.; Farrell, J. D.; Oakley, M. T.; Wales, D. J.; Johnston, R. L. *J. Chem. Theory Comput.* **2015**, *11*, 2377-2384.
- [38] Abascal, J. L. F.; Vega, C. *J. Chem. Phys.* **2005**, *123*, 234505.
- [39] Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111-5116.
- [40] Hagberg, A. A.; Schult, D. A.; Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*; Pasadena, CA USA, **2008**.
- [41] Rega, N.; Brancato, G.; Petrone, A.; Caruso, P.; Barone, V. *J. Chem. Phys.* **2011**, *134*, 074504.
- [42] Wheeler, R. A.; Dong, H.; Boesch, S. E. *ChemPhysChem* **2003**, *4*, 382-384.
- [43] Brooks, B. R.; Janezic, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522-1542.
- [44] Levy, R. M.; la Luz Rojas, O. D.; Friesner, R. A. *J. Phys. Chem.* **1984**, *88*, 4233-4238.
- [45] Pejov, L.; Spangberg, D.; Hermansson, K. *J. Phys. Chem. A* **2005**, *109*, 5144-5152.
- [46] Nonella, M.; Mathias, G.; Eichinger, M.; ; Tavan, P. *J. Phys. Chem. B* **2003**, *107*, 316-322.
- [47] Bastida, A.; Soler, M. A.; Zuniga, J.; Requena, A.; Kalstein, A.; Fernandez-Alberti, S. *J. Phys. Chem. A* **2010**, *114*, 11450-11461.
- [48] Kohanoff, J. *Compu. Mater. Sci.* **1994**, *2*, 221 - 232.
- [49] Martinez, M.; Gaignot, M.-P.; Borgis, D.; Vuilleumier, R. *J. Chem. Phys.* **2006**, *125*, 144106.
- [50] Bowman, J. M.; Zhang, X.; Brown, A. *J. Chem. Phys.* **2003**, *119*, 646.
- [51] Kaledin, M.; Brown, A.; Kaledin, A. L.; Bowman, J. M. *J. Chem. Phys.* **2004**, *121*, 5646-5653.
- [52] Chin, W.; Piuze, F.; Dognon, J.-P.; Dimicoli, I.; Mons, M. *J. Chem. Phys.* **2005**, *123*, 084301.
- [53] Jaeqx, S.; Oomens, J.; Cimas, A.; Gaignot, M.-P.; Rijs, A. M. *Angew. Chem. Int. Ed.* **2014**, *53*, 3663-3666.
- [54] Marx, D. *ChemPhysChem* **2006**, *7*, 1848-1870.
- [55] Oomens, J.; Steill, J. D.; Redlich, B. *J. Am. Chem. Soc.* **2009**, *131*, 4310-4319.

- [56] Balaj, O.; Semrouni, D.; Steinmetz, V.; Nicol, E.; Clavaguéra, C.; Ohanessian, G. *Chem. Eur. J.* **2012**, *18*, 4583-4592.
- [57] Marinica, D. C.; Grégoire, G.; Desfrancois, C.; Schermann, J. P.; Borgis, D.; Gaigeot, M. P. *J. Phys. Chem. A* **2006**, *110*, 8802-8810.
- [58] Gaigeot, M.-P.; Cimas, A.; Seydou, M.; Kim, J.-Y.; Lee, S.; Schermann, J.-P. *J. Am. Chem. Soc.* **2010**, *132*, 18067-18077.
- [59] Dopieralski, P. D.; Latajka, Z.; Olovsson, I. *Acta Crystallographica Section B* **2010**, *66*, 222-228.
- [60] Wang, S.; MacKay, L.; Lamoureux, G. *J. Chem. Theory. Comput.* **2014**, *10*, 2881-2890.
- [61] Wu, X.; Thiel, W.; Pezeshki, S.; Lin, H. *J. Chem. Theory. Comput.* **2013**, *9*, 2672-2686.
- [62] van Duin, A. C. T.; Zou, C.; Joshi, K.; Bryantsev, V.; Goddard, W. A. CHAPTER 6 A Reaxff Reactive Force-field for Proton Transfer Reactions in Bulk Water and its Applications to Heterogeneous Catalysis. In *Computational Catalysis*; The Royal Society of Chemistry : **2014**.
- [63] Goldblum, A. *Journal of Molecular Structure : THEOCHEM* **1988**, *179*, 153 - 163.
- [64] Lobaugh, J.; Voth, G. A. *J. Chem. Phys.* **1996**, *104*, 2056-2069.
- [65] Vuilleumier, R.; Borgis, D. *J. Mol. Struct.* **1997**, *436-437*, 555 - 565.
- [66] Maupin, C. M.; Wong, K. F.; Soudackov, A. V.; Kim, S.; Voth, G. A. *J. Phys. Chem. A* **2006**, *110*, 631-639.
- [67] Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549-2564.
- [68] Wu, J. C. *Development of Accurate and Efficient Models for Biological Molecules*, Thesis, The University of Texas at Austin, **2011**.
- [69] Elking, D. M.; Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A.; Pedersen, L. G. *J. Chem. Theory Comput.* **2010**, *6*, 190-202.
- [70] Chaudret, R.; Gresh, N.; Narth, C.; Lagardère, L.; Darden, T. A.; Cisneros, G. A.; Piquemal, J.-P. *J. Phys. Chem. A* **2014**, *118*, 7598-7612.
- [71] Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. *J. Phys. Chem. A* **2003**, *107*, 10353-10359.
- [72] Narth, C.; Lagardère, L.; Polack, E.; Gresh, N.; Wang, Q.; Bell, D. R.; Rackers, J. A.; Ponder, J. W.; Ren, P. Y.; Piquemal, J.-P. *J. Comput. Chem.* **2016**, *37*, 494-506.
- [73] Ponder, J. W. "personal communication", **2016**.
- [74] Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827-7843.

Annexe A

Article

Vibrational mode assignment of finite temperature infrared spectra using the AMOEBA polarizable force field



Cite this: *Phys. Chem. Chem. Phys.*, 2015, 17, 25968

Vibrational mode assignment of finite temperature infrared spectra using the AMOEBA polarizable force field†

Florian Thauhay,^a Jean-Pierre Dognon,^b Gilles Ohanessian^a and Carine Clavaguéra*^a

The calculation of infrared spectra by molecular dynamics simulations based on the AMOEBA polarizable force field has recently been demonstrated [Semrouni *et al.*, *J. Chem. Theory Comput.*, 2014, **10**, 3190]. While this approach allows access to temperature and anharmonicity effects, band assignment requires additional tools, which we describe in this paper. The Driven Molecular Dynamics approach, originally developed by Bowman, Kaledin *et al.* [Bowman *et al.* *J. Chem. Phys.*, 2003, **119**, 646, Kaledin *et al.* *J. Chem. Phys.*, 2004, **121**, 5646] has been adapted and associated with AMOEBA. Its advantages and limitations are described. The IR spectrum of the Ac-Phe-Ala-NH₂ model peptide is analyzed in detail. In addition to differentiation of conformations by reproducing frequency shifts due to non-covalent interactions, DMD allows visualizing the temperature-dependent vibrational modes.

Received 18th April 2015,
Accepted 17th July 2015

DOI: 10.1039/c5cp02270c

www.rsc.org/pccp

1 Introduction

Gas phase infrared spectroscopy was developed and applied to biologically relevant molecules with the aim to decipher their structures based on the sensitivity of the frequencies to the environment of the chemical groups.^{1–5} Both double resonance and photodissociation techniques were applied to both neutral and ionic species in the last decade. Most of the time, detailed interpretation of such experiments involves theoretical calculations. The infrared spectra are usually computed using static harmonic quantum mechanics calculations and band assignments rely on the main components of normal modes. The main drawback of this approach is the use of scaling factors to compensate for anharmonicity and the fact that calculations are restricted to 0 K, even though temperature effects are known to be significant. Currently, several approaches exist to account for anharmonicity in Hessian based non-dynamical methods.^{6–9} They are accurate but generally limited to the study of small-size systems. Molecular dynamics (MD) simulations have been established as the method of choice to include directly in the calculation temperature and anharmonic effects such as frequency shifts, band broadening or band combinations.^{10–15}

Born–Oppenheimer (BOMD) and Car–Parrinello (CPMD) molecular dynamics have been successfully applied to gas-phase and solvated biomolecules, however they are limited to short propagation times of tens of picoseconds.^{16–22}

We have recently proposed the extension of the polarizable AMOEBA force field^{23–25} to infrared spectroscopy of gas phase peptide models.²⁶ This approach was based on the successful combination of such a force field, that includes refined electrostatics and an explicit polarization term, and the Fourier transform of the dipole–dipole time correlation function during a finite temperature molecular dynamics. This method allows to reproduce spectroscopic signatures that reflect the role of the environment due to hydrogen bonds or solvent molecules. Long-time simulations are accessible by classical molecular dynamics providing the possibility to see conformational dynamics signatures directly on the IR spectra. However when the system under study becomes of larger size, band assignments based on chemical intuition or static quantum chemistry calculations become non-straightforward, and sometimes irrelevant.

Indeed, the dynamic spectra identification is a recurrent challenge and various methods have been proposed for vibrational mode assignments.^{13,27–48} Wheeler *et al.* have suggested principal mode analysis (PMA)^{11,27} in which the average mass-weighted normal modes and associated frequencies are calculated from a classical MD trajectory by evaluating the eigenvectors and eigenvalues of the covariance matrix associated with the mass-weighted displacements of the molecular coordinates from their equilibrium positions.^{28–30} The simplicity of this approach makes it attractive, although several studies have

^a Laboratoire de Chimie Moléculaire, Ecole Polytechnique, CNRS, 91128 Palaiseau Cedex, France. E-mail: carine.clavaguera@polytechnique.edu

^b CEA/Saclay, DSM/IRAMIS/NIMBE, CNRS, UMR 3685, Laboratoire de Chimie Moléculaire et de Catalyse pour l'Energie, 91191 Gif-sur-Yvette, France

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5cp02270c

pointed out some limitations, *e.g.* when the potential energy surface of the molecule is anharmonic or when the thermal equilibrium is not guaranteed.^{49–51}

The instantaneous normal mode analysis (INMA) and its variants^{31–36} are obtained by diagonalizing the Hessian matrix at each instantaneous configuration of the molecule. The INMs provide an instantaneous decoupled second-order description of the vibrational motions of the molecule at the corresponding time-dependent configuration. The INMs therefore evolve with time, which complicates their identification.³⁷ Furthermore, INMA can be very time-consuming on large systems.

The vibrational spectrum can also be obtained by the Kohanoff's procedure³⁸ that allows to extract the eigenfrequencies and eigenmodes from the analysis of MD trajectories without having to calculate the second derivatives matrix.³⁹ A first global estimation of the frequencies is followed by the determination of the eigenvectors through a least squares fit of the trajectory including orthogonality constraints. Then, the trajectory is projected onto each of the normal modes. At that point, each projected trajectory contains mainly one frequency component, which is re-estimated. Starting with this framework, several methods were developed these last years.⁴⁰ The standard procedure to estimate the vibrational frequencies is based on the Fourier transformed velocity autocorrelation function.

Vuilleumier *et al.* proposed a method for obtaining normal modes based on a localization criterion for the Fourier transformed velocity time-correlation functions of the effective modes.¹³ There exists a choice of the localization function for which the method becomes equivalent to PMA based on covariance matrix diagonalization. A proper choice of the localization function leads to a method with a strong analogy with the usual normal mode analysis of equilibrium structures, where the thermal averaged Hessian replaces the Hessian. The Hessian itself is never calculated.

Bowman *et al.* and Kaledin *et al.* developed a method named Driven Molecular Dynamics (DMD) to perform a full normal-mode analysis, that does not require a calculation of the Hessian.^{41–43} They showed that normal modes can be obtained using a straightforward MD approach. The method employs an external harmonic driving term that can be used to scan the spectrum in a continuous wave fashion to determine resonant absorptions, that for weak signals are the normal-mode frequencies. The molecular motions, induced by driving at resonant frequencies, are (by definition) the vibrational modes.^{41,42,44}

Finally, digital filtering methods may be used to selectively enhance or suppress the vibrational motion in a MD simulation solely on the basis of frequency.^{45–47}

Among these approaches, we have chosen the driven molecular dynamics method which provides a direct way to understand molecular vibrational modes at a relatively low computational cost and is able to study nonharmonic motion and mode coupling. One important advantage of DMD is the capability to describe easily low frequency vibrational modes in the far IR region.⁵² They define the motions of large scale conformational changes which generally occur along the torsional degrees of freedom. Because of the long-range and collective nature of

these vibrational modes, they are highly sensitive to molecular structure and its chemical environment, and thus provide a fingerprint of the structure.⁵³ Moreover, these modes can be used to determine the mechanical anharmonicity in such large amplitude motions. In this context, the frequency driven approach is able to monitor the anharmonicity degree as a function of the temperature.^{52,54,55} In the following sections, we briefly describe the DMD algorithm and its implementation in the current version of the TINKER program.⁵⁶ Then we discuss its coupling with a previous implementation of finite temperature IR spectra by molecular dynamics simulations. The computed frequencies are used directly as input for DMD to assign vibrational modes. Specific tools were developed for this purpose. They are described here and applied to *N*-methyl-acetamide, a peptidic fragment model. Finally, an application on the Ac-Phe-Ala-NH₂ dipeptide is detailed using the full chain of methods necessary to carry out in order to obtain band assignments of finite temperature IR spectra.

2 Computational strategy

2.1 Quantum chemistry calculations

Geometry optimizations were performed using density functional theory (DFT) with different functionals, *i.e.* the pure GGA dispersion-corrected functional B97-D,⁵⁷ the hybrid functional B3LYP⁵⁸ and the global-hybrid meta exchange-GGA functional M06,⁵⁹ associated with the 6-311+G(d,p) basis set. IR spectra computed at the DFT/M06 level were previously found to be consistent with experimental results on peptides.^{60,61} Some energetics were computed also with Dunning's correlation consistent basis set cc-pVTZ. All the calculations were carried out with the GAUSSIAN program package.⁶²

2.2 The AMOEBA force field and its parameters

The AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) polarizable force field has been used successfully to reproduce structures and relative energies of peptides.^{25,63,64} One of the main differences with first generation potentials is its more complex electrostatic model, including permanent charge, dipole and quadrupole moments on each atom which are derived from quantum mechanical calculations.²⁵ Another valuable feature is the inclusion of many-body polarization effects which are explicitly treated using a self-consistent atomic dipole polarization procedure.⁶⁵ Repulsion–dispersion interactions between pairs of nonbonded atoms are represented by a buffered 14-7 potential.⁶⁶ The intramolecular valence terms consist in bond stretchings, angle bendings, and torsions with the MM3 force field energy form.⁶⁷ Additional terms are used such as out-of-plane bending and valence crossing terms to model the coupling between bonds and angles. Except for multipoles, the AMOEBA-BIO-09 set of parameters available in TINKER 6 was used for the current study.⁵⁶

It was shown previously that structure changes, and even more IR spectra, are very sensitive to weak interactions, such as intra or intermolecular hydrogen bonds, in MD simulations.²⁶

In AMOEBA, thanks to the multipolar expansion of the electrostatic energy, one can expect a good representation of hydrogen bonds. Ponder *et al.* studied the importance of multipoles moments and intramolecular polarization on conformational energies of polar molecules.⁶³ They conclude cautiously on the use of small model systems to derive intermolecular parameters for larger systems. From the distributed multipole analysis (DMA) method developed by Stone,^{68,69} a set of multipoles was derived for the whole dipeptide studied in Section 3 and assigned to each atomic site. The effect of the method, the basis set and the reference structure to extract multipoles on the computed infrared spectrum of the dipeptide studied in Section 3 are presented in ESI.†

2.3 Dipole autocorrelation function Fourier transform approach (DACF)

Infrared spectra are most often computed within the double harmonic approximation. Accordingly, vibrational frequencies at energy minima are directly derived from second derivatives of the potential energy with respect to molecular coordinates. *A posteriori*, some scaling factors can be applied to the frequencies in order take into account partly anharmonicity and finite temperature. A better way to include these effects is to perform dynamics simulations, in classical, Born–Oppenheimer or Car–Parrinello framework, and then calculate the Fourier transform of the dipole auto correlation function (DACF) to obtain IR vibrational spectra. We have proposed previously such calculations of the power spectrum using the AMOEBA force field and the TINKER program.²⁶ In this study, the DACF procedure was usually performed as follows. MD trajectories were propagated using the Velocity–Verlet algorithm, with a time step of 0.1 fs. All trajectories began with an equilibration phase, in the canonical ensemble, to reach the equilibrium temperature. The temperature was maintained by the Nosé–Hoover thermostat which was found previously to be the most appropriate.²⁶ The length of the equilibration phase depends on the temperature to attain, usually between 10 and 50 ps. At the end, atomic coordinates were stored and then three simulations, of 200 ps each, were performed to record the dipole moment along the trajectory. These trajectories started with the same initial geometry but atomic velocities and kinetic energies were distributed randomly. Infrared spectra were calculated by the DACF procedure and then averaged over the three simulations.

2.4 Vibrational mode analysis based on driven molecular dynamics

2.4.1 Method. The AMOEBA-DACF approach is able to determine accurately finite temperature vibrational spectra including anharmonic effects.²⁶ However, the assignment of bands in MD simulation of absorption spectra remains a tricky task, requiring sophisticated vibrational mode analysis. Some methods are available from the literature such as principal mode analysis (PMA), normal mode analysis (NMA) or MD trajectory filtering methods. Bowman *et al.* proposed a novel approach to perform a mode analysis that does not require a calculation of the Hessian.^{41,43}

The method implements an external, harmonic driving term that can be used to scan a frequency range and to determine resonant absorptions. This approach, called Driven Molecular Dynamics (DMD), provides a viable alternative to the standard Hessian-based normal mode analysis. It was introduced in a previous version of the TINKER program by Kaledin *et al.*^{42,43} In this work, starting from their kernel, we performed developments for an implementation of DMD into the TINKER 6 and 7 software and for vibrational mode analysis.

The Hamiltonian for driven molecular dynamics of a molecular system of N atoms is given by:

$$H(\mathbf{p}, \mathbf{q}, t) = \sum_i \frac{p_i^2}{2m_i} + V(\mathbf{q}) + U(t)$$

where \mathbf{q} and \mathbf{p} represent the $3N$ atomic coordinates and conjugate momenta, respectively, and V is the molecular potential. The driving force depends only on internuclear distances. It is included at each step of the dynamics *via* the driving term $U(t)$ as follows:

$$U(t) = \sum_{i,j} \lambda_{ij} r_{ij} \sin(\omega_n t)$$

In this equation, λ_{ij} are the coupling constants (the driving parameters), r_{ij} the internuclear distances, and ω_n the normal-mode pulsation obtained from a DACF calculation. When this force is felt by a vibrator at a pulsation very close to ω_n , the system should respond to the driving force by executing motion that is a normal mode (resonance) and with an absorption of energy. At non-resonant frequencies, the absorbed energy is small and oscillatory with time. On resonance, it increases quickly. At resonant frequencies the molecular motions induced by a weak driven force closely correspond to the normal-modes, while harder driving induces nonharmonic motion and leads to mode-coupling and nonharmonic shifts which are missing in standard normal-mode analysis.⁵² In a previous study, the issue of the choice of the driving parameter λ determining the amplitudes of the atomic motion is discussed.⁴² When λ is chosen small, the molecular motions are of small amplitude, in the harmonic limit. For large λ values, DMD allows the study of coupled anharmonic motion. For peptides in gas phase, the typical values for lambda were found to be between 1×10^{-6} and 2×10^{-5} Hartree Bohr⁻¹.

2.4.2 Limitations and improvements. DMD is a method to perform the calculation of the IR spectra with a full normal-mode analysis that does not require a calculation of the Hessian. The calculation of all eigenvalues and corresponding eigenvectors of the Hessian would scale approximately as $O(N^3)$ which is infeasible for large number (N) of atoms. The DMD approach requires $O(N)$ gradients of the potential per time step. A trajectory of N_{steps} scales as $O(N \times N_{\text{steps}})$ and the scan of N_{scan} frequencies scales as $O(N \times N_{\text{steps}} \times N_{\text{scan}})$. As stressed by Bowman *et al.*,⁴¹ the scaling of DMD becomes competitive compared with standard Hessian-based method only for systems including more than 1000 atoms. In the context of the present study, the frequencies are already known from DACF MD simulations. DMD is used only for the assignments of vibrational modes

and scales as $O(N \times N_{\text{steps}})$ for each mode of interest. Moreover, the intensities derived from absorbed energies are not really comparable to experimental intensities. Conversely, the power spectrum obtained from DACF simulations can be compared to the experimental absorption spectrum since it is derived from the dipole moment and allows a direct access to fundamental and combination bands. Furthermore, the method was proven to be accurate for frequency positions when associated to a polarizable force field such as AMOEBA.²⁶ Nevertheless, DMD has unique capability: it can selectively excite a mode on a frequency criterium.

It is within this framework that we have developed DMD based tools for vibrational mode analysis based on the DACF frequencies. As previously mentioned, the two spectra are not equivalent for the intensities.⁴² Accordingly, a strong infrared intensity does not ensure the capability of the internal coordinate to absorb energy.⁵⁵ At a resonant frequency, a DMD trajectory shows oscillations of atoms around their equilibrium positions with an amplitude proportional to the absorbed energy.⁵⁵ For example, Fig. 1 displays displacements of C=O and N-H internal coordinates of *N*-methyl-acetamide for two resonant driving frequencies corresponding to each stretching mode. In the bottom part, the N-H stretching mode is excited at its resonant frequency of 3518 cm^{-1} with light DMD conditions (*i.e.* 5000 steps, 0.1 fs by step, $\lambda = 5 \times 10^{-7}$ Hartree Bohr⁻¹). It can be seen that the system is monotonically absorbing energy with increasing simulation time with negligible absorption in the C=O. In the top part, at the C=O stretching mode

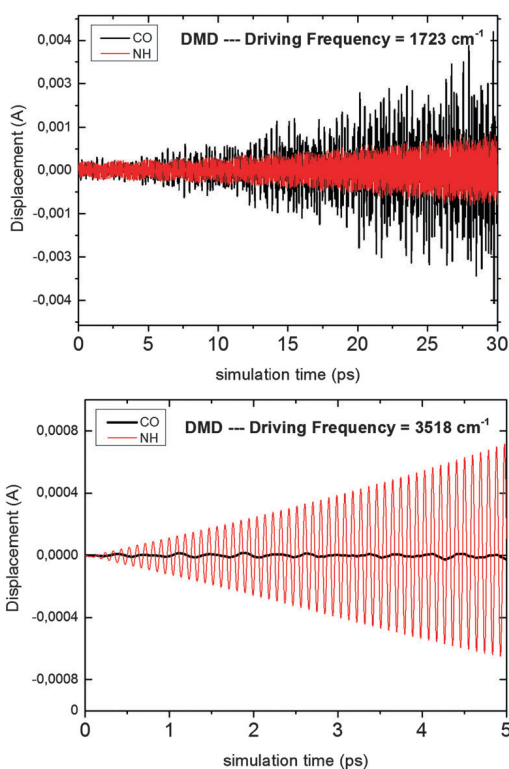


Fig. 1 Monitoring the C=O and N-H *N*-methyl-acetamide stretch modes by DMD simulations. See text for DMD conditions.

resonant frequency of 1723 cm^{-1} , it is much less obvious to define the resonant coordinate on a single criterion of oscillation amplitude. In order to obtain clearly visible C=O oscillations, the conditions of the DMD simulation have to be more specific, *i.e.* 30 000 steps, 0.2 fs by step, $\lambda = 2 \times 10^{-5}$ Hartree Bohr⁻¹ in this case. The latter example also emphasizes that a filtering of motions, based on frequency and not on amplitude, is required to observe a specific normal mode.

In the case of the model system H_5O_2^+ , Bowman *et al.* showed that for the lowest frequency mode (178 cm^{-1}), the coordinate displacement after a DMD simulation can be used to define the associated normal mode with an excellent agreement with the Hessian approach.⁴¹ For large-size and flexible systems, such as proteins or hydrated ions, an issue can be the “resolving power” of the method, *i.e.*, the ability to distinguish modes that are very close in frequency. It may be necessary to drive for longer times to separate modes with nearly the same frequency. Driving for longer times must be done with care by choosing smaller coupling constants because, as the energy absorbed increases, anharmonic motions will be observed.⁴¹

Increasing the temperature implies frequency shifts due to anharmonicity effects. It is difficult to correlate anharmonicity arising from temperature (DACF) or absorbed energy (DMD), because the mechanisms of energy absorption are very different and the normal mode response is very dependent on the potential energy surface of the system. The position of the N-H stretch maximum intensity in the *N*-methyl-acetamide molecule can be collected from both simulations with the idea to calibrate the amplitude of the driving force to reproduce the observed shifts with temperature (see Fig. S2 in ESI†). For example, at 200 K, the maximum amplitude is observed close to 3495 cm^{-1} . However, DMD simulation conditions depend strongly on the interrogated modes, and of the population of the modes at very close frequencies. Consequently, competitions can occur presenting a challenge to the DMD approach. We will detail later our procedure for normal modes analysis in a dipeptide case, depending on the frequency range and with the aim to study coupled anharmonic motions.

2.4.3 Tools implementation. The main idea is to catch the atomic motions characteristic of a vibrational mode during the DMD simulation at ω_n , and to display vibrations both as displacement vectors and in animated form in a similar way as conventional Hessian-based analysis. The characteristic motions must be collected when the energy absorbed by the mode under study is near its maximum. The mode will be described by a difference of nuclear positions between two steps during the dynamics, carefully chosen. In-between these two steps, at non-resonant frequencies, the absorbed energy is small and oscillatory with time and can be viewed as “noise” fluctuations in the visualization process. To get a clear picture of the normal mode it is desirable to reduce the noise.

In order to provide band assignments of a MD spectrum, the following procedure was implemented:

1 Monitor oscillation amplitudes of internal coordinates during the DMD, and calculate their oscillation frequencies. The structure was first energy minimized with TINKER software until the

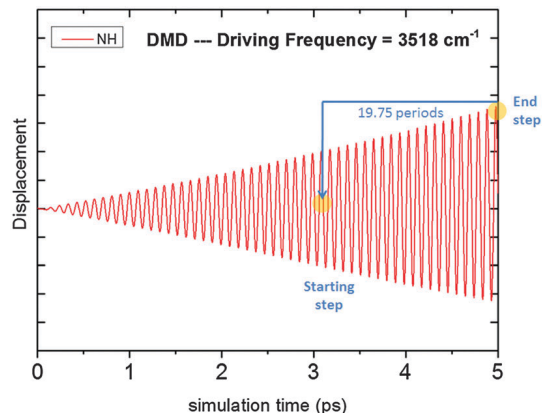


Fig. 2 Choice of end and starting steps for the description of normal modes from DMD simulations. Here, the N–H stretch of *N*-methyl-acetamide. The driving frequency at 3518 cm^{-1} is obtained from DACF spectrum at 1 K.

RMS gradient was less than $10^{-5}\text{ kcal mol}^{-1}$. Starting from the system initially at rest at this stationary point, 5000 steps of DMD simulation are enough (e.g., a total dynamics of 0.5 or 1 ps, depending on the time step) without an equilibration phase. From the molecular motions corresponding to each resonant driving frequency (e.g. Fig. 2), the frequency of the corresponding mode can be calculated. For a better numerical accuracy, this is carried out over the 20 last periods.

2 *The internal coordinate which vibrates with the pulsation closest to ω_n (driving pulsation) is identified as the reference resonant coordinate and is by definition the vibrational mode.* We focus on frequency rather than magnitude because some modes absorb more energy, even off-resonance. Therefore, a mode with a strong IR absorption is not necessarily the one with maximal absorbed energy in a DMD process. This is particularly important when the frequency of the normal mode is not known accurately (e.g. for a high temperature spectrum).

3 *Deduction of the time steps that will be used for the description of the motion, i.e. a starting point and an ending point.* These steps are defined through the internal coordinate selected in the previous step. We start by choosing the end step as the one for which the oscillation amplitude of this coordinate is maximal (see Fig. 2). The starting step is associated to an internal coordinate close to its equilibrium value, i.e. with a displacement close to zero. This is observed every $\frac{1}{4}$ period. As 20 periods is the reference time for the frequency calculation, the starting step is deduced from $20 - \frac{1}{4}$ period. The associated time is the equivalent to a reversal of $\frac{n}{4}$ period with a value set to $n = 79$.

4 *Removing the non-resonant motions.* This “filtering” process is also based on a frequency criterion on internal coordinates. For all resonant internal coordinates, the motion (between starting and ending steps) for the associated cartesian coordinates is kept. For all other coordinates, displacement is set to zero. The selectivity of the filter is a powerful tool to isolate a specific motion or to visualize collective motions in a frequency range.

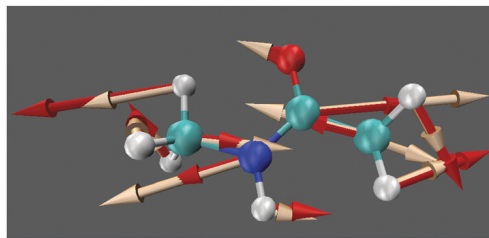


Fig. 3 Amide II mode of *N*-methyl-acetamide (1551 cm^{-1}): in white, normal mode from static calculation and in red, from DMD simulation (10 000 steps, 0.1 fs by step, $\lambda = 10^{-6}$ Hartree Bohr $^{-1}$, filter = 5 cm^{-1}).

In Fig. 3, an example of an amide II mode analysis is provided from DMD-based tools (in red) on the *N*-methyl-acetamide molecule.

The results from Hessian diagonalization is also shown in white. All the components are reproduced. Intensity differences are mainly explained by different energy absorption mechanisms, anharmonicity in DMD calculation, and inevitable small geometric differences between optimized and starting step structures.

These current DMD-based analysis tools emphasize the capability to collect normal modes dynamically. The coupling with the DACF procedure allows the assignment of finite temperature spectra, inside or outside harmonic condition framework. The filter criteria in frequency and amplitude can be associated to different display formats allowing alternative views of the movement associated to a given frequency. The next section will be devoted to an application of these DMD assignment tools to the finite temperature IR spectrum of the Ac-Phe-Ala-NH₂ dipeptide.

3 Results and discussion

The protected peptides series of Ac-Xxx-Phe-NH₂ and Ac-Phe-Xxx-NH₂ (with Xxx = Gly, Ala, Val and Pro) have been studied by Chin *et al.* using IR/UV double resonance spectroscopy and DFT quantum chemistry calculations.^{70,71} More recently, Jaeqx *et al.* have investigated the structures and infrared spectra of Ac-Phe-Ala-NH₂ and Ac-Phe-Gly-NH₂ by IR-UV ion-dip spectroscopy (IR-IDS) and Born–Oppenheimer molecular dynamics simulations.⁷² These model peptides are flexible enough to form various non-covalent interactions leading to different chemical environments of the CO and NH groups, and thus to specific spectroscopic signatures. Based on this available set of results, we have also considered Ac-Phe-Ala-NH₂ as a good candidate to assess DMD-based assignments of DACF spectra.

3.1 Selection of conformations

Replica Exchange Molecular Dynamics simulations were performed using AMOEBA on Ac-Phe-Ala-NH₂ (see ESI[†]). Among a collection of generated conformations, ten low-energy structures have been selected for quantum chemistry calculations and are presented in Fig. 4. The high flexibility of the peptide allows various intramolecular non-covalent interactions, and the main ones are provided in Fig. 5. These interactions often

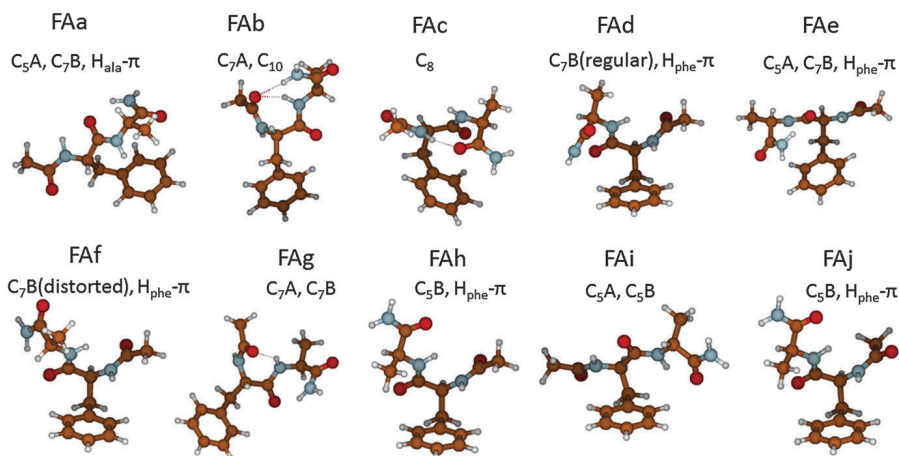


Fig. 4 Low-energy structures of FA at the B3LYP/6-311+G(d,p) level associated to the main non-covalent interactions for each structure.

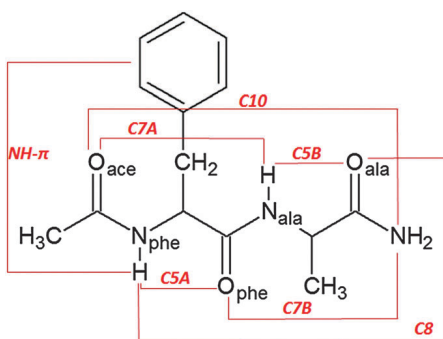


Fig. 5 Nomenclature of non-covalent interactions in FA structures.

correspond to hydrogen bonds between an amine or amide group and a neighboring oxygen in a C_5 or C_7 motif. Furthermore, C_8 and C_{10} interactions are possible between more distant groups. Another stabilizing effect is the $N-H \cdots \pi$ interaction that can exist between an amide hydrogen and the aromatic ring.

The relative energies of the ten conformations were computed at several levels of calculation. Details about the DFT results are presented in Table S2 in ESI†. The energetic order can be impacted by the functional used, however the differences are rather small and all methods yield FAa as the most stable structure, similarly than in previous studies.^{70,72} This conformation is stabilized by one C_5 , one C_7B and one $H_{Ala}-\pi$ interaction.

The main objective of this paper is the normal mode analysis of finite temperature spectra. This implies that the followed structures remain stable during the dynamics at a given temperature. At 50 K and even more at 200 K, barriers between conformations can be crossed and the barrier heights have been estimated around 5 kcal mol^{-1} . Among the ten starting structures, five structures, *i.e.* FAa, FAd, FAe, FAf and FAh, keep their original conformation (see ESI†). Consequently, the following assignments of spectra will be discussed only for these structures.

3.2 Infrared spectra and normal mode analysis

3.2.1 N-H stretching modes. Assignments have been proposed previously by comparing the experimental spectrum

to harmonic frequency calculations on the FAa structure at the B3LYP/6-31+G(d) level of theory (with a scaling factor of 0.96).⁷⁰ The 50 K spectra and the frequency maxima for the FAa, FAd, FAe, FAf and FAh conformations are compared to experimental values in Fig. 6. These five structures remain stable during a 200 ps simulation at 50 K. There are 4 N-H stretching contributions in the peptide, *i.e.* peptidic N-H of alanine and phenylalanine and 2 N-H from the NH_2 group. The latter are symmetric and antisymmetric combinations if the environment of both N-H is similar but they are more local modes otherwise.

Mode assignments can also be based on DMD simulations. The $N-H_{Ala}$ frequency is only well reproduced in FAa and FAh (exp: 3445 cm^{-1}) because it is involved in a weak interaction, $H_{Ala}-\pi$ in FAa and C_5B in FAh. In other structures, where it is free, the computed frequency is overestimated. FAa, FAd, FAF and FAh structures reproduce well the $N-H_{Phe}$ frequency position (exp: 3422 cm^{-1}). The environment is similar for FAd, FAF and FAh, in a $H_{Phe}-\pi$ weak interaction. In FAa, it is also involved in a rather weak interaction (see below). However in FAe, the N-H is involved in two interactions ($H_{Phe}-\pi$ and C_5A), and the frequency is red-shifted in comparison with experiments. NH_2 stretchings are well described in FAa, FAe and FAF structures because it is involved in a C_7B interaction in each case. In FAd, this interaction exists in the starting structure but it is not preserved during the simulation and the amine group prefers to make a π interaction.

To conclude, FAa is the better structure to reproduce the N-H stretching modes in agreement with previous studies. The average error on the frequencies is *ca.* 5 cm^{-1} in comparison with experiments.⁷⁰ However, five bands, instead of four, can be distinguished in the FAa spectrum due to change in the N-H environment during the dynamics. In FAa, the $C=O_{Phe}$ is shared between two hydrogen bonding interactions with the $N-H_{Phe}$ and one H_{NH_2} (Fig. 7). This competition weakens each hydrogen bond. In FAF, this competition does not exist because the $C=O_{Ala}$ interacts only with one H_{NH_2} . Therefore, hydrogen bonding should be stronger for FAF. Indeed, the average length of the $O \cdots H$ hydrogen bonds was found at *ca.* 2.07 \AA for FAa and 2.00 \AA for FAF, during a 5 ps simulation at 50 K. In FAa,

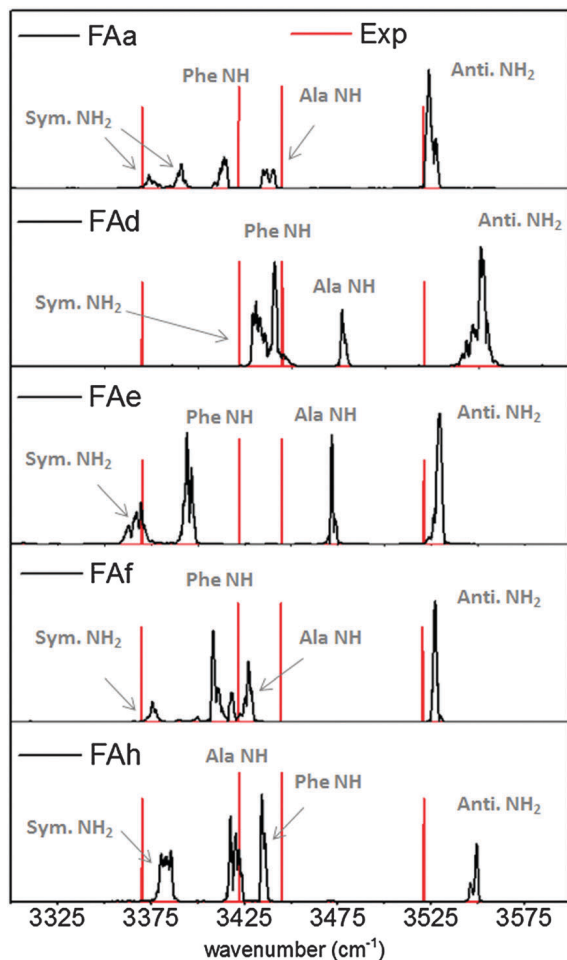


Fig. 6 DACF spectra in the 3300–3600 cm^{-1} range for the 5 stable structures during dynamics simulations at 50 K. Experimental frequencies are provided as red bars.

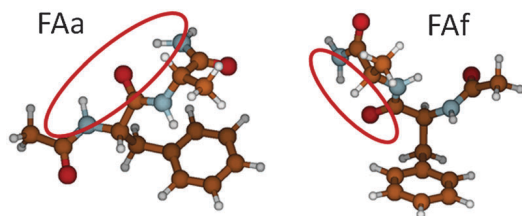


Fig. 7 Hydrogen bonds in FAa and FAf structures.

the band at 3392 cm^{-1} is a representative case of a major competition, and the band at 3375 cm^{-1} corresponds to the loss of the competitive interaction, that strengthens back the remaining one.

If static calculations are usually suitable to assign localized frequency modes, DMD allows to follow modes associated to frequency shifts with increasing temperature and to take into account coupling between modes. For example in Fig. 8, the frequency of the driving force is obtained from DACF spectra for two conformations, 3527 cm^{-1} for FAf and 3549 cm^{-1} for FAh, that correspond to the frequencies of the antisymmetric NH_2 stretch mode. A value of $\lambda = 10^{-6} \text{ Hartree Bohr}^{-1}$ was chosen to

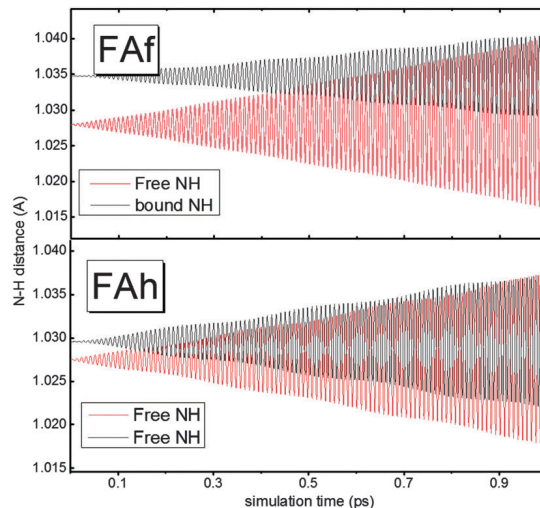


Fig. 8 N–H mode coupling in antisymmetric NH_2 stretchings for FAf and FAh structures. DMD conditions: 10 000 steps of 0.1 fs, $\lambda = 10^{-6} \text{ Hartree Bohr}^{-1}$.

stay in the regime of small oscillations. The two N–H have a similar environment in FAh whereas in FAf, the C_7B interaction decouples the two modes. In following variations of the N–H distances during a short simulation, DMD is able to highlight the strong coupling in FAh and the more local behaviour of NH_2 stretching in FAf.

3.2.2 Fingerprint range. The IR spectrum of Ac-Phe-Ala- NH_2 was recorded in the $100\text{--}1850 \text{ cm}^{-1}$ range using the FELIX free electron laser.⁷² DACF spectra have been computed at 1 (not shown), 50 and 200 K to compare with experiments (Fig. 9). It is difficult to correlate strictly experimental and calculated temperature. For example in the case of low-temperature experimental spectra, Gaigeot *et al.* recommended to simulate the spectra with a higher temperature to speed up the dynamics.⁷² Carboniere *et al.* discussed the limits of *ab initio* MD at finite temperature to take into account all anharmonicity effects that can be recorded in static calculations.¹⁴ In this study, experimental band widths at *ca.* 15 K are close to those calculated

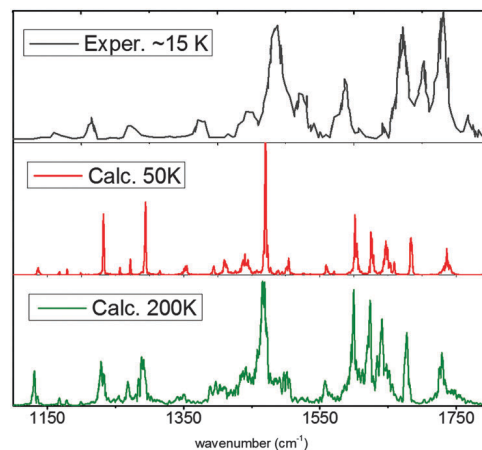


Fig. 9 Experimental⁷² and DACF spectra of FAa in the $1100\text{--}1800 \text{ cm}^{-1}$ range at 50 and 200 K.

Table 1 DMD Assignments for 1 K and 200 K DACF spectra for the FAa structure in the 1160–1800 cm^{-1} range, in comparison with experimental data⁷² and static AMOEBA calculations

Exp.	Hessian-based		DMD-based			
			DACF 1 K	DACF 200 K		
1160	1180, 1199	Def. CH_2 , bend $\text{CH}(\text{cycle})$	1179	Def. cycle		
1215	1233	Str. $\text{H}_2\text{C}-\text{C}(\text{cycle})$	1233	Bend CH cycle, str. $\text{H}_2\text{C}-\text{C}(\text{cycle})$	1233	Bend CH , str. $\text{H}_2\text{C}-\text{C}(\text{cycle})$
1273	1273, 1295	Amide III	1273, 1295	Amide III	1288	Amide III
1371	1354	Bend (CCN, CNC, $\text{NH}(\text{Ala})$)	1354	Bend CCN(Phe, Ala)	1352	Bend (CCN, CNC)
	1413, 1417	Umb. $\text{CH}_3(\text{Ac})$, str. CC, bend CH_2	1413	Bend ($\text{CH}_3(\text{Ac})$, CH_2 , CH cycle)	1398	Bend ($\text{NH}(\text{Phe}, \text{Ala})$, CH)
1442	1444, 1446	Umb. $\text{CH}_3(\text{Ac})$, str. CC, $\text{CN}(\text{Phe}, \text{NH}_2)$	1444	Bend CH aliph., str. CN	1437	Umb. $\text{CH}_3(\text{Ac})$, bend ($\text{NH}(\text{Ala})$, $\text{CH}(\text{aliph.})$)
	1471	Bend CH aliph.	1471	Bend $\text{CH}_3(\text{Phe})$, CH cycle	1465	Bend CH aliph., NH , str. CC
1488	1489, 1502	Bend $\text{CH}_3(\text{Phe})$			1501	Bend CH_2
1519	1505, 1528	Bend CH_2 , def. cycle	1528	Bend CH_3	1557	Umb. $\text{CH}_3(\text{Phe})$, bend $\text{NH}(\text{Ala})$
1575	1560	Umb. $\text{CH}_3(\text{Phe})$	1560	Umb. $\text{CH}_3(\text{Phe})$	1599	Umb. CH_3 , bend (NH_2 , CH_2 , $\text{NH}(\text{Ala})$, CH aliph.)
1586	1604	Str. CN, bend $\text{NH}(\text{Phe}, \text{Ala})$	1604	Str. (CC, CN)	1624	Bend NH , str. CO
	1628	Bend NH , str. CO	1628	Bend NH , str. CO	1642	Sym (bend NH_2 , str. $\text{CO}(\text{Ala})$), str. CC cycle
1642	1649	Sym (bend NH_2 , str. $\text{CO}(\text{Ala})$), str. CC cycle	1649, 1662	Sym (bend NH_2 , str. $\text{CO}(\text{Ala})$)	1642	Sym (bend NH_2 , str. $\text{CO}(\text{Ala})$), str. CC cycle
1672	1685	Sym str. $\text{CO}(\text{Ac}, \text{Phe})$	1687	Str. $\text{CO}(\text{Ala}, \text{Phe})$,	1678	Str. CN(Phe, Ala), $\text{CO}(\text{Phe})$
1703	1687	Anti str. $\text{CO}(\text{Ac}, \text{Phe})$				
1732	1737	Anti (bend NH_2 , str. $\text{CO}(\text{Ala})$)	1735	Anti (bend NH_2 , str. $\text{CO}(\text{Ala})$)	1730	Anti (bend NH_2 , str. $\text{CO}(\text{Ala})$)

at 200 K, such as the pattern around 1400–1500 cm^{-1} in Fig. 9. However, frequencies are not really affected by the temperature of the simulation.

Assignments for DACF frequencies in the fingerprint range are proposed in Table 1. They have been obtained in two ways, Hessian diagonalization at the AMOEBA optimized geometry and normal mode analysis from DMD simulations. For DMD attributions, modes are obtained first for a 1 K spectrum to check its capability to visualize modes similar to the ones obtained by a static calculations, and second for a 200 K DACF spectrum to visualize changes in modes induced by anharmonicity and temperature. The parameters used for DMD assignments are detailed in ESI.†

Modes in the $\text{C}=\text{O}$ stretch region are very close to those obtained at the static DFT/M06 level (scaled by 0.9613), with on one hand, the coupling between NH_2 bending and CO_{Ala} stretch, and on the other hand the coupling between $\text{C}=\text{O}_{\text{Phe}}$ and $\text{C}=\text{O}_{\text{Ace}}$ stretchings. In the 1650–1750 cm^{-1} range, three bands are observed experimentally, while four bands have been calculated, CO_{Phe} and CO_{Ala} stretchings (coupled), and a symmetric and antisymmetric combination of CO_{Ala} stretch and NH_2 bend modes (see Fig. 10 for symmetric combination). These two last components are still found when the temperature is increased to 200 K, while the coupling between $\text{C}=\text{O}_{\text{Phe}}$ and $\text{C}=\text{O}_{\text{Ace}}$ stretch modes decreases with temperature. In 1 K DACF spectrum, both contributions at 1649 and 1662 cm^{-1} have been assigned to the same symmetric contribution of NH_2 bending and CO_{Ala} stretching, which reflect the previously mentioned competition between the $\text{CO}_{\text{Ala}} \cdots \text{H}_{\text{NH}_2}$ and $\text{CO}_{\text{Ala}} \cdots \text{H}_{\text{Phe}}$ hydrogen bonds.

The experimental pattern around 1400–1550 cm^{-1} is assigned to C–H bends and amide II bands with C–N contributions at low temperature, and an additional contribution of N–H bends at 200 K. Other N–H bendings, at 1522 and 1543 cm^{-1} are in good agreement with experiments. The bands between 1400 and

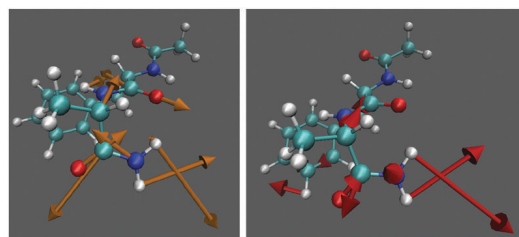


Fig. 10 Assignments from static (left, orange arrows) and from DMD calculations (right, red arrows) for 1649 cm^{-1} normal mode defined as symmetric combination of NH_2 bending and $\text{C}=\text{O}_{\text{Ala}}$ stretching.

1500 cm^{-1} are assigned to C–H bends. Furthermore, one advantage of DMD is to capture anharmonic movements, for example, allowed by increasing the temperature. The intense peak at 1599 cm^{-1} in the 200 K spectrum corresponds to the predominant contributions of CH_3 umbrella modes and NH_2 bending modes (Fig. 11, left part). The other intense band is attributed to C–H with large CH_2 and CH_3 bend contributions (Fig. 11, right part).

Around 1250–1270 cm^{-1} , DMD assignments are C–N stretchings, while DFT attributions correspond to collective N–H bends.

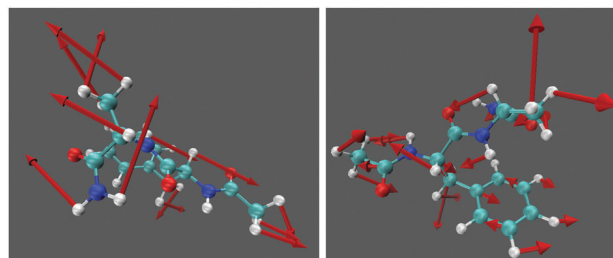


Fig. 11 Assignments from DMD calculation for 1599 cm^{-1} normal mode defined as CH_3 umbrella and NH_2 bending (left) and 1465 cm^{-1} (right) normal mode for C–H bend contributions.

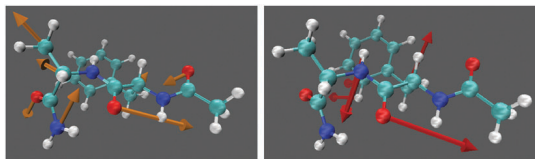


Fig. 12 Assignments from static (left, orange arrows) and from DMD calculations (right, red arrows) for 351 cm^{-1} normal mode which corresponds to NCO(Phe) bending.

NH_2 rocking is calculated by DFT at 1164 cm^{-1} , while in dynamics, it is found at 1073 cm^{-1} . Experimental intensities in this frequency range are very low, and make it difficult to conclude on the position of this band. DMD can also be used to observe soft modes in the far IR range where motions are essentially collective and a large part of the atoms vibrates with energy absorption. For instance, a frequency at 351 cm^{-1} corresponds to the NCO(Phe) bend mode as a major contribution (using an amplitude filter), see Fig. 12.

4 Conclusions

We have discussed the implementation of driven molecular dynamics in the latest version of the TINKER program using the AMOEBA polarizable force field. Powerful DMD-based analysis tools allow assigning and visualizing normal modes. In a previous work, we demonstrated that the AMOEBA-DACF approach is able to determine accurately finite temperature vibrational spectra including anharmonic effects. Good agreement with experiments is illustrated again here on the Ac-Phe-Ala- NH_2 dipeptide, through extensive analysis of the conformations (see ESI[†]) and the IR spectra (AMOEBA-DACF) with an identification of normal modes (AMOEBA-DACF-DMD). This approach is intrinsically easy to extend to large molecules. It may thus provide a route to calculation and interpretation of IR spectra when the currently used methods, *i.e.* quantum mechanical based on either Hessian or molecular dynamics, become out of reach.

Acknowledgements

The authors gratefully acknowledge Prof. M. Kaledin for having provided us with the software corresponding to the original implementation of DMD into TINKER 4.1.

References

- E. Garand, M. Z. Kamrath, P. A. Jordan, A. B. Wolk, C. M. Leavitt, A. B. McCoy, S. J. Miller and M. A. Johnson, *Science*, 2012, **335**, 694–698.
- E. Gloaguen and M. Mons, *Topics in Current Chemistry*, Springer, Berlin Heidelberg, 2015, pp. 1–46.
- Y. M. E. Fung, T. Besson, J. Lemaire, P. Matre and R. A. Zubarev, *Angew. Chem., Int. Ed.*, 2009, **48**, 8340–8342.
- P. Kupser, K. Pagel, J. Oomens, N. Polfer, B. Koks, G. Meijer and G. v. Helden, *J. Am. Chem. Soc.*, 2010, **132**, 2085–2093.
- O. Balaj, D. Semrouni, V. Steinmetz, E. Nicol, C. Clavaguera and G. Ohanessian, *Chem. – Eur. J.*, 2012, **18**, 4583–4592.
- R. B. Gerber, G. M. Chaban, B. Brauer and Y. Miller, *First-principles calculations of anharmonic vibrational spectroscopy of large molecules*, Elsevier, Amsterdam, 2005, ch. 9, pp. 165–194.
- V. Barone, *J. Chem. Phys.*, 2005, **122**, 014108.
- P. Carbonniere, T. Lucca, C. Pouchan, N. Rega and V. Barone, *J. Comput. Chem.*, 2005, **26**, 384–388.
- D. M. Benoit, *Front. Biosci.*, 2009, **14**, 4229–4241.
- H.-B. Xie, L. Jin, S. Rudić, J. P. Simons and R. B. Gerber, *J. Phys. Chem. B*, 2012, **116**, 4851–4859.
- N. Rega, G. Brancato, A. Petrone, P. Caruso and V. Barone, *J. Chem. Phys.*, 2011, **134**, 074504.
- N. Rega, *Theor. Chem. Acc.*, 2006, **116**, 347–354.
- M. Martinez, M.-P. Gaigeot, D. Borgis and R. Vuilleumier, *J. Chem. Phys.*, 2006, **125**, 144106.
- P. Carbonniere, A. Dargelos, I. Ciofini, C. Adamo and C. Pouchan, *Phys. Chem. Chem. Phys.*, 2009, **11**, 4375–4384.
- Q. Wan, L. Spanu, G. A. Galli and F. Gygi, *J. Chem. Theory Comput.*, 2013, **9**, 4124–4130.
- M.-P. Gaigeot, R. Vuilleumier, M. Sprik and D. Borgis, *J. Chem. Theory Comput.*, 2005, **1**, 772–789.
- M.-P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3336–3359.
- F. Ingrosso, G. Monard, M. Hamdi Farag, A. Bastida and M. F. Ruiz-López, *J. Chem. Theory Comput.*, 2011, **7**, 1840–1849.
- M. Rossi, V. Blum, P. Kupser, G. von Helden, F. Bierau, K. Pagel, G. Meijer and M. Scheffler, *J. Phys. Chem. Lett.*, 2010, **1**, 3465–3470.
- P. Carbonniere, S. Thicoipe, T. Very and X. Assfeld, *Int. J. Quantum Chem.*, 2012, **112**, 2221–2230.
- M. Thomas, M. Brehm, R. Fligg, P. Vohringer and B. Kirchner, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6608–6622.
- D. C. Marinica, G. Grégoire, C. Desfrancois, J. P. Schermann, D. Borgis and M.-P. Gaigeot, *J. Phys. Chem. A*, 2006, **110**, 8802–8810.
- P. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, **107**, 5933–5947.
- A. Grossfield, P. Ren and J. W. Ponder, *J. Am. Chem. Soc.*, 2003, **125**, 15671–15682.
- J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon, *J. Phys. Chem. B*, 2010, **114**, 2549–2564.
- D. Semrouni, A. Sharma, J.-P. Dognon, G. Ohanessian and C. Clavaguera, *J. Chem. Theory Comput.*, 2014, **10**, 3190–3199.
- R. A. Wheeler, H. Dong and S. E. Boesch, *ChemPhysChem*, 2003, **4**, 382–384.
- R. A. Wheeler and H. Dong, *ChemPhysChem*, 2003, **4**, 1227–1230.
- B. R. Brooks, D. Janežič and M. Karplus, *J. Comput. Chem.*, 1995, **16**, 1522–1542.
- R. M. Levy, O. De la Luz Rojas and R. A. Friesner, *J. Phys. Chem.*, 1984, **88**, 4233–4238.
- R. M. Strat, *Acc. Chem. Res.*, 1995, **28**, 201–207.
- M. Buchner, B. M. Ladanyi and R. M. Strat, *J. Chem. Phys.*, 1992, **97**, 8522–8535.

- 33 M. Cho, G. R. Fleming, S. Saito, I. Ohmine and R. M. Stratt, *J. Chem. Phys.*, 1994, **100**, 6672–6683.
- 34 R. E. Larsen and R. M. Stratt, *J. Chem. Phys.*, 1999, **110**, 1036–1052.
- 35 M. Nonella, G. Mathias, M. Eichinger and P. Tavan, *J. Phys. Chem. B*, 2003, **107**, 316–322.
- 36 L. Pejov, D. Spångberg and K. Hermansson, *J. Phys. Chem. A*, 2005, **109**, 5144–5152.
- 37 A. Bastida, M. A. Soler, J. Zúñiga, A. Requena, A. Kalstein and S. Fernández-Alberti, *J. Phys. Chem. A*, 2010, **114**, 11450–11461.
- 38 J. Kohanoff, *Comput. Mater. Sci.*, 1994, **2**, 221–232.
- 39 G. Onida, W. Andreoni, J. Kohanoff and M. Parrinello, *Chem. Phys. Lett.*, 1994, **219**, 1–7.
- 40 M. C. González Lebrero, L. L. Perissinotti and D. A. Estrin, *J. Phys. Chem. A*, 2005, **109**, 9598–9604.
- 41 J. M. Bowman, X. Zhang and A. Brown, *J. Chem. Phys.*, 2003, **119**, 646–650.
- 42 M. Kaledin, A. Brown, A. L. Kaledin and J. M. Bowman, *J. Chem. Phys.*, 2004, **121**, 5646–5653.
- 43 A. Brown, M. Kaledin, J. Bowman and A. Kaledin, *Driven Molecular Dynamics for Normal Modes of Biomolecules without the Hessian, and Beyond*, Chapman and Hall/CRC, 2005, pp. 281–300.
- 44 M. Kaledin, A. L. Kaledin and J. M. Bowman, *J. Phys. Chem. A*, 2006, **110**, 2933–2939.
- 45 P. Dauber-Osguthorpe and D. J. Osguthorpe, *J. Am. Chem. Soc.*, 1990, **112**, 7921–7935.
- 46 P. Dauber-Osguthorpe and D. J. Osguthorpe, *Biochemistry*, 1990, **29**, 8223–8228.
- 47 S. C. Phillips, M. T. Swain, A. P. Wiley, J. W. Essex and C. M. Edge, *J. Phys. Chem. B*, 2003, **107**, 2098–2110.
- 48 J. Sun, G. Niehues, H. Forbert, D. Decka, G. Schwaab, D. Marx and M. Havenith, *J. Am. Chem. Soc.*, 2014, **136**, 5031–5038.
- 49 M. Schmitz and P. Tavan, *J. Chem. Phys.*, 2004, **121**, 12233–12246.
- 50 M. Schmitz and P. Tavan, *J. Chem. Phys.*, 2004, **121**, 12247–12258.
- 51 F. Calvo, P. Parneix and N.-T. Van-Oanh, *J. Chem. Phys.*, 2010, **133**, 074303.
- 52 M. Kaledin and D. T. Adedeji, *J. Phys. Chem. A*, 2015, **119**, 1875–1884.
- 53 H. H. Mantsch and D. Naumann, *J. Mol. Struct.*, 2010, **964**, 1–4.
- 54 M. Kaledin, A. L. Kaledin, J. M. Bowman, J. Ding and K. D. Jordan, *J. Phys. Chem. A*, 2009, **113**, 7671–7677.
- 55 M. Kaledin and C. A. Wood, *J. Chem. Theory Comput.*, 2010, **6**, 2525–2535.
- 56 J. W. Ponder, *TINKER - Software Tools for Molecular Design (version 6)*, <http://dasher.wustl.edu/tinker>, 2014.
- 57 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799.
- 58 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 59 Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **125**, 194101.
- 60 K. Joshi, D. Semrouni, G. Ohanessian and C. Clavaguera, *J. Phys. Chem. B*, 2012, **116**, 483–490.
- 61 J. K. Martens, I. Compagnon, E. Nicol, T. B. McMahon, C. Clavaguera and G. Ohanessian, *J. Phys. Chem. Lett.*, 2012, **3**, 3320–3324.
- 62 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision D.01*, Gaussian Inc., Wallingford CT, 2009.
- 63 T. D. Rasmussen, P. Ren, J. W. Ponder and F. Jensen, *Int. J. Quantum Chem.*, 2007, **107**, 1390–1395.
- 64 D. Semrouni, G. Ohanessian and C. Clavaguera, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3450–3462.
- 65 P. Y. Ren and J. W. Ponder, *J. Comput. Chem.*, 2002, **23**, 1497–1506.
- 66 T. A. Halgren, *J. Am. Chem. Soc.*, 1992, **114**, 7827–7843.
- 67 N. L. Allinger, Y. H. Yuh and J. H. Lii, *J. Am. Chem. Soc.*, 1989, **111**, 8551–8566.
- 68 A. J. Stone, *Chem. Phys. Lett.*, 1981, **83**, 233–239.
- 69 A. J. Stone and M. Alderton, *Mol. Phys.*, 1985, **56**, 1047–1064.
- 70 W. Chin, F. Piuze, J.-P. Dognon, I. Dimicoli and M. Mons, *J. Chem. Phys.*, 2005, **123**, 084301.
- 71 W. Chin, J.-P. Dognon, C. Canuel, F. Piuze, I. Dimicoli, M. Mons, I. Compagnon, G. von Helden and G. Meijer, *J. Chem. Phys.*, 2005, **122**, 054317.
- 72 S. Jaque, J. Oomens, A. Cimas, M.-P. Gaigeot and A. M. Rijs, *Angew. Chem., Int. Ed.*, 2014, **53**, 3663–3666.

Annexe B

Article

Empirical valence-bond models based on polarizable force fields for infrared spectroscopy

Chapter 5

Empirical valence-bond models based on polarizable force fields for infrared spectroscopy

Florian Thunay, Florent Calvo, Gilles Ohanessian and
Carine Clavaguéra

Keywords: infrared spectroscopy, proton transfer, potentiel energy surface exploration, polarizable force field, molecular dynamics, finite temperature effects

5.1 Introduction

Proton transfer (PT) through hydrogen bond networks plays a major role in chemical and biochemical processes [1]. For instance, a major part of biochemical reactions operates under controlled pH conditions, because the protonation state of the chemical groups involved is critical. Furthermore, proton transport in biological media and through membranes governs cellular activity to a large extent. In addition, organization at the molecular level strongly influences physical and chemical properties of the solvent but also the solute itself. This influence can be indirect through the environmental conditions being imposed, but also more direct through the hydrogen bond network and the possible exchange of mobile protons. In recent years the number of works focused on the study of hydrogen bond networks and its capacity for charge transportation has dramatically increased [2, 3, 4].

Infrared (IR) spectroscopy is a powerful tool to probe the structure of hydrogen-bonded compounds, especially biomolecules. In proteins, the fingerprint range (i.e. 1000–2000 cm^{-1}) is particularly informative. The so-called amide I and II bands (corresponding to the C=O stretching and CNH bending modes, respectively) found in this region are very sensitive to hydrogen bond networks. Furthermore, charged groups such as NH_3^+ and CO_2^- make strong hydrogen bonds that also give rise to characteristic bands [5]. The best appraisal of the intrinsic impact of hydrogen bonds on vibrational frequencies can be gained in the gas phase, whereby in absence of the environment the local effects can all be appreciated. While conventional absorption spectroscopy cannot be directly used in gas phase, specific techniques such as infrared multiple photon dissociation (IRMPD) have been developed. The IRMPD method has been found to be particularly suited to probe the structures of gaseous ions [6, 7, 8], providing a wealth of insights into the intrinsic properties of biomolecular ions in their local environment.

Despite the large amount of experimental data available for infrared signatures of biomolecules in the gas phase, theoretical calculations remain an essential tool to assist interpretation of spectra as they provide fundamental knowledge about the role and the relative magnitude of hydrogen bonds in such species. Usually, energetic and spectroscopic properties are obtained applying quantum mechanical (QM) methods on static conformations. These calculations are typically (but not always) based on density functional theory (DFT), which allows systems of biological interest such as peptides to be studied at a relatively low computational cost. However, in many situations the agreement between experimental and computed spectra is poor and corrections to account for an-

harmonicities (which can be noticeably important at room temperature) have to be included [9, 10, 11, 12, 13]. Anharmonicities are particularly important in proton transfer, and therefore one needs to go far beyond the conventional QM approach in order to include dynamical effects. Several computational approaches have been previously developed to address the spectroscopic manifestations of PT [14, 15, 16, 17, 18, 19, 20, 21, 22].

Early attempts at modeling PT in molecular systems include the pioneering work of Goldblum [23], which relies on highly parametrized potential energy surfaces. In this work, the semi-empirical MNDO method was adapted to consider proton migration between two OH^- anions, using an analytical function of the two states to weight the proton evolution along the interoxygen distance [23]. A similar two-state function was also used in the so-called “hydrogen dynamics” or HYDYN method, which allows PT processes to be modeled both with nonpolarizable and polarizable force fields. The method was used to study structural and dynamical properties of small protonated water clusters, such as Eigen and Zundel complexes [24]. This multistate description has become widespread with the development of the empirical valence-bond model (EVB) by Warshel and co-workers [25, 26]. The EVB approach has been extended by several groups for the study of proton transport in water, most notably by Voth and coworkers [14] and by Borgis *et al.* [21]. Using molecular dynamics (MD) simulations based on a multistate EVB (MS-EVB) model, Voth *et al.* interpreted a broad absorption as a characteristic infrared signature of the strong hydrogen bond [15]. More recently, this approach has been extended to incorporate nuclear quantum effects in the description of proton diffusion, highlighting the concept of presolvation in which multiple weak hydrogen bonds on the hydronium ion facilitate PT [16]. The use of EVB for exploring reaction dynamics is reviewed in Chapter 4 [27].

With the development of first-principle molecular dynamics methods, either of the Born-Oppenheimer (BOMD) or Car-Parrinello (CPMD) types, it was possible to explicitly model PT processes in molecular systems. These approaches combine electronic structure methods with a classical or quantum mechanical description of nuclear motion. CPMD simulations have notably been used to study PT in protonated dialanine [17] and protonated nicotine in water [28]. Proton transfer dynamics in the $(\text{HCO}_3^-)_2$ dimer, and the corresponding vibrational spectrum, have also been studied by CPMD and path-integral molecular dynamics simulations [18]. Quantum dynamical methods have also been employed to treat smaller size systems containing only a few tens of degrees of freedom. For instance, the multiconfiguration time-dependent Hartree (MCTDH)

approach has been applied to model the Zundel cation and was found to fairly reproduce its experimental IR spectrum [19], despite many Fermi resonances. The MCTDH method was also successfully applied to model the vibrational spectrum of malonaldehyde [20].

These calculations, although very accurate, are computationally expensive and therefore limited in size and time scales. This has motivated the use of semi-empirical methods to address the behavior of PT in water clusters using approaches such as AM1 or PM3 [29, 30]. Semi-empirical methods have also been used to describe the diabatic states in EVB models [31]. Alternatively, fully QM methods on parametrized potential energy surfaces have been applied to model PT processes in (bio)molecules. Hybrid approaches combining classical MD and QM derived proton hopping have also been developed to simulate proton transport in extended systems [32]. QM/MM methods have been applied to model an excess proton in water and enzymatic activity [33, 34]. Finally, reactive force fields such as ReaxFF [35, 36] (for details see Chapter 1), have also been applied to treat PT reactions in the context of heterogeneous catalysis [37].

In this chapter, we describe an accurate computational framework for modeling the IR spectra of biomolecules undergoing proton transfer in the gas phase. Briefly, our model combines the AMOEBA (atomic multipole optimized energetics for biomolecular simulation) polarizable force field [38, 39], as implemented in the Tinker program [40], with an EVB approach. AMOEBA includes an explicit self-consistent polarization term and a multipole expansion treatment of the electrostatic effects [38], both ingredients appearing necessary to reproduce accurate IR spectra of anharmonic systems, especially those bearing electrically charged sites. [41, 42].

We first present the experimental motivation behind the development of our framework, which is focused on the aspartate molecule, and proceed by showing the limitations of standard QM approaches, as well as the simulations with non reactive force fields, for reproducing IR spectra. Subsequently we introduce the ingredients of the force field that allow proton exchange between the two anionic sites to be accounted for within the EVB framework. The success of the model for reproducing the experimentally measured spectrum is discussed in the light of the underlying dynamics of the proton. The dynamical and spectroscopic pictures are related to one another and evolve jointly as a function of temperature. Possible extensions and perspectives of the EVB-AMOEBA approach in the context of IR spectroscopy of biomolecules in the gas phase are finally discussed.

5.2 Infrared spectra of aspartate and non-reactive calculations

Aspartate and glutamate are amino acids bearing a carboxylic acid function in their side chain, which at biological pH is deprotonated. As a single unit, these molecules have two carboxylate anions and a single proton that can be bound to either carboxylic group or bridging between them in a sharing configuration (**Fig. 5.1**). Thus, they are potential candidates for a dynamically shared proton configuration in the gas phase. This was first demonstrated by Oomens *et al.* [43] using IRMPD. Herein, we take this case to illustrate our implementation of the EVB methodology, which we have devised for simulating infrared spectra of gaseous ions. However, before discussing the performance of the EVB approach, we show how standard QM approaches compare to experimental measurements.

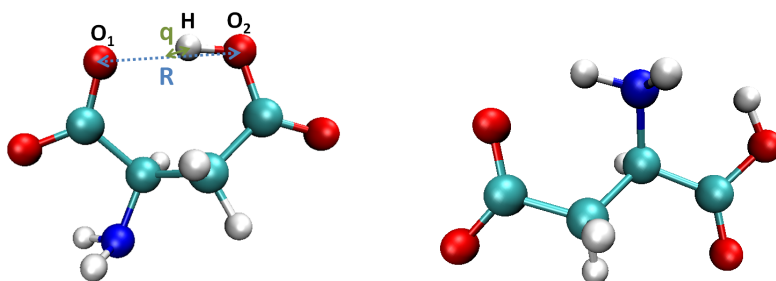


Figure 5.1: Most stable structures of aspartate: bent conformation (left) with definitions of the R and q distances, and linear conformation (right).

5.2.1 Experimental approach

Oomens *et al.* measured the IR spectra of different amino acids using IRMPD. It was found that in the fingerprint region a sharp difference between deprotonated Trp, Tyr, Phe, Cys and Ser on one hand, and Asp and Glu on the other hand is observed [43]. In the former case, they were assigned to carboxylate signatures near 1630 and 1330 cm^{-1} , arising from the antisymmetric and symmetric OCO stretching modes, respectively. In the second case, the Asp and Glu spectra were found to be dominated by a very broad absorption feature extending from ca. $1300\text{--}1700\text{ cm}^{-1}$, which was attributed to dynamical proton sharing between the two carboxylate groups. This unusually broad band was not reproduced by harmonic spectrum calculations. Using the same IRMPD technique, we recorded these spectra again at room temperature in the $1100\text{--}1800\text{ cm}^{-1}$ range

(**Fig. 5.2**, top), and the relatively similar experimental setup at CLIO (Centre Laser Infrarouge d’Orsay, France).

The IR spectra were recorded using a free-electron laser (FEL) coupled to a 7T FT-ICR mass spectrometer and obtained by plotting the fragmentation efficiency, F_{eff} as a function of wavelength. F_{eff} is defined as $F_{\text{eff}} = -\log[I_p/(I_p + \sum I_{\text{frag}})]$, in which I_p and I_{frag} are the parent and fragment ion intensities, respectively. By optimizing carefully experimental parameters, we were able to record a better resolved spectrum (for experimental details please, see ref. [43, 44]). The spectrum displayed in **Fig. 5.2** is clearly different from those of other deprotonated amino acids, with a main broad feature extending from 1470 to 1680 cm^{-1} , which is likely to be a signature of C=O or OCO stretching when a proton is shared between the two anions. On the red side of this band, a shoulder appears near 1450 cm^{-1} .

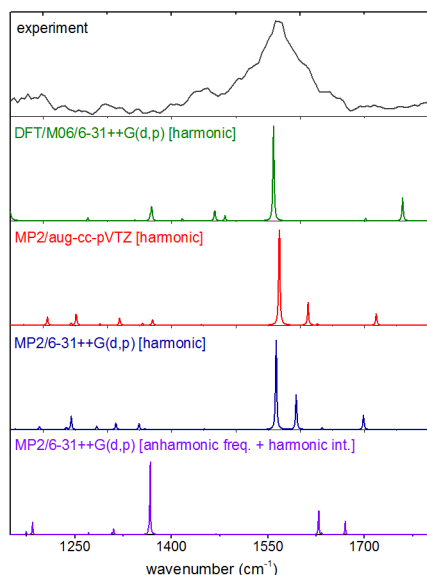


Figure 5.2: IRMPD spectrum measured for aspartate and predicted by static QM calculations using different methods, at the harmonic levels of DFT/M06/6-31++G(d,p), MP2/aug-cc-pVTZ and MP2/6-31++G(d,p), and using anharmonic frequencies at the MP2/6-31++G(d,p) level but with harmonic intensities.

5.2.2 Quantum chemical calculations

Most of our finite temperature simulations rely on the use of the AMOEBA polarizable force field [38, 39], which is exclusively parametrized on *ab initio* data. To extend this approach within the EVB framework for the study of PT processes, we aim to follow the same lines, i.e to parametrize the coupling term between the diabatic states involved in the proton sharing using electronic structure data. It is important to keep in mind that the intrinsic performance of the QM approach should be also assessed with respect to experiments and within its limits of feasibility, where assumptions such as the harmonic approximations for both the potential energy and the dipole moment surface are used. The static IR spectra of Asp were thus calculated for two different conformers using DFT and the modern M06 functional, as well as at the post-Hartree-Fock MP2 level using the Gaussian09 software package [45].

Among the two conformers considered for Asp, the bent conformer is the most stable structure at all levels of theory tested (by 7.4 kcal/mol at the MP2/aug-cc-pVTZ level of theory). For this conformer, two different isomers exist, which differ in the location of the proton (see **Fig. 5.1**). In the most stable isomer, the proton is bound to the side chain, i.e. an oxygen of the COO^- group farthest away from the amine. The geometry slightly depends on the level of theory, the OO distance varying from 2.46 Å at the DFT/M06/6-31++G(d,p) level to 2.44 Å at the MP2/aug-cc-pVTZ level, the shortest distance of the proton to an oxygen being equal to 1.08 Å for both methods.

As usual after the computation of the IR spectra [46, 47], the harmonic frequencies were scaled by a factor of 0.98 for MP2/aug-cc-pVTZ, 0.96 for MP2/6-31++G(d,p) and 1.03 for M06/6-31++G(d,p) to reproduce the intensity maximum of the experimental spectrum (**Fig. 5.2**). As can be seen from **Fig. 5.2** (top), the C=O stretching modes in the harmonic spectra cannot account for the broad band experimentally observed. Furthermore, there is no computed frequency at the MP2 level that corresponds to the experimental band at 1450 cm^{-1} . These results are not improved when calculating anharmonic frequencies at the MP2/6-31++G(d,p) level of theory. Furthermore, the anharmonic intensities predict very intense bands around 1300 and 1400 cm^{-1} , which overshadow the rest of the spectrum. This means that QM approaches, even when anharmonicity is included, cannot reproduce the signature of the shared proton experimentally observed in aspartate.

5.2.3 Finite temperature IR spectra based on AMOEBA

The AMOEBA force field

A direct way to include both anharmonicities and dynamical effects in the computed IR spectrum consists of performing MD simulations. The AMOEBA polarizable force field was selected for its ability to reproduce various physicochemical properties in both the gas and condensed phases. It has been successfully used to reproduce structures and relative energies of peptides [48, 38, 49], and also to model the hydration process of various multicharged ions [50, 51, 52, 53]. One of the main differences with first-generation force fields such as AMBER or CHARMM lies in its more realistic description of the nonbonded intermolecular terms (*vide infra*):

$$\Delta E_{\text{int}} = \Delta E_{\text{vdw}} + \Delta E_{\text{elec}} + \Delta E_{\text{pol}}$$

In particular, the electrostatic interactions (E_{elec}) originate from the interaction of atomic multipoles (up to quadrupoles) derived from QM calculations [38] in order to ensure a greater accuracy than the Coulomb interaction based on point charges used in standard force fields and valid at first order only. Another valuable feature of AMOEBA is the inclusion of many-body polarization effects (E_{pol}) in which the induced dipoles are treated self-consistently [54, 38]. The combination of such a refined electrostatic term and an explicit treatment of polarization effects leads to a good reproduction of the electrostatic potential and an accurate representation of the hydrogen bonds both in their strength and directionality [48, 41]. The van der Waals term (E_{vdw}) takes into account exchange-repulsion and dispersion interactions between pairs of nonbonded atoms using a buffered 14-7 potential [55], in contrast to the 12-6 one usually used by standard force fields. This potential has the advantage to reproduce simultaneously series of *ab initio* results in the gas phase and liquid properties on noble gases and diatomic species. In addition, the bonding terms, consisting in bond stretchings, angle bendings, torsions and stretch-bend couplings, are all derived from the MM3 force field [56]. Except for multipoles, we used the AMOEBAbio-09 set of parameters available in TINKER 7 [40].

Infrared spectra from molecular dynamics simulations

As mentioned before, theoretical IR spectra are most often obtained from harmonic QM calculations. *A posteriori*, and as was actually shown in **Fig. 5.2**, some scaling factors can be applied to partially account for anharmonicities and temperature effects. However, a more rigorous way to include these effects is to

perform simulations at finite temperature MD using potential energy and dipole moment surfaces that are either classically parameterized [57, 58, 41] or use an explicit description of electronic structure [10, 12]. From the trajectories the IR vibrational spectrum $\mathcal{I}(\omega)$ is obtained by Fourier transforming the dipole moment time autocorrelation function (DACF), a standard harmonic weight being applied to yield a quantity comparable to an absorption intensity:[10]

$$\mathcal{I}(\omega) \propto \omega [1 - \exp(-\beta\hbar\omega)] \int e^{-i\omega t} \langle \vec{\mu}(t) \cdot \vec{\mu}(0) \rangle dt,$$

where we have denoted $\beta = 1/k_{\text{B}}T$ with k_{B} the Boltzmann constant, $\vec{\mu}(t)$ the dipole moment vector at time t , and $\langle \cdot \rangle$ an average over the time origin.

This methodology has recently been shown to yield accurate IR spectra for model peptides when compared against experimental measurements [41]. In this example, the DACF procedure was performed as follows:

- All trajectories are initiated with an equilibration phase in the canonical ensemble to reach the desired temperature, equilibration being reached using a Nosé-Hoover thermostat. The equations of motion are integrated using the velocity Verlet algorithm;
- At the end of the equilibration period, atomic coordinates are stored and three independent simulations of 200 ps each are performed at constant total energy, again using the velocity Verlet integrator. The dipole moment vector $\vec{\mu}(t)$ is recorded every time step (0.1 fs) along the trajectory;
- The time autocorrelation function $\langle \vec{\mu}(t) \cdot \vec{\mu}(0) \rangle$ is calculated from the time series of $\vec{\mu}(t)$ and Fourier transformed to yield the infrared spectrum in the 1000–2000 cm^{-1} frequency range;
- The final IR spectrum is averaged over the contributions from the independent trajectories (in this case three).

Role of the multipoles

A key aspect of AMOEBA is the use of a set of multipoles for the electrostatic contribution to the potential energy. In many applications, such as configurational sampling, MD at thermal equilibrium or for determining solvation energies, the dependence of the computed properties on the multipolar component is usually disregarded. However, in the context of IR spectroscopy, this contribution can be important in the calculation of the dipole moment and the IR frequencies. Furthermore, the multipolar description improves the calculation

of the infrared intensities, as was notably shown in the case of the Ace-Phe-NH₂ peptide [42].

The original version of AMOEBA suitable for biomolecules and employed here (AMOEBA_{bio}-09)[38] includes a standard multipolar set, which we denote as O. In our study we have considered new multipoles, extracted from *ab initio* calculations at the MP2/aug-cc-pVTZ//MP2/cc-pVTZ level of theory. The multipoles were obtained from the *ab initio* electron density using the distributed multipole analysis method [59]. Furthermore, atomic dipoles and quadrupoles were refined on the electrostatic potential while keeping the charges fixed [39]. In this optimization process, the density was evaluated around each atom with a fixed radius r usually taken as 0.65 Å for all atoms except hydrogen, for which either $r = r_{\text{H}} = 0.325$ Å or $r_{\text{H}} = 0.31$ Å have been recommended [48, 41, 60]. Using these values, we have thus obtained two alternative sets of multipoles for the aspartate molecule in its most stable configuration, which we denote as set I and set II. Likewise, two corresponding sets were obtained for the other configuration in which the proton has transferred to the other carboxylate. Only one isomer exists at the MP2 level (**Fig. 5.1**), protonation of the other site leading to the transfer back to the initial site. However, the two isomers differ when Asp is modelled using a non reactive force field because two distinct sets of parameters are required.

In order to get an energetic reference property that should be reproduced by the force field, we used the difference in MP2 single-point energies between the AMOEBA equilibrium configurations of the two isomers. The performance of these multipolar sets against MP2 data is illustrated in **Table 5.1**, and the IR spectra obtained at 300 K with the AMOEBA force field and the three sets of multipoles are represented in **Fig. 5.3**. Overall, the two spectra with reparametrized multipoles are very similar to each other, but they differ in the 1200–1600 cm⁻¹ wavelength range from the spectrum obtained using the original set of multipoles. While these results confirm the importance of the multipolar electrostatic description on IR spectra, they do not allow for a clear distinction between the reparametrized sets I and II based on experimental comparison.

In the absence of spectroscopic discrimination, we turn to energetics and notice that the original set of multipoles leads to a larger error in the relative stabilities between the two isomers. In contrast, the two new sets of multipoles significantly improve this situation. Following previous studies and based on this result, we chose the set II of multipole parameters, using a value $r_{\text{H}} = 0.31$ Å (**Table 5.2.3**).

The IR spectra obtained from AMOEBA for the most stable conformer are

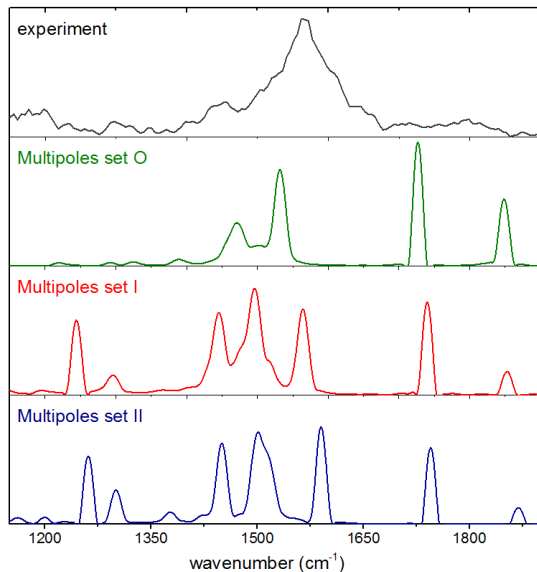


Figure 5.3: Non-reactive AMOEBA spectra for the aspartate molecule obtained at 300 K for the three sets of multipoles O, I, and II as defined in the text.

Multipolar set	original (O)	new (I)	new (II)
r_H (Å)	0.65	0.325	0.31
ΔE (kcal/mol)			
MP2/aug-cc-pVTZ	-2.29	-2.30	-2.05
AMOEBA	-11.09	-1.31	-1.15

Table 5.1: Relative energies of aspartate at the MP2/cc-pVTZ level and with the AMOEBA force field using the original multipole set (O) and two newly proposed versions I and II differing by the value of r_H where the reference electrostatic potential is evaluated around each hydrogen atom during the adjustment process.

clearly unable to account for the experimental features in this spectral range. In particular, the C=O stretches remain excessively resolved and shifted to the blue below 1600 cm^{-1} and above 1750 cm^{-1} . Before invoking proton sharing, it is useful to consider additional possible causes for such discrepancies, namely the presence of the other isomer differing in the protonated site, as well as temperature effects that are notorious in broadening and shifting the vibrational peaks [61]. AMOEBA simulations with multipole set II were thus repeated

starting from the other conformer and at different temperatures (100, 200 and 400 K). The calculated spectra are shown in **Fig. 5.4**. Both the consideration of another isomer and the inclusion of temperature effects are insufficient to reach experimental agreement.

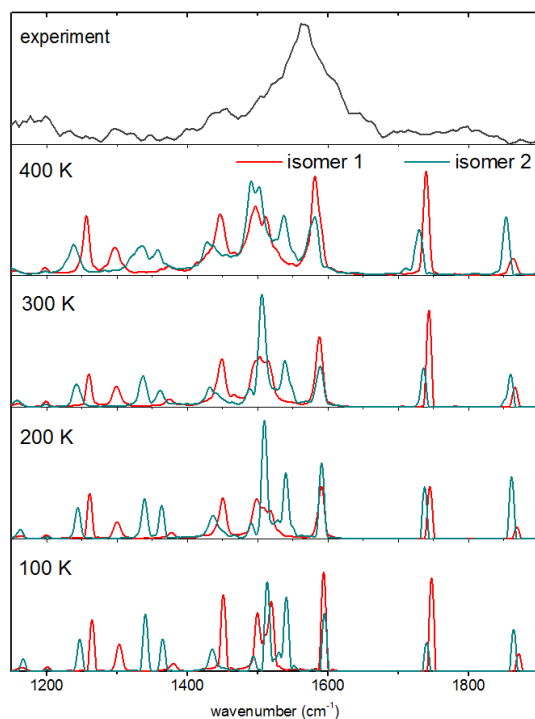


Figure 5.4: Finite temperature spectra obtained from the non-reactive AMOEBA model with multipole set II and for the two isomers differing in the protonated site, at four temperatures. The experimental spectrum is shown in the upper panel.

5.3 Empirical valence-bond modeling of proton transfer

The failure of both static QM methods and classical MD using a non-reactive force field to reproduce the experimental IR spectra suggests that alternative approaches are necessary in order to account for the essential missing feature, i.e. the dynamical signature of proton sharing between the two carboxylate sites. In principle, such a property should be computed using *ab initio* MD [17] or classical but reactive force fields [37]. However, given the computational cost or the extensive parameterization used in these approaches, the EVB framework [26, 15] combined with the AMOEBA polarizable force field was chosen instead. The approach turns out to be highly accurate to reproduce the missing spectral characteristics, giving thus a great insight into their dynamical origin.

5.3.1 2-state EVB model

The systems under consideration here are amino acids with two carboxylic groups sharing a proton (**Fig. 5.1**). The EVB methodology was used to represent the ground state energy of the delocalized system as the linear combination of two diabatic valence states ψ_i , each of them i having the proton only between oxygens of the two COO^- groups facing each other:

$$\Psi = g_1\psi_1 + g_2\psi_2 \quad (5.1)$$

Strictly speaking, the molecules contain four carboxylate oxygen atoms that correspond to four possible protonation sites. However, MD simulations indicate that the rotation of a COO^- group is a rare event considering the timescale of the simulations. This mere observation is confirmed by evaluating the barrier for rotation, which for the present force field is found to be 13 kcal/mol, in the same range than *ab initio* values. Therefore, a two-state description can be safely used.

AMOEBA is used to describe the potential energy V_i of each diabatic state in which the proton is covalently bound to either of the two oxygens participating to the exchange, and only hydrogen-bonded to the other oxygen. In matrix representation, the 2×2 Hamiltonian can be written with V_{12} being the coupling between the two diabatic states allowing proton sharing. For the present 2-state

model the lowest eigenvalue of the Hamiltonian explicitly reads

$$\begin{aligned} V_{\text{EVB}} &= \frac{1}{2} \left[V_1 + V_2 - \sqrt{(V_1 - V_2)^2 + 4V_{12}^2} \right] \\ &= \frac{1}{2} (V_1 + V_2 - D), \\ \text{with } D &= [(V_1 - V_2)^2 + 4V_{12}^2]^{1/2} \end{aligned} \quad (5.2)$$

Similarly the excited state is given by $\frac{1}{2}(V_1 + V_2 + D)$. The components (g_1, g_2) of the eigenvector corresponding to the EVB state on the (V_1, V_2) basis are as follows:

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} [1 + (V_{\text{EVB}} - V_1)^2/V_{12}^2]^{-1/2} \\ (1 - g_1^2)^{1/2} \end{pmatrix} \quad (5.3)$$

with the eigenvector associated to the excited state being $(-g_2, g_1)$.

The trajectories on the EVB surface are generated by propagating the equations of motion using the forces deriving from the potential V_{EVB} :

$$\frac{\partial V_{\text{EVB}}}{\partial x} = \frac{1}{2} \left(\frac{\partial V_1}{\partial x} + \frac{\partial V_2}{\partial x} \right) - \frac{1}{2D} \left[(V_1 - V_2) \left(\frac{\partial V_1}{\partial x} - \frac{\partial V_2}{\partial x} \right) + 4V_{12} \frac{\partial V_{12}}{\partial x} \right], \quad (5.4)$$

which is equivalent to the expression obtained from the Hellmann-Feynman theorem:

$$\frac{\partial V_{\text{EVB}}}{\partial x} = g_1^2 \frac{\partial V_1}{\partial x} + g_2^2 \frac{\partial V_2}{\partial x} + 2g_1 g_2 \frac{\partial V_{12}}{\partial x}. \quad (5.5)$$

Finally, physical observables such as the dipole moment $\vec{\mu}$ are obtained by linear combination over their values on the diabatic states, e.g.

$$\vec{\mu}_{\text{EVB}} = g_1^2 \vec{\mu}_1 + g_2^2 \vec{\mu}_2. \quad (5.6)$$

Implementation of EVB model with AMOEBA

The two-state EVB model used here relies on three ingredients V_1 , V_2 and V_{12} all based on analytical representations of the corresponding energies. For V_1 and V_2 , the valence terms are derived from the MM3 force field and are expressed using the following form [56]:

$$U_{\text{short-range}} = U_{\text{bond}} + U_{\text{angle}} + U_{b\theta} + U_{\text{oop}} + U_{\text{torsion}}, \quad (5.7)$$

where the various terms on the right-hand side denote the bond stretching, bending, bond-bending coupling, out-of-plane bending, and dihedral torsion terms, respectively. In this work, special attention was paid to the stretching potential between the proton and its covalently bound oxygen, a term which is expanded to fourth order as:

$$U_{\text{bond}}(\ell) = K_s(\ell - \ell_0)^2 [1 + C(\ell - \ell_0) + Q(\ell - \ell_0)^2], \quad (5.8)$$

and for which K_s is the stretching constant, C and Q two factors for the cubic and quartic terms, respectively. In conventional MM3 and AMOEBA these two terms adopt values chosen to match the variations of a Morse potential with the same equilibrium distance, frequency and well depth, and read $C = -2.55 \text{ \AA}^{-1}$ and $Q = (7/12)2.55^2 \text{ \AA}^{-2}$ independently of the nature of the atoms involved in the bond [38]. This single-parameter limitation is problematic for the present situation, in which the covalent bond experienced by the sharing proton is more anharmonic than a regular bond. Hence we have considered giving more flexibility to the potential by adjusting the parameters C and Q in the specific case of the OH bonds involving the mobile proton.

Coupling between diabatic states

The off-diagonal coupling term V_{12} is used to represent the mixing between the two diabatic structures that is necessary to allow for proton exchange. A broad variety of expressions have been proposed in the literature to model this term, usually motivated by computational efficiency. Typically, the coupling potential for a proton shared by two atomic sites depends on the distance R between these two sites [14, 62, 63] and the distance q of the proton [62, 63] to the center of the line separating these sites (**Fig. 5.1**),

$$q = \left\| \vec{r}_{\text{H}} - \frac{1}{2}(\vec{R}_{\text{O}_1} + \vec{R}_{\text{O}_2}) \right\|.$$

In the case of protonated water, Vuilleumier and Borgis used a simple form for $V_{12} = A \exp(-\alpha R^2 - \gamma q^2)$, with A , α and γ three parameters [62].

For the present problem, the form chosen is inspired by the expression proposed by Vuilleumier and Borgis, but with a smooth decay when $q \geq 0.4 \text{ \AA}$. It still employs both R and q variables as well but a slightly different form for the off-diagonal term,

$$V_{12} = A \frac{\exp(-\alpha R)}{1 + \gamma q^2}, \quad (5.9)$$

with A , α and γ to be adjusted.

Compared to other forms available in the literature such as those advocated by Voth and coauthors [15, 63], the present form for the coupling is deliberately simple, for two reasons. First, we mainly rely on the accuracy of the AMOEBA force field to describe the main part of the proton motion. Indeed, the role of the coupling is not to govern the simulation, but just to connect the two valence states together. It can be noted that the form does not depend on the $\widehat{\text{O}_1\text{O}_2\text{H}}$ angle, which could be necessary to account for the deviation from linearity of the hydrogen bond. To some extent we expect the presence of the multipoles to

account for a satisfactory orientation of the bond. However, in order to further justify the expression above we have performed MP2/cc-pVTZ calculations of the $\widehat{O_1O_2H}$ angle around the proton-sharing configurations, varying the distance R and evaluating this angle without constraining the geometry except along this coordinate. The results of these calculations, shown in **Fig. 5.5**, indicate that the angle is always below 10° in the relevant range of O-O distance in the 2.4-2.9 Å range, which corresponds to an usual hydrogen bond, thereby providing a better justification for the use of R and q only as the useful coordinates.

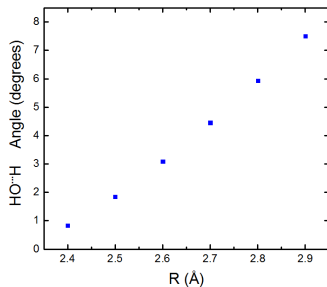


Figure 5.5: Optimal $\widehat{O_1O_2H}$ angle between the vectors $\overrightarrow{O_1O_2}$ and $\overrightarrow{O_2H}$ as a function of the O_1O_2 distance, as obtained from MP2/cc-pVTZ calculations without constraint except on O_1O_2 during the geometry optimization.

A second justification for the relatively simple form of the off-diagonal EVB coupling elements is to be found in the hopefully better transferability to chemically similar systems without having to go through further parametrization. We also note that no additional properties such as dissociation energies or pKa, as previously considered by several authors [25, 14, 64] were

included here in the adjustment process.

The parameters of the EVB model were optimized to reproduce *ab initio* energy surfaces initially obtained for the adipic acid $(CH_2)_4(COOH)_2$, a molecule similar to aspartate [65], and found to be satisfactorily transferable to aspartate. The *ab initio* calculations employed MP2 theory and the cc-pVTZ basis set, they were limited to ranges in R and q relevant to PT between two nearby oxygen atoms, leading to $A = 420$ kcal/mol, $\alpha = 0.745 \text{ \AA}^{-1}$ and $\gamma = 5.35 \text{ \AA}^{-2}$. The EVB parameters we obtained by standard least-square fitting of the EVB-AMOEBA potential energy surface against QM calculations, using a zero-temperature Monte Carlo procedure in the space of parameters and periodically decreasing the maximum amplitude of the random moves.

A slight further adjustment was made in the case of aspartate based on

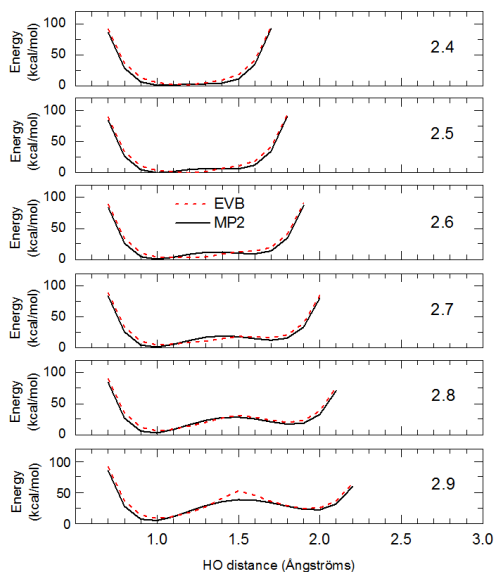


Figure 5.6: MP2/cc-pVTZ and EVB-AMOEBAs potential energy surfaces obtained along the coordinates R and r_{OH} causing PT between the two carboxylate sites.

reference MP2/cc-pVTZ calculations performed by constrained geometry optimizations for each R distance, followed by a scan to fix the position of the hydrogen, leading to 96 energy points. The same geometries were kept to compute the AMOEBA energies with both V_1 and V_2 potentials for the two protonation sites, respectively. Based on these calculations the parameter A was changed to $A = 370$ kcal/mol. Moreover, the cubic parameter C of the U_{bond} term was adjusted to -2.60 \AA^{-1} in order to achieve a better curvature of the surface for larger R distances.

The performance of the EVB model for aspartate can be gauged from **Fig. 5.6** where several potential energy profiles are shown following the R and r_{OH} coordinates. The level of agreement is overall very satisfactory, the EVB potential appearing capable of reproducing the relative energies of the two minima and the energy barrier along the proton transfer coordinate to within 4 kcal/mol at 2.7 \AA .

This EVB model was implemented in Tinker 7 [40] and the scheme in **Fig. 5.7** presents the procedure. The main challenge was to coexist two force fields at each step of the dynamics while limiting the additional computational cost with respect to the original implementation of AMOEBA. In order to avoid

I/O bottlenecks, the data necessary to compute V_1 and V_2 energies and the corresponding forces are stored in memory in the initial step. A supplementary cost is due to the parameter attribution on each atom. At each time step of the dynamics, V_1 , V_2 and their gradients are computed, followed by the calculation of the EVB coupling term and its gradient. Then, velocities and accelerations are computed to generate the next step, and the dipole moment is recorded. For the next step, the order of the two isomers forming the two states is reversed to limit the initialization procedure to one state only. For a 50 ps simulation, the EVB-AMOEBA simulation is approximately 1.6 times longer than the standard AMOEBA simulation.

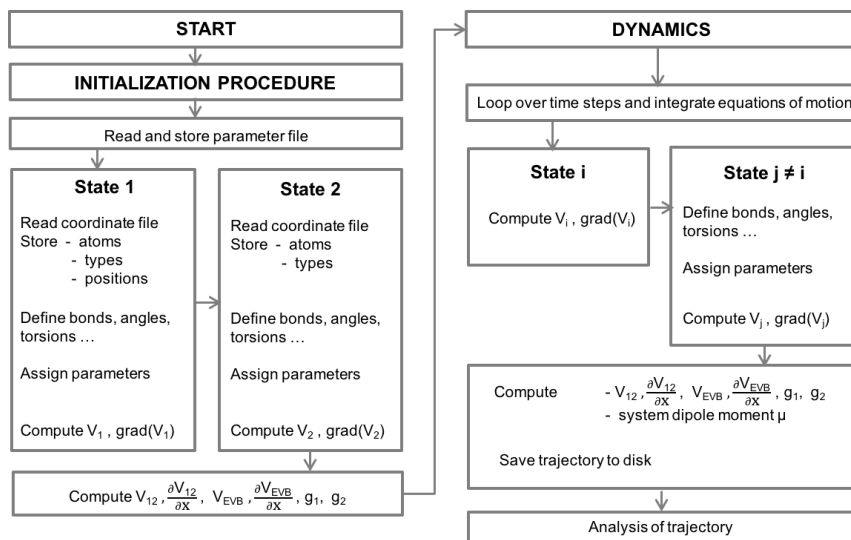


Figure 5.7: Implementation of the EVB model in Tinker 7.

5.3.2 Dynamics under the EVB-AMOEBA potential

MD simulations have been repeated with the form of the EVB potential and the associated parameters described above. The time evolution of several properties is depicted in **Fig. 5.8** along a typical trajectory at 300 K. In this trajectory, the proton is mainly localized over either of the carboxylate sites, as shown by the corresponding EVB weight g_i being close to 1. The analysis of the trajectory highlights that this occurs about 95% of the time. Occasionally, the proton can be trapped between the two sites, as occurs in the portion between 6 and 8 ps [**Fig. 5.8(a)**]. Proton sharing is manifested by the two coefficients g_i

being both close to $1/\sqrt{2}$, and under such conditions the geometric variables R and q appear to experience much smoother variations, which indicates a much more rigid conformation with the proton lying near the middle of the OO bond (small q values). The mechanism of PT is also visualized in **Fig. 5.9** where three snapshots of the molecule before, during and after PT are shown, as occurring after about 1.5 ps in **Fig. 5.8**.

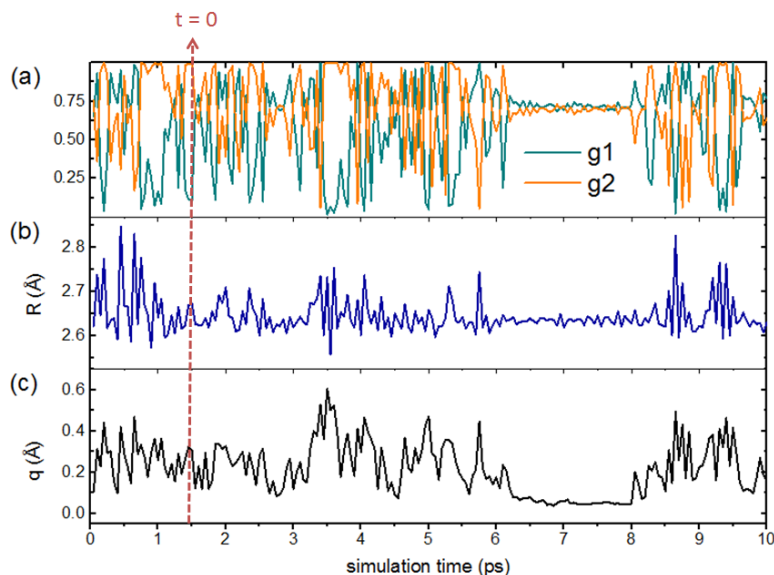


Figure 5.8: 10-ps EVB-AMOEB simulation at 300 K, (a) evolution of g_1 (green), g_2 (orange), (b) R and q distances (in Å). Short-time averaging over 50 fs windows has been used to improve visual quality.

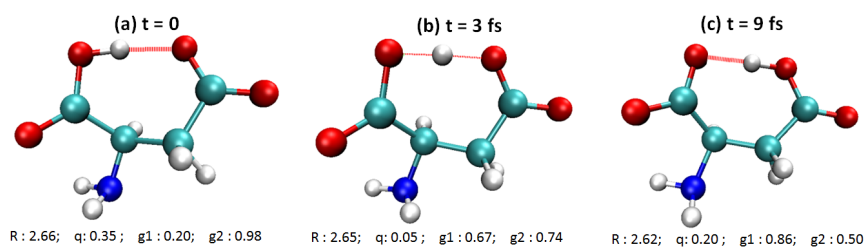


Figure 5.9: Proton exchange during an EVB-AMOEB MD simulation. The geometrical parameters R and q , the EVB weights g_1 and g_2 and the time frames are indicated.

Looking in more details into how the geometric parameters vary during such a reactive trajectory, the distance R between the protonated sites fluctuates between 2.5 and 2.9 Å and the proton exchange appears to be fast as the g_i coefficients oscillate between 0 and 1, unless the proton is locked in the shared position. The distributions of the R and q distances visited during 10 EVB trajectories of 50 ps each are represented in **Fig. 5.10(a)** and **(b)**, respectively. The O-O distance is spread out over and below the static equilibrium value of 2.68 Å, such a spreading being of course caused by the finite temperature of the simulation. More interestingly, the average R distance appears at a lower value closer to 2.62 Å and also corresponding to the distance of the proton-shared configuration in **Fig. 5.8(b)**. The distribution of q distances appears bimodal, which is the expected behavior in presence of proton sharing. Here the distribution extends on both sides of the equilibrium which only corresponds to the proton being on the energetically lowest carboxylate site.

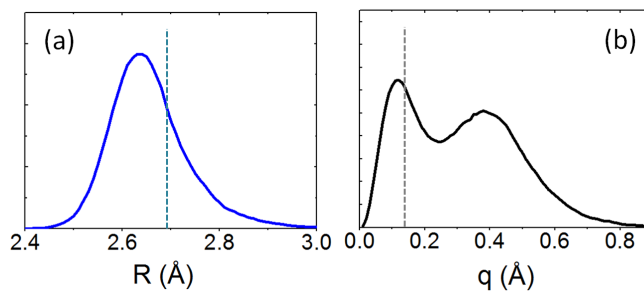


Figure 5.10: Distance distributions for (a) R and (b) q , as averaged on 10 simulations of 50 ps each at 300 K.

5.3.3 Infrared spectra with the EVB-AMOEBA approach

Before discussing the spectroscopic manifestations of proton sharing as predicted by the present model, we consider a simpler situation in which proton transfer cannot operate, i.e. where the molecule lies in its linear (extended) conformation with the two carboxylate sites opposite to each other. In this case the coupling element of the EVB matrix is extremely small and the two basis functions are effectively uncoupled, the proton being stuck at its initial site. The infrared spectrum obtained at 100 K for the linear conformer of aspartate with the EVB model and the non reactive AMOEBA model applied here with the same set of multipoles are barely distinguishable from each other (data not shown). More importantly, and even considering the colder conditions

of these calculated spectra, they deviate qualitatively from the experimentally measured spectrum. Incidentally this allows us to discard the linear conformer from further spectroscopic consideration.

The IR spectrum obtained for the bent conformer from the EVB method under the conditions discussed in the previous section is shown in **Fig. 5.11** together with the experimental spectrum. We have also reported the spectra obtained from the same EVB approach but with the other two sets of multipole parameters denoted as O and I, all at 300 K and using 10 independent trajectories of 50 ps each. The main change in comparison with the non-reactive spectra is that CO modes bring closer to provide a clear signature of proton exchange with a broad band similar to the experimental one. However, the spectral features clearly exhibit differences depending on the choice of the multipolar set, and the equilibrium distributions of R and q distances obtained with the three parametrizations have been represented in **Fig. 5.12** to assist interpretation of these results. The spectrum obtained with the original set of multipoles from AMOEBAbio-09 predicts a CO stretching band excessively shifted to the blue. The distribution of q distances is largely unimodal, the shoulder on the larger values indicating minor PT.

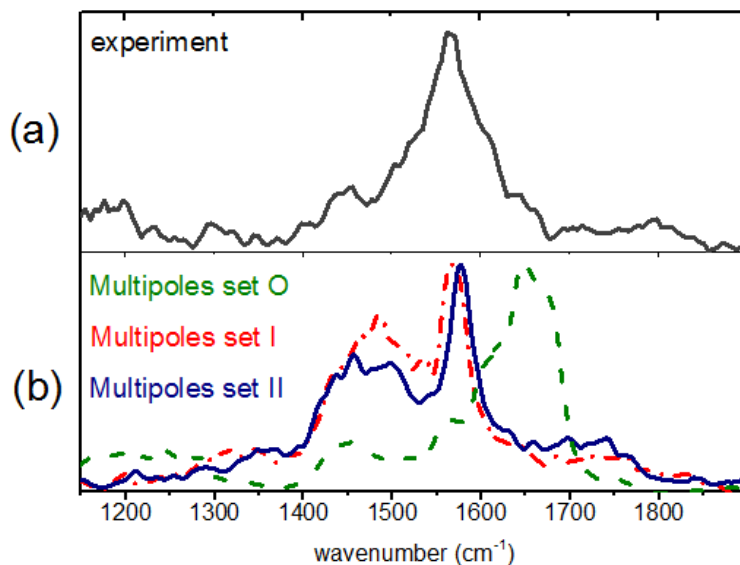


Figure 5.11: (a) Experimental (black line) and (b) calculated IR spectra at 300 K, as obtained from the EVB approach with the three multipole parameter sets O, I, and II.

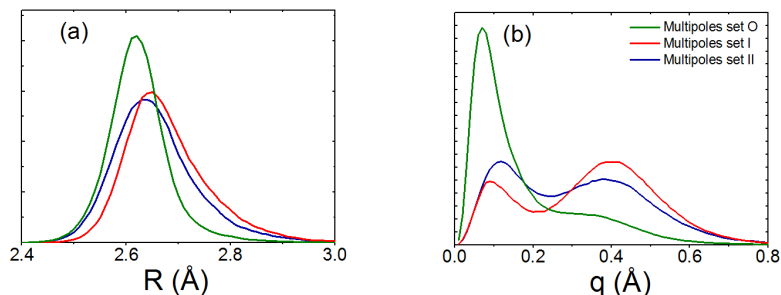


Figure 5.12: Equilibrium distributions of (a) R and (b) q , as obtained from the EVB simulations at 300 K using the three multipolar sets O, I, and II.

In contrast, the multipolar sets I and II obtained from taking $r_H = 0.325$ Å and $r_H = 0.31$ Å, respectively, in the adjustment process predict a much more balanced bimodal distribution in q , suggesting a much stronger importance of proton-exchange configurations. Both spectra match very well the intensity maximum near 1575 cm^{-1} and give also a broad contribution around 1450 cm^{-1} which is undoubtedly present in the measured spectrum although with lower relative intensity.

Having validated the present EVB-AMOEBA model for spectroscopic characterization at room temperature, we have investigated the dynamics of the system under other conditions. Temperature plays a key role in PT, and in the classical description of nuclear motion it is expected to facilitate such transfer and possibly promote proton sharing. In **Fig. 5.13** we have represented the IR spectra obtained from the EVB model with multipole set II at 50, 100, 300 and 400 K, all from 10 trajectories of 50 ps each. As anticipated, the spectra at low temperature appear much more resolved, with the peaks increasingly broadening and shifting as temperature increases.

At low temperature $T < 100$ K, the proton is shared between the two sites at $R = 2.6$ Å and small values of q , which can be directly visualized on the distributions of these distances in **Fig. 5.14(a)** and **(b)**. To confirm that the proton is indeed shared by the two carboxylate sites we have also calculated the distributions of the EVB weights g_1 and g_2 . These are shown in **Fig. 5.14(c)** and **(d)**, and clearly exhibit highly probable regions where g_1 and g_2 are both close to $1/\sqrt{2}$, indicating nearly equal delocalization on both sites.

A transition to unequally shared proton on both sides of the barrier occurs already at 100 K, where the distribution of q becomes bimodal, the EVB weights localising on more extreme values close to 0 and 1. The spectroscopic signature

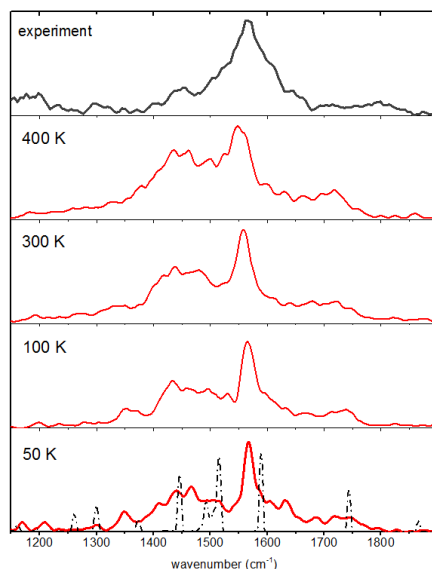


Figure 5.13: EVB-AMOEBA spectra as a function of temperature between 50 and 400 K, with the experimental spectrum being reported again in the upper panel. The non-reactive AMOEBA spectrum at low temperatures is also superimposed in dot-dashed black lines.

of this transition is only quite modest in **Fig. 5.13**, the direct temperature effects being notably difficult to disentangle from the pure contributions of proton hopping versus proton sharing. At this temperature the EVB spectrum significantly differs from the non-reactive spectra obtained with AMOEBA only, and it could not be accounted for either by superimposing the non-reactive spectra obtained with the proton on each of the sites.

As temperature increases, the proton remains in a mixed state where it constantly hops between the two carboxylate sites. The contributions to the C=O stretch modes are shifted to the red until they reach the experimental position at 300 K, before showing excessive shift and broadening above this temperature. These results show that the dynamical transfer of the proton on both sides of the barrier is essential in reproducing the experimental broadening of the IR spectrum in this range. Furthermore, and although such spectral features are intimately related to the motion of the proton between the two carboxylate sites, they do not directly provide evidence for the nature of this proton motion. Only through a detailed analysis of the trajectories, we can interpret the proton behavior as evolving from an equally shared configuration stabilized by

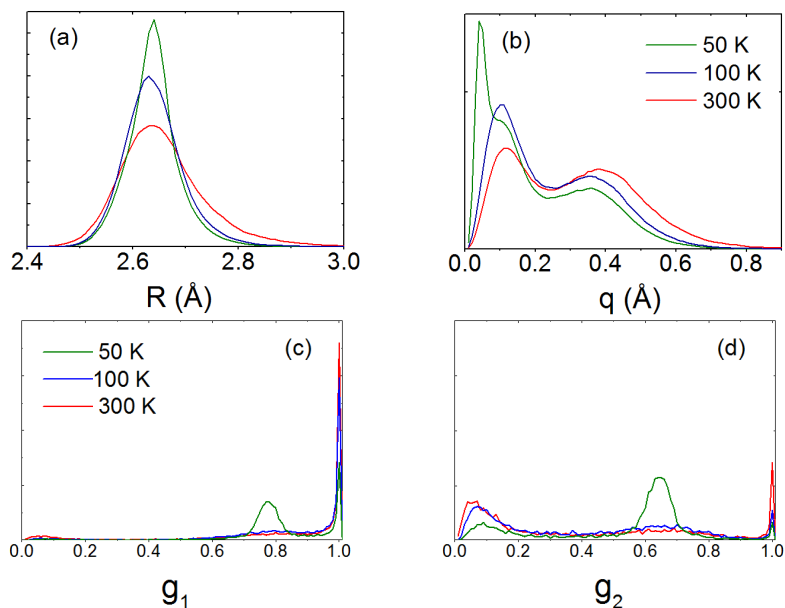


Figure 5.14: Equilibrium distributions of (a) R and (b) q obtained at 50, 100 and 300 K, and distributions of the EVB weights (c) g_1 and (d) g_2 , as obtained at 50, 100 and 300 K.

resonance mechanism at low temperature to a constant hopping motion between the carboxylate sites through an intermediate barrier at 100 K and above. Such a qualitative change in the motion of the mobile proton originates from the thermal motion of the molecular backbone: as temperature increases, the R distance enlarges [Fig. 5.14(a)], which results in a stronger barrier to proton delocalization [Fig. 5.6] and a change in the nature of proton motion.

5.4 Concluding remarks

In this chapter, we discussed our implementation of a two-state EVB model for the computation of IR spectra, in particular for systems containing a signature of proton motion in flexible molecules. In this model we used the AMOEBA polarizable force field to describe the potential energy of the diabatic states and a simple form of the coupling potential, following here previous works by Vuilleumier and Borgis for protonated water [62]. As a representative example, the EVB-AMOEBA model was applied to Asp deprotonated amino acid, for which an experimental gas-phase IR spectrum was newly recorded. The ex-

perimental spectrum provides a clear spectroscopic signature of a dynamically shared proton between two carboxylate groups, which was not possible to reproduce by standard QM and MD approaches. A careful parameterization from *ab initio* calculations, used to obtain the multipoles for the AMOEBA potential and the off-diagonal coupling term in the EVB model, was necessary to get a consistent dynamical picture of the proton sharing and an accurate spectral signature.

The work described in this chapter represents the first step in the extension of the polarizable AMOEBA force field to address reactivity in relation with characteristic spectroscopic features. Beyond this first successful assessment, further developments are in progress in order to extend the methodology to a general two-state model in which the protonation sites may differ, and subsequently to develop a multi-state EVB model. While there is no fundamental restriction to the application of polarizable force fields as the building block of a multistate EVB model, technicalities and the significant computational cost associated with the multipolar expansion will require efficient parallelization algorithms in order to address time scales compatible with low-frequency motions and slow rearrangements. The ultimate goal will be to extend the force field capabilities so they include proton transfer reactions in peptides and in protein active sites in relation with experimental infrared signatures. Finally, the model will be used to study ionic chemistry in large water droplets in the presence of acidic or basic impurities such as sulfuric acid, and their infrared spectroscopy.

5.5 Acknowledgments

FT is grateful to Ecole polytechnique for PhD fellowship. Implementation of EVB model was supported by travel grants from EMIE GDR 3533 funded by CNRS. Use of the FT-ICR mass spectrometer at Orsay was supported by the "Spectrométrie de masse FT-ICR à très haut champ" research infrastructure funded by the CNRS Institute of Chemistry. We are grateful to Edith Nicol for carrying out the IRMPD experiments on deprotonated Asp and to the CLIO team for support.

Bibliography

- [1] Marx, D. (2006) Proton Transfer 200 Years after von Grotthuss: Insights from Ab Initio Simulations. *ChemPhysChem*, **7** (9), 1848–1870.
- [2] Tuckerman, M.E., Marx, D., and Parrinello, M. (2002) The nature and transport mechanism of hydrated hydroxide ions in aqueous solution. *Nature*, **417** (6892), 925–929.
- [3] Wraight, C.A. (2006) Chance and design—Proton transfer in water, channels and bioenergetic proteins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1757** (8), 886–912.
- [4] Knight, C. and Voth, G.A. (2012) The Curious Case of the Hydrated Proton. *Accounts of Chemical Research*, **45** (1), 101–109.
- [5] Semrouni, D., Balaj, O., Calvo, F., Correia, C., Clavaguéra, C., and Ohanessian, G. (2010) Structure of sodiated octa-glycine: IRMPD spectroscopy and molecular modeling. *J. Am. Soc. Mass Spectrom.*, **21**, 728–738.
- [6] Maître, P., Caër, S.L., Simon, A., Jones, W., Lemaire, J., Mestdagh, H., Heninger, M., Mauclaire, G., Boissel, P., Prazeres, R., Glotin, F., and Ortega, J.M. (2003) Ultrasensitive spectroscopy of ionic reactive intermediates in the gas phase performed with the first coupling of an IR FEL with an FTICR-MS. *Nucl. Instrum. Methods Phys. Res., Sect. A*, **507** (1-2), 541 – 546.
- [7] MacAleese, L. and Maître, P. (2007) Infrared spectroscopy of organometallic ions in the gas phase: From model to real world complexes. *Mass Spectrom. Rev.*, **26** (4), 583–605.
- [8] Polfer, N.C. and Oomens, J. (2009) Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass Spectrom. Rev.*, **28** (3), 468–494.

- [9] Gaigeot, M.P., Vuilleumier, R., Sprik, M., and Borgis, D. (2005) Infrared Spectroscopy of N-Methylacetamide Revisited by ab Initio Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, **1** (5), 772–789.
- [10] Gaigeot, M.P. (2010) Theoretical spectroscopy of floppy peptides at room temperature. A DFTMD perspective: gas and aqueous phase. *Phys. Chem. Chem. Phys.*, **12** (14), 3336–3359.
- [11] Ingrosso, F., Monard, G., Hamdi Farag, M., Bastida, A., and Ruiz-López, M.F. (2011) Importance of Polarization and Charge Transfer Effects to Model the Infrared Spectra of Peptides in Solution. *J. Chem. Theory Comput.*, **7** (6), 1840–1849.
- [12] Rossi, M., Blum, V., Kupser, P., von Helden, G., Bierau, F., Pagel, K., Meijer, G., and Scheffler, M. (2010) Secondary Structure of Ac-Ala(n)-LysH(+) Polyalanine Peptides (n=5, 10, 15) in Vacuo: Helical or Not? *J. Phys. Chem. Lett.*, **1** (24), 3465–3470.
- [13] Thomas, M., Brehm, M., Fligg, R., Vohringer, P., and Kirchner, B. (2013) Computing vibrational spectra from ab initio molecular dynamics. *Phys. Chem. Chem. Phys.*, **15** (18), 6608–6622.
- [14] Lobaugh, J. and Voth, G.A. (1996) The quantum dynamics of an excess proton in water. *The Journal of Chemical Physics*, **104** (5), 2056–2069.
- [15] Schmitt, U.W. and Voth, G.A. (1999) The computer simulation of proton transport in water. *The Journal of Chemical Physics*, **111** (20), 9361–9381.
- [16] Biswas, R., Tse, Y.L.S., Tokmakoff, A., and Voth, G.A. (2016) Role of Pre-solvation and Anharmonicity in Aqueous Phase Hydrated Proton Solvation and Transport. *The Journal of Physical Chemistry B*, **120** (8), 1793–1804.
- [17] Marinica, D.C., Grégoire, G., Desfrancois, C., Schermann, J.P., Borgis, D., and Gaigeot, M.P. (2006) Ab Initio Molecular Dynamics of Protonated Dialanine and Comparison to Infrared Multiphoton Dissociation Experiments. *The Journal of Physical Chemistry A*, **110** (28), 8802–8810.
- [18] Dopieralski, P.D., Latajka, Z., and Olovsson, I. (2010) Proton-transfer dynamics in the HCO₃ dimer of KHCO₃ from Car-Parrinello and path-integrals molecular dynamics calculations. *Acta Crystallographica Section B*, **66** (2), 222–228.

- [19] Vendrell, O., Gatti, F., and Meyer, H.D. (2007) Dynamics and Infrared Spectroscopy of the Protonated Water Dimer. *Angewandte Chemie International Edition*, **46** (36), 6918–6921.
- [20] Hammer, T. and Manthe, U. (2011) Intramolecular proton transfer in malonaldehyde: Accurate multilayer multi-configurational time-dependent Hartree calculations. *The Journal of Chemical Physics*, **134** (22), 224305.
- [21] Vuilleumier, R. and Borgis, D. (1999) Transport and spectroscopy of the hydrated proton: A molecular dynamics study. *J. Chem. Phys.*, **111** (9), 4251–4266.
- [22] Marciante, M. and Calvo, F. (2014) Modelling infrared action spectra of protonated water clusters. *Molecular Simulation*, **40** (1-3), 176–184.
- [23] Goldblum, A. (1988) Calculation of proton transfers in hydrogen bonding interactions with semi-empirical MNDO/H. *Journal of Molecular Structure: THEOCHEM*, **179** (1), 153 – 163.
- [24] Wolf, M.G. and Groenhof, G. (2014) Explicit proton transfer in classical molecular dynamics simulations. *Journal of Computational Chemistry*, **35** (8), 657–671.
- [25] Warshel, A. and Weiss, R.M. (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes. *Journal of the American Chemical Society*, **102** (20), 6218–6226.
- [26] Warshel, A. (2002) Molecular Dynamics Simulations of Biological Reactions. *Accounts of Chemical Research*, **35** (6), 385–395.
- [27] Harvey, J.N., O’Connor, M., , and Glowacki, D.R. (2016) *Theory and Applications of the Empirical Valence Bond Approach: From Physical Chemistry to Chemical Biology*, John Wiley & Sons, Ltd., chap. Empirical Valence Bond Methods for Exploring Reaction Dynamics in the Gas Phase and in Solution.
- [28] Gaigeot, M.P., Cimas, A., Seydou, M., Kim, J.Y., Lee, S., and Schermann, J.P. (2010) Proton Transfer from the Inactive Gas-Phase Nicotine Structure to the Bioactive Aqueous-Phase Structure. *Journal of the American Chemical Society*, **132** (51), 18067–18077.
- [29] Wang, S., MacKay, L., and Lamoureux, G. (2014) Development of Semiempirical Models for Proton Transfer Reactions in Water. *Journal of Chemical Theory and Computation*, **10** (8), 2881–2890.

- [30] Wu, X., Thiel, W., Pezeshki, S., and Lin, H. (2013) Specific Reaction Path Hamiltonian for Proton Transfer in Water: Reparameterized Semiempirical Models. *Journal of Chemical Theory and Computation*, **9** (6), 2672–2686.
- [31] Vener, M., Rostov, I., Soudackov, A., and Basilevsky, M. (2000) Semiempirical modeling free energy surfaces for proton transfer in polar aprotic solvents. *Chemical Physics*, **254** (2–3), 249 – 265.
- [32] Lill, M.A. and Helms, V. (2001) Molecular dynamics simulation of proton transport with quantum mechanically derived proton hopping rates (QHOP MD). *The Journal of Chemical Physics*, **115** (17), 7993–8005.
- [33] Pezeshki, S. and Lin, H. (2015) Adaptive-Partitioning QM/MM for Molecular Dynamics Simulations: 4. Proton Hopping in Bulk Water. *Journal of Chemical Theory and Computation*, **11** (6), 2398–2411.
- [34] Riccardi, D., Yang, S., and Cui, Q. (2010) Proton transfer function of carbonic anhydrase: Insights from QM/MM simulations. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1804** (2), 342 – 351.
- [35] van Duin, A.C.T., Dasgupta, S., Lorant, F., and Goddard-III, W.A. (2001) ReaxFF: A Reactive Force Field for Hydrocarbons. *The Journal of Physical Chemistry A*, **105** (41), 9396–9409.
- [36] Nagy, T. and Meuwly, M. (2016) *Theory and Applications of the Empirical Valence Bond Approach: From Physical Chemistry to Chemical Biology*, John Wiley & Sons, Ltd., chap. Modelling Chemical Reactions Using Empirical Force Fields.
- [37] van Duin, A.C.T., Zou, C., Joshi, K., Bryantsev, V., and Goddard, W.A. (2014) CHAPTER 6 A Reaxff Reactive Force-field for Proton Transfer Reactions in Bulk Water and its Applications to Heterogeneous Catalysis, in *Computational Catalysis*, The Royal Society of Chemistry, pp. 223–243.
- [38] Ponder, J.W., Wu, C., Ren, P., Pande, V.S., Chodera, J.D., Schnieders, M.J., Haque, I., Mobley, D.L., Lambrecht, D.S., DiStasio, R.A., Head-Gordon, M., Clark, G.N.I., Johnson, M.E., and Head-Gordon, T. (2010) Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B*, **114** (8), 2549–2564.
- [39] Ren, P., Wu, C., and Ponder, J.W. (2011) Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.*, **7** (10), 3143–3161.

- [40] Ponder, J.W. (2015), TINKER - Software Tools for Molecular Design (version 7), <http://dasher.wustl.edu/tinker>.
- [41] Semrouni, D., Sharma, A., Dognon, J.P., Ohanessian, G., and Clavaguéra, C. (2014) Finite Temperature Infrared Spectra from Polarizable Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, **10** (8), 3190–3199.
- [42] Thaunay, F., Dognon, J.P., Ohanessian, G., and Clavaguera, C. (2015) Vibrational mode assignment of finite temperature infrared spectra using the AMOEBA polarizable force field. *Physical Chemistry Chemical Physics*, **17** (39), 25 968–25 977.
- [43] Oomens, J., Steill, J.D., and Redlich, B. (2009) Gas-Phase IR Spectroscopy of Deprotonated Amino Acids. *Journal of the American Chemical Society*, **131** (12), 4310–4319.
- [44] Balaj, O., Semrouni, D., Steinmetz, V., Nicol, E., Clavaguéra, C., and Ohanessian, G. (2012) Structure of Sodiated Polyglycines. *Chem. Eur. J.*, **18**, 4583–4592.
- [45] 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., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M., Knox, J.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., and Fox, D.J., Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT 2009.
- [46] Merrick, J.P., Moran, D., and Radom, L. (2007) An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A*, **111** (45), 11 683–11 700.
- [47] Semrouni, D., Clavaguéra, C., Dognon, J.P., and Ohanessian, G. (2010)

- Assessment of density functionals for predicting the infrared spectrum of sodiated octa-glycine. *Int. J. Mass Spectrom.*, **297** (1–3), 152–161.
- [48] Rasmussen, T.D., Ren, P., Ponder, J.W., and Jensen, F. (2007) Force field modeling of conformational energies: Importance of multipole moments and intramolecular polarization. *Int. J. Quantum Chem.*, **107** (6), 1390–1395.
- [49] Semrouni, D., Ohanessian, G., and Clavaguera, C. (2010) Structural, energetic and dynamical properties of sodiated oligoglycines: relevance of a polarizable force field. *Phys. Chem. Chem. Phys.*, **12** (14), 3450–3462.
- [50] Grossfield, A., Ren, P., and Ponder, J.W. (2003) Ion solvation thermodynamics from simulation with a polarizable force field. *J. Am. Chem. Soc.*, **125** (50), 15 671–15 682.
- [51] Piquemal, J.P., Perera, L., Cisneros, G.A., Ren, P., Pedersen, L.G., and Darden, T.A. (2006) Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure. *J. Chem. Phys.*, **125** (5), 054511.
- [52] Marjolin, A., Gourlaouen, C., Clavaguéra, C., Ren, P., Wu, J., Gresh, N., Dognon, J.P., and Piquemal, J.P. (2012) Toward accurate solvation dynamics of lanthanides and actinides in water using polarizable force fields: from gas-phase energetics to hydration free energies. *Theo. Chem. Acc.*, **131**, 1–14.
- [53] Semrouni, D., Isley, W.C.I., Clavaguéra, C., Dognon, J.P., Cramer, C.J., and Gagliardi, L. (2013) Ab Initio Extension of the AMOEBA Polarizable Force Field to Fe^{2+} . *J. Chem. Theory Comput.*, **9** (7), 3062–3071.
- [54] Ren, P. and Ponder, J.W. (2002) Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.*, **23** (16), 1497–1506.
- [55] Halgren, T.A. (1992) The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.*, **114** (20), 7827–7843.
- [56] Allinger, N.L., Yuh, Y.H., and Lii, J.H. (1989) Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.*, **111** (23), 8551–8566.
- [57] Fanourgakis, G.S. and Xantheas, S.S. (2008) Development of transferable interaction potentials for water. V. Extension of the flexible, polarizable,

- Thole-type model potential (TTM3-F, v. 3.0) to describe the vibrational spectra of water clusters and liquid water. *J. Chem. Phys.*, **128** (7), 074 506.
- [58] Schultheis, V., Reichold, R., Schropp, B., and Tavan, P. (2008) A Polarizable Force Field for Computing the Infrared Spectra of the Polypeptide Backbone. *J. Phys. Chem. B*, **112** (39), 12 217–12 230.
- [59] Stone, A.J. (1981) Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.*, **83**, 233–239.
- [60] Semrouni, D., Cramer, C.J., and Gagliardi, L. (2014) AMOEBA force field parameterization of the azabenzenes. *Theoretical Chemistry Accounts*, **134**, 1–11.
- [61] Marx, D. and Hutter, J. (2009) *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, Cambridge University Press, Cambridge.
- [62] Vuilleumier, R. and Borgis, D. (1997) Molecular dynamics of an excess proton in water using a non-additive valence bond force field. *Journal of Molecular Structure*, **436–437**, 555 – 565. Structure, Properties and Dynamics of Molecular Systems.
- [63] Maupin, C.M., Wong, K.F., Soudackov, A.V., Kim, S., and Voth, G.A. (2006) A Multistate Empirical Valence Bond Description of Protonatable Amino Acids. *The Journal of Physical Chemistry A*, **110** (2), 631–639.
- [64] Schlegel, H.B., , and Sonnenberg, J.L. (2006) Empirical Valence-Bond Models for Reactive Potential Energy Surfaces Using Distributed Gaussians. *Journal of Chemical Theory and Computation*, **2** (4), 905–911.
- [65] Thauay, F., Calvo, F., Ohanessian, G., and Clavaguéra, C. work in progress.

Deuxième partie

Applications

Chapitre 4

Dynamique des ions hydratés

L'étude de la solvation d'ions dans l'eau représente un vaste domaine de recherche, et trouve son implication dans de nombreux processus chimiques et biologiques [1]. L'effet des ions est souvent illustré par l'exemple des travaux de Hofmeister et al. [2], qui dans les années 1880, ont mesuré la capacité de sels à précipiter l'albumine d'oeuf de poule en solution aqueuse. La 'série de Hofmeister', qui classe les ions en fonction de leur effet sur la solubilité de protéines, est devenue célèbre grâce à sa remarquable reproductivité sur d'autres propriétés en solution, telles que les tensions de surface, la stabilité colloïdale, ou la sélectivité chromatographique [3, 4]. Il est largement admis que la série reflète la capacité des ions à réorganiser le réseau de liaisons hydrogènes [5–7], à plus ou moins longue distance. L'hydratation des ions est un processus exothermique, au cours duquel de l'énergie est libérée grâce à la formation d'interactions électrostatiques entre l'ion et les molécules d'eau. Celles-ci se réarrangent en sphères de solvation, dont l'organisation (leur nombre et leurs tailles) est très dépendante de l'ion. La compréhension des effets structuraux au delà de la première couche de solvation constitue toujours un défi, par le comportement dynamique de l'eau liquide, la présence de contre-ions, et des propriétés qui peuvent être différentes en solution ou aux interfaces. Une étude simplifiée du phénomène est permise par les expériences de microsolvatation [8, 9], qui élimine la contribution du contre-ion et permet, par la sélection en taille de l'agrégat, d'observer une évolution des propriétés de l'ion hydraté. En contrepartie, d'autres questions sont posées, comme la forme que prennent ces agrégats, et la propension de l'ion à y être fortement solvaté, ou au contraire à migrer à sa surface. Dans ces problématiques, la modélisation est un outil de choix. Pour les petits agrégats, les calculs *ab initio* statiques prédisent les structures les plus stables [10, 11] tandis que la dynamique moléculaire classique ou quantique est employée pour les plus grandes tailles, les études aux interfaces liquide/vapeur, ou en solution [12, 13].

Dans ces travaux de thèse, l'objectif principal est de fournir des outils à la modélisation de spectres IR d'ions gazeux ou hydratés, beaucoup reposant sur la dynamique moléculaire et le champ de forces AMOEBA. Le prérequis à l'utilisation du champ de forces dans une problématique d'attribution de spectres IR à température finie, est d'établir sa capacité à reproduire les comportements dynamiques des espèces modélisées. En particulier, pour un ion solvaté, AMOEBA doit pouvoir rendre compte de sa capacité à organiser le réseau de molécules d'eau autour de lui sous la forme de sphères de solvation. Cette capacité peut être évaluée par différents paramètres, tels que le nombre et la mobilité des molécules d'eau en 1^{ère} ou en 2^{ème} sphères. Dans le cas d'ions solvatés dans des nanogouttes, s'ajoute l'étude du comportement privilégié de l'ion, fortement solvaté au sein de l'agrégat, ou au contraire migrant à sa surface.

La capacité du champ de forces polarisable AMOEBA à reproduire ces comportements a été évaluée au cours de simulations de dynamique moléculaire. Les ions étudiés, La^{3+} , Ca^{2+} , Li^+ , Na^+ , K^+ , Cs^+ , F^- , Cl^- , Br^- et SO_4^{2-} , ont fait l'objet d'une calibration ces dernières années [14–17] associés à la version 03 du champ de forces de l'eau de Ren et Ponder [18]. Dans la philosophie AMOEBA, les paramètres des ions sont extraits de calculs *ab initio* [17, 19] visant

à reproduire certaines propriétés, telles que la distance ion-eau, ou les énergies d'interaction avec un petit nombre de molécules d'eau. La première section de ce chapitre décrit l'étude de la dynamique de ces ions dans une nanogoutte de 100 molécules d'eau, avec la version 03 du champ de forces de l'eau et conclut sur sa capacité à reproduire toutes les propriétés mentionnées précédemment. Récemment, les auteurs du champ de forces ont proposé un nouveau jeu de paramètres de l'eau, AMOEBA14, obtenu par une nouvelle procédure de calibration [20] et visant à reproduire de nombreuses propriétés de la phase liquide. AMOEBA14 améliore sensiblement la spectroscopie de l'eau pure, qui était mal reproduite par AMOEBA03 (détails en **chapitre 5**). Pour une application à l'hydratation d'ions, cette version appelle de nouveaux jeux de paramètres pour les ions, qui n'ont pas encore été publiés à ce jour. Avant de se focaliser sur la spectroscopie, on souhaite vérifier que les comportements dynamiques, qui sont bien reproduits en AMOEBA03, restent correctement traités avec AMOEBA14. Cette étude est l'objet de la 2^{ème} section de ce chapitre.

4.1 Dynamique des ions dans une nanogoutte de 100 H₂O (AMOEBA03)

La publication est en fin de rédaction et donc insérée directement dans le manuscrit.

Dynamics of ions in a water drop using the AMOEBA polarizable force field

Florian Thaunay, Gilles Ohanessian, Carine Clavaguéra

LCM, CNRS, Ecole polytechnique, Université Paris Saclay, 91128 Palaiseau, France

Abstract

Various ions carrying a charge from -2 to +3 have been introduced in a spherical drop of 100 water molecules. The behavior of the ions in the aggregate has been followed by molecular dynamics simulations associated to the AMOEBA polarizable force field which is suited for multiply-charged ions. For the whole series, the structural properties of the water organization around the ion agree with available experimental data. The computed residence times of the water molecules in the first coordination shell are consistent with experiments for the cation series and are proposed as a general trend for the anions.

Keywords: Ion hydration, molecular dynamics, polarizable force field, coordination number, residence time, water cluster

Email addresses: gilles.ohanessian@polytechnique.edu (Gilles Ohanessian),
carine.clavaguera@polytechnique.edu (Carine Clavaguéra)

1. Introduction

Understanding the dynamics of simple ions dissolved in water may have wide applications in inorganic chemistry, biological environment or atmospheric chemistry [1]. The approaches to study the ion solvation mechanisms are multiple, both experimental and theoretical. In solution, an ion deconstructs locally the water network, to the benefit of several layers of solvation. The competition between water-ion and water-water interactions reflects the ability of the ion to organize molecules around it. These interactions have been studied for a long time and are symbolized by the Hofmeister series which turned out to reflect trends in several properties of hydrated ions, besides their ability to precipitate proteins [2]. Nowadays, the effect of ions on the structuration of water can be measured by different spectroscopic techniques such as NMR, neutron and x-ray diffraction, x-ray absorption spectroscopy, infrared and Raman spectroscopy. Gas-phase experiments are also available to study a single ion in an isolated water drop composed of tens to hundreds molecules without the effect of the environment such as counterions. Indeed, mass spectrometry was used coupled with an electrospray ionization source [3] or threshold collision-induced dissociation [4] to obtain hydration energies for a broad set of hydrated ions. Furthermore, vibrational predissociation spectra have been obtained for water clusters of various sizes that incorporate diverse ions [5–8].

From a theoretical point of view, micro-hydrated ions have been generally considered using static quantum chemical calculations[9]. However, a dynamical behavior is only obtained from molecular dynamics (MD) simulations that can be either DFT MD [10, 11] or classical MD [12]. Both methods have been used in condensed phase or at the liquid/vapor interfaces to compute ion hydration properties [13]. Average structural properties and water residence around the ion usually require simulations of in the nanosecond range to get reliable values. Consequently, DFT MD cannot provide an appropriate description due to the limitation in the simulation time. MD simulations based on classical force fields have been widely used to model ion hydration and have been improved to treat multicharged ions [12, 14–16].

The divalent cations Mg^{2+} and Ca^{2+} , as well as Na^+ , are known to be very hydrated ions with a strong structuring effect on at least their two first solvation shells [12, 13, 17] whereas for larger monovalent cations (such as K^+ and Cs^+) or polyatomic ions (such as tetramethylammonium), water-water interactions

are often stronger than ion-water interactions [18].

The behavior of anions has been largely debated both by experimentalists and theoretists. Experiments on very small clusters ($n \leq 6$) established the presence of Cl^- , Br^- and I^- at the surface [5, 19]. *Ab initio* calculations on small clusters predicted F^- inside and the other halides at the surface [13, 20]. The differences in the solvation process for F^- and Cl^- in small (4 to 20 molecules) water clusters have been studied by classical MD indicating a fully solvated F^- vs. only a partially solvated Cl^- [21, 22]. Conversely, Dang et al. concluded on the presence of both F^- and Cl^- at the surface of small water clusters with 1 to 6 molecules by polarizable MD simulations [23, 24].

For larger systems, particularly in the case of bulk liquid/vapor interfaces, the behavior of halides may be different because the curvature of the surface can play an important role and promote complete solvation of the ions. Tobias et al. combined results from experiments and MD simulations on NaCl aerosols [25]. Experimentally, they radiated aqueous NaCl aerosol to create OH radicals, and then, they observed chlorine production at the interface, suggesting a wide presence of chloride ions close to the surface. In MD simulations, they used different force fields and a wide range of system sizes, and concluded that Cl^- remains mainly close to the interface. Then, they carried out numerous polarizable MD simulations to compare the mobility of various anions and cations [26]. They concluded that the smallest ions, i.e. F^- , Na^+ , OH^- , move away from the surface while the largest ones (H_3O^+ , NO_3^- , Cl^- , Br^- , I^-) remain at the surface, with a more pronounced effect for Br^- and I^- .

In several computational studies, the influence of the force field on the results has been pointed out, especially the role of an explicit inclusion of many-body effects in the simulations [12, 27–29]. For example, a strong energetic stabilization due to polarization effects has been highlighted for halide hydration [30]. We propose here a systematic study of various ions in a water drop of 100 molecules by classical molecular dynamics based on the polarizable AMOEBA force field [31]. The ions considered, La^{3+} , Ca^{2+} , Li^+ , Na^+ , K^+ , Cs^+ , F^- , Cl^- , Br^- , SO_4^{2-} , are both cations and anions, carrying a charge ranging from -2 to +3. The same force field, that includes an explicit polarization term and a high level electrostatic representation, is used for all cases to provide a coherent

overview of their structural and dynamics properties. The behavior of the ions in the cluster is followed during relatively long simulations and the residence time of the water molecules in the first coordination shell is evaluated through a model of persisting coordination as proposed in the literature [13].

2. Simulation details

In recent years, the AMOEBA polarizable force field has been successfully used to determine the structural and thermodynamic properties of various ions in solution such as K^+ , Na^+ and Cl^- in liquid water and formamide [32], halides at the water/vapor interface [33], hydration of divalent cations (Ca^{2+} , Mg^{2+} [16], Zn^{2+} [15] and Fe^{2+} [14]) and lanthanide and actinide cations [29]. The two major features of the AMOEBA force field are an explicit inclusion of polarization effects using a self-consistent atomic dipole polarization procedure, and a multipolar expansion up to quadrupoles on each atom to represent the electrostatic interactions.

In 2003, Ren and Ponder released the 03-version of their AMOEBA polarizable force field for water [31]. This water model was able to reproduce various gas phase data of water clusters and liquid phase properties of pure water. More recently, the authors proposed a new set of parameters for water, called AMOEBA14, obtained from an automated optimization procedure, in order to enforce the agreement with a variety of liquid phase properties [34]. In this study, we used available ion parameters from the literature which have been obtained mainly to reproduce ion-water interactions using the AMOEBA03 water force field. Accordingly, the simulations have been performed with this 03-version. For the intermolecular parameters, each atom has 2 parameters for the buffered 14-7 repulsion-dispersion term, an atomic charge and polarizability. Furthermore, a polarization-damping scheme is used via a smeared charge distribution as proposed by Thol e [31]. It was proven previously that this damping factor, initially at 0.39 for water and single charged ions, should be decreased for multicharged ions [15, 16, 29]. The parameters used are provided in **Table S1** in Supplementary Material. In addition in the case of the sulfate ion, a multipole distribution was used for S and O atoms.

The simulations were carried out with the TINKER 6 software [35]. The trajectories were propagated using the velocity-Verlet algorithm with a 1 femtosecond time step. The minimum simulation time was 10 ns. The temperature was maintained at 200 K using the Nose-Hoover thermostat. The initial geometry of the drop with 100 water molecules was taken as that of an experimentally observed symmetrical cluster [36]. Simulations were started from two different initial configurations, with the ion (-II to +III charged) inside, i.e. structure called *Ion-Int*, or on the surface, i.e. *Ion-Ext*, of the water cluster. In the first case, the ion was positioned at the center of the cage, followed by an AMOEBA geometry optimization. For the second case, one needs realistic starting structures, i.e. with a representative water network, for both positive and negative ions. During the simulations with F^- and Ca^{2+} initially inside the drop, the ions approached the surface and provided adequate starting structures for an *Ion-Ext* simulation. We noted that the AMOEBA geometry optimization can deform the initial structure, even for the *Ion-Int* structure, especially when the ion is large and multiply-charged (**Fig. 1**). In order to avoid evaporation of water molecules during the dynamics at 200 K, a spherical van der Waals potential encloses the cluster with a 12 Å radius. In practice, the molecules almost completely ignore the wall, but they return to the cluster if they evaporate. At this temperature, evaporation occurs on average less than once per simulation.

The ion coordination numbers were obtained from the radial distribution function of ion-O pairs. They can be compared directly to experimental values to check cluster model ability to reproduce the competitive interactions such as ion-water and water-water interactions. Residence times of a water molecule in the first coordination sphere of the ion were computed from the model of persisting coordination proposed by Impey *et al.* [37]. A parameter $t^* = 2$ ps was used to consider a temporary migration of a water molecule from the first coordination sphere.

3. Results and Discussion

Dynamical behavior

For each ion, the results are gathered in a series of six graphs, as shown in **Fig. 2, 4** and **5** for La^{3+} , Cs^+ and Cl^- , respectively. The figures concerning

the other ions are available in Supplementary Material (**Fig. S2-S8**). The time evolution of various properties is followed starting with *Ion-Ext* and *Ion-Int* structures in the left and right columns, respectively. The upper panel displays the time evolution of the distance between the ion and the center of mass of the drop. The initial distance varies depending on the localization of the ion and the result of the initial geometry optimization. This distance cannot confirm by itself if the ion is present at the surface or within the cluster. However, in the cases where the drop remains essentially spherical, one can easily correlate distance ranges with ion positions. Indeed, the diameter of the spherical cluster being ca. 17 Å, the distances to the center of mass can be naturally distributed as follows: $\leq 5\text{-}6$ Å for a complete double solvation shell with the ion inside the drop, ca. 6 Å for an incomplete second shell, and ca. 7-8 Å for an ion at the surface. The middle and lower panels in the figures monitor the coordination number (first neighbours) and the number of second neighbours of the ions, respectively, as a function of time. The boundaries of the coordination spheres are obtained from the radial distribution functions computed over the simulations (**Fig. S1**).

La^{3+} and Ca^{2+} carry the highest positive charge, and they have a similar structuring behavior of the water drop for the simulations starting both from an ion at the surface or in the center of the cluster (**Fig. 2** and **S2**). The high charge leads to strong interaction with the molecules in the first shell and the cation clearly favors an internal position. **Table 1** presents the mean number of water molecules in the first and second shells obtained by the simulations, together with experimental values when available. For La^{3+} , the CN remains at 9 with an average La-O distance of ca. 2.51 Å which both agree with the experimental data (9.0 and 2.52 Å, respectively [38]). This very strong effect in the first shell was only observed for lanthanum but was already highlighted by Bowron and al. to explain the sharp break in the aqueous solvent structure (in the bulk for LaCl_3 molar solutions) by the high coordination number of 9 and the tricapped trigonal prism arrangement around La^{3+} [39]. The experimental CN of Ca^{2+} is known to be strongly dependent on the concentration with an average value of ca. 8 for low concentration [40] in agreement with the computed CN of 8.0. The structuring effect is also present in the second solvation shell for both cations, with a limited variation of the number of water molecules between 16 and 20 that can be compared to the experimental number of water

molecules of ca. 17 for Ca^{2+} [41] and 18 for La^{3+} [38]. The final structures for both calcium simulations remain relatively spherical (**Fig. 3**). This observation correlates with the Hofmeister series in which Ca^{2+} is one of the most chaotropic ions.

In comparison with the previous cases, the alkali metal ions are considered to be relatively poorly solvated, and their coordination strongly depends on the size of the cation and the concentration of the solution. Li^+ is the smallest cation of the series and despite its single charge, it remains solvated by two shells of water in the major part of the simulations. When the starting point corresponds to the ion outside the drop, the fluctuation of the distance between the ion and the center of mass of the aggregate is small. For the inside starting point case, the ion spends few nanoseconds outside before stabilizing inside the drop. The mean CN is 4.4 (see fluctuations between 4 and 5 first neighbours in **Fig. S3**) with an average Li-O distance of 2.04 Å in the first shell. The second sphere is also well structured with ca. 14 water molecules for the simulation time during which the cation is stable inside the drop. These results agree well with the available experimental data for the first coordination sphere and a probable existence of a second shell [42].

With increasing cation size, going from Na^+ to K^+ and then Cs^+ , a progressive reduction of the structuring power of the ion is observed during the simulations as expected in the Hofmeister series. Na^+ has also a global solvating behavior. When the starting point corresponds to the ion at the surface of the cluster, 2 ns are necessary for the cation to migrate inside. Furthermore, Na^+ remains mainly solvated as for the inside starting point simulation. The mean CN is 5.9 in agreement with experimental data in solution, with a Na-O distance of 2.45 Å in the first coordination sphere, and small fluctuations in the number of second neighbours around 17 water molecules. For K^+ and Cs^+ , the depopulation of the second shell, and even of the first shell, is observed during both types of simulation with large fluctuations of the numbers of first and second neighbours around the cation (**Fig. 4**). The mean CN are ca. 6.2 and 7.9 for K^+ and Cs^+ , respectively, computed from the inside starting point simulations. These values are in the experimental range (**Table 1**).

In the anions series, the sulfate ion is fully solvated with an average of 12

neighbours in first coordination shell organized with 3 water molecules hydrogen bound to each of sulfate oxygens. The simulations indicate a mean $O_w-O_{sulfate}$ distance of ca. 3.85 Å. In contrast, the second shell is sometimes partly depopulated, which may be an effect of the size of the SO_4^{2-} ion with respect to the size of the cluster.

Then, the series of the halogen anions (F^- , Cl^- , Br^-) has been confined in the drop. F^- , the smallest anion, is mainly solvated by two complete solvation shells inside the cluster despite few hundreds of picoseconds (over 10 nanoseconds) spent at the surface with a depopulated second shell. The related effect is a fluctuation of the coordination number with an average value of 5.9 and a mean F-O distance of ca. 2.84 Å. With chloride, we observe the first behavior of an ion positioning at the surface of the cluster (**Fig. 5**), suggesting that water-water are stronger than ion-water interactions. Its high polarization favors its position at the surface with partly filled solvation spheres and fluctuations in the number of first and second neighbours (4 to 6 and 8 to 15, respectively). The radial distribution function of the chloride ion highlights an important number of water exchanges between the two solvation spheres when Cl^- is inside the drop, and a stability of the water network when it is at the surface. Br^- has a similar behavior with a quick migration of the anion at the surface of the aggregate and an average CN of 5.0 was obtained for both anions.

In summary, small atomic anions are mainly solvated while the bigger prefer to be at the surface. However, SO_4^{2-} is different, with a preference to be far from the interface. Its behavior is explained by the large energetic penalty associated to the partial desolvation of the doubly charged anion at the surface [43].

Water exchanges

The residence times of a water molecule in the first coordination sphere of the various ions are provided in **Table 2**. This property can be used to quantify the strength of the ion-ligand interactions and to estimate the ion structuring ability. Experimental values are given for comparison. They are generally obtained by NMR using methods based on relaxation time measurements and ^{17}O labeling, or by quasi-elastic neutron scattering. While the data for the cations are relatively easily available, experimental data for anions are much scarce. In this study, the comparison with experiments may be sometimes difficult due to the cluster approach. Furthermore, the temperature was set at 200 K for two

reasons. First, working on clusters, evaporation of water can occur at higher temperatures and we try to minimize the use of the van der Waals wall. Second, it is known that in MD simulations, the mobility of water molecules is more important than experimentally at the same temperature [44] and the correspondence between experimental and computed temperature is questionable [45]. For these reasons, we mainly discuss the results qualitatively by checking orders of magnitude and trends in a series.

For metal cations, the experimental residence time spans 20 orders of magnitude and is mainly observed to be a function of the ionic radius and of the ion charge [46]. The values extracted from the simulations are consistent with experimental data. For Ca^{2+} and La^{3+} , the residence time is ca. few nanoseconds and consequently, the simulations showed only few exchanges and no exchange, respectively. Regarding the monovalent cations, the residence time decreases from Li^+ to Cs^+ , the latter being the most labile of the series. In this case, the computed time from the inside starting point is twice as small as that for the starting point at the cluster surface due to the large fluctuation in the number of first neighbours (**Fig. 4**).

For anions, experimental data come from femtosecond mid-infrared nonlinear spectroscopy, that allows to obtain reorientation time of water molecules in the first sphere and can be used as a proof of rigidity of the first shell [47]. The dynamics is faster from I^- to Br^- and Cl^- with times in the range of tens of picoseconds [48]. A decrease of the residence time was obtained from MD, using a simple point charge potential for water and a Lennard-Jones potential for ions, and was explained by a geometric criterion based on the size of the anion [49]. Laage and Hynes performed MD simulations in periodic boundary conditions using both polarizable and non polarizable models for water, and proposed several models to determine water residence time for halides from 5 to 10 ps with the polarizable model [50]. The residence times extracted from the present cluster approach simulations are larger in magnitude but the order is preserved from F^- to Br^- . In the Cl^- and Br^- cases, the times obtained are very different in the two simulations. As they naturally prefer to be at the cluster surface, the organization of the water molecules around the ion and their mobility are indeed disturbed in comparison with bulk experiments or simulations.

4. Conclusions

Molecular dynamics simulations of ions in a water drop were carried out with the AMOEBA polarizable force field to capture their propensity to reside inside the cluster or to migrate from the center to the surface and conversely. The 03-version of the parameters for water was used to be consistent with the ion parameters available in the literature. For the cation series, the multiply-charged and small monovalent ions favor an inside position whereas the larger monovalent ions prefer staying at the surface of the drop. The most structuring ions present a clear water organization around the cation with two and even three water shells. The structural properties such as numbers of first and second neighbors and ion-water distances are well reproduced in comparison with experiments. For the +3 and +2 cations, no or few exchanges have been observed during the simulations between water molecules in the first and second shells. Furthermore, the residence time decreases as the size of the monovalent cation increases. Regarding the anions, the results reproduce both the structuration of the solvent around the ion and the trend of the biggest ones to stay at the surface of the drop. This partial desolvation can be correlated to the ion size and polarizability. Residence times of the water molecules in the first shell have been proposed for the anion series. To conclude, the AMOEBA force field used without any reparametrization is able to provide accurate structural and dynamical properties for the whole series of the various ions under study.

Acknowledgements

F.T. thanks Ecole Polytechnique for a PhD fellowship. This work was granted access to the HPC resources of [CCRT/ CINES/ IDRIS] under the allocation c2015077128 made by GENCI (Grand Equipement National de Calcul Intensif).

References

- [1] D. J. Tobias, J. C. Hemminger, Getting Specific About Specific Ion Effects, *Science* 319 (5867) (2008) 1197–1198.
- [2] W. Kunz, J. Henle, B. Ninham, Zur Lehre von der Wirkung der Salze (about the science of the effect of salts): Franz Hofmeister’s historical papers, *Current Opinion in Colloid and Interface Science* 9 (1-2) (2004) 19 – 37.
- [3] A. T. Blades, P. Jayaweera, M. G. Ikononou, P. Kebarle, Ion-molecule clusters involving doubly charged metal ions (M^{2+}), *Int. J. Mass Spectrom. Ion Process* 102 (1990) 251 – 267.
- [4] T. E. Cooper, D. R. Carl, P. B. Armentrout, Hydration Energies of Zinc(II): Threshold Collision-Induced Dissociation Experiments and Theoretical Studies, *J. Phys. Chem. A* 113 (49) (2009) 13727–13741.
- [5] P. Ayotte, C. G. Bailey, G. H. Weddle, M. A. Johnson, Vibrational Spectroscopy of Small $Br^-(H_2O)_n$ and $I^-(H_2O)_n$ Clusters: Infrared Characterization of the Ionic Hydrogen Bond, *J. Phys. Chem. A* 102 (18) (1998) 3067–3071.
- [6] J. S. Prell, J. T. O’Brien, E. R. Williams, Structural and Electric Field Effects of Ions in Aqueous Nanodrops, *J. Am. Chem. Soc.* 133 (13) (2011) 4810–4818.
- [7] J. T. O’Brien, E. R. Williams, Effects of Ions on Hydrogen-Bonding Water Networks in Large Aqueous Nanodrops, *J. Am. Chem. Soc.* 134 (24) (2012) 10228–10236.
- [8] X.-B. Wang, A. P. Sergeeva, J. Yang, X.-P. Xing, A. I. Boldyrev, L.-S. Wang, Photoelectron Spectroscopy of Cold Hydrated Sulfate Clusters, $SO_4^{2-}(H_2O)_n$ ($n=4-7$): Temperature-Dependent Isomer Populations, *J. Phys. Chem. A* 113 (19) (2009) 5567–5576.
- [9] D. S. Lambrecht, L. McCaslin, S. S. Xantheas, E. Epifanovsky, M. Head-Gordon, Refined energetic ordering for sulfate–water ($n=3-6$) clusters using high-level electronic structure calculations, *Mol. Phys.* 110 (19-20) (2012) 2513–2521.

- [10] E. Cauët, S. Bogatko, J. H. Weare, J. L. Fulton, G. K. Schenter, E. J. Bylaska, Structure and dynamics of the hydration shells of the Zn^{2+} ion from ab initio molecular dynamics and combined ab initio and classical molecular dynamics simulations, *J. Chem. Phys.* 132 (19) (2010) 194502.
- [11] Y. Zhao, H. Li, X. C. Zeng, First-Principles Molecular Dynamics Simulation of Atmospherically Relevant Anion Solvation in Supercooled Water Droplet, *J. Am. Chem. Soc.* 135 (41) (2013) 15549–15558.
- [12] S. Tazi, J. J. Molina, B. Rotenberg, P. Turq, R. Vuilleumier, M. Salanne, A transferable ab initio based force field for aqueous ions, *J. Chem. Phys.* 136 (11) (2012) 114507.
- [13] T.-M. Chang, L. X. Dang, Recent Advances in Molecular Simulations of Ion Solvation at Liquid Interfaces, *Chem. Rev.* 106 (4) (2006) 1305–1322.
- [14] D. Semrouni, W. C. Isley, C. Clavaguéra, J.-P. Dognon, C. J. Cramer, L. Gagliardi, Ab Initio Extension of the AMOEBA Polarizable Force Field to Fe^{2+} , *J. Chem. Theo. Comput.* 9 (7) (2013) 3062–3071.
- [15] J. C. Wu, J.-P. Piquemal, R. Chaudret, P. Reinhardt, P. Ren, Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field, *J. Chem. Theo. Comput.* 6 (7) (2010) 2059–2070.
- [16] J.-P. Piquemal, L. Perera, G. A. Cisneros, P. Ren, L. G. Pedersen, T. A. Darden, Towards accurate solvation dynamics of divalent cations in water using the polarizable AMOEBA force field: From energetics to structure, *J. Chem. Phys.* 125 (5) (2006) 054511.
- [17] K. J. Tielrooij, N. Garcia-Araez, M. Bonn, H. J. Bakker, Cooperativity in Ion Hydration, *Science* 328 (5981) (2010) 1006–1009.
- [18] J. S. Prell, E. R. Williams, Structures of Thermal, Mass-Selected Water Clusters Probed with Hydrophobic Ion Tags and Infrared Photodissociation Spectroscopy, *J. Am. Chem. Soc.* 131 (11) (2009) 4110–4119.
- [19] J.-H. Choi, K. T. Kuwata, Y.-B. Cao, M. Okumura, Vibrational Spectroscopy of the $\text{Cl}^-(\text{H}_2\text{O})_n$ Anionic Clusters, $n = 1-5$, *J. Phys. Chem. A* 102 (3) (1998) 503–507.

- [20] J. W. Caldwell, P. A. Kollman, Ab initio calculations on aquated chloride ($\text{Cl}^-(\text{H}_2\text{O})_{14}$) clusters: comparison with the results from molecular dynamics simulations, *J. Phys. Chem.* 96 (21) (1992) 8249–8251.
- [21] L. Perera, M. L. Berkowitz, Structure and dynamics of $\text{Cl}^-(\text{H}_2\text{O})_{20}$ clusters: The effect of the polarizability and the charge of the ion, *J. Chem. Phys.* 96 (11) (1992) 8288–8294.
- [22] L. S. Sremaniak, L. Perera, M. L. Berkowitz, Enthalpies of formation and stabilization energies of $\text{Br}^-(\text{H}_2\text{O})_n$ ($n=1,2, \dots, 15$) clusters. Comparisons between molecular dynamics computer simulations and experiment, *Chem. Phys. Lett.* 218 (4) (1994) 377 – 382.
- [23] L. X. Dang, J. E. Rice, J. Caldwell, P. A. Kollman, Ion solvation in polarizable water: molecular dynamics simulations, *J. Am. Chem. Soc.* 113 (7) (1991) 2481–2486.
- [24] L. X. Dang, Development of nonadditive intermolecular potentials using molecular dynamics: Solvation of Li^+ and F^- ions in polarizable water, *J. Chem. Phys.* 96 (9) (1992) 6970–6977.
- [25] E. M. Knipping, M. J. Lakin, K. L. Foster, P. Jungwirth, D. J. Tobias, R. B. Gerber, D. Dabdub, B. J. Finlayson-Pitts, Experiments and Simulations of Ion-Enhanced Interfacial Chemistry on Aqueous NaCl Aerosols, *Science* 288 (5464) (2000) 301–306.
- [26] E. C. Brown, M. Mucha, P. Jungwirth, D. J. Tobias, Structure and Vibrational Spectroscopy of Salt Water/Air Interfaces: Predictions from Classical Molecular Dynamics Simulations, *J. Phys. Chem. B* 109 (16) (2005) 7934–7940.
- [27] C. Caleman, J. S. Hub, P. J. van Maaren, D. van der Spoel, Atomistic simulation of ion solvation in water explains surface preference of halides, *Proceedings of the National Academy of Sciences* 108 (17) (2011) 6838–6842.
- [28] D. Hagberg, S. Brdarski, G. Karlström, On the Solvation of Ions in Small Water Droplets, *J. Phys. Chem. B* 109 (9) (2005) 4111–4117.

- [29] A. Marjolin, C. Gourlaouen, C. Clavaguéra, P. Y. Ren, J.-P. Piquemal, J.-P. Dognon, Hydration gibbs free energies of open and closed shell trivalent lanthanide and actinide cations from polarizable molecular dynamics, *J. Mol. Mod.* 20 (2014) 2471.
- [30] L. Perera, M. L. Berkowitz, Many-body effects in molecular dynamics simulations of $\text{Na}^+(\text{H}_2\text{O})_n$ and $\text{Cl}^-(\text{H}_2\text{O})_n$ clusters, *J. Chem. Phys.* 95 (3) (1991) 1954–1963.
- [31] P. Ren, J. W. Ponder, Polarizable atomic multipole water model for molecular mechanics simulation, *J. Phys. Chem. B* 107 (24) (2003) 5933–5947.
- [32] A. Grossfield, P. Ren, J. W. Ponder, Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field, *J. Am. Chem. Soc.* 125 (50) (2003) 15671–15682.
- [33] L. Tůma, D. Jeníček, P. Jungwirth, Propensity of heavier halides for the water/vapor interface revisited using the Amoeba force field, *Chem. Phys. Lett.* 411 (1–3) (2005) 70–74.
- [34] M. L. Laury, L.-P. Wang, V. S. Pande, T. Head-Gordon, J. W. Ponder, Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model, *J. Phys. Chem. B* 119 (29) (2015) 9423–9437.
- [35] J. W. Ponder, TINKER - Software Tools for Molecular Design, version 6, Washington University School of Medicine, Saint Louis, U.S.A. (2013).
URL <http://dasher.wustl.edu/tinker>
- [36] A. Müller, H. Bögge, E. Diemann, Structure of a cavity-encapsulated nanodrop of water, *Inorg. Chem. Comm.* 6 (1) (2003) 52–53.
- [37] R. W. Impey, P. A. Madden, I. R. McDonald, Hydration and mobility of ions in solution, *J. Phys. Chem.* 87 (25) (1983) 5071–5083.
- [38] P. R. Smirnov, V. N. Trostin, Structural parameters of the nearest surrounding of lanthanide ions in aqueous solutions of their salts, *Russ. J. Gen. Chem.* 82 (3) (2012) 360–378.
- [39] S. Diaz-Moreno, S. Ramos, D. T. Bowron, Solvation Structure and Ion Complexation of La^{3+} in a 1 Molal Aqueous Solution of Lanthanum Chloride, *J. Phys. Chem. A* 115 (24) (2011) 6575–6581.

- [40] Y. S. Badyal, A. C. Barnes, G. J. Cuello, J. M. Simonson, Understanding the Effects of Concentration on the Solvation Structure of Ca^{2+} in Aqueous Solution. II: Insights into Longer Range Order from Neutron Diffraction Isotope Substitution, *J. Phys. Chem. A* 108 (52) (2004) 11819–11827.
- [41] F. Jalilehvand, D. Spångberg, P. Lindqvist-Reis, K. Hermansson, I. Persson, M. Sandström, Hydration of the Calcium Ion. An EXAFS, Large-Angle X-ray Scattering, and Molecular Dynamics Simulation Study, *J. Am. Chem. Soc.* 123 (3) (2001) 431–441.
- [42] J. Mähler, I. Persson, A Study of the Hydration of the Alkali Metal Ions in Aqueous Solution, *Inorg. Chem.* 51 (1) (2012) 425–438.
- [43] P. Jungwirth, J. E. Curtis, D. J. Tobias, Polarizability and aqueous solvation of the sulfate dianion, *Chem. Phys. Lett.* 367 (5-6) (2003) 704–710.
- [44] H. Ohtaki, T. Radnai, Structure and dynamics of hydrated ions, *Chem. Rev.* 93 (3) (1993) 1157–1204.
- [45] P. Carbonnière, A. Dargelos, I. Ciofini, C. Adamo, C. Pouchan, Vibrational analysis of glycine radical: a comparative ab initio static and dynamic study, *Phys. Chem. Chem. Phys.* 11 (21) (2009) 4375–4384.
- [46] L. Helm, A. E. Merbach, Inorganic and Bioinorganic Solvent Exchange Mechanisms, *Chem. Rev.* 105 (6) (2005) 1923–1960.
- [47] M. F. Kropman, H. J. Bakker, Dynamics of Water Molecules in Aqueous Solvation Shells, *Science* 291 (5511) (2001) 2118–2120.
- [48] P. S. Salmon, W. S. Howells, R. Mills, The dynamics of water molecules in ionic solution. II. Quasi-elastic neutron scattering and tracer diffusion studies of the proton and ion dynamics in concentrated Ni^{2+} , Cu^{2+} and Nd^{3+} aqueous solutions, *J. Phys. C: Solid State Phys.* 20 (34) (1987) 5727.
- [49] S. Chowdhuri, A. Chandra, Dynamics of Halide Ion, Water Hydrogen Bonds in Aqueous Solutions: Dependence on Ion Size and Temperature, *J. Phys. Chem. B* 110 (19) (2006) 9674–9680.
- [50] D. Laage, J. T. Hynes, On the Residence Time for Water in a Solute Hydration Shell: Application to Aqueous Halide Solutions, *J. Phys. Chem. B* 112 (26) (2008) 7697–7701.

- [51] A. T. Blades, , P. Kebarle, Sequential Hydration Energies of the Sulfate Ion, from Determinations of the Equilibrium Constants for the Gas-Phase Reactions: $\text{SO}_4(\text{H}_2\text{O})_n^{2-} = \text{SO}_4(\text{H}_2\text{O})_{n-1}^{2-} + \text{H}_2\text{O}$, *J. Phys. Chem. A* 109 (37) (2005) 8293–8298.
- [52] R. L. Wong, E. R. Williams, Dissociation of $\text{SO}_4^{2-}(\text{H}_2\text{O})_n$ Clusters, $n = 3-17$, *J. Phys. Chem. A* 107 (50) (2003) 10976–10983.
- [53] M. Eigen, Fast elementary steps in chemical reaction mechanisms, *Pure and Applied Chemistry* 6 (1963) 97–116.
- [54] D. T. Richens, *The Chemistry of Aqua Ions: Synthesis, Structure and Reactivity: A Tour Through the Periodic Table of the Elements*, Wiley, ISBN: 978-0-471-97058-3, 1997.

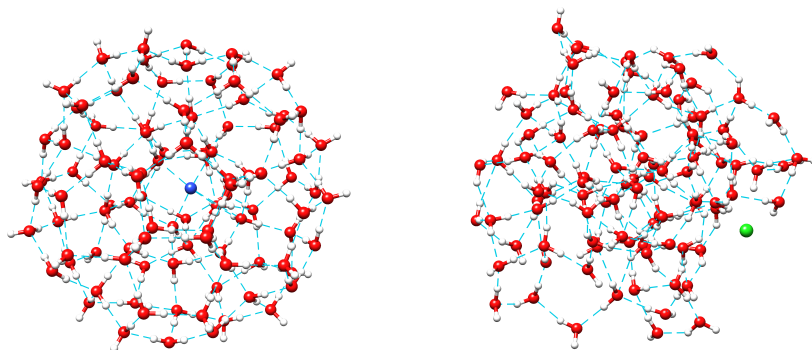


Figure 1: On the left side, *Ion-Int* optimized starting structure for Na^+ , and on the right side, *Ion-Ext* optimized starting structure for Cl^- .

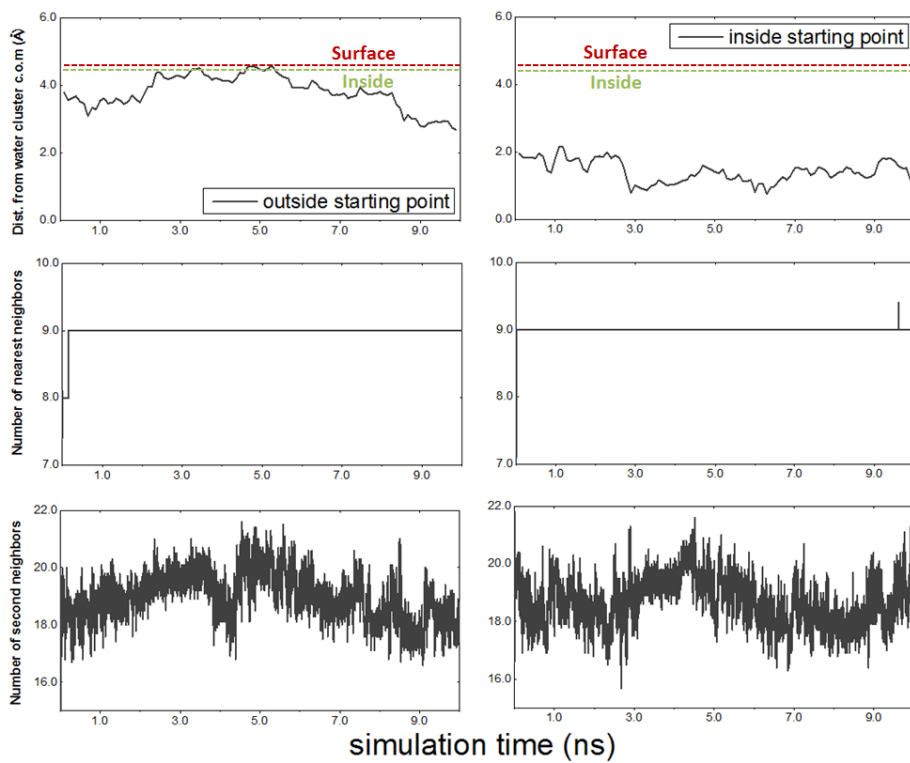


Figure 2: Time evolution of the ion-center of mass distance, numbers of first and second neighbours for the La^{3+} water drop for an outside starting point (left) and an inside starting point (right). See text for details.

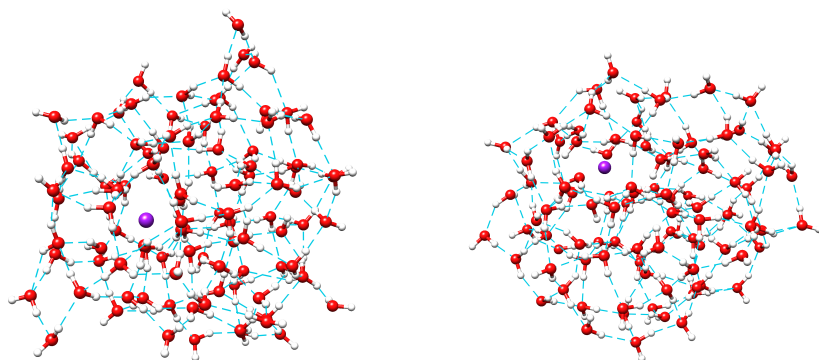


Figure 3: Snapshots of final structures from different simulations for the Ca²⁺ water drop.

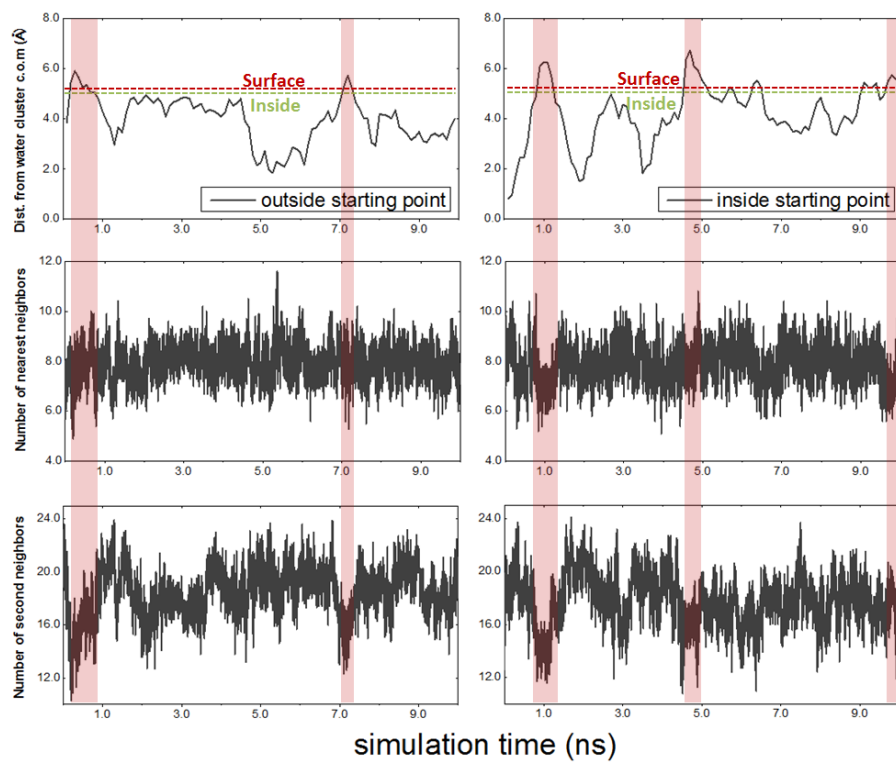


Figure 4: Time evolution of the ion-center of mass distance, numbers of first and second neighbours for the Cs^+ water drop for an outside starting point (left) and an inside starting point (right). See text for details.

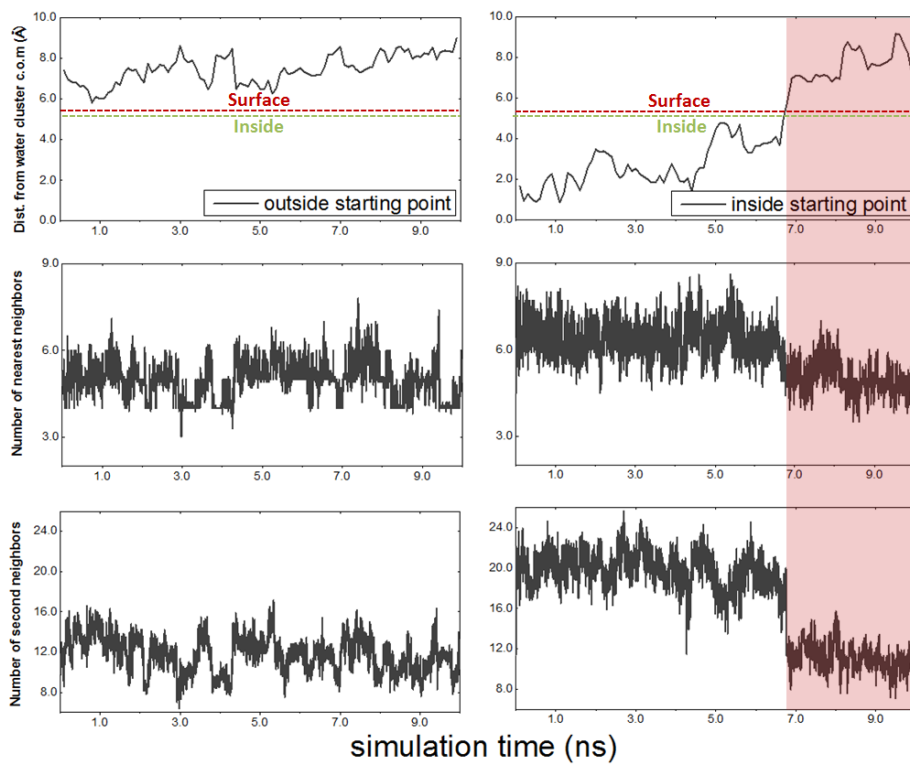


Figure 5: Time evolution of the ion-center of mass distance, numbers of first and second neighbours for the Cl^- water drop for an outside starting point (left) and an inside starting point (right). See text for details.

Table 1: Coordination numbers and number of water molecules in the second shell in comparison with available experimental data.

Ions	EXP		AMOEBA			
	1st Sh.	2nd Sh.	1st Sh.		2nd Sh.	
			Av.	Var.	Av.	Var.
La ³⁺	9.0 ^(a)	18	9.0	9.0	18.8	17.1-20.6
Ca ²⁺	8.0 ^(b)	17	8.0	7.6-8.3	17.7	15.7-19.9
Li ⁺	4 ^(c)	12	4.4	3.4-5.4	13.9	11.4-16.1
Na ⁺	4-8 ^(c)		5.9	5.1-6.5	16.8	13.9-19.6
K ⁺	4-6 ^(c)		6.2	5.5-7.4	16.2	12.6-19.9
Cs ⁺	6-8 ^(c)		7.9	6.4-9.4	18.1	13.9-22.2
F ⁻	4-6 ^(c)		5.9	5.2-6.6	17.9	13.0-23.0
Cl ⁻	5-6 ^(c)		5.0	3.7-6.2	11.8	8.3-15.2
Br ⁻	6 ^(c)		5.0	4-6	11.7	8.0-15.4
SO ₄ ²⁻	12 ^(d)		12.2	11.2-13.1	24.7	18.7-29.5

(a) [38]

(b) [16, 41]

(c) [44]

(d) [51, 52]

Table 2: Water molecule residence times (in ns) in the first solvation shell of the ions.

Ions	Experiments	AMOEBA <i>Int/Ext</i>
F ⁻		1.71 / 1.17
Cl ⁻		0.43 / 0.75
Br ⁻		0.08 / 0.64
Li ⁺	2.37 [46]	2.60 / 3.20
Na ⁺	1.25 [46]	3.08 / 2.35
K ⁺	0.81 [46]	0.66 / 0.58
Cs ⁺	0.28 [46], 0.2 [53], 0.1 [54]	0.27 / 0.44
Ca ²⁺	2.96 [46], ~ 10 [54]	few exchanges
La ³⁺	4.8 [46], ~ 10 [54]	no exchange

4.2 Validation du nouveau jeu de paramètres de l'eau AMOEBA14 pour la dynamique des ions solvatés

Cette étude n'a pas l'objectif d'exhaustivité et de prédiction de celle réalisée en AMOEBA03. Sans accord de version entre les paramètres des ions et de l'eau, ce modèle n'a aucune légitimité pour proposer des données calculées ou pour valider un comportement particulier. L'unique objectif est d'identifier un éventuel biais dans les comportements dynamiques des ions dans les agrégats, avant de valider l'utilisation de la version 14 du modèle de l'eau pour la spectroscopie des ions hydratés.

4.2.1 Comportement dynamique des ions

Nous nous sommes limités à envisager la situation la plus favorable pour chaque ion, i.e en surface de l'agrégat pour Br^- et Cl^- et à l'intérieur pour tous les autres. Comme précédemment, les trajectoires sont propagées par l'algorithme Verlet-Vitesse avec un pas de temps de 1 fs. La température est maintenue à 200 K en utilisant le thermostat Nosé-Hoover. Les géométries de départ sont les mêmes que dans l'étude précédente (détails de construction dans l'article en section 1.1). Un potentiel de Lennard-Jones (LJ), sphérique, entoure les agrégats et pare à une éventuelle évaporation de molécule d'eau. Ce mur de LJ doit être suffisamment large pour être ignoré par les molécules d'eau en surface de l'agrégat. Un rayon de 12 Å est suffisant pour tous les ions, à l'exception de La^{3+} , pour lequel des structures distordues étaient observées. Dans ce cas, le rayon du mur a été étendu à 18 Å. Chaque simulation dure 10 ns. Les distances entre l'ion et le centre de masse de l'agrégat d'eau sont collectées à chaque pas, puis sont moyennées sur des périodes de 100 ps (**Fig 4.2**). Le nombre de molécules d'eau de 1^{ère} et 2^{ème} sphères est également calculé à chaque pas et fait apparaître des écarts ponctuels aux coordinences moyennes. Pour limiter le nombre de graphiques, nous ne présentons que l'analyse de ces données temporelles. Pour tous les ions, à l'exception des deux anions de surface, nous reportons un dépeuplement partiel d'une couche par une bande de couleur, rouge pour la 1^{ère} sphère de solvatation et bleue pour la 2^{ème}.

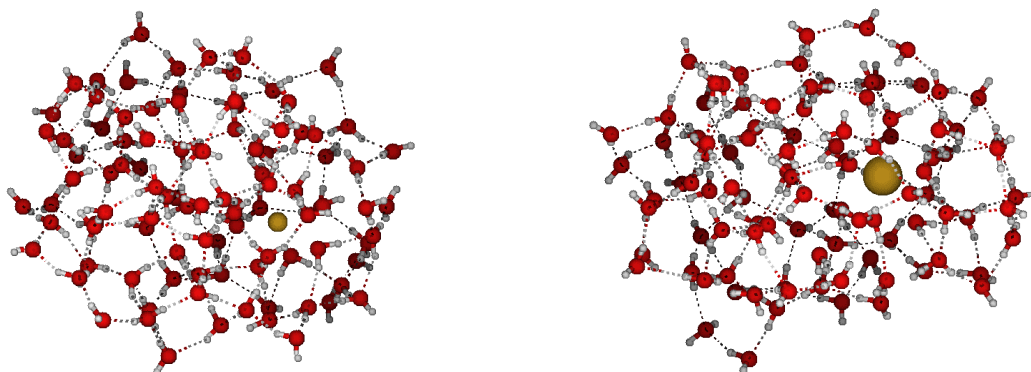


FIGURE 4.1 – Structures finales des simulations des nanogouttes de Ca^{2+} (à gauche), et La^{3+} (à droite).

Tous les cations restent majoritairement au sein des agrégats. Les ions La^{3+} et Ca^{2+} sont fortement hydratés tout au long de la simulation. A l'issue des 10 ns, si la taille du mur est choisie suffisamment grande pour ne pas être ressentie par les molécules en surface, les agrégats gardent une forme relativement sphérique (**Fig 4.1**). Cette obligation d'augmenter la taille du mur pour La^{3+} témoigne que l'ion, par son volume et sa forte coordination en 1^{ère} sphère, fait gonfler l'agrégat, ce qui pourrait justifier les ruptures dans la structure du solvant qu'a observées Bowron pour des solutions molaires LaCl_3 [21]. On observe un peu plus de fluctuations pour les cations monovalents. Li^+ se déplace assez rapidement en périphérie de l'agrégat (mais conserve une 1^{ère} sphère complète), comportement qui peut être expliqué par la taille de la cavité de l'agrégat expérimental qui sert de structure de départ [22], peu

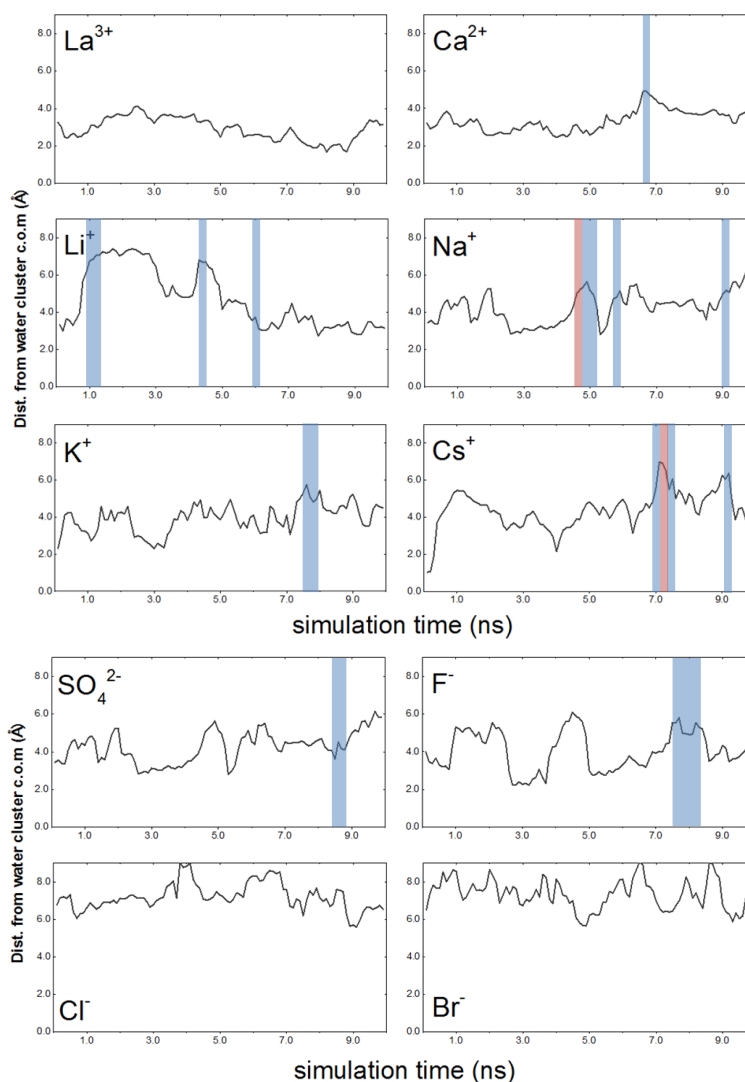


FIGURE 4.2 – Suivi de la distance de l'ion au centre de masse de l'agrégat pour un ion dans une nanogoutte de 100 molécules d'eau. Tous les ions sont initialement à l'intérieur de l'agrégat, à l'exception de Br^- et Cl^- qui sont à sa surface.

adaptée au faible rayon de l'ion, qui éprouve des difficultés à construire ses sphères de solvation. On remarque d'ailleurs qu'à partir de 5 ns, l'ion est pleinement hydraté, a retrouvé une position plus centrale dans l'agrégat, et ne montre plus de mobilité. L'ion Na^+ a en moyenne six $1^{\text{ère}}$ s voisins, mais montre un rapide épisode de dépeuplement de la $1^{\text{ère}}$ sphère (4 voisins vers 4.8 ns). A l'inverse, le potassium est fortement solvaté pendant toute la dynamique alors qu'en AMOEBA03, il montrait quelques abaissements de coordinence. Enfin le césium montre deux épisodes de désolvation partielle, en accord avec les observations en AMOEBA03 d'un cation un peu plus faiblement solvaté que les autres cations monovalents. Le comportement des anions est également reproduit à l'identique par AMOEBA14 : les ions F^- et SO_4^{2-} sont bien hydratés, tandis que Cl^- et Br^- restent en surface de l'agrégat (**Fig 4.2**).

4.2.2 Influence des paramètres

Une des modifications majeures entre les deux versions du champ de forces de l'eau concerne les paramètres de van der Waals (vdW), pour lesquels une augmentation du diamètre et du puits de l'oxygène, associée à une

TABLE 4.1 – Comparatif des paramètres de vdW pour l'eau dans AMOEBA03 et AMOEBA14.

paramètres	unités	AMOEBA03	AMOEBA14
Diamètre de vdW de O	Å	3.405	3.5791
Puits de vdW de O	kcal/mol	0.11	0.1512
Diamètre de vdW de H	Å	2.655	2.1176
Puits de vdW de H	kcal/mol	0.0135	0.0105
facteur de réduction de vdW de H	–	0.91	0.8028

diminution de ceux de l'hydrogène, rapproche le modèle d'une situation à un seul site de répulsion-dispersion, centré sur l'oxygène (**Tab 4.1**). Les changements de paramètres agissent à deux niveaux : d'une part la distance de référence ion-eau est modifiée, avec un impact plus important sur les interactions impliquant les hydrogènes de l'eau (et donc ressenti par les anions) (**Fig 4.3**, à gauche); d'autre part sur le potentiel de vdW à la géométrie d'équilibre, avec une stabilisation importante des interactions cation-eau (**Fig 4.3**, à droite).

Cet effet est directement ressenti sur les distances moyennes au sein des deux sphères de solvation. Nous utilisons comme critère d'évaluation la distance entre l'ion et l'oxygène de la molécule d'eau (distance M-O), qui n'est pas directement assimilable à la distance $M^- \dots HOH$ pour les anions, mais permet d'aboutir aux mêmes conclusions. Pour les anions, la distance moyenne M-O diminue de 0.1 Å, à la fois en 1^{ère} et 2^{ème} sphères, entre les versions 03 et 14. L'effet est inverse pour les cations (+0.16 Å en 1^{ère} et +0.1 Å en 2^{ème} sphère) (**Tab 4.2**). Un exemple est donné en **Fig 4.4**, pour les ions Cl^- (-0.07 Å en 1^{ère} et 2^{ème} sphère) et La^{3+} (+0.15 Å en 1^{ère} et +0.1 Å en 2^{ème} sphère). Pour chacun, la fonction de distribution radiale illustre que l'effet est très localisé sur les 2 premières sphères de solvation.

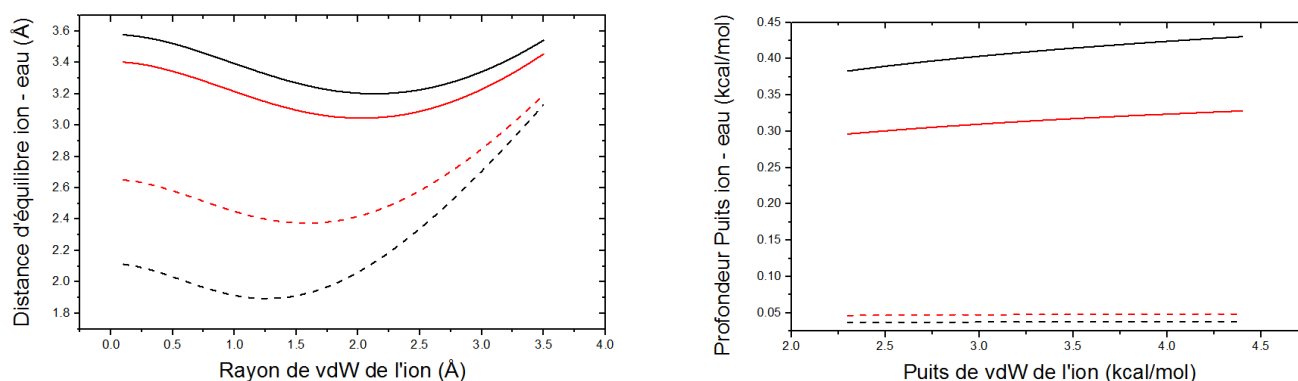


FIGURE 4.3 – Evolution des distances (à gauche) et des profondeurs de puits (à droite) de référence d'une interaction de vdW entre un ion (défini par ses rayon et puits de vdW) et un oxygène (trait plein) ou un hydrogène (trait pointillé) d'une molécule d'eau, en fonction du champ de forces AMOEBA03 (rouge) ou AMOEBA14 (noir).

Cette déviation est ressentie dans la coordination de l'ion. En moyenne, le changement de champ de forces fait diminuer le nombre de molécules d'eau dans les sphères de solvation des anions (-0.5 H_2O en 1^{ère} sphère et -0.7 en 2^{ème}) et augmenter la coordination des cations (+0.5 H_2O en 1^{ère} sphère et +0.6 en 2^{ème}) (**Tab 4.2**). Tous les ions ne sont pas uniformément affectés, puisqu'on observe peu d'effets sur les 1^{ères} sphères de solvation des anions de surface (à corrélérer avec leur faible coordination et leur présence en surface qui limite les géométries d'interactions), et pratiquement aucun sur les cations multichargés (pour lesquels l'effet de la charge est prépondérant, surtout en 1^{ère} sphère). L'effet le plus notable est observé sur les cations monochargés Cs^+ , K^+ et Li^+ , pour lesquels on avait déjà observé avec AMOEBA03 des fluctuations rapides de leur coordination en 1^{ère} sphère. Na^+ fait figure d'exception, car stabilisé dans un site octaédrique, l'augmentation des distances au sein de la 1^{ère} sphère n'est pas suffisante pour lui faire accepter de façon significativement différente une 7^{ème} molécule d'eau.

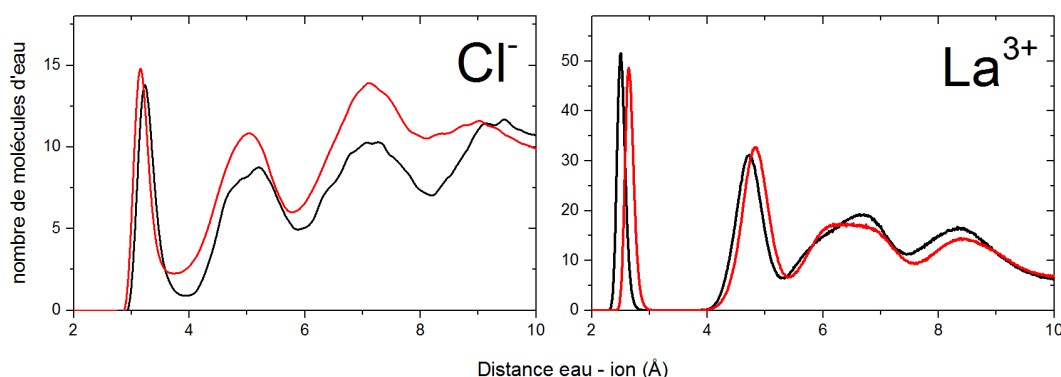


FIGURE 4.4 – Fonction de distribution radiale $\text{Cl}^- - \text{O}$ (départ en surface) et $\text{La}^{3+} - \text{O}$ (départ au centre de l'agrégat) pour 10 ns de simulation avec AMOEB03 (noir) ou AMOEB14 (rouge).

TABLE 4.2 – Comparaison des distances moyennes entre l'ion et l'oxygène d'une molécule d'eau, et des coordinations en 1^{ères} et 2^{ndes} sphère, pour les versions AMOEB03 et AMOEB14 du champ de forces de l'eau. Les données sont extraites de simulations de 10 ns à 200 K, pour les agrégats $\text{M}(\text{H}_2\text{O})_{100}^x$, $x=-2$ à $+3$.

	Distance M-O				Coordination			
	1 ^{ère} Sphère		2 ^{ème} Sphère		1 ^{ère} Sphère		2 ^{ème} Sphère	
	AM03	AM14	AM03	AM14	AM03	AM14	AM03	AM14
La^{3+} (int)	2.51	2.66	4.74	4.84	9	9	18.8	19.6
Ca^{2+} (int)	2.44	2.59	4.63	4.73	8	8	17.7	19.5
Li^+ (int)	2.04	2.29	4.10	4.25	4.4	5.2	13.9	14.8
Na^+ (int)	2.45	2.61	4.43	4.5	5.9	6.0	16.8	16.5
K^+ (int)	2.85	2.98	4.52	4.66	6.2	7.1	16.2	17.3
Cs^+ (int)	3.37	3.48	4.83	4.86	7.9	8.7	18.1	16.8
F^- (int)	2.84	2.73	4.67	4.52	5.9	5.4	17.9	17.5
Cl^- (ext)	3.29	3.22	5.12	5.05	5.0	5.0	11.8	12.5
Br^- (ext)	3.47	3.40	5.28	5.15	5.0	4.8	11.7	11.2
SO_4^{2-} (int)	3.85	3.70	5.59	5.51	12.2	11.5	24.7	25.6

4.2.3 Temps de résidence

Le temps de résidence des molécules d'eau dans la 1^{ère} sphère de solvation des ions est évalué à travers le modèle de coordination persistente proposé par Impey et al. [23]. Ce modèle autorise des excursions momentanées de molécules à l'extérieur de la 1^{ère} sphère, si elles sont inférieures à un temps t^* , généralement pris égal à 2 ps. C'est une propriété sensible, souvent mal reproduite par les champs de forces non polarisables. Cela peut expliquer que l'inhomogénéité de version des paramètres des ions et de l'eau se ressent plus pour cette propriété.

L'utilisation d'AMOEB14 (**Tab 4.3**) modifie des temps de résidences. La^{3+} et Ca^{2+} ont toujours une dynamique d'échanges trop lente pour calculer un temps de résidence à partir d'une simulation de 10 ns. Les cations monovalents perdent un peu de la stabilité de leur 1^{ère} sphère, dans un accord logique avec l'augmentation de la distance ion-eau et l'augmentation de la coordinence. Les écarts sont assez faibles pour rester en accord avec les données expérimentales. La mobilité au sein de la 1^{ère} sphère de Na^+ est même mieux reproduite avec la version 14. A l'inverse, les anions en surface gagne en stabilité, pour des raisons analogues. Le cas de F^- est un peu particulier, puisqu'il devient plus mobile avec AMOEB14, ce qui peut être attribué à sa coordinence de 5.9 en AMOEB03 qui devient en moyenne 5.4 en AMOEB14 et l'éloigne souvent d'une géométrie octaédrique.

Le comportement du sulfate est notre principale source de préoccupation, puisque l'étude de la spectroscopie de l'ion hydraté dans des agrégats de taille croissante occupe une partie des chapitres suivants. Dans une nanogoutte de 100 molécules d'eau, AMOEB03 lui prédit en moyenne 12.2 voisins, et la simulation AMOEB14 fournit la

TABLE 4.3 – Temps de résidence des molécules d’eau, (en ns) dans la première sphère de solvation des ions.

Ions	Expériences	AMOEBA 03 <i>Int/Ext</i>	AMOEBA 14
F ⁻		1.71 / 1.17	0.94
Cl ⁻		0.43 / 0.75	1.11
Br ⁻		0.08 / 0.64	0.48
Li ⁺	2.37 [24]	2.60 / 3.20	1.49
Na ⁺	1.25 [24]	3.08 / 2.35	1.31
K ⁺	0.81 [24]	0.66 / 0.58	0.47
Cs ⁺	0.28 [24], 0.2 [25], 0.1 [6]	0.27 / 0.44	0.28
Ca ²⁺	2.96 [24], ~ 10 [6]	peu d’échanges	peu d’échanges
La ³⁺	4.8 [24], ~ 10 [6]	pas d’échange	pas d’échange

valeur de 11.5, qui reste en accord avec la valeur expérimentale de 12. Au sein de la 2^{ème} sphère, les deux versions prédisent en moyenne 25 à 26 molécules d’eau, ce qui laisse envisager un effet négligeable sur son comportement. Cette étude valide donc la version 14 du champ de forces pour reproduire les comportements de l’ion sulfate hydraté dans des agrégats de grande taille. Il n’a pas été fait mention de la capacité du champ de forces à reproduire les énergies d’interactions ion-eau, mais plusieurs exemples seront traités en **chapitre 5**, comme l’évaluation des énergies relatives de l’ion sulfate hydraté dans de petits agrégats d’eau, en fonction des versions du champ de forces de l’eau.

Bibliographie

- [1] Robinson, G. W.; Zhu, S.-B.; Singh, S.; Evans, M. W. *Water in Biology, Chemistry and Physics. Experimental Overviews and Computational Methodologies*; World Scientific, Singapore : **1996**.
- [2] Kunz, W.; Henle, J.; Ninham, B. *Curr. Opin. Colloid Interface Sci.* **2004**, *9*, 19 - 37.
- [3] Collins, K. D.; Washabaugh, M. W. *Quart. Rev. Biophys.* **1985**, *18*, 323–422.
- [4] Lopez-Leon, T.; Jodar-Reyes, A. B.; Bastos-Gonzalez, D.; Ortega-Vinuesa, J. L. *J. Phys. Chem. B* **2003**, *107*, 5696-5708.
- [5] Conway, B. E. *Ionic hydration in chemistry and biophysics*; Elsevier : Amsterdam : **1981**.
- [6] Richens, D. T. *The Chemistry of Aqua Ions : Synthesis, Structure and Reactivity : A Tour Through the Periodic Table of the Elements*; Wiley : **1997**.
- [7] O'Brien, J. T.; Williams, E. R. *J. Am. Chem. Soc.* **2012**, *134*, 10228–10236.
- [8] O'Brien, J. T.; Prell, J. S.; Bush, M. F.; Williams, E. R. *J. Am. Chem. Soc.* **2010**, *132*, 8248-8249.
- [9] Asmis, K. R.; Neumark, D. M. *Acc. Chem. Res.* **2012**, *45*, 43-52.
- [10] Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1992**, *96*, 8249-8251.
- [11] Lambrecht, D. S.; McCaslin, L.; Xantheas, S. S.; Epifanovsky, E.; Head-Gordon, M. *Mol. Phys.* **2012**, *110*, 2513–2521.
- [12] Gopalakrishnan, S.; Jungwirth, P.; Tobias, D. J.; Allen, H. C. *J. Phys. Chem. B* **2005**, *109*, 8861-8872.
- [13] Brown, E. C.; Mucha, M.; Jungwirth, P.; Tobias, D. J. *J. Phys. Chem. B* **2005**, *109*, 7934-7940.
- [14] Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- [15] Grossfield, A.; Ren, P.; Ponder, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 15671–15682.
- [16] Grossfield, A. *J. Chem. Phys.* **2005**, *122*, 14671-14682.
- [17] Piquemal, J.-P.; Perera, L.; Cisneros, G. A.; Ren, P.; Pedersen, L. G.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*, 054511.
- [18] Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.
- [19] Clavaguéra, C.; Calvo, F.; Dognon, J.-P. *J. Chem. Phys.* **2006**, *124*, 074505.
- [20] Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. *J. Phys. Chem. B* **2015**, *119*, 9423-9437.
- [21] Daaz-Moreno, S.; Ramos, S.; Bowron, D. T. *J. Phys. Chem. A* **2011**, *115*, 6575-6581.
- [22] Müller, A.; Bögge, H.; Diemann, E. *Inorg. Chem. Commun.* **2003**, *6*, 52-53.
- [23] Impey, R. W.; Madden, P. A.; McDonald, I. R. *J. Phys. Chem.* **1983**, *87*, 5071–5083.
- [24] Helm, L.; Merbach, A. E. *Chem. Rev.* **2005**, *105*, 1923–1960.
- [25] Eigen, M. *Pure Appl. Chem.* **1963**, *6*, 97-116.

Chapitre 5

Spectroscopie des ions hydratés ($n \geq 36$)

L'eau liquide est structurée par un réseau tétraédrique de liaisons hydrogènes. La caractérisation de l'eau liquide, sous forme d'agrégats $(\text{H}_2\text{O})_n$, de $n = 2$ à plusieurs dizaines, ou encore d'aérosols [1], a été et continue d'être le champ de nombreuses études depuis des décennies [2–8]. La présence d'un ion, hydrophile ou hydrophobe, au sein de ce réseau, perturbe cette architecture, par la présence d'interactions compétitives ou coopératives entre l'eau et le soluté. En particulier, l'hydratation du dianion sulfate SO_4^{2-} est très étudiée, du fait de son intérêt dans des contextes environnementaux ou biologiques [9–11]. Si les premières études sur $\text{SO}_4^{2-}(\text{H}_2\text{O})_n$ se sont concentrées sur les petites tailles $n = 3-7$ (voir refs 12-21 de [12], **annexe D**), des données expérimentales sont maintenant disponibles jusqu'à $n = 250$ [13, 14]. Les méthodes d'ionisation par électrospray (ESI) permettent de former des sulfates hydratés dans des agrégats d'eau de tailles contrôlées. La discrimination en taille est permise par la sélection en masse dans un piège à ions, mais aussi par l'évaporation de l'eau à partir des ions piégés, obtenue par dissociation induite par collisions (CID) ou par photo-dissociation IR (IRPD, IRMPD si plusieurs photons sont impliqués, ou BIRD si le spectromètre joue le rôle d'un corps noir).

Williams réalise de nombreuses séries de spectres IRPD sur des agrégats solvatés [14–16]. En particulier, pour le sulfate, il propose des spectres pour une série de tailles $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$, avec $36 \leq n \leq 80$ [13], reproduite en **Fig 5.1**. Dans la gamme des élongations O-H, un large massif couvre la zone $3100-3700 \text{ cm}^{-1}$, et s'élargit lorsque la taille de l'agrégat augmente. Dans le même temps, le maximum d'intensité se déplace vers le rouge, de $\sim 3480 \text{ cm}^{-1}$ pour $n = 36$, à $\sim 3400 \text{ cm}^{-1}$ pour $n = 80$, fréquence qui correspond quasiment au maximum d'absorption de l'eau liquide et qui pourrait indiquer des molécules d'eau ressentant un environnement semblable à l'eau pure. De plus, une bande fine apparaît autour de 3710 cm^{-1} , à partir de $n = 47-50$, rendant compte de l'apparition de OH libres à la surface de l'agrégat. Notons que, dans les agrégats d'eau pure, la présence de OH libres de surface est détectée dès les plus petites tailles [4, 17].

Les signatures expérimentales fournissent des indices sur le pouvoir structurant de l'ion et sur la distance à laquelle il cesse d'être ressenti. Mais seule une étude théorique, combinant dynamique et spectroscopie, peut apporter un modèle du comportement de l'ion et de son environnement.

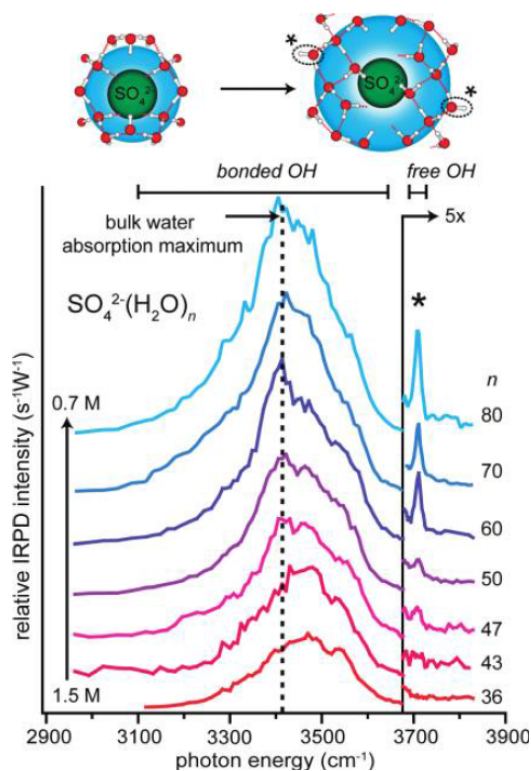


FIGURE 5.1 – Spectres IRPD de $[SO_4(H_2O)_n]^{2-}$ avec $36 \leq n \leq 80$, mesurés à 130 K. Expérimentalement, la zone des élongations O-H se compose d'un large massif de 3100 à 3700 cm^{-1} , qui s'élargit et dont le maximum se déplace vers le rouge lorsque la taille de l'agrégat augmente. De plus, une bande fine apparaît autour de 3710 cm^{-1} , à partir de $n = 47-50$, rendant compte de la présence d'OH libres à la surface de l'agrégat. Figure 1 de l'article [13].

5.1 Spectroscopie IR des bandes d'élongation OH. Application aux sulfates

Cette étude suit une progression chronologique de notre travail, avec les premiers résultats utilisant le modèle AMOEBA03 pour l'eau, et le constat de son incapacité à reproduire l'ensemble du massif des OH liés, puis une solution partielle pour augmenter sa flexibilité de réponse à l'environnement. Nous avons introduit en **chapitre 2** les modifications de la version AMOEBA14 des paramètres de l'eau, et validé sa capacité à reproduire la dynamique des ions hydratés en **chapitre 4**. A présent, nous testons le champ de forces sur un domaine beaucoup plus sensible qu'est la spectroscopie IR. Cette version se comporte intrinsèquement mieux que la précédente, mais demande des ajustements. Nous définissons les modifications de paramètres et les évolutions du modèle permettant de reproduire de façon satisfaisante les spectres des ions sulfates hydratés dans des agrégats de grande taille ($n = 36-100$). Une fois le modèle validé, nous discutons les informations qu'apporte la dynamique, notamment via le couplage entre le DACF et l'analyse dynamique des modes de vibration par DMD. Enfin, nous généralisons l'utilisation du DACF à l'étude des bandes d'élongation OH d'une large série d'ions hydratés, en fonction de leurs charges (-II à +III) et tailles ($n = 36, 100$).

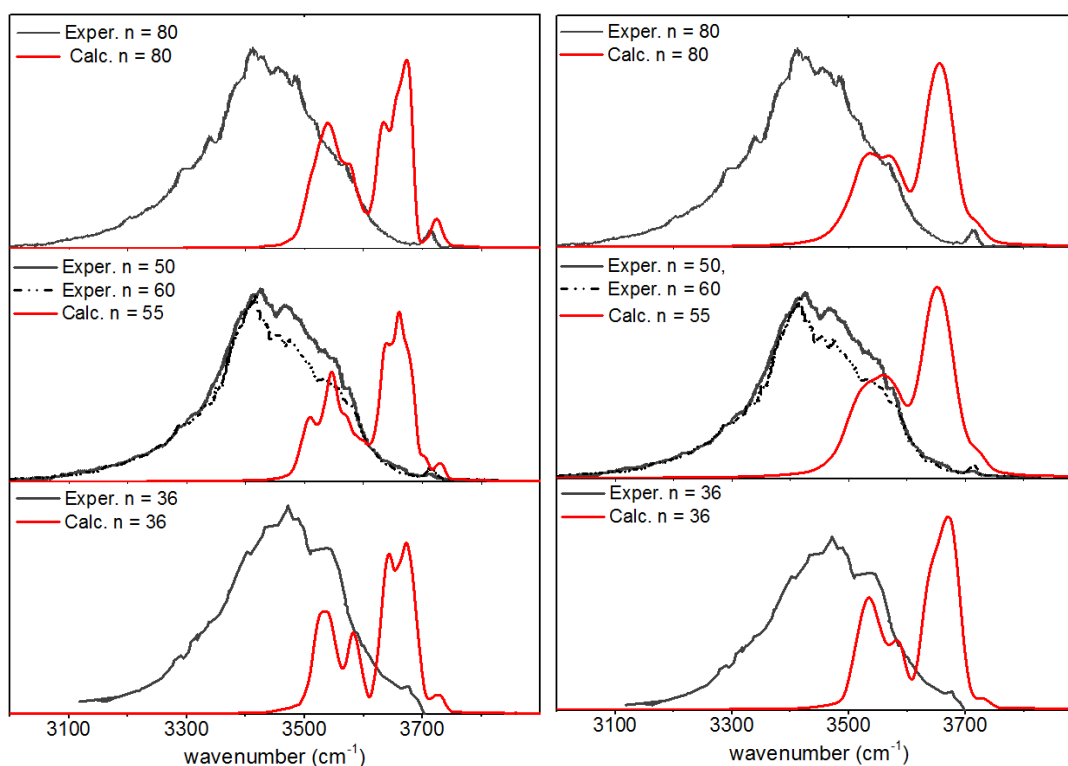


FIGURE 5.2 – Spectres DACF (courbes rouges) à 100 K (à gauche) et 200 K (à droite) pour les agrégats $[SO_4(H_2O)_n]^{2-}$, avec $n = 36$ (en bas), 55 (au centre) et 80 (en haut) molécules d'eau, comparés aux spectres expérimentaux de Williams (en noir) pour $n = 36, 50, 60$ (traits pointillés) et 80 à 130 K [14].

5.1.1 Position des massifs liés et apparition des bandes de OH libres en fonction de la taille (AMOEBAA03)

Toutes les simulations de dynamique DACF-AMOEBAA sont réalisées dans des conditions identiques, sauf précision particulière. Pour chaque taille d'agrégat $[SO_4(H_2O)_n]^{2-}$ avec $n = 36, 55, 80$ et 100, des structures initiales sont équilibrées [18] via des simulations de dynamique moléculaire de 2 à 3 ns, avec un pas d'intégration de 1 fs. Les équations du mouvement sont propagées via l'algorithme Verlet-Vitesse. Ces simulations sont réalisées dans l'ensemble canonique à 300 K en utilisant le thermostat Nosé-Hoover. Ces conditions permettent un bon équilibre des géométries des structures initiales. Pour calculer les spectres, les simulations sont effectuées en deux temps. Une première phase d'équilibration en température dans l'ensemble canonique, entre 10 et 50 ps, en fonction de la température à atteindre, puis 200 ps, avec un pas d'intégration de 0.1 fs, dans l'ensemble microcanonique, en enregistrant le moment dipolaire à chaque pas. Chaque spectre présenté est une moyenne sur 3 simulations.

Dans la série d'agrégats de taille croissante $[SO_4(H_2O)_n]^{2-}$ avec $n = 36, 55$ et 80, les spectres DACF à 100 K (**Fig 5.2, à gauche**) rendent compte de l'apparition d'une bande fine autour de 3720 cm^{-1} , de taille croissante avec la taille de l'agrégat. La position des elongations des OH libres est plutôt satisfaisante, 10 cm^{-1} déplacée vers le bleu par rapport à l'expérience (exp. : 3713 cm^{-1} , calc. : 3722 cm^{-1}). A 200 K (**Fig 5.2, à droite**), la bande perd en structuration, prenant la forme d'un simple épaulement, visible dès $n = 36$. La présence de OH libres sur le spectre DACF de $n = 36$ ne peut être discutée sur la base d'un seul spectre calculé. Elle peut naturellement résulter du modèle (effet structurant du sulfate insuffisant), mais également n'être qu'un artefact du choix de la structure de départ.

Les spectres de Williams [14] pour le sulfate, ou de Buck [19] pour les agrégats d'eau pure, montrent que les massifs des OH liés s'élargissent avec la taille de l'agrégat et que le maximum d'intensité se déplace vers le rouge. Les

raisons en sont la multiplication des couplages qui provoque des dédoublements de raies, et le nombre croissant d'interactions pour une molécule d'eau (jusqu'à être deux fois donneuse et deux fois acceptrice de liaison H) qui amoindrit la force des vibrateurs internes et déplace leurs fréquences vers le rouge. Dans les spectres calculés AMOEBA03, les massifs des OH liés sont très peu dépendants de la taille de l'agrégat. Ils sont divisés en deux bandes centrées sur les fréquences de la molécule isolée, ce qui indique très clairement un manque de flexibilité du modèle aux perturbations de l'environnement. Ce problème est récurrent dans de nombreux champs de forces (SPC/E, TIP4P, SPC/Fw, RWK2, AMOEBA) et a déjà été décrit par Burnham et Xantheas [20].

De nombreuses équipes sont spécialisées dans la modélisation de propriétés de l'eau et sur le développement de champs de forces de l'eau [21–23]. Il ne s'agissait pas pour nous de chercher à modifier conceptuellement le champ de forces, mais de proposer une solution simple à ce problème de description des massifs de OH en liaison hydrogène. Basiquement, les effets relatifs des composantes intra et intermoléculaires semblent être déséquilibrés au profit de l'intramoléculaire, et induisent un manque de flexibilité du modèle. Pour mieux traduire cette compétition, deux voies semblaient pouvoir être suivies :

- Influencer sur les interactions entre atomes non liés. Burnham et Xantheas [20] proposent un modèle de charges fluctuantes pour l'eau pure s'appuyant sur des surfaces d'énergie potentielle et de moment dipolaire du moment calculées par Partridge et Schwenke [24]. La présence de l'ion complique cette approche, mais nous discuterons l'implémentation d'un modèle de charges fluctuantes dans la section consacrée à AMOEBA14.
- Augmenter la flexibilité de l'intramoléculaire, via la composante d'élongation OH. Par la facilité de sa mise en oeuvre, c'est cette voie que nous avons explorée.

5.1.2 Modification de la constante d'anharmonicité du potentiel (AMOEBA03)

L'absence de déplacement vers le rouge de l'ensemble des élongations pourrait indiquer un manque de flexibilité des vibrateurs OH face à de multiples interactions. On rappelle l'expression du potentiel d'élongation dans AMOEBA, qui dérive directement du potentiel MM3 [25, 26].

$$E_b = K_b(b - b_0)^2[1 - cbnd(b - b_0) + qbnd(b - b_0)^2] \quad (5.1)$$

$$E_b = K_b(b - b_0)^2[1 - 2.55(b - b_0) + (7/12)2.55^2(b - b_0)^2] \quad (5.2)$$

avec K_b la constante d'élongation, $cbnd$ et $qbnd$ respectivement des facteurs de proportionnalité pour les termes cubique et quartique (indépendants du système étudié), b la longueur de liaison et b_0 sa valeur de référence. Le terme cubique $cbnd$ contrôle l'anharmonicité de la liaison et le terme d'ordre 4 $qbnd$ évite les situations de dissociation aux distances trop éloignées de la référence.

Nous avons cherché à recalibrer ce potentiel d'élongation pour les solvates avec 36 à 100 molécules d'eau, afin d'induire un déplacement vers les basses fréquences pour les molécules à multiples interactions, tout en perturbant au minimum les vibrateurs qui interagissent faiblement et dont les fréquences sont globalement bien traduites. Nous avons choisi d'apporter la flexibilité dont le potentiel semble manquer en agissant uniquement sur $cbnd$. En effet, la valeur de K_b , qui fixe la position du puits et le comportement harmonique (dans l'approximation de faibles déplacements autour de la distance d'équilibre) est bon, comme le montrent les fréquences des élongations symétrique et antisymétrique pour la molécule d'eau isolée (exp. : 3657 et 3756 cm^{-1} ; DACF : 3655 et 3753 cm^{-1}). Nous prévoyions que la correction à apporter à l'anharmonicité serait dépendante de l'ion et de la taille de l'agrégat.

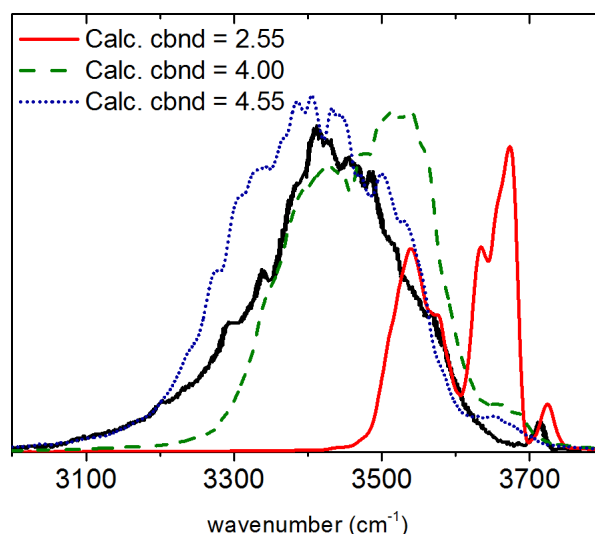


FIGURE 5.3 – Influence de la modification du facteur d'anharmonicité $cbnd$ sur la capacité des spectres DACF de $[SO_4(H_2O)_{80}]^{2-}$, à 100 K, à reproduire le spectre expérimental à 130 K (noir). Le spectre expérimental est reproduit de [13].

Nous avons validé une nouvelle valeur de $cbnd$ par ajustement de spectres DACF sur les spectres expérimentaux pour $n = 55$ et 80 , ces agrégats étant assez gros pour contenir des molécules d'eau impliquées dans tous les types d'interactions. Un exemple de l'influence de $cbnd$ sur le spectre calculé de $[SO_4(H_2O)_{80}]^{2-}$ est présenté en **Fig 5.3**. La flexibilité nouvelle des vibrateurs autorise un élargissement du massif et un déplacement général vers les basses fréquences. Toutefois, la bande de OH libres est elle aussi affectée, même à 100 K. Une valeur de $cbnd = 4.0 \text{ \AA}^{-1}$ a été validée. Elle permet un bon accord général du massif pour $n = 55$ et 80 , pour deux températures calculées 100 K et 200 K englobant la température expérimentale de 130 K (**Fig 5.4**), et reproduit bien la queue vers les basses fréquences. En contrepartie, la forme des bandes de OH libres perd en structure. On observe, suivant les cas, une bande ou un changement de pente, autour de 3700 cm^{-1} , ce qui ne permet plus de discuter l'apparition nette d'une bande avec l'augmentation de la taille de l'agrégat, (**Fig 5.5**) et encore moins de rendre compte de l'augmentation de la proportion d'OH libres avec la taille croissante de l'agrégat. De plus, son maximum d'intensité est déplacé vers le rouge d'environ 50 cm^{-1} .

Ce modèle, perturbant même les molécules d'eau les plus faiblement liées, montre la nécessité d'une réponse spécifique à l'environnement de chaque molécule, ce qui ne peut être apporté qu'au travers des interactions non liées. De plus, la correction à apporter à $cbnd$ pour reproduire le massif des OH liés peut être assez importante, dans des environnements neutres ou faiblement polarisants, comme pour les agrégats d'eau pure (**Fig 5.6**). Enfin, la correction variant avec les ions étudiés, toute comparaison de comportement deviendrait délicate.

5.1.3 Validation du modèle AMOEBA14 de l'eau pour la spectroscopie

En **chapitre 4**, nous avons montré que, bien que les termes de vdW des ions aient été calibrés sur la version 03 du modèle de l'eau, les comportements dynamiques sont globalement reproduits quand on les utilise sans changement avec AMOEBA14.

Concernant la spectroscopie, aucune donnée spécifique n'accompagne la présentation du nouveau modèle de l'eau [27], à l'exception de l'information que les termes intramoléculaires de AMOEBA03 ont été conservés, ce

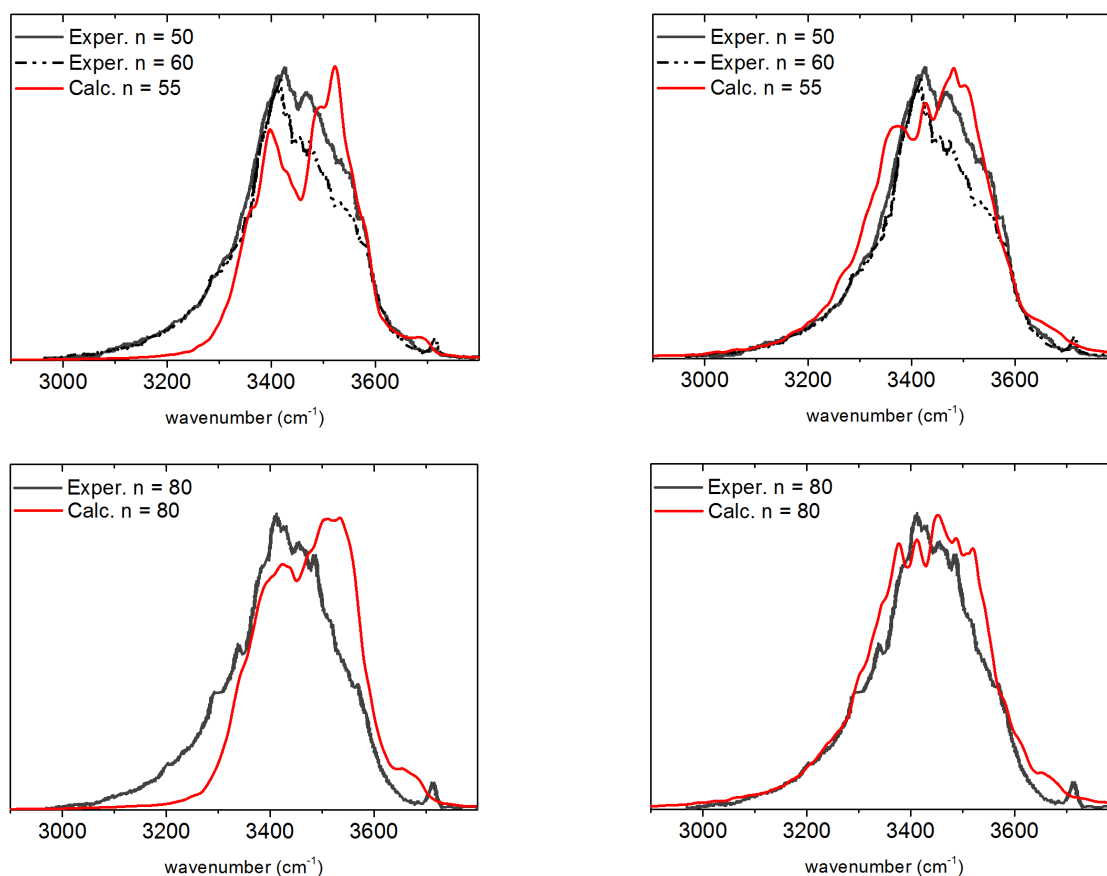


FIGURE 5.4 – Spectres IR d’ion sulfate hydraté par 55 (en haut) ou 80 (en bas) molécules d’eau. A gauche, la température calculée est de 100 K, et à droite de 200 K. La constante $cbnd$ est fixée à 4.0 \AA^{-1} . Les spectres expérimentaux, enregistrés à 130 K, sont issus de [14].

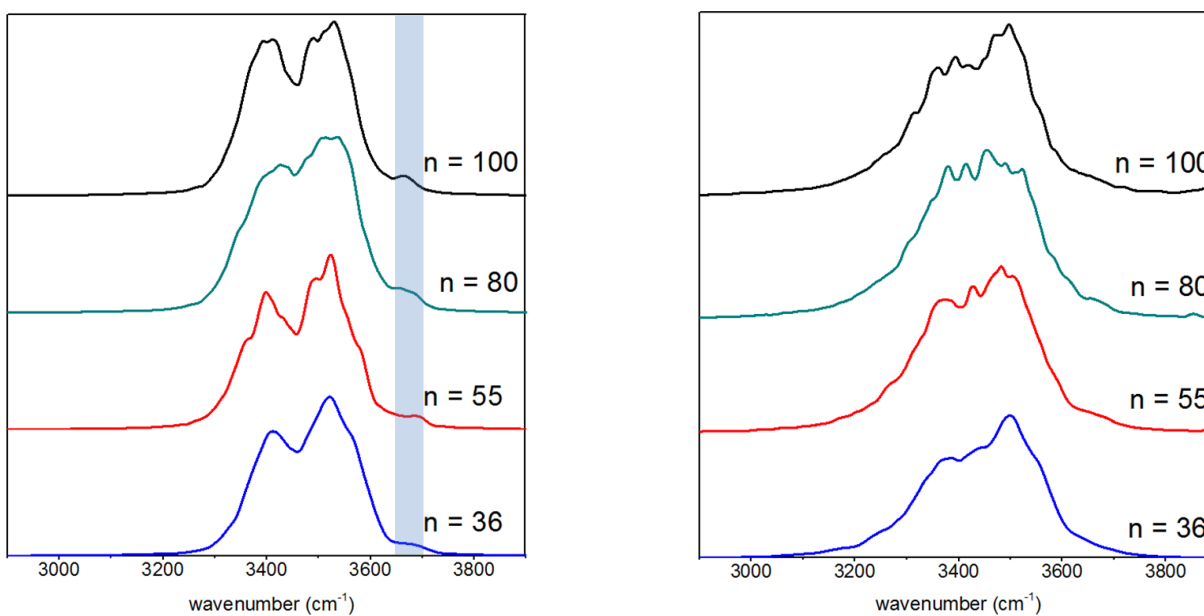


FIGURE 5.5 – Spectres IR d’ion sulfate hydraté dans des agrégats de tailles croissantes ($n = 36, 55, 80$ et 100), calculés pour deux températures 100 K (à gauche) et 200 K (à droite), $cbnd$ fixée à 4.0 \AA^{-1} .

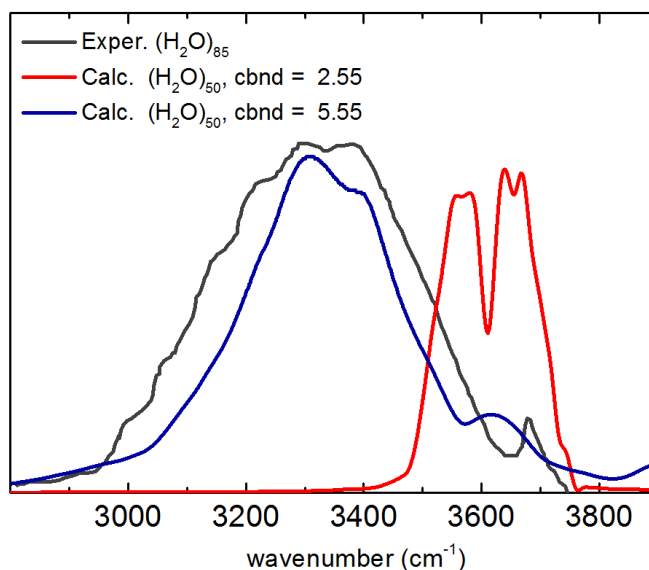


FIGURE 5.6 – Spectres calculés à 200 K pour des agrégats de 50 molécules d'eau, en employant une *cbnd* à 2.55 (rouge) ou à 5.55 (bleu). Les allures sont comparées qualitativement au spectre expérimental pour 85 molécules d'eau [19] (noir). L'accord général du massif principal n'est obtenu qu'au détriment de la bande de OH libres.

qui devrait garantir une absence de changement sur le spectre du monomère. Nous avons vérifié les fréquences du monomère et du dimère de l'eau par DACF. Comme attendu, pour le monomère (**Tab 5.1**) les faibles écarts de fréquences sont uniquement dûs au choix des auteurs d'utiliser nos corrections des paramètres intramoléculaires proposées pour les calculs de fréquences statiques, au lieu des dynamiques [28]. Pour le dimère (**Tab 5.2**), l'accord est encore bon, $\sim 20 \text{ cm}^{-1}$ d'écart en moyenne. On note que les fréquences DACF à 1 K ne nécessitent pas le recours à un facteur d'échelle, à la différence des fréquences statiques *ab initio*.

TABLE 5.1 – Pour le monomère H_2O , les fréquences DACF à 1 K, pour AMOEBA03 et AMOEBA14, sont comparées aux fréquences expérimentales

Monomère	EXP	AMOEBA03	AMOEBA14
pliage	1595	1595	1594
élong. sym.	3657	3658	3656
élong. anti.	3756	3756	3755

TABLE 5.2 – Pour le dimère $(\text{H}_2\text{O})_2$, écart entre les fréquences DACF (AMOEBA14) à 1 K ou MP2 ((c) [29]), et la moyenne d'une sélection de données exp. ((a) [30], (b) [31]).

Dimère	EXP				MP2 (c) aug-cc-pVTZ	AMOEBA14 DACF - 1 K
	Ar ^(a)	N ₂ ^(a)	He ^(b)	Moyenne		
pliage accept.	1593	1601		1597	27	21
pliage don.	1611	1619		1615	40	27
élong. sym. don.	3574	3550	3545	3556	192	12
élong. sym. accept.	3634	3627	3600	3620	218	20
élong. anti. don.	3709	3699	3714	3707	227	20
élong. anti. accept.	3726	3715	3730	3724	231	11

En augmentant la taille de l'agrégat d'eau, les fréquences du pliage se décalent rapidement vers le bleu, pour conduire à une situation analogue à celle présentée dans iAMOEBA [32] (**Fig 5.7**), pour lequel la fréquence du

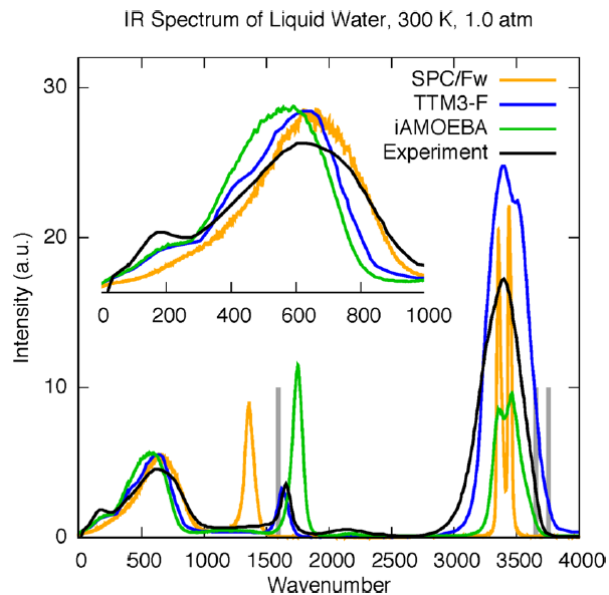


FIGURE 5.7 – Spectre IR de l’eau liquide reproduit par plusieurs champs de forces. iAMOEBAs (en vert) surestime de 100 cm^{-1} la gamme de fréquences du pliage de l’eau. Figure 4 publiée dans l’article [32]

pliage de l’eau liquide est surestimée de plus de 100 cm^{-1} . Nous avons choisi deux agrégats d’eau pure, de 21 et 50 molécules, afin de calibrer une nouvelle constante de pliage de l’eau. Nous utilisons dans la suite la valeur de $K_\theta = 43.65\text{ kcal.mol}^{-1}.\text{degre}^{-1}$ au lieu de la valeur de $48.98\text{ kcal.mol}^{-1}.\text{degre}^{-1}$ de référence. Nous avons pu observer que cette modification n’influence pas significativement les autres massifs, et notamment pas le massif des élongations (qui pourrait être affecté au travers du terme croisé couplant le pliage et l’élongation).

5.1.4 Position des massifs liés et apparition des bandes de OH libres en fonction de la taille (AMOEBAs14)

Les spectres DACF pour $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$ avec $n = 36, 55, 80, 100$, ont été calculés dans les mêmes conditions que précédemment (**Fig 5.8**). De manière générale, les allures sont bien plus satisfaisantes qu’avec AMOEBAs03. La résolution et les intensités des bandes des OH libres sont en bon accord avec l’expérience, bien que déplacées vers le rouge de 40 cm^{-1} en moyenne. Sur les petites tailles et à basse température, le spectre calculé est assez dépendant de la structure de départ. Pour en limiter l’impact, les spectres de $n = 36$ sont moyennés sur 4 structures initiales (soit $4 \times 3 \times 200\text{ ps}$). Ainsi, à 100 K , il ne semble apparaître un OH libre qu’épisodiquement, plus en accord avec le spectre expérimental. Les massifs des OH liés sont plutôt bien centrés autour de 3450 cm^{-1} , et sont plus larges qu’en AMOEBAs03 (300 au lieu de 200 cm^{-1}), mais pas encore suffisamment pour décrire l’intégralité des massifs expérimentaux. La forme des massifs, en double bosse, même si elle est moins marquée que précédemment, n’est pas vraiment satisfaisante. Seul $n = 36$ présente une allure très proche de l’expérience, qui sera expliquée par les outils d’attribution dynamique des modes normaux.

Il n’est guère encourageant de devoir intervenir sur le modèle de l’eau, à peine ses nouveaux paramètres publiés. Avant de nous y résoudre, nous avons vérifié la cohérence des paramètres du sulfate, ce qui nous a conduit à recalibrer la bande d’élongation S=O et à extraire un nouveau jeu de multipôles pour le sulfate. La justification et les détails de cette calibration sont présentés en **annexe C**, mais confirment que le problème de la forme des massifs,

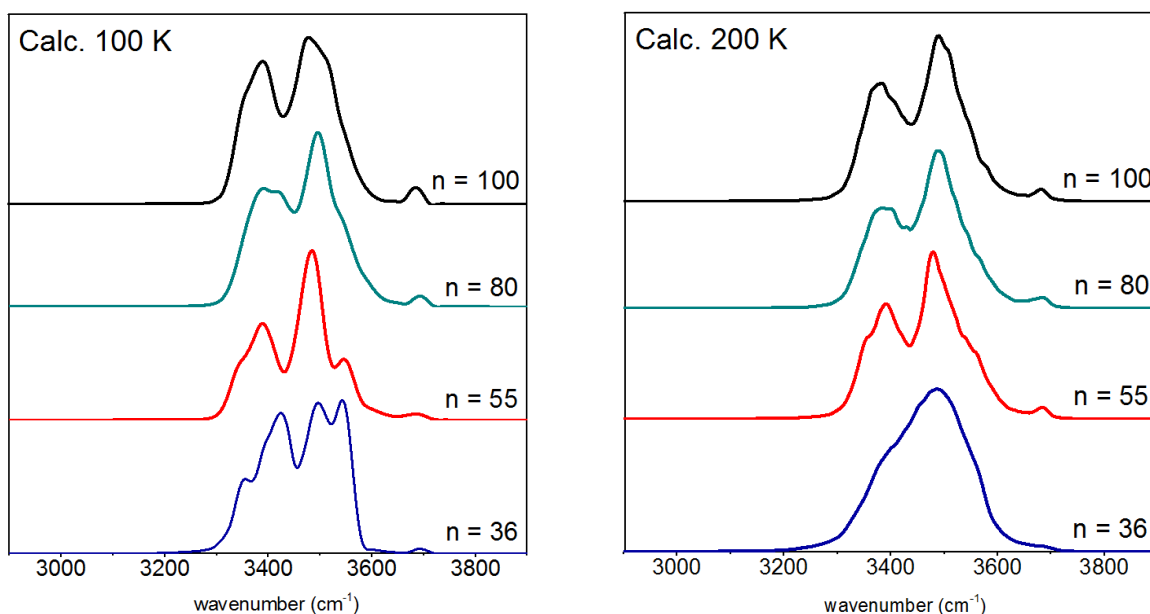


FIGURE 5.8 – Spectres IR d’ion sulfate hydraté dans des agrégats de tailles croissantes ($n = 36$ à 100) pour deux températures 100 K et 200 K, en utilisant le modèle AMOEBA14 de l’eau.

propre au modèle de l’eau, ne saurait être résolu uniquement via des compensations apportées par l’ion.

5.1.5 Modèle à charges fluctuantes

Nous avons déjà évoqué le modèle de charges fluctuantes de Burnham et Xantheas [33], appliqué au champ de forces TTM2-F pour l’eau pure. Le modèle est ajusté sur les surfaces d’énergie potentielle et de moment dipolaire calculées par Partridge et Schwenke [24] pour l’eau pure. Les auteurs postulent que la mauvaise description par les champs de forces classiques est due au fait que les charges sont fixes et indépendantes de la géométrie de la molécule. Sans redistribution des charges lors du changement de géométrie, il serait par exemple impossible de traduire l’augmentation de l’angle \widehat{HOH} dans des agrégats de taille importante. Xantheas et Fanourgakis [20] proposent une version TTM3-F dans laquelle la redistribution des charges atomiques permet au moment dipolaire de varier de façon non linéaire avec les changements de géométrie. Le modèle est testé sur sa capacité à reproduire des spectres calculés au niveau MP2/aug-cc-pVTZ d’agrégats $(\text{H}_2\text{O})_n$ jusqu’à $n = 20$. Si l’accord avec l’*ab initio* n’est pas parfait, le modèle démontre clairement sa capacité à proposer des bandes distribuées dans toute la gamme $3100\text{-}3800\text{ cm}^{-1}$, attestant d’une grande flexibilité de réponse à l’environnement.

Ponder [34] note que, comparée aux énergies d’induction SAPT, l’énergie de polarisation AMOEBA est généralement trop faible. La différence pourrait facilement être attribuée au transfert de charge et une exploration de méthodes permettant de l’inclure dans le modèle d’AMOEBa est en cours. Un modèle de charges fluctuantes est une façon d’introduire de la polarisation mais AMOEBA l’inclut par un calcul explicite, avec toute la directionnalité offerte par les multipôles. Il ne s’agissait donc pas de se substituer à ce modèle, mais uniquement d’apporter une contribution supplémentaire. Tenter d’ajuster le modèle de charges fluctuantes uniquement sur des surfaces du monomère de l’eau, sans prendre en compte les interactions spécifiques avec les ions (dont on cherche à étudier

une large série), n'aurait pas de sens. Il semble beaucoup plus direct de calibrer le modèle sur un jeu de données spectroscopiques, et d'en mesurer la sensibilité à la variation de l'ion ou de la taille de l'agrégat. Qualitativement, la forme choisie est celle proposée par Xantheas et Fanourgakis. Dans leur modèle, la variation des charges prend la forme :

$$\tilde{q}^{H_1} = q^{H_1} + d_r(r_{OH_1} - r_e) + d_\theta(\theta - \theta_e), \quad (5.3)$$

$$\tilde{q}^{H_2} = q^{H_2} + d_r(r_{OH_2} - r_e) + d_\theta(\theta - \theta_e), \quad (5.4)$$

$$\tilde{q}^O = -\tilde{q}^{H_1} - \tilde{q}^{H_2}, \quad (5.5)$$

où \tilde{q}^{H_1} et \tilde{q}^{H_2} sont les charges fluctuantes des hydrogènes H_1 et H_2 , et dépendent des charges fixes q^{H_1} et q^{H_2} , et de l'écart à la longueur OH de référence r_e et à l'angle θ_e de référence. Deux paramètres d_r et d_θ pondèrent l'influence des changements de géométrie sur les charges. La charge effective sur l'oxygène est déduite de façon à garder chaque molécule neutre. Qualitativement, nous attendons du modèle qu'il affecte significativement les géométries loin de l'équilibre, en préservant les vibrateurs faiblement perturbés, comme les OH libres. Nous proposons donc une dépendance quadratique des charges en fonction des distances et angles. Les valeurs des paramètres d'ajustement visent à reproduire l'ensemble des massifs pour les sulfates à $n = 36-100$.

En notant $\Delta r_i = (r_{OH_i} - r_e)$ et $\Delta\theta = (\theta - \theta_e)$, le modèle devient :

$$\tilde{q}^{H_i} = q^{H_i} + 0.6(1.1\Delta r_i^2 + 0.05\Delta r_i) + (1.1 \times 10^{-4}\Delta\theta^2 + 5 \times 10^{-4}\Delta\theta), \quad (5.6)$$

$$\tilde{q}^O = -\tilde{q}^{H_1} - \tilde{q}^{H_2}, \quad (5.7)$$

On peut noter que le modèle de charges fluctuantes est basé sur les variations de la géométrie intramoléculaire et que les effets d'environnement ne sont, encore une fois, traités que de façon indirecte.

TABLE 5.3 – Comparaison des variations de longueurs OH, angles \widehat{HOH} et charges par rapport aux valeurs de référence, dans un modèle à charges statiques ou fluctuantes, appliqué au système $[SO_4(H_2O)_{80}]^{2-}$.

	charges statiques	charges fluctuantes
Δr (Å)	[-0.05 :+0.1]	[-0.07 :+0.13]
$\Delta \theta$ (deg.)	[-15 :+10]	[-17 :+13]
$\Delta \tilde{q}^H$ (e)	0	[-0.05 :+0.06]
$\Delta \tilde{q}^O$ (e)	0	[-0.11 :+0.11]

Le modèle induit une majoration des variations maximales de géométries de l'ordre de 20 à 30 % par rapport au modèle à charges fixes, comme présenté dans le **Tab 5.3**. On notera que la valeur de l'angle \widehat{HOH} au cours des simulations est en moyenne plus faible que sa valeur de référence, et que le modèle de charges fluctuantes n'améliore que très faiblement la situation. La gamme de variation de charges des hydrogènes reste dans les limites des variations illustrées pour la calibration de la surface de moment dipolaire de l'eau par Xantheas [20], et est assumée comme physiquement représentative. Nous nous sommes posé la question du cumul des deux modèles de polarisation au sein du champ de forces, et la possibilité du double comptage d'une contribution. Le modèle de polarisation d'AMOEBa n'induit pas de polarisation directe intramoléculaire dans l'eau, les 3 atomes d'une molécule étant par définition dans un même groupe de polarisation, aucune contribution directe (de type $\alpha_i T_\alpha^{ij} M_j$) intramoléculaire n'est calculée. A l'inverse, le modèle de charges fluctuantes conserve la neutralité des molécules, mais autorise des variations de charges au sein d'une molécule. Le calcul auto-cohérent de la polarisation est réalisé à chaque pas de simulation, avec un jeu de charges *a priori* différent de celui du pas précédent, mais qui n'évolue pas au cours du processus itératif. Les variations de distances r_{OH} , d'angles θ , et de charges effectives \tilde{q}^{H_1} , \tilde{q}^{H_2} et

\bar{q}^O , ont été suivies pendant des simulations de 200 ps pour nous assurer que leur gamme de variations n'évolue pas au cours de la dynamique (ce qui serait une indication d'un biais du modèle).

Pour $n = 36$, le modèle à charges fluctuantes induit principalement un élargissement du massif vers le rouge, permettant de reproduire efficacement les basses fréquences jusqu'à 3300 cm^{-1} (**Fig 5.9**). Pour $n = 80$, le massif est élargi et légèrement déplacé vers le rouge. La position et la largeur de la bande des OH libres est gardée intacte, la variation d'intensité peut être attribuée à l'échantillonnage. La forme en double bosse du massif des OH liés ne disparaît pas, même si, avec l'élargissement du massif le rouge, l'effet est atténué.

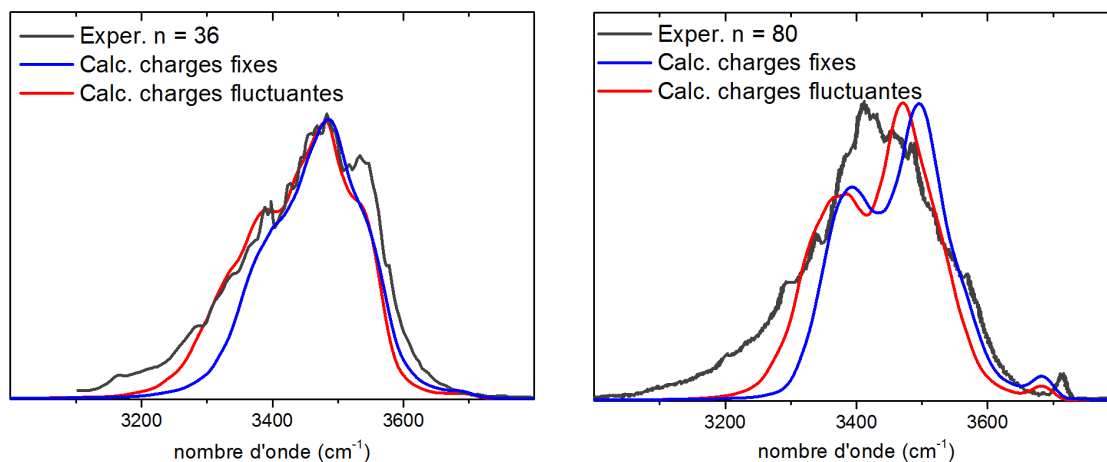


FIGURE 5.9 – Spectres DACF à 200 K de $[SO_4(H_2O)_{36}]^{2-}$ (à gauche) et $[SO_4(H_2O)_{80}]^{2-}$ (à droite), dans la gamme des élongations O-H, calculés à partir d'un modèle à charges fixes ou fluctuantes. Les spectres expérimentaux à 130 K sont reproduits de [13].

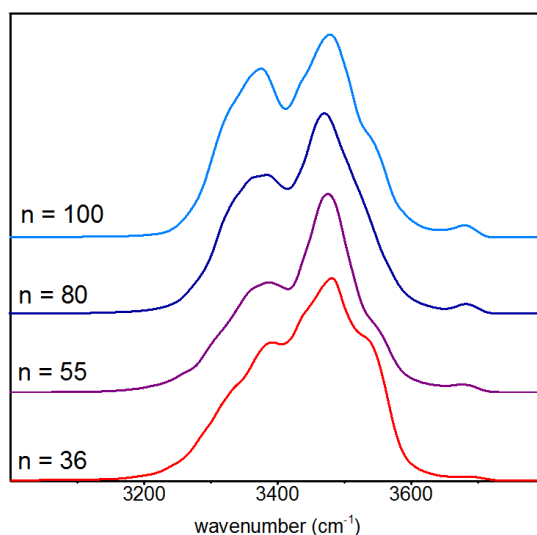


FIGURE 5.10 – Spectres de $[SO_4(H_2O)_n]^{2-}$ de taille croissante, avec $n = 36$ à 100, centrés sur la gamme des élongations O-H, calculés à 200 K, à partir du modèle à charge fluctuantes.

Sur la série $n = 36-100$ (**Fig 5.10**), on assiste à une augmentation de l'intensité des contributions dans le rouge avec la taille de l'agrégat, en même temps que l'apparition progressive d'une bande de OH libres. Qualitativement, la nouvelle contribution dans le rouge peut être attribuée à l'émergence d'une troisième puis d'une quatrième sphères de solvation, pour lesquelles l'influence de l'ion sulfate s'affaiblit. On peut relier ces résultats aux spectres expérimentaux de Buck [19] pour l'eau et de Williams [13,14], pour les sulfates hydratés, qui montrent très clairement

qu'à nombre de molécules d'eau équivalent, le massif de l'eau s'élargit beaucoup plus dans le rouge (**Fig 5.11**). Sur les spectres du sulfate, on note un déplacement progressif vers le rouge de la bande de OH libres avec l'augmentation de la taille de l'agrégat et la diminution de l'effet de l'ion central, qui reste non négligeable pour 250 molécules d'eau, indiquant des effets structurants à assez longue distance.

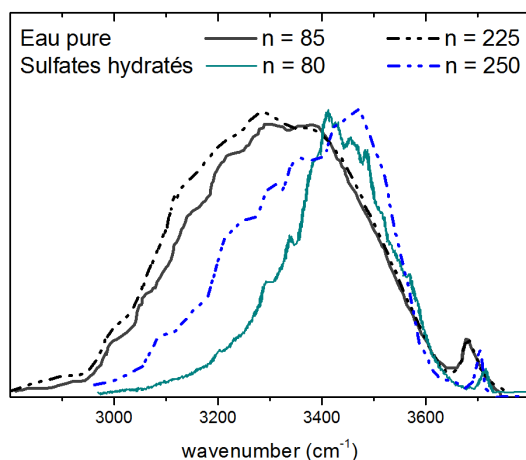


FIGURE 5.11 – Spectres expérimentaux d'agrégats d'eau pure [19] et de sulfates hydratés [13,14] sur la gamme des elongations O-H.

5.1.6 Validation du modèle sur les agrégats d'eau pure

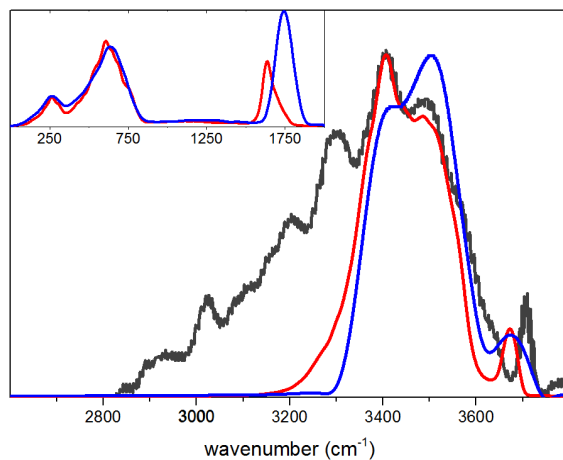


FIGURE 5.12 – Spectres DACF d'agrégats d'eau pure $(\text{H}_2\text{O})_{50}$ calculés à 200 K, avec les paramètres initiaux de AMOEBA14 (bleu), et avec les modifications sur la constante de pliage \overline{HOH} et le modèle à charges fluctuantes (rouge), comparés au spectre expérimental de $(\text{H}_2\text{O})_{48}$ [19] (noir).

La modification de la constante de pliage de l'eau repositionne correctement le mode (**Fig 5.12**), sans modification du spectre dans l'IR lointain ou dans la gamme des elongations O-H. Le modèle à charges fluctuantes améliore la description des hautes fréquences du massif des elongations O-H, et augmente la résolution de la bande des OH libres. L'accord est très correct sur la gamme $3300\text{-}3600\text{ cm}^{-1}$ et un élargissement de 50 cm^{-1} vers le rouge est observé. Cependant, le modèle ne reproduit toujours pas la gamme $2850\text{-}3200\text{ cm}^{-1}$, qui devrait être peuplée par les

modes les plus anharmoniques. La calibration du modèle de charges fluctantes sur des agrégats d'eau améliorerait probablement leur description, de même que la prise en compte des effets de transfert de charge. Certains des modes de combinaison, tels que l'harmonique du mode de pliage, pourraient également contribuer. Ce dernier apparaît bien vers 3200 cm^{-1} mais dans une intensité très faible. Des calculs GVPT2 sur de petits agrégats $[\text{SO}_4(\text{H}_2\text{O})_9]^{2-}$, confirment la faible intensité de ces contributions.

5.1.7 Attribution des modes normaux de vibration : DMD sur $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$, $n = 36$ et 80

Les spectres de dynamique, obtenus par la fonction d'autocorrelation du dipôle, n'offrent pas d'accès direct aux modes normaux de vibration. Les outils développés autour de la DMD, détaillés au **chapitre 3**, permettent d'accéder à ces modes de façon dynamique. Pour un ion hydraté par plusieurs sphères de solvation, la réponse des vibrateurs OH de l'eau fluctue au gré de la dynamique de réarrangement de l'agrégat. La DMD capture des modes normaux pendant des simulations très courtes (généralement 200 fs), signature d'une structure pendant la dynamique, mais qui ne peut prétendre rendre compte à elle seule de l'ensemble des bandes d'un spectre calculé de 600 ps. Afin de relier chaque type de vibrateur ou d'environnement à sa gamme de fréquences d'absorption, il peut être nécessaire d'appliquer la procédure DMD à un jeu plus ou moins important de structures.

$[\text{SO}_4(\text{H}_2\text{O})_{36}]^{2-}$

Pour attribuer le spectre DACF à 200 K de $[\text{SO}_4(\text{H}_2\text{O})_{36}]^{2-}$, 3 structures ont été extraites de la dynamique DACF puis réoptimisées avec AMOEBA (**Fig 5.13**). Ces structures sont différentes par leurs énergies relatives et leurs géométries. La structure W36-A possède 12 molécules en 1^{ère} sphère de solvation alors que les structures W36-B et W36-C en ont seulement 11. Toutes ont 18 molécules en 2^{ème} sphère, alors qu'il a été observé pour $[\text{SO}_4(\text{H}_2\text{O})_{100}]^{2-}$ que la seconde sphère est complète avec ~ 25 molécules. W36-B est la seule à posséder une molécule d'eau en 4^{ème} sphère et W36-B et W36-C ont chacune une OH libre. Dans les deux cas, la molécule concernée est une fois donneuse et une fois acceptrice de liaison hydrogène (AD) et leur environnement est semblable (chacune donne une liaison hydrogène vers une molécule (AADD) et en reçoit une d'une molécule (ADD)).

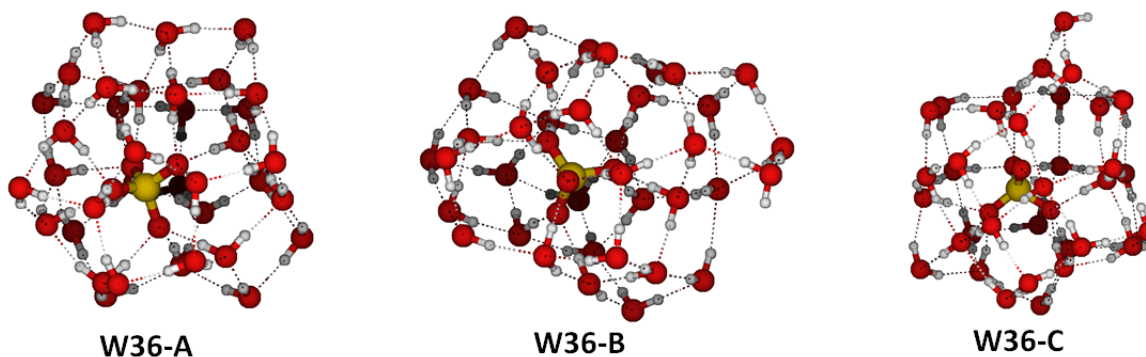


FIGURE 5.13 – 3 structures extraites d'une dynamique DACF. Après optimisation avec AMOEBA, la structure A est la plus basse énergie. Les structures B et C sont respectivement 1 et 2.9 kcal.mol⁻¹ plus hautes en énergie.

Pourtant, le vibrateur libre de W36-C résonne dans la gamme $3722\text{-}3726\text{ cm}^{-1}$, très déplacée dans le bleu, mais acceptable du fait de l'environnement (AD) et l'autre vibrateur de la molécule affiche une résonance à 3448

cm^{-1} . Pour W36-B, le vibrateur libre résonne dans la gamme $3658\text{-}3663 \text{ cm}^{-1}$, et le vibrateur lié à 3484 cm^{-1} . Pour expliquer ces différences, il suffit de suivre les oscillations des vibrateurs au cours d'une simulation DMD où la fréquence d'entraînement correspond à la fréquence de résonance du vibrateur libre. En **Fig 5.14**, on suit, à gauche, les deux oscillateurs de la molécule (AD) de W36-B, à la fréquence de 3663 cm^{-1} , et à droite, ceux de W36-C à la fréquence de 3726 cm^{-1} . Deux points sont facilement identifiables. La longueur moyenne du vibrateur lié dans W36-C est notablement plus importante que dans W36-B, et l'inverse est observé pour les OH libres (de l'ordre de 0.5 %). On peut corrélérer un écart important à la longueur OH de référence à la présence d'une liaison hydrogène forte, qui, au travers du couplage de Urey-Bradley, induit le rétrécissement du vibrateur libre. Le second point est la perturbation née de la présence des vibrateurs des molécules proches. Si l'élongation libre de W36-C croît linéairement sans contrainte, on observe qu'il n'en est pas de même pour W36-B. L'environnement autour du vibrateur lié est changeant et déforme son élongation. Il faut rappeler ici le mode d'action de la force d'entraînement, qui est appliquée sur chaque atome, à chaque pas de la dynamique, mais avec une intensité dépendante du temps et de l'environnement du vibrateur. Si cette dépendance est responsable du phénomène de résonance, cet apport d'énergie induit également des fluctuations au sein des vibrateurs, un peu à la manière d'un thermostat (comme décrit en **chapitre 3**, le choix de conditions d'entraînement plus strictes vise en partie à limiter ces fluctuations). Dans le cas de W36-B, la perturbation de l'environnement du OH lié se répercute sur le OH libre. Notamment, dans une première phase (les 50 premières fs), les oscillations du vibrateur lié sont fortes (car perturbé par son environnement proche) et inhibent celles du vibrateur libre.

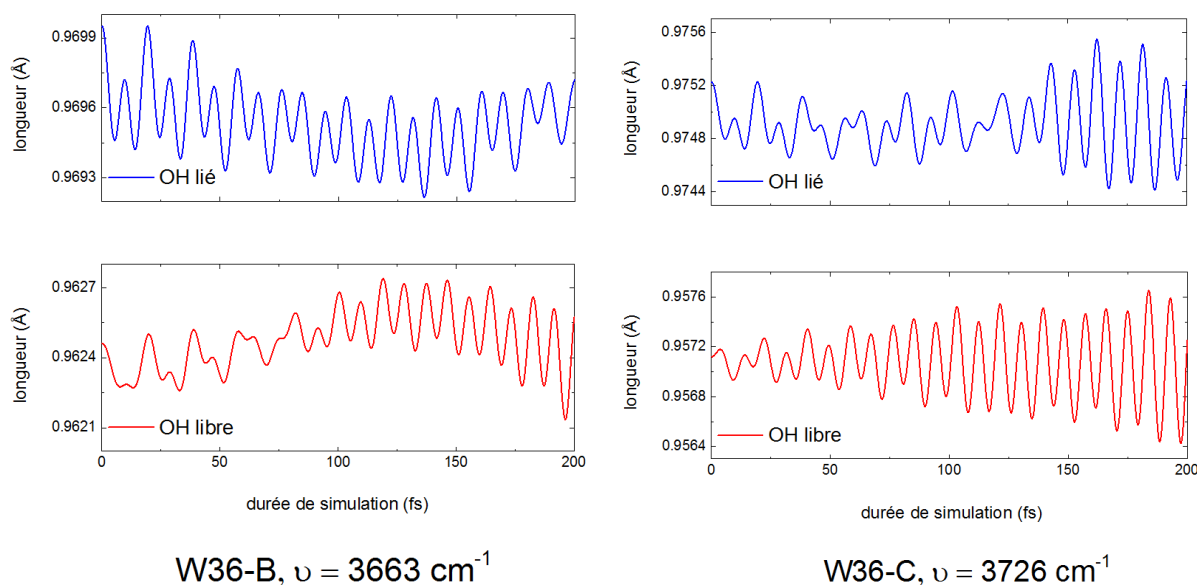


FIGURE 5.14 – Suivi des élongations de deux molécules d'eau présentant une OH libre. A gauche, à partir d'une simulation DMD à $\nu = 3663 \text{ cm}^{-1}$ sur la structure W36-B, et à droite, $\nu = 3726 \text{ cm}^{-1}$ sur la structure W36-C.

L'ensemble des attributions faites en utilisant des simulations DMD est présenté sur la **Fig 5.15** et aide à justifier la forme générale du massif des OH liés en DACF, en très bon accord avec l'expérience. Dans le massif des OH liés, on observe une grande concentration de bandes sous le maximum DACF (autour de $3450\text{-}3500 \text{ cm}^{-1}$), majoritairement des molécules de 1^{ère} ou de 2^{ème} sphères. Autour de ce maximum, la distribution de modes est assez homogène, du fait de la présence d'environnements très différents :

- Molécules de 1^{ère} sphère vers un oxygène du sulfate de coordinance 2 ou 3.
- Molécules de 2^{ème} sphère liées à ces deux types de molécules de 1^{ère} sphère.
- Molécules de 2^{ème} sphère AADD ou ADD.
- Molécules de 2^{ème} sphère uniquement liées à des molécules de 1^{ère} et 2^{ème} sphère ou également à des molécules

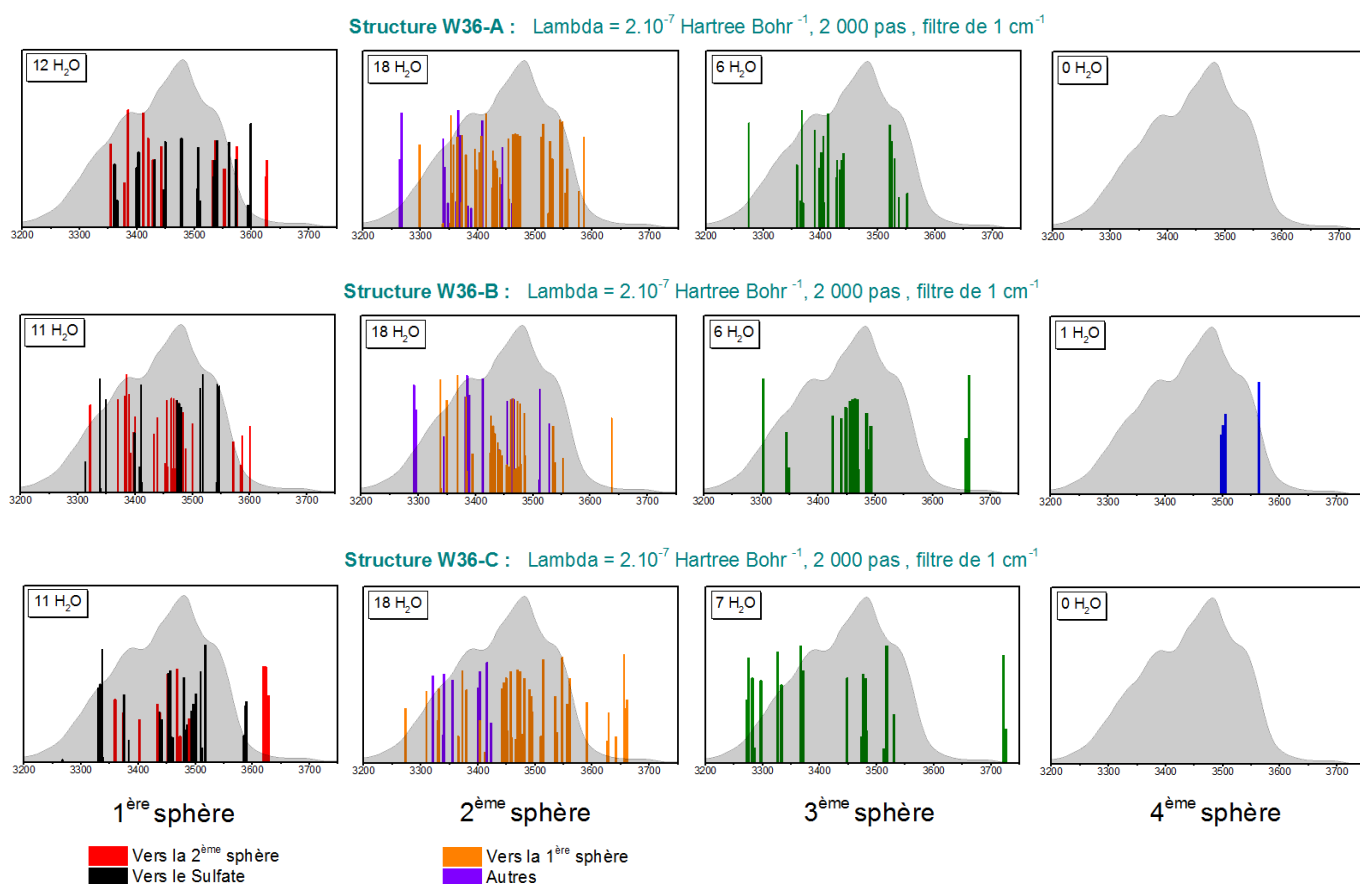
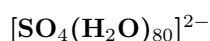


FIGURE 5.15 – Attribution dynamique de modes normaux pour $[SO_4(H_2O)_{36}]^{2-}$ par DMD à partir de 3 structures stables extraites d’une dynamique AMOEBA. Les fréquences de résonance DMD sont superposées au spectre DACF à 200 K (gris).

de 3^{ème} sphère.

– Molécules de 3^{ème} sphère.

Ces molécules sont caractérisées par des déplacements en lien avec leurs interactions particulières, et il n’y a pas à cette taille un type de molécules nettement prépondérant sur tous les autres. De plus, la répartition en fréquences des différents types de vibrateurs, qui est très nette pour les plus petits agrégats (voir $n = 24-25$ en **chapitre 6**), l’est beaucoup moins devant la diversité des environnements. Des calculs de fréquences statiques DFT/ ω B97-XD/6-311++G(d,p), réalisés sur quelques structures stables, attribuent les bandes les plus rouges aux vibrateurs de 1^{ère} sphère dirigés vers la 2^{ème} et aux vibrateurs de 2^{ème} sphère dirigés vers la 2^{ème} ou la 3^{ème}. La DMD est plutôt en accord avec la DFT concernant les contributions de la 2^{ème} sphère, mais est beaucoup plus mitigée sur l’ordre des contributions de 1^{ère} sphère, qu’elle prévoit très dépendant de leur environnement.



Pour $n = 80$, il y a 160 vibrateurs OH dont les fréquences en DACF s’échelonnent dans la gamme 3225-3725 cm⁻¹. Une recherche de structures de basse énergie a été réalisée avec l’outil REMD de TIREX. 16 répliques ont été générées. La distribution des températures a été choisie avec une progression géométrique dans une gamme restreinte (100-350 K), afin d’éviter l’évaporation de molécules en surface. Chaque trajectoire a été propagée avec un pas de temps de 1 fs et la température a été contrôlée avec le thermostat Nosé-Hoover. Des échanges entre

répliques de températures adjacentes ont été proposées toutes les 10 ps, et les configurations ont été sauvegardées tous les 5 ps. Le temps total de simulation de chaque réplique est de 2 ns. Les structures ont ensuite été optimisées avec AMOEBA et les deux structures les plus stables ont été sélectionnées pour une analyse DMD (**Fig 5.16**), pour laquelle les conditions précédentes ont été gardées, $\lambda = 2.10^{-7}$ Hartree.Bohr $^{-1}$ et $n_{step} = 2000$, à l'exception du filtre en fréquence abaissé à 0.5 cm^{-1} , devant la multiplication des réponses des sphères extérieures. La structure W80-A est la structure la plus stable à l'issue de la procédure REMD. Ses deux premières sphères de solvation sont pleines puisqu'elle compte 12 molécules en 1^{ère} sphère et 25 en seconde. Elle possède également 28 molécules en 3^{ème} sphère et 15 en 4^{ème} sphère, et 6 OH libres, également répartis entre ses deux sphères extérieures. Pour chacun de ces OH libres, la fréquence de plus forte résonance est inscrite dans la gamme $3690\text{-}3699 \text{ cm}^{-1}$, mais d'autres résonances plus faibles sont observées sur toute la gamme $3686\text{-}3708 \text{ cm}^{-1}$. La structure W80-B est également une structure stable issue du REMD, $0.5 \text{ kcal.mol}^{-1}$ plus haute en énergie que W80-A. Elle possède 11 molécules en 1^{ère} sphère, 25 en 2^{ème}, 26 en 3^{ème} et 18 en 4^{ème}. Elle présente également 6 OH libres, dont 5 en 4^{ème} sphère. Leurs fréquences de résonance sont réparties sur la gamme $3672\text{-}3722 \text{ cm}^{-1}$.

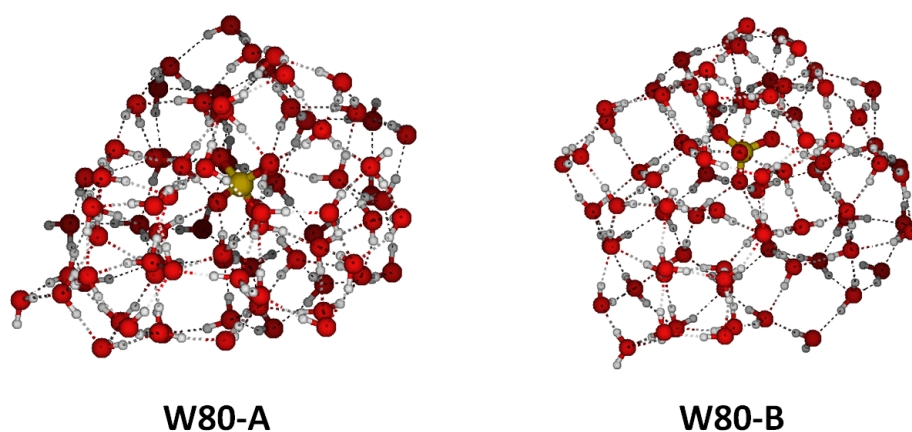


FIGURE 5.16 – Structures initiales de l'analyse DMD. A gauche, W80-A, à droite W80-B

Le spectre DACF pour $n = 80$ affiche une forme en double bosse qu'il n'avait pas pour $n = 36$. La première explication est l'homogénéisation des environnements. Dans les deux agrégats W80-A et W80-B, les deux premières sphères sont pleines ou quasiment et la 2^{ème} sphère n'est pas du tout en surface (car entourée de 43 ou 44 molécules d'eau). La moitié des molécules de l'agrégat est en 3^{ème} ou 4^{ème} sphère, et cette proportion devient de $\sim 2/3$ dans un agrégat de 100 molécules d'eau. Ces molécules, à l'exception des quelques unes présentant des OH libres, ont des environnements assez similaires. Les vibrateurs d'une molécule sont généralement très couplés et conduisent à des combinaisons symétriques et antisymétriques, les premières majoritairement pour $\nu \leq 3350 \text{ cm}^{-1}$ tandis que les secondes apparaissent lorsque $\nu \geq 3425 \text{ cm}^{-1}$. Elles expliquent l'apparition progressive d'un massif en double bosse pour $n = 80$ puis 100, les modèles de polarisation et de charges fluctuantes manquant de flexibilité pour discriminer ces environnements proches.

L'analyse de répartition des modes par DMD a l'avantage d'offrir une image globale des gammes de résonance des différents vibrateurs mais la comparaison directe avec les intensités DACF peut être source d'erreurs. Par exemple, la structure W80-B (**Fig 5.17**) fait apparaître de multiples contributions de 4^{ème} sphère entre 3410 et 3425 cm^{-1} , qui pourraient faire penser que de nombreux vibrateurs résonnent à cette gamme qui est pourtant un minimum local sur le spectre DACF. Mais cet ensemble de bandes est en fait majoritairement issu du même vibrateur, dont la flexibilité lui permet de résonner sur une gamme assez étendue de fréquences. Cet exemple souligne le choix de représentation des résonances. On aurait pu ne faire apparaître, pour chaque vibrateur, que la fréquence de plus forte résonance, mais on perd alors l'information de capacité du mode à résonner sur une gamme

de fréquences plus ou moins large. On a préféré traiter uniformément toutes les résonances, uniquement par un filtre de fréquences. La contrepartie est d'être prudent en comparant les intensités DACF à la densité de modes DMD, d'autant qu'une telle comparaison requiert un échantillonnage de structures plus important que celui présenté ici.

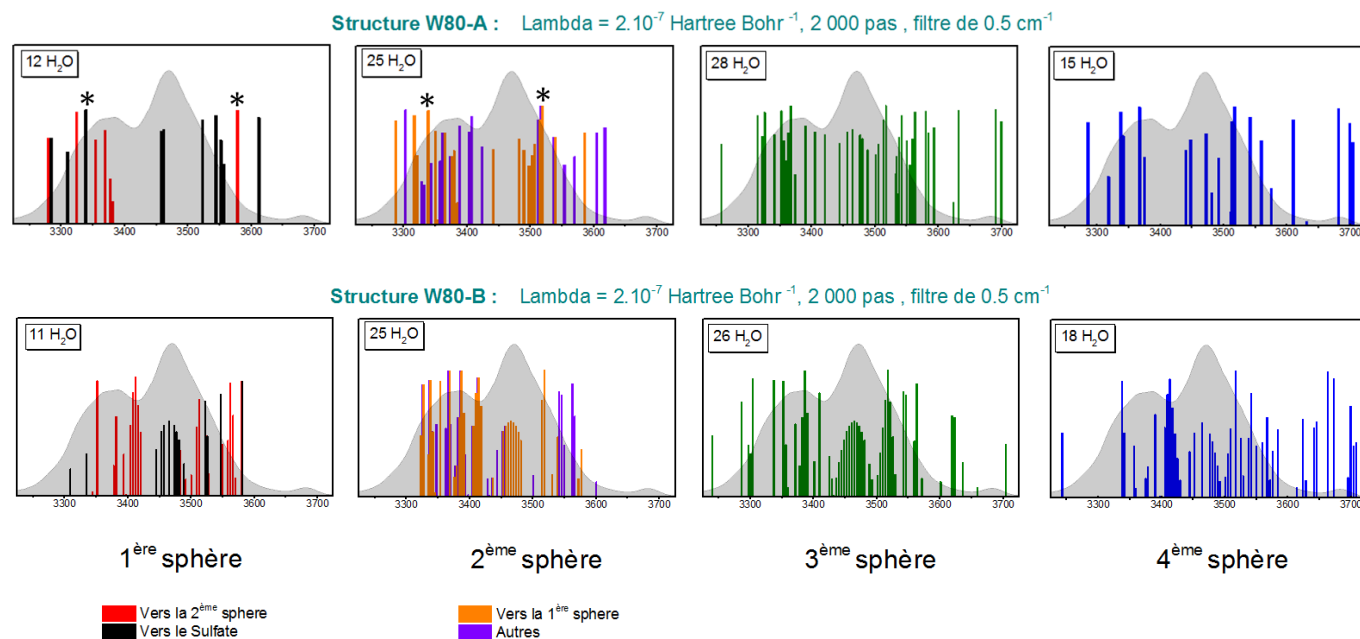


FIGURE 5.17 – Attribution dynamique de modes normaux pour $[SO_4(H_2O)_{80}]^{2-}$ par DMD à partir de 2 structures stables extraites de REMD. Les fréquences de résonance DMD sont superposées au spectre DACF à 200 K (gris).

Nous avons suivi les oscillations de plusieurs vibreurs de 1^{ère} et 2^{ème} sphères de solvation, au cours de simulations DMD de 0.2 ps. En **Fig 5.17**, nous avons pointé (symboles *) 4 fortes résonances traduisant des situations différentes. Pour chacune, les oscillations des deux vibreurs d'une molécule d'eau sont représentées en **Fig 5.18**. Le premier vibreur (légende en noir) est celui qui a été relevé par l'outil DMD. Est indiquée la différence entre sa fréquence de vibration et la fréquence d'entraînement. Le cas le plus simple est le (a) : les deux vibreurs de la molécule de 1^{ère} sphère ont des longueurs initiales proches ($\sim 2,5 \cdot 10^{-4}$ Å). Ils voient leur amplitude d'oscillation augmenter fortement et linéairement, dans un parfait couplage antisymétrique. En (b) : l'environnement est analogue, mais la fréquence observée correspond à un couplage symétrique des deux élongations. On n'observe pas d'augmentation des amplitudes d'oscillation, mais le mouvement symétrique est logiquement moins efficace, puisqu'on ne peut pas compter sur la mobilité de l'oxygène pour accompagner le mouvement des hydrogènes. Les oscillations du vibreur lié par liaison H au sulfate sont plus faibles, en accord avec un environnement moins flexible et une plus forte liaison H. En (c), vibrations d'une molécule de 2^{ème} sphère dans un couplage symétrique standard. Les environnements des deux vibreurs (l'un vers la 1^{ère} sphère, l'autre vers une autre molécule de 2^{ème} sphère) semblent assez proches, au vu des amplitudes d'oscillations respectives. On note à la fois les longueurs moyennes et les amplitudes d'oscillations, plus fortes qu'en 1^{ère} sphère. Enfin, en (d), la molécule de 2^{ème} sphère est liée à une molécule de 1^{ère} et une de 3^{ème} sphère. On observe ici la complexité de l'organisation dans l'agrégat et la grande mobilité dans l'environnement de la 3^{ème} sphère. Dans un tel environnement dynamique, au-delà d'une définition de mode local ou couplé, on pourrait introduire la notion de durée de vie du couplage, inversement proportionnelle à la mobilité de l'environnement.

Il est intéressant d'observer comment se répartissent les modes de 1^{ère} sphère avec l'augmentation de la taille de l'agrégat. En effet, le problème en 1^{ère} sphère peut-être vu comme une compétition entre d'un côté l'interaction avec le sulfate, et de l'autre l'interaction avec le reste de l'agrégat. L'interaction la plus forte induit le plus fort

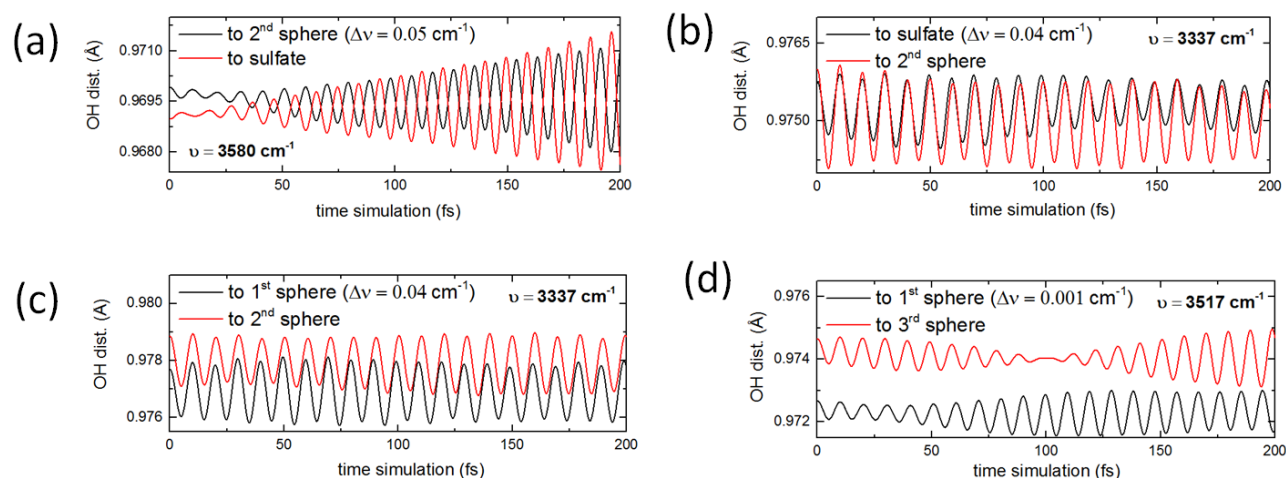


FIGURE 5.18 – W80-A : suivi des élongations O-H à leur fréquence de résonance, et du second vibrateur de la molécule d'eau. En 1^{ère} ligne ((a) et (b)), molécules de 1^{ère} sphère et en 2^{ème} ligne ((c) et (d)), de 2^{ème} sphère.

déplacement dans le rouge. Si les environnements des deux vibreurs sont très différents, leurs modes seront localisés. Si la compétition est équilibrée, on observe des modes de combinaisons symétriques et antisymétriques des élongations des deux vibreurs de la molécule. Un modèle qualitatif de la compétition en fonction de la taille de l'agrégat peut être proposé :

- Dans les petits agrégats ($n = 6$ par exemple), toutes les molécules sont en 1^{ère} sphère. Sauf exception, toutes ont une composante d'élongation vers le sulfate, et une vers une autre molécule d'eau. La liaison hydrogène avec le sulfate est notablement plus forte et le vibrateur OH lié au sulfate a toujours la fréquence la plus déplacée dans le rouge (discuté en détail au **chapitre 6**).
- En augmentant la taille de l'agrégat ($n = 24-25$, **chapitre 6**), le sulfate obtient une coordinence en 1^{ère} sphère entre 9 et 12, soit entre 2.25 et 3 liaisons hydrogènes par oxygène du sulfate. Dans les structures les plus stables, toutes les molécules de 2^{ème} sphère sont (ADD), c'est à dire que leurs deux hydrogènes et leur oxygène sont chacun une fois en interaction avec une molécule d'eau de 1^{ère} sphère. Notamment l'oxygène est donc une seule fois accepteur, et son interaction avec un hydrogène de 1^{ère} sphère est plus forte que l'interaction SO...H. On observe une très nette inversion de l'ordre d'apparition des modes dans le rouge, à la fois en DMD et par des calculs statiques DFT.
- En augmentant toujours la taille de l'agrégat ($n = 36$ puis 80), les molécules de 2^{ème} sphère deviennent progressivement toutes (AADD), chacune des liaisons hydrogène devient ainsi plus faible. Dans le même temps, l'environnement du sulfate ne varie pas et on assiste à un équilibre entre les environnements des deux vibreurs de 1^{ère} sphère.

Avec l'augmentation du nombre de vibreurs naissent des possibilités de couplages et les vibreurs de 1^{ère} sphère peuvent résonner sur l'ensemble de la gamme de fréquences du massif de OH liés.

La méthode DACF-AMOEBA a été employée pour reproduire la signature des élongations O-H d'ions sulfates hydratés dans des agrégats. Il est intéressant d'observer que la nouvelle version des paramètres de l'eau, dont la procédure de calibration oriente plutôt à reproduire des propriétés thermodynamiques de l'eau liquide, améliore grandement la spectroscopie de l'eau pure ou en solvant d'un ion. Toutefois, le fait que la paramétrisation ne fasse pas intervenir de données spectroscopiques est clairement ressenti et de nombreux exemples ont été présentés ici, tels que les mauvaises positions de la bande de pliage de l'eau ou d'élongation du sulfate. Le modèle de polarisation d'AMOEBA ne lui apporte pas toute la flexibilité nécessaire pour décrire la multitude des interactions au sein

des agrégats chargés. Nous avons proposé un modèle de charges fluctuantes, qui améliore la description des fortes anharmonicités. La calibration de ce modèle pourrait s'appuyer sur une gamme plus complète de taille d'agrégats, et serait renforcée par la paramétrisation de l'ion sulfate pour la version 14 du champ de forces de l'eau. Le DACF couplé à AMOEBA est un outil très performant pour calculer des spectres dynamiques à température finie. Sa faiblesse, l'attribution des modes normaux, est déjà bien comblée par les outils de la DMD. La méthode a permis l'attribution de toute la gamme des élongations O-H pour des agrégats contenant jusqu'à 80 molécules d'eau, et nous permet de mieux comprendre comment la force des interactions au sein des diverses sphères de solvation évolue avec l'augmentation de la taille de l'agrégat. Reste que pour les grands systèmes, il est difficile d'observer une image nette de chacune des sphères (et notamment de la première) tout en permettant des couplages, tels qu'ils apparaissent avec la température. Un développement futur pourrait consister à exciter sélectivement certaines des molécules, par exemple grâce à des paramètres d'entraînement dépendant de la sphère de solvation.

5.2 Spectroscopie sur les bandes d'élongation OH. Dépendance à l'ion

L'étude sur l'ion sulfate hydraté a montré que les massifs spectroscopiques évoluent assez fortement avec la taille de l'agrégat. La présence de 2 à 4 sphères de solvation entraîne une rééquilibration entre les interactions concurrentes eau-eau et eau-sulfate. Ainsi, les positions relatives de chaque type de bandes évoluent avec la construction d'un agrégat de plus en plus gros. Mais quel est l'effet d'un changement d'ion sur ce schéma de réorganisation ? Pour obtenir des indices, nous cherchons à reproduire les signatures spectrales de différents ions de charge -II à +III solvatés dans des agrégats de tailles $n = 36$ et 100 , par la même procédure DACF que précédemment, et à les confronter à une série de spectres expérimentaux.

5.2.1 Spectres expérimentaux d'ions hydratés $M(H_2O)_n$, $n = 36$ et 250

Williams et al. ont réalisé des spectres IRPD de nano-gouttelettes contenant un ion et 35-37 [16] ou ~ 250 molécules d'eau [14], pour la gamme des élongations O-H. Les spectres sont reproduits en **Fig 5.19**. Pour $n = 35-37$, Williams observe que le massif des OH liés prend la forme d'une seule large bande, assez similaire au spectre de l'eau pure, en accord avec des preuves théoriques et expérimentales [35, 36] que les énergies de liaison dans les agrégats de ces tailles sont assez indépendantes de l'état de charge. Le maximum d'intensité varie en fonction de l'ion, d'autant plus déplacé vers le rouge que l'ion est chargé positivement, avec 3480 cm^{-1} pour SO_4^{2-} et 3400 cm^{-1} pour La^{3+} , qui est également la fréquence d'intensité maximale pour l'eau pure en phase liquide. Pour $n \sim 250$, les massifs s'élargissent naturellement et les allures deviennent plus semblables, l'effet de l'ion étant statistiquement ressenti par une plus faible proportion des molécules d'eau. Ne subsiste, sur le massif des OH liés, qu'une dissymétrie plus marquée pour les anions et le sodium, avec un peuplement plus faible de la gamme $3150-3350\text{ cm}^{-1}$ (effet surtout marqué pour le sulfate).

La gamme des OH libres témoigne de la capacité de l'ion à organiser son réseau de liaisons hydrogène, autorisant ou empêchant la libération d'un nombre plus ou moins important de vibreurs OH en surface, dont on sait qu'ils existent pour les agrégats d'eau pure (**Fig 5.20**). Les déplacements vers le rouge en fonction du degré de charge positive suivent la même tendance que dans le massif des OH liés, avec une amplitude de variation moindre (pour 250 molécules d'eau : SO_4^{2-} : 3705 cm^{-1} , La^{3+} : 3700 cm^{-1}), attribuée à la distance plus importante entre l'ion et les molécules concernées.

Pour certains ions ($n \sim 250$: La^{3+} , $n = 35-37$: TBA^+ , TPA^+ , Na^+ , Li^+ , Ba^{2+} , Ca^{2+} , Mg^{2+} et Tm^{3+}), un pic additionnel (ou parfois simplement un épaulement), déplacé vers le bleu d'environ 20 cm^{-1} par rapport à la bande

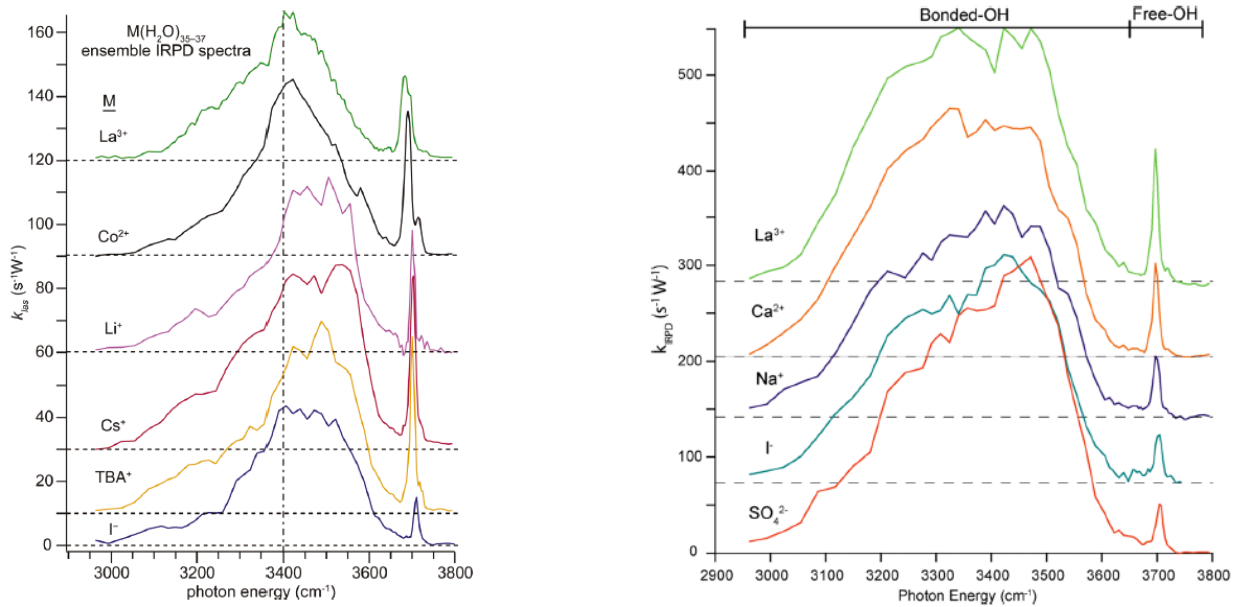


FIGURE 5.19 – Ensemble de spectres IRPD de $M(\text{H}_2\text{O})_{35-37}$ (à gauche) et $M(\text{H}_2\text{O})_{\sim 250}$ (à droite), à 133 K, dans la région des elongations O-H. Spectres publiés dans les figures 1 de [16] et 2 de [14].

principale de OH libres, peut être observé. Il n'est visible pour La^{3+} qu'à $n \sim 250$ et, à l'inverse, qu'à $n = 36$ pour Na^+ et Ca^{2+} . Williams l'identifie comme la contribution de molécules une fois donneuses et une fois acceptrices (AD), alors que la bande principale est attribuée aux molécules deux fois acceptrices et une fois donneuses (AAD). Dans une autre étude, il discute l'apparition de la cristallinité dans les agrégats de lanthane hydraté de taille croissante ($n = 50-550$) [15]. Ce petit pic y est également observable sur les petites tailles ($n = 50-250$) et décroît avec l'augmentation de taille, mais n'est pas discuté. Williams souligne que le rapport des massifs OH libres / OH liés suit généralement une loi en $1/r$, exception faite du plus petit agrégat de la série ($n = 50$), pour lequel l'intensité sous la bande de OH libres est anormalement forte, ce qu'il justifie par l'influence forte du champ électrique de l'ion qui oriente des liaisons OH additionnelles vers l'extérieur. Pour $n = 36$, cette bande n'apparaît pas, mais un épaulement à 3694 cm^{-1} est attribué à la répercussion de la coordinence de 9 en 1^{ère} sphère qui créerait différents environnements de OH libres (non observé pour Tm^{3+} , de coordinence 8 et qui présente un spectre très similaire à celui de La^{3+} pour $n = 20$). En effet, le lanthane est connu pour s'associer très fortement aux molécules de 1^{ère} sphère et à structurer fortement le reste du réseau de molécules d'eau [37], ce que nous avons également observé en dynamique moléculaire polarisable (**chapitre 3**). De très fortes interactions à courte distance peuvent expliquer l'absence de bande (AD) pour $n = 36$. Des agrégats non sphériques justifient la présence d'environnements différents pour les liaisons hydrogène et l'apparition de la bande (AD) à 3720 cm^{-1} à partir de $n = 50$.

5.2.2 Spectres DACF d'ions hydratés $M(\text{H}_2\text{O})_n$, $n = 36$ et 100

Pour $n = 36$, le DACF reproduit le déplacement progressif vers le rouge du massif des OH liés (**Fig 5.21**), lorsque l'ion devient plus positif. La forme générale des massifs est en bon accord avec l'expérience, excepté pour le lanthane, trop large, mais également très dépendant du point de départ de la simulation, du fait de la stabilisation importante du réseau de molécules d'eau. La discussion des effets des ions sur les massifs liés est difficile du fait des nombreux effets qui entrent en compétition : la présence de l'ion en surface ou au centre de l'agrégat, le caractère plus ou moins sphérique de l'agrégat (et l'écart à l'ergodicité de la simulation qu'il peut entraîner), le rayon d'effet de l'ion, fonction de sa charge et de sa coordinence et enfin le faible nombre de spectres expérimentaux pour comparaison. Pour les bandes de OH libres, l'ordre expérimental, du rouge vers le bleu, est $\text{La}^{3+} < \text{Ca}^{2+} < \text{Li}^+ <$

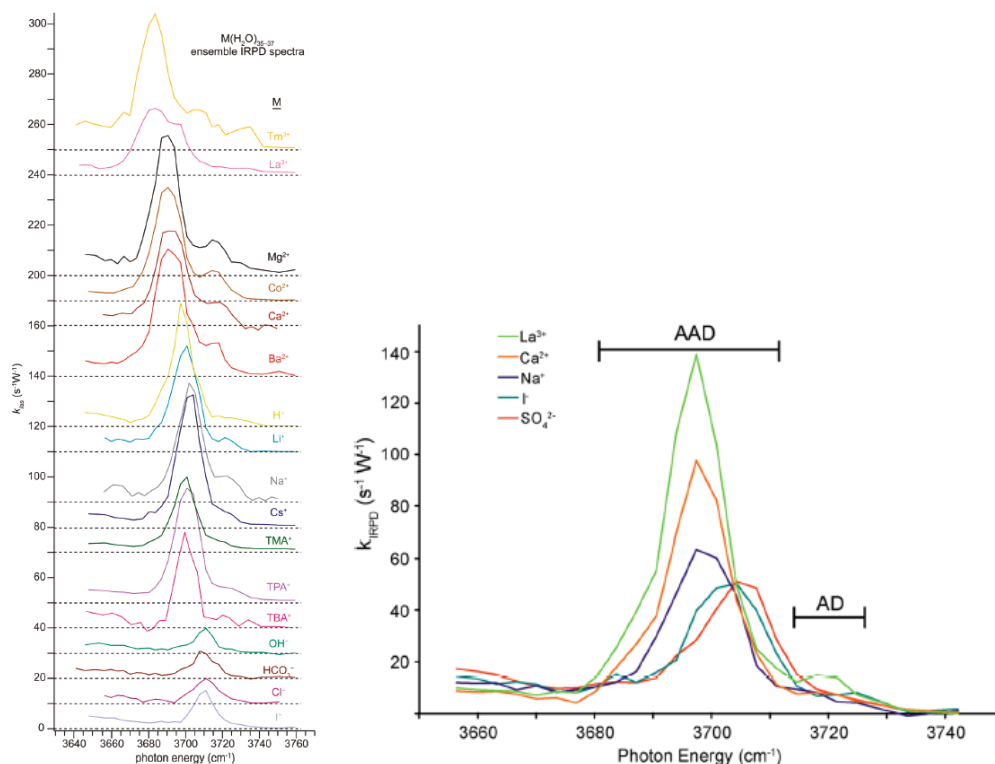


FIGURE 5.20 – Ensemble de spectres IRPD de $M(\text{H}_2\text{O})_{35-37}$ (à gauche) et $M(\text{H}_2\text{O})_{\sim 250}$ (à droite), à 133 K, dans la région des elongations des OH libres. Spectres publiés dans les figures 2 de [16] et 3 de [14].

$\text{Cs}^+ = \text{Na}^+ < \text{Cl}^- = \text{I}^-$. En DACF (**Fig 5.22**), l'accord est très convenable puisque le calcium et le lanthane ont les bandes les plus vers le rouge, les cations monochargés ayant des positions très similaires. À l'exception du sulfate qui ne présente pas de OH libre pour cette taille, les anions ont des déplacements très voisins, plus dans le bleu que les cations. En termes d'intensités relatives, les faibles intensités des anions sont bien reproduites, de même que l'intensité plus faible de La^{3+} par rapport aux autres cations, du fait de sa bande OH libre élargie. Pour $n = 100$, la situation est intermédiaire entre $n = 36$ et $n = 250$; la dissymétrie observée expérimentalement à $n = 250$ sur les massifs des OH liés des anions et du sodium, ainsi que des massifs plus symétriques pour le calcium et le lanthane, est clairement retrouvée avec le DACF (**Fig 5.23**). Pour le massif des OH libres, expérimentalement, la bande se déplace vers le rouge et augmente en intensité lorsque la charge augmente algébriquement. La tendance est bien traduite par le DACF à 200 K (**Fig 5.24**), mais pas les positions absolues (trop dans le rouge) et l'importance des déplacements est un peu surestimée (30 cm^{-1} au sein de la gamme calculée, contre $15\text{-}20 \text{ cm}^{-1}$ expérimentalement).

La méthode DACF-AMOEBa est capable de reproduire les grandes tendances de déplacements de bandes (massifs OH liés ou bandes de OH libres) en fonction de l'ion solvaté et de la taille de l'agrégat. Les déplacements observés expérimentalement sont souvent faibles, parfois de l'ordre de quelques cm^{-1} et mettraient en échec beaucoup de modèles théoriques. Toutefois, AMOEBa n'a pas encore montré tout son potentiel dans ces problématiques. La description de la spectroscopie de l'eau a été grandement améliorée par l'introduction du nouveau champ de forces de l'eau AMOEBa14. Des améliorations sont à prévoir, notamment pour dépasser le modèle actuel de polarisation et inclure les effets de transfert de charge. Une calibration des paramètres des ions pour AMOEBa14 permettra d'améliorer la description de la dynamique des ions. Une fois cette étape franchie, la DMD pourra être utilisée pour observer comment les elongations OH se distribuent en fonction des ions.

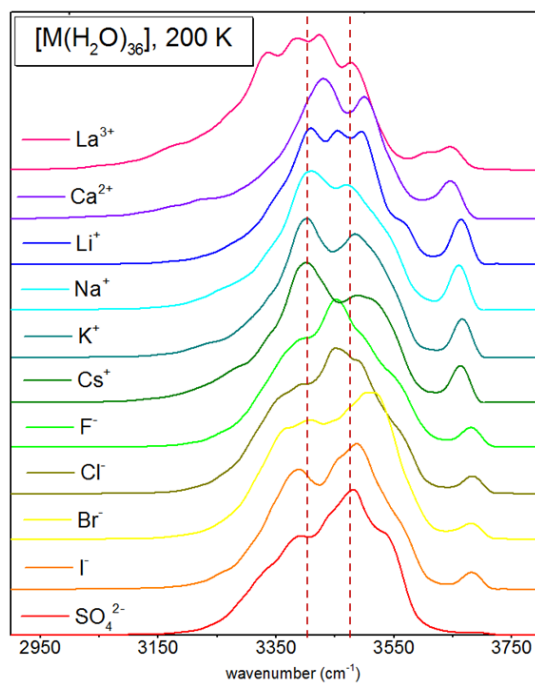


FIGURE 5.21 – Ensemble de spectres DACF à 200 K pour $[M(H_2O)_{36}]^x$, $x = -II$ à $+III$, dans la région des elongations O-H. Deux repères (traits en pointillés) sont placés au maxima d'intensité des spectres du sulfate et du lanthane, pour faciliter l'observation du déplacement du maximum dans le rouge vers les charges positives.

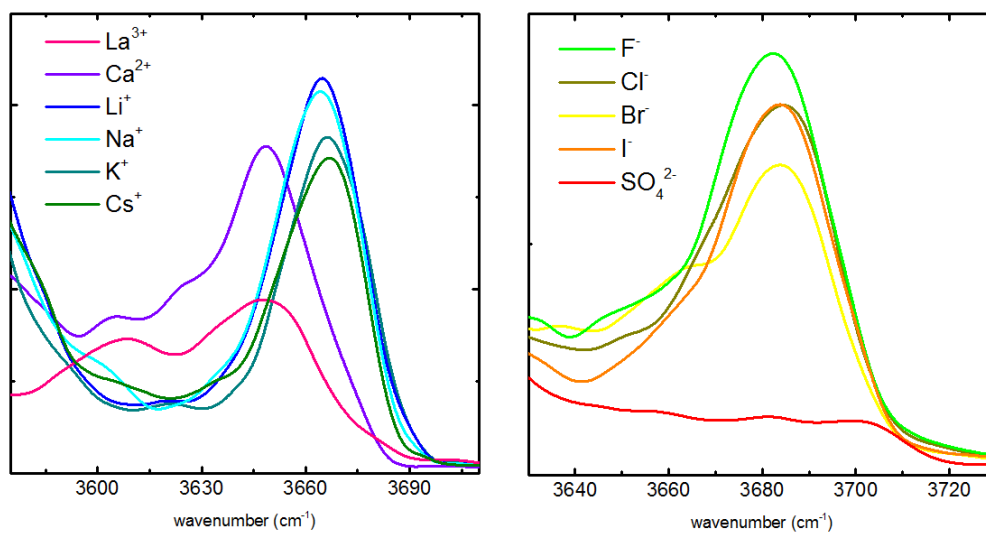


FIGURE 5.22 – Ensemble de spectres DACF à 200 K pour $[M(H_2O)_{36}]^x$, $x = -II$ à $+III$, dans la région des OH libres, pour les cations (à gauche) et les anions (à droite)

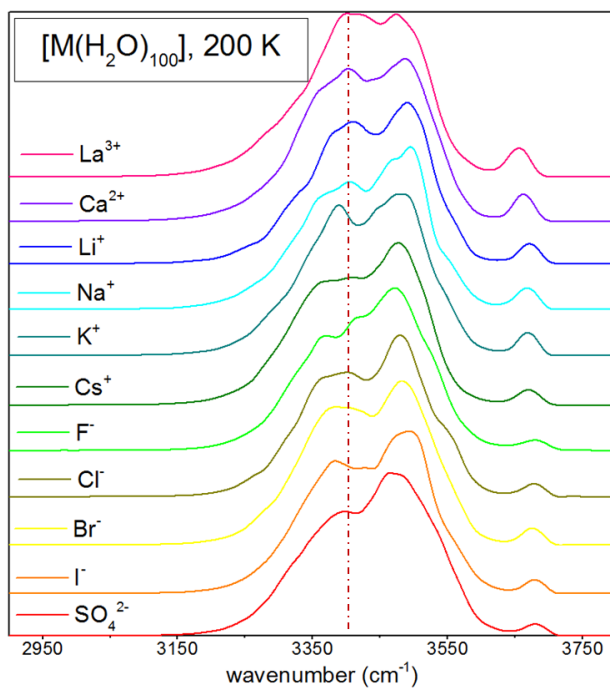


FIGURE 5.23 – Ensemble de spectres DACF à 200 K pour $[M(H_2O)_{100}]^x$, $x = -II$ à $+III$, dans la région des elongations O-H. Un repère (traits en pointillés) est placé au maximum de l'intensité de l'eau pure en phase liquide

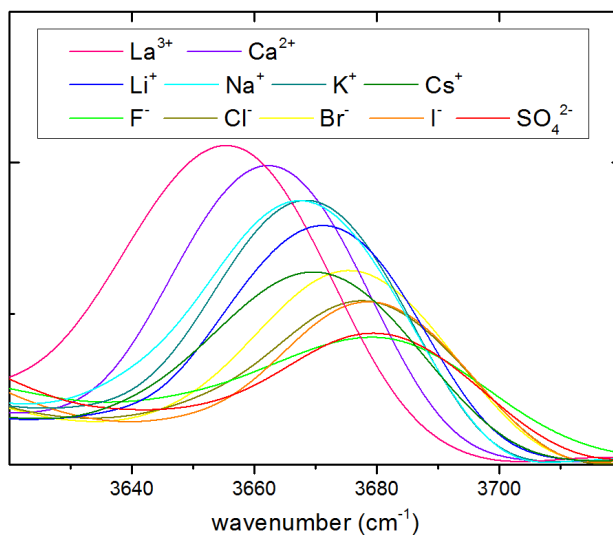


FIGURE 5.24 – Ensemble de spectres DACF à 200 K pour $[M(H_2O)_{100}]^x$, $x = -II$ à $+III$, dans la région des OH libres

Bibliographie

- [1] Devlin, J. P.; Joyce, C.; Buch, V. *J. Phys. Chem. A* **2000**, *104*, 1974-1977.
- [2] Sadlej, J.; Buch, V.; Kazimirski, J. K.; Buck, U. *J. Phys. Chem. A* **1999**, *103*, 4933-4947.
- [3] Liu, K.; Brown, M. G.; ; Saykally, R. J. *J. Phys. Chem. A* **1997**, *101*, 8995-9010.
- [4] Pradzynski, C. C.; Dierking, C. W.; Zurheide, F.; Forck, R. M.; Buck, U.; Zeuch, T.; Xantheas, S. S. *Phys. Chem. Chem. Phys.* **2014**, *16*, 26691-26696.
- [5] Keutsch, F. N.; Saykally, R. J. *Proc. Natl. Acad. Sci.* **2001**, *98*, 10533-10540.
- [6] Bruderemann, J.; Lohbrandt, P.; Buck, U.; Buch, V. *J. Chem. Phys.* **2000**, *112*, 11038.
- [7] Huang, J.; Bartell, L. S. *J. Phys. Chem.* **1995**, *99*, 3924-3931.
- [8] Abu-samha, M.; Borve, K. J.; Winkler, M.; Harnes, J.; Saethre, L. J.; Lindblad, A.; Bergersen, H.; Ohrwall, G.; Bjorneholm, O.; Svensson, S. *J. Phys. B At. Mol. Opt.* **2009**, *42*, 055201.
- [9] Berndt, T.; Stratmann, F.; Brasel, S.; Heintzenberg, J.; Laaksonen, A.; Kulmala, M. *Atmos. Chem. Phys.* **2008**, *8*, 6365-6374.
- [10] Laaksonen, A. *et al. Atmos. Chem. Phys.* **2008**, *8*, 7255-7264.
- [11] Ramanathan, V.; Crutzen, P.; Kiehl, J.; Rosenfeld, D. *Science* **2001**, *294*, 2119-2124.
- [12] Thauhay, F.; Clavaguéra, C.; Ohanessian, G. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25935-25945.
- [13] O'Brien, J. T.; Prell, J. S.; Bush, M. F.; Williams, E. R. *J. Am. Chem. Soc.* **2010**, *132*, 8248-8249.
- [14] O'Brien, J. T.; Williams, E. R. *J. Am. Chem. Soc.* **2012**, *134*, 10228-10236.
- [15] Cooper, R. J.; DiTucci, M. J.; Chang, T. M.; Williams, E. R. *J. Am. Chem. Soc.* **2016**, *138*, 96-99.
- [16] Prell, J. S.; O'Brien, J. T.; Williams, E. R. *J. Am. Chem. Soc.* **2011**, *133*, 4810-4818.
- [17] Buck, U.; Ettischer, I.; Melzer, M.; Buch, V.; Sadlej, J. *Phys. Rev. Lett.* **1998**, *80*, 2578.
- [18] Jana, C. *Modelling the structure and IR spectra of micro-hydrated ions*, Thesis, Ecole Polytechnique, Palaiseau, France, **2015**.
- [19] Pradzynski, C. C.; Forck, R. M.; Zeuch, T.; Slavicek, P.; Buck, U. *Science* **2012**, *337*, 1529-1532.
- [20] Fanourgakis, G. S.; Xantheas, S. S. *J. Chem. Phys.* **2008**, *128*, 074506.
- [21] Szalewicz, K.; Leforestier, C.; van der Avoird, A. *Chem. Phys. Lett.* **2009**, *482*, 1 - 14.
- [22] Bukowski, R.; Szalewicz, K.; Groenenboom, G. C.; van der Avoird, A. *Science* **2007**, *315*, 1249-1252.
- [23] Fanourgakis, G. S.; Xantheas, S. S. *J. Phys. Chem. A* **2006**, *110*, 4100-4106.
- [24] Partridge, H.; Schwenke, D. W. *J. Chem. Phys.* **1997**, *106*, 4618-4639.
- [25] Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551-8566.
- [26] Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8566-8575.

- [27] Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. *J. Phys. Chem. B* **2015**, *119*, 9423-9437.
- [28] Semrouni, D. *Energétique et spectroscopie de polypeptides par dynamique moléculaire : champ de force de seconde génération et chimie quantique*, Thesis, Ecole Polytechnique, Palaiseau, France, **2010**.
- [29] Xantheas, S. S.; Dunning, T. H. *J. Chem. Phys.* **1993**, *99*, 8774.
- [30] Bentwood, R.; Barnes, A.; Orville-Thomas, W. *J. Mol. Spectrosc.* **1980**, *84*, 391 - 404.
- [31] Page, R. H.; Frey, J. G.; Shen, Y.-R.; Lee, Y.-T. *Chem. Phys. Lett.* **1984**, *106*, 373-376.
- [32] Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. *J. Phys. Chem. B* **2013**, *117*, 9956-9972.
- [33] Burnham, C. J.; Xantheas, S. S. *J. Chem. Phys.* **2002**, *116*, 5115.
- [34] Ponder, J. W. "personal communication", **2016**.
- [35] Donald, W. A.; Leib, R. D.; Demireva, M.; Negru, B.; Neumark, D. M.; Williams, E. R. *J. Phys. Chem. A* **2011**, *115*, 2-12.
- [36] Donald, W. A.; Williams, E. R. *J. Phys. Chem. A* **2008**, *112*, 3515-3522.
- [37] Daaz-Moreno, S.; Ramos, S.; Bowron, D. T. *J. Phys. Chem. A* **2011**, *115*, 6575-6581.

Annexe C

Paramétrisation de l'ion sulfate

Dans les spectres IRMPD de Zhou et al. [1], enregistrés dans la gamme 500-1800 cm^{-1} , lorsque le sulfate est hydraté dans de petits agrégats ($n = 3-24$), la bande d'élongation SO est large de 100 cm^{-1} , centrée sur $\sim 1100 \text{ cm}^{-1}$. AMOEBA (indépendamment de la version du modèle de l'eau) la reproduit à 1280 cm^{-1} . Nous avons ajusté la constante d'élongation de 600 à 400 $\text{kcal.mol}^{-1}.\text{\AA}^{-2}$ pour calibrer le centre de la bande à 1100 cm^{-1} .

Les multipôles du sulfate, et notamment les charges portées par le soufre (+0.86816 e) et les oxygènes (-0.71704 e), semblent étonnement faibles. A titre de comparaison, McCammon [2] dérive le potentiel du sulfate dans 30 molécules d'eau à partir d'une SEP au niveau MP4SDTQ et obtient deux jeux de charges (S,O) à (+2.4 e, -1.1 e) et (+2.0 e, -1.0 e). Le premier jeu de charges est notamment réutilisé par Jonhston [3] dans sa recherche de configurations pour $n=3-50$.

Les détails de la paramétrisation du sulfate dans AMOEBA n'ont pas été publiés, mais la calibration des paramètres des ions fait généralement appel à de petits agrégats, l'ion dans 1 à 6 molécules d'eau [4]. Même s'il est peu concevable que moins de 3 molécules d'eau aient été utilisées, la stabilisation de l'ion vis-à-vis du détachement d'électron en requérant au minimum 3 [5], nous avons extrait des charges pour le sulfate isolé ou entouré de 6 molécules d'eau. Les charges ont été obtenues soit d'une extraction par DMA [6], soit par analyse de population NBO, avec plusieurs méthodes/bases aussi diverses que B3LYP/6-311+G(d,p) et MP2/aug-cc-pVTZ. Une partie des résultats sont proposés dans le **Tab 5.4**.

TABLE 5.4 – Détermination de charges atomiques pour le sulfate. La direction z est définie par l'axe d'ordre 2 de la molécule d'eau, et dz est la composante du dipôle atomique le long de cet axe. Pour les deux méthodes d'extraction (DMA et NBO) et les deux systèmes, les charges sont extraites au niveau MP2/aug-cc-pVTZ sur les géométries B3LYP/6-311+G(d,p)

		AMOEBA [7]	SO_4^{2-}		$[\text{SO}_4(\text{H}_2\text{O})_6]^{2-}$	
			DMA	NBO	DMA	NBO
Soufre	q	0.87	2.07	2.45	2.13	2.52
	dz	0	0		0	
Oxygène	q	-0.72	-1.02	-1.11	-1.04	-1.09
	dz	0.59	0.02		0.10	

Tous ces calculs attribuent au soufre une charge partielle comprise entre +2 et +2.5 e, et aux oxygènes une charge partielle voisine de -1 e. Nous proposons donc un nouveau jeu de multipôles pour le sulfate. Le système $[\text{SO}_4(\text{H}_2\text{O})_6]^{2-}$, dans sa structure tétraédrique (pour des questions de symétrie) est optimisé au niveau B3LYP/6-311+G(d,p). Les multipôles sont extraits par la procédure DMA à partir de la densité électronique calculée au

niveau MP2/aug-cc-pVTZ. Puis, en contraignant les charges, le reste des multipôles est optimisé sur le potentiel électrostatique au niveau MP2/aug-cc-pVTZ. Il est à noter que les multipôles des molécules d'eau ne sont pas rigoureusement les mêmes, bien qu'assez proches, que ceux du modèle AMOEBA14. Toutefois, la même extraction de multipôles sur le sulfate isolé donne des charges très similaires (avec un écart de l'ordre du pourcent), ce qui indique un effet direct relativement faible de l'environnement, mais favorise une bonne orientation du reste des multipôles.

Pour $[\text{SO}_4(\text{H}_2\text{O})_{55}]^{2-}$, (Fig 5.25), les différences entre les spectres calculés avec les anciens ou les nouveaux jeux de multipôles sont vraiment minimales. Il est probable que les faibles charges dans le jeu initial soient en partie compensées par la composante dz du dipôle atomique porté par l'oxygène (0.59 dans le jeu initial contre 0.1 dans notre extraction DMA).

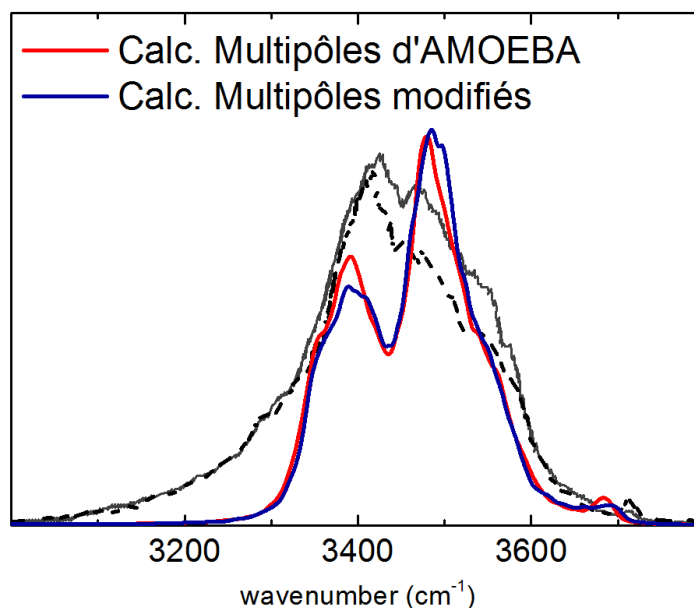


FIGURE 5.25 – Spectres DACF de $[\text{SO}_4(\text{H}_2\text{O})_{55}]^{2-}$, calculés à 200 K, à partir de deux jeux de multipôles pour le sulfate, celui de référence dans AMOEBA (rouge) et les nouveaux multipôles extraits (bleu). Les spectres calculés sont comparés aux spectres expérimentaux pour $n = 50$ (gris, trait plein) et $n = 60$ (noir, trait pointillé) [8].

Bibliographie

- [1] Zhou, J.; Santambrogio, G.; Brümmer, M.; Moore, D. T.; Woste, L.; Meijer, G.; Neumark, D. M.; Asmis, K. R. *J. Chem. Phys.* **2006**, *125*, 111102.
- [2] Cannon, W. R.; Pettitt, B. M.; McCammon, J. A. *J. Phys. Chem.* **1994**, *98*, 6225-6230.
- [3] Smeeton, L. C.; Farrell, J. D.; Oakley, M. T.; Wales, D. J.; Johnston, R. L. *J. Chem. Theory Comput.* **2015**, *11*, 2377-2384.
- [4] Grossfield, A.; Ren, P.; Ponder, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 15671–15682.
- [5] Wang, X.-B.; Nicholas, J. B.; Wang, L.-S. *J. Chem. Phys.* **2000**, *113*, 10837-10840.
- [6] Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233 - 239.
- [7] Lambrecht, D. S.; Clark, G. N. I.; Head-Gordon, T.; Head-Gordon, M. *J. Phys. Chem. A* **2011**, *115*, 11438–11454.
- [8] O'Brien, J. T.; Williams, E. R. *J. Am. Chem. Soc.* **2012**, *134*, 10228–10236.

Chapitre 6

Energie et Spectroscopie de l'ion sulfate hydraté dans de petits agrégats d'eau

Deux équipes ont produit des spectres IR expérimentaux d'ions sulfates hydratés : celles de Williams à Berkeley et d'Asmis à Leipzig. Les ions sont formés par ESI puis transférés dans la cellule à ultra-basse pression d'un spectromètre de masse FT-ICR. Dans les deux cas, cette cellule est refroidie, mais à des températures nominales différentes : 17 K pour Asmis, 133 K pour Williams [1], ce qui a des conséquences importantes sur les spectres qui en résultent puisque les dynamiques accessibles à ces températures sont très différentes. Les ions piégés sont irradiés par des lasers IR, mais à des longueurs d'onde différentes : dans le cas de Williams, il s'agit d'un laser de table OPO/OPA, ce qui permet la spectroscopie dans la gamme 2500-3800 cm^{-1} . Dans le cas d'Asmis, il s'agit du laser à électrons libres FELIX, qui lui donne accès à la gamme 600-2000 cm^{-1} . Plus récemment, cette équipe a également enregistré des spectres d'élongation O-H, toujours à 17 K, mais aucun spectre n'est encore publié.

6.0.3 Résultats expérimentaux de l'équipe de Williams

L'équipe de Williams a développé un dispositif expérimental [1] qui inclut une source d'ions solvatés comprenant de quelques à quelques centaines de molécules, un spectromètre de masse FT-ICR et un laser OPO/OPA, permettant l'acquisition de spectres IRPD d'ions solvatés dans des agrégats d'eau de taille variable, sur la gamme des élongations O-H ($\sim 2500\text{-}3800 \text{ cm}^{-1}$). En absence d'irradiation laser, une dissociation est déjà observée, par radiation infrarouge du corps noir (Blackbody Infrared Radiation Dissociation (BIRD)), et associée à une constante de vitesse de dissociation k_{BIRD} . Sous irradiation laser, la constante de photodissociation du précurseur, k_{IRPD} , est mesurée pour chaque fréquence du laser, et correspond à $k_{IRPD} = k_{BIRD} + k_{Laser}$, où k_{Laser} est la constante de dissociation due au laser et définit la vraie intensité IRPD. Avec ce dispositif, ils ont récemment enregistré deux séries de spectres IRPD d'ions sulfates hydratés $[SO_4(H_2O)_n]^{2-}$, avec $n = 9\text{-}13$, et avec $n = 22\text{-}25$. Il s'agit d'un travail en collaboration, dans lequel nous réalisons la modélisation. Ces spectres, présentés en **Fig 6.1** et **Fig 6.2**, ne sont donc pas encore publiés.

L'évolution des spectres avec l'augmentation de la taille de l'agrégat fournit des indices quant à l'environnement du sulfate et la présence d'interactions spécifiques. Le profil pour $n = 9$ présente une double bande très marquée, avec un creux vers 3360 cm^{-1} . Avec l'augmentation de la taille de l'agrégat, de $n = 10\text{-}13$, l'intensité relative de la bande de plus basse fréquence diminue, au point que $n = 13$ ne présente plus qu'un large épaulement. Nos recherches de structures stables pour $n = 12$ et 13 [3] nous permettent d'émettre une première hypothèse. Plusieurs structures ont été identifiées, pour lesquelles le sulfate présente une coordinence de 9 en 1^{ère} sphère, nettement plus stables que des structures dites 'à couche fermée', telle que la structure de coordinence 12 [4, 5]. La présence de molécules

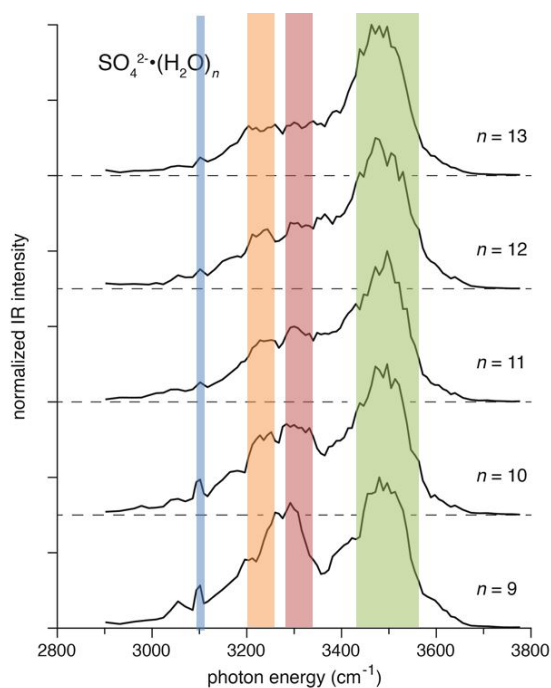


FIGURE 6.1 – Spectres IRPD à 133 K de $[SO_4(H_2O)_n]^{2-}$, $n = 9-13$, récemment enregistrés par l'équipe de Williams [2].

en 2^{de} sphère de solvation pour ces tailles explique les profils très semblables pour $n = 12$ et 13 et différents de $n = 9$, si on peut valider que ces molécules vibrent essentiellement dans la gamme de fréquences $3400-3700\text{ cm}^{-1}$. $n = 11$ a un profil très similaire à $n = 12$ (bandes orange et rouge), avec simplement le massif dans le bleu à peine moins large, ce qui pourrait impliquer un environnement assez semblable autour du sulfate et la perte d'une molécule de 2^{de} sphère. $n = 10$ est intermédiaire, présentant ces mêmes bandes (orange et rouge sur le **Fig 6.1**), mais également la petite résonance très anharmonique vers 3100 cm^{-1} observée pour $n = 9$, l'ensemble laissant imaginer un mélange de structures de type $n = 9$ et $n = 11-13$. Cette bande vers 3100 cm^{-1} est présente pour les autres tailles aussi, seulement d'intensité relative moindre.

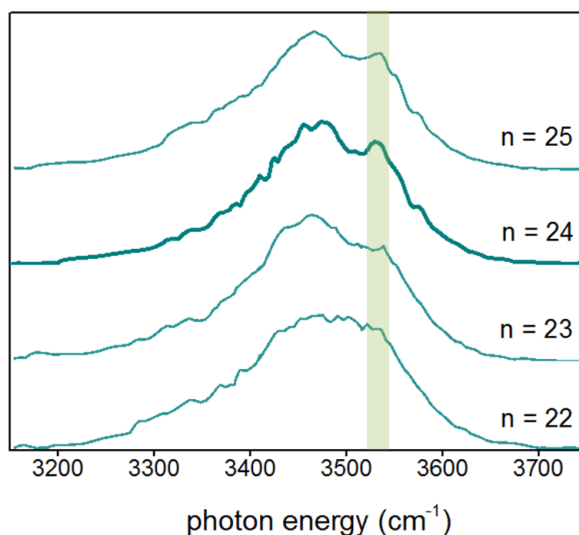


FIGURE 6.2 – Spectres IRPD à 133 K de $[SO_4(H_2O)_n]^{2-}$, $n = 22-25$, récemment enregistrés par l'équipe de Williams [2].

Sur la série $n = 22-25$ (**Fig 6.2**), les allures des spectres IRPD sont beaucoup plus similaires. La 1^{ère} sphère est probablement complète et l'apparition d'une bande (couleur verte) vers 3530 cm^{-1} peut être attribuée à un environnement particulier de molécules de 2^{ème} sphère, que seuls les calculs pourront définir sans ambiguïté.

D'autres informations sont fournies par l'expérience, au cours de l'étape de formation des agrégats dans le spectromètre de masse. A partir de nanogoutelettes chargées, la source d'ions ESI permet la formation d'agrégats de taille variable, par évaporations successives de molécules d'eau. L'abondance des fragments donne des indices sur leur stabilité relative. En effet, une taille n donnée peut être abondante, soit par sa propre stabilité, soit parce qu'il est plus facile d'arracher une molécule d'eau dans l'agrégat de taille $n + 1$. Le même type d'informations peut être obtenu par la mesure des constantes de dissociation BIRD (Blackbody Infrared Radiative Dissociation), si la dissociation est réalisée thermiquement et non par un laser. Ces deux sources d'informations, issues des expériences de Williams (abondances ESI en **Fig 6.3** et constantes de vitesses de dissociation BIRD en **Fig 6.4**), permettent de supposer que $n = 12, 24$ et 36 seraient plus stables que les tailles voisines, ce que Williams qualifie de 'nombres magiques'. Wong et Williams [4] avaient déjà identifié, par mesure de k_{BIRD} à 21°C sur les agrégats $n = 7-17$, que $n = 12$ pourrait être plus stable que $n = 11$ et 13 . Blades et Kebarle [6], dans des conditions un peu différentes, n'ont pas observé cette tendance, mais le justifient par les plus hautes températures de leurs expériences, qui engendrent une plus forte distribution de structures et notamment des structures plus faciles à dépeupler. Parmi les petites tailles, l'image d'une structure $n = 12$ très stable, à même d'engendrer des structures stables similaires pour $n = 11$ et 13 , par simple retrait ou ajout d'une molécule d'eau, se dessine. Elle est d'ailleurs déjà partiellement validée, puisque la structure la plus stable calculée pour $n = 13$ [3] est la structure la plus stable de $n = 12$ à laquelle une molécule a été rajoutée en 2^{nde} sphère. A l'inverse, sur le spectre ESI de Williams (**Fig 6.3**), l'abondance de $n = 9$ est un minimum, possiblement moins stable que $n = 8$ ou 10 . Sur la série $n = 22-26$, la constante de vitesse de dissociation BIRD (**Fig 6.4**) fournit l'information que la forte abondance ESI de $n = 24$ est majoritairement due à la facilité d'arracher une molécule d'eau de l'agrégat de taille $n = 25$. A l'inverse, $n = 23$ et 24 semblent être de stabilités similaires.

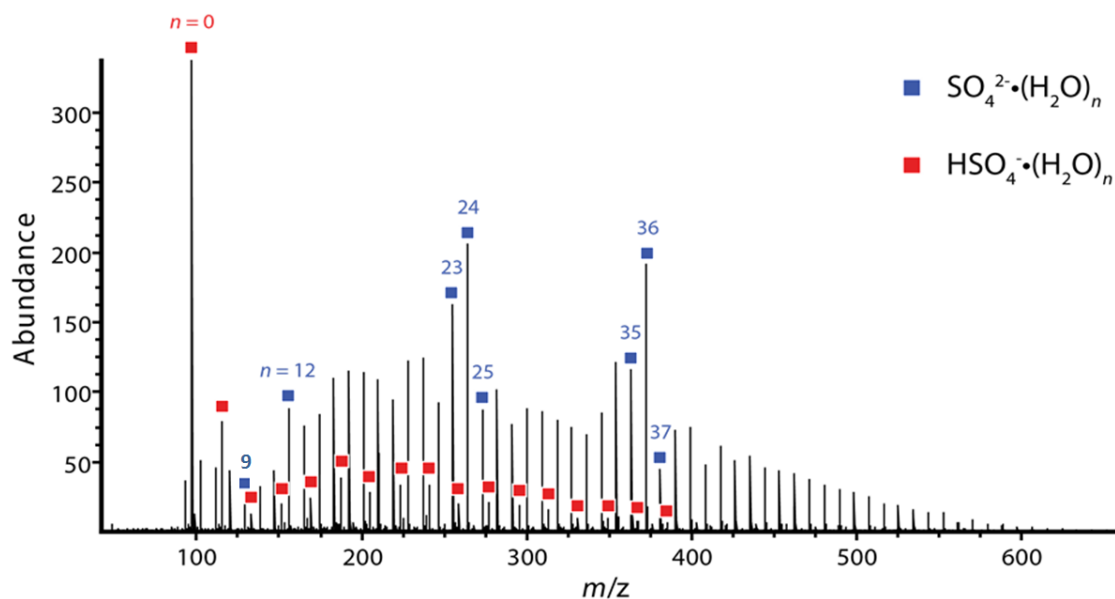


FIGURE 6.3 – Spectre ESI de $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$ et abondances relatives des fragments. Figure réalisée par Williams et coll. [2].

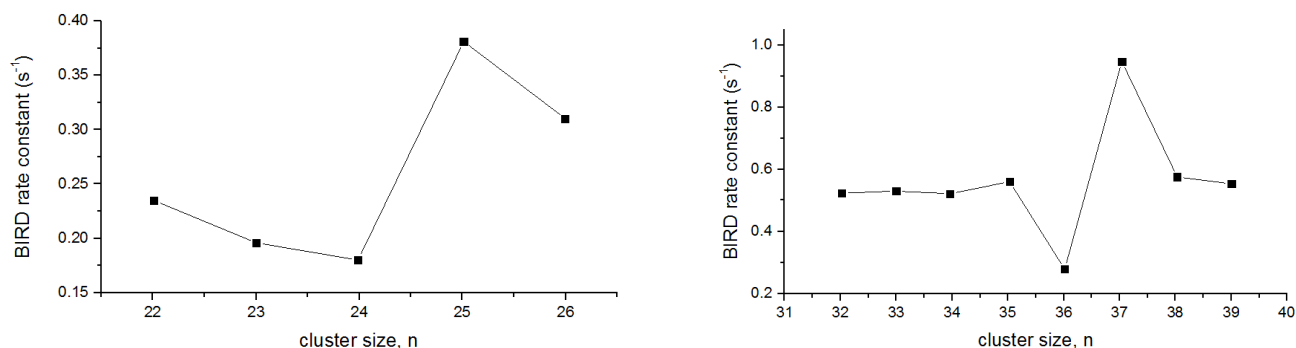


FIGURE 6.4 – Constante de vitesse de dissociation BIRD de $[SO_4(H_2O)_n]^{2-}$ dans les gammes $n = 22-26$ et $32-39$. Figures réalisées par Williams et coll. [2].

6.0.4 Stratégie de calculs

Comme nous l'avons vu au chapitre précédent, le champ de forces AMOEBA a montré sa capacité à reproduire la structure, la dynamique et la spectroscopie des ions solvatés $M(H_2O)_n^x$, pour $x = -2$ à $+3$ et $n = 36-100$. Pour les petits solvates, l'approche de référence reste la mécanique quantique statique, généralement dans la double approximation harmonique. Le plus souvent, plusieurs structures sont énergétiquement compétitives. Si elles sont également cinétiquement stables, alors la durée de transition structurale est négligeable devant les durées de vie des structures de basse énergie. Dans ce cas, l'approche consistant à décrire un spectre expérimental par une combinaison de spectres monoconformationnels semble justifiable. A l'inverse, si les états de transitions sont facilement accessibles, le système peut passer un temps non négligeable loin des minima, et il est nécessaire d'adopter une approche incluant les effets dynamiques. Cette hypothèse de stabilité cinétique a été vérifiée pour la structure la plus stable de $n = 12$. La première étape consiste alors à identifier les structures de basse énergie. Dans ce travail, la recherche globale de minima est réalisée par les outils de la dynamique moléculaire, puis des optimisations locales par DFT ou MP2 permettent d'obtenir les géométries pertinentes. A ces géométries, nous calculons les spectres statiques et les grandeurs thermodynamiques dépendantes de la température, afin de proposer que le spectre à la température expérimentale peut être obtenu par une combinaison des spectres statiques calculés à 0 K, pondérés par leurs énergies suivant une population de Maxwell-Boltzmann. Pour ces différentes étapes (optimisation locale, calcul des énergies relatives, spectres statiques), la précision des calculs est très dépendante du niveau de théorie utilisé, qui lui-même est limité par la taille du système. Sur la série $n = 6-13$, un niveau théorique uniforme a été validé. Les géométries ont été optimisées au niveau M11/6-311++G(d,p), à la suite des travaux de calibration de méthodes par Head-Gordon et al. [7, 8], sur les tailles $n = 3-7$. Pour les énergies relatives, la récente méthode DLPNO-CCSD(T) [9, 10] a été associée à la base aug-cc-pVTZ. Ce niveau de calcul a été validé dans l'étude de l'hydratation de $[SO_4(H_2O)_n]^{2-}$ avec $n = 12$ et 13 [3] (**annexe D**). Uniquement pour $n = 6$ et 9 , des optimisations de géométrie aux niveaux MP2/cc-pVTZ ou RI-MP2/cc-pVTZ, suivies par un calcul d'énergie au niveau MP2/aug-cc-pVTZ, ont été réalisées sur des structures préalablement validées par DFT, afin d'observer l'influence de la géométrie et de la méthode sur les spectres calculés. Le coût calculatoire associé à MP2 ou RI-MP2 exclut de les employer pour une étude systématique ou pour les tailles supérieures.

Sur la série $n = 23-25$, les méthodes post-HF ne sont plus accessibles, et la DFT devient le niveau de référence. La fonctionnelle M11 est mise en difficulté pour décrire les énergies relatives de ces agrégats dans lesquels les interactions eau-eau deviennent prépondérantes. Truhlar et al. [11] ont analysé les capacités de 61 fonctionnelles à reproduire les énergies relatives et de liaisons des agrégats $(H_2O)_{16}$ et $(H_2O)_{17}$, comparées au niveau de référence CCSD(T). Ils ont identifié la fonctionnelle corrigée de la dispersion $\omega B97XD$ comme la fonctionnelle capable de reproduire avec précision les deux propriétés. Nous avons adopté le niveau $\omega B97XD/cc-pVTZ$ pour les optimisations de géométrie et les calculs de fréquences à ces tailles.

L'approche quantique statique est également comparée à une approche par dynamique AMOEBA. Il n'est pas attendu du champ de forces, dont les paramètres du sulfate ont été calibrés pour AMOEBA03, et ceux de l'eau pour la spectroscopie des gros agrégats, de reproduire avec la même précision que la mécanique quantique, les énergies relatives des conformères, ou la force des interactions spécifiques. Toutefois, si les énergies relatives sont acceptables, le DACF peut prétendre reproduire un spectre de mélange, incluant les effets d'anharmonicité dynamique, et ainsi rendre compte de l'évolution du spectre en température, lorsque les interconversions entre géométries deviennent plus fréquentes.

$[SO_4(H_2O)_6]^{2-}$ a été très étudié et ses structures stables sont identifiées [7, 8, 12–14]. Ses isomères peuvent être classés en fonction de leur nombre de liaisons sulfate-eau et eau-eau et conduisent à 5 familles distinctes [7]. Williams a proposé une attribution du spectre expérimental [12] sur la base des spectres statiques d'un isomère de chacune de ces familles. Ce système nous sert d'exemple pour pointer la difficulté des modèles théoriques à attribuer de façon univoque les bandes expérimentales à des structures, et nous permet de détailler la méthodologie employée pour les tailles supérieures ($n = 9, 12, 13, 23-25$).

6.1 Méthodologie : $[SO_4(H_2O)_6]^{2-}$ comme système modèle

6.1.1 Problématique : Description d'un spectre multiconfigurationnel par une approche statique

Williams [12] attribue le spectre expérimental de photodissociation de $[SO_4(H_2O)_6]^{2-}$ (**Fig 6.5**) sur la base de calculs statiques B3LYP/aug-cc-pVDZ sur 5 structures de basse énergie, confirmées ensuite par Head-Gordon [7], et qui interviennent également dans l'étude de Wang et al. [13]. Sur la base des énergies et propriétés thermodynamiques (E_{ZPE} , H_{130K} , S_{130K}) calculées (**Tab 6.1**), il conclut que toutes les structures sont énergétiquement compétitives à l'exception de la forme tétraédrique (T_d), mais tempère sur la dépendance au niveau de théorie utilisé. En outre, plusieurs facteurs légitiment de conserver la structure T_d dans l'étude : d'une part l'étude de Wang [13], qui lui attribue une stabilité comparable aux autres géométries, et d'autre part celle de Zhou [15], dont les spectres IRMPD sur la gamme des empreintes digitales, et notamment l'absence de bande à 865 cm^{-1} , lui font attribuer le spectre expérimental à une structure tétraédrique. D'un point de vue qualitatif, les spectres calculés par Williams couvrent la gamme expérimentale et la structure T_d semble effectivement, sur la seule base des spectres, la candidate idéale à l'attribution de la bande vers 3575 cm^{-1} .

Toutefois, les énergies relatives ne permettent pas de valider cette hypothèse. La distribution de Maxwell-Boltzmann permet de calculer les populations relatives N_i et N_j de deux niveaux d'énergie i et j d'un système donné, à une température T , en fonction de leur énergies E_i et E_j (k_B est la constante de Boltzmann) :

$$\frac{N_i}{N_j} = \exp\left(\frac{-E_i + E_j}{k_B T}\right) \quad (6.1)$$

Par exemple, pour un mélange de 2 structures, à 130 K, la plus énergétique contribue à 1 % de la population totale si l'écart entre les deux est inférieur à $2.5\text{ kJ}\cdot\text{mol}^{-1}$ (**Fig 6.6**). Ainsi, à cette température, sur la base des enthalpies libres (G), seule la structure C_3 serait peuplée. De plus, la conclusion sur les contributions relatives des structures à la population totale est totalement dépendante des corrections apportées à l'énergie (E_{elec} , ZPE , ΔH_{130K} ou ΔG_{130K}), aucune ne permettant de justifier que la structure T_d est responsable de la bande vers 3575 cm^{-1} .

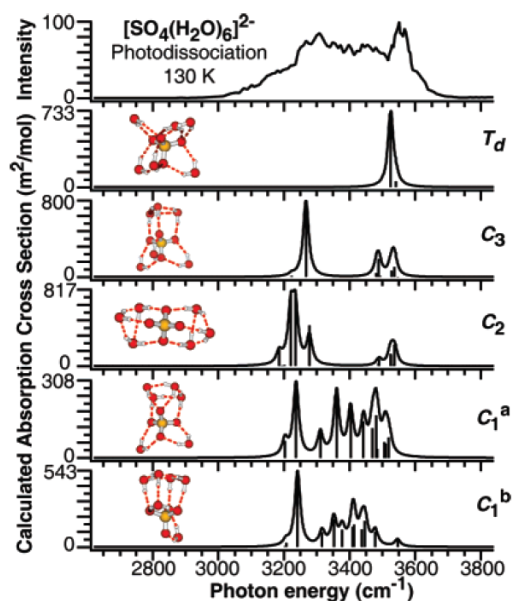


FIGURE 6.5 – Spectre de photodissociation de $[SO_4(H_2O)_6]^{2-}$ à 130 K, et spectres calculés au niveau B3LYP/aug-cc-pVDZ, dans l'approximation harmonique et avec un facteur d'échelle de 0.962 sur les fréquences. Figure 1 issue de [12].

TABLE 6.1 – Energies relatives (en $\text{kJ}\cdot\text{mol}^{-1}$) de 5 conformères de $[SO_4(H_2O)_6]^{2-}$ au niveau B3LYP/aug-cc-pVDZ. Données issues de [12].

	T_d	C_3	C_2	C_1^a	C_1^b
E_{elec}	20.5	4.9	0.2	3.8	0.0
+ ZPE	8.4	0.0	3.1	3.3	3.2
+ ΔH_{130K}	9.8	0.0	1.4	2.3	1.2
- $T \Delta S_{130K}$	8.5	0.0	5.8	4.1	5.4

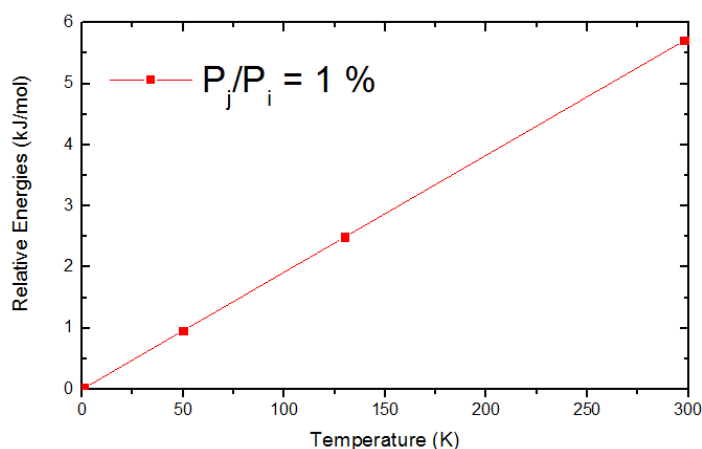


FIGURE 6.6 – Critère de Boltzmann : énergie relative entre deux structures pour que la plus haute en énergie contribue pour 1 % de la population totale.

6.1.2 Choix d'une méthode pour reproduire la géométrie et la spectroscopie des petits agrégats d'eau incorporant un ion sulfate

L'exemple précédent montre les difficultés rencontrées pour rendre compte de la stabilité relative des isomères de sulfates hydratés dans de petits agrégats d'eau, et justifie le nombre important d'études théoriques sur la gamme

de tailles $n = 3-7$ (voir refs 12-21 de [3]). Les calculs de Head-Gordon et al. [8] fournissent une image assez nette des diverses structures de basse énergie, mais pointent également les incertitudes sur les calculs de ZPE, et la forte dépendance des stabilités relatives aux contributions vibrationnelles. Ils comparent également les capacités de plusieurs fonctionnelles de la densité et de la méthode MP2 [7], à reproduire à la fois les énergies relatives et les énergies de dissociation de petits agrégats ($n = 3-6$). MP2 est jugée la méthode la plus performante pour reproduire les énergies *ab initio* de référence au niveau CCSD(T)/CBS*. Parmi les fonctionnelles, 3 doubles hybrides (XYG3, ω B97X-2 et XYGJ-OS), et l'hybride méta-GGA M11 occupent les premières places.

Ces benchmarks n'incluent pas de données spectroscopiques. Les spectres de $[SO_4(H_2O)_6]^{2-}$ calculés par Williams sont une bonne illustration du fait que l'approximation harmonique échoue à décrire les élongations fortement anharmoniques, à l'origine des intensités en dessous de 3200 cm^{-1} . La méthode choisie doit être capable de décrire la géométrie induite par la présence de fortes liaisons hydrogènes et reproduire les déplacements IR associés. Nous comparons 3 fonctionnelles, B3LYP, M06 et M11, associées à la base 6-311++G(d,p) (notée b1), ainsi que la méthode MP2, associée à deux bases, b1 et cc-pVTZ (notée b2). La structure C_1^a a été choisie comme modèle d'étude, car elle combine toutes les interactions possibles à cette taille : sulfate-eau (mono- et bi-dentate) et eau-eau (**Fig 6.7**). Le premier constat est la décorrélation importante entre la géométrie et le spectre calculé. Les géométries sont globalement peu dépendantes du niveau de théorie, les liaisons hydrogènes monodentates étant calculées, toutes méthodes confondues, entre 1.8 et 1.85 Å et, environ 0.1 Å plus courtes que les bidentates, pénalisées par l'ouverture de l'angle \widehat{HOH} , et les liaisons entre molécules d'eau, plus faibles. Les longueurs des liaisons intramoléculaires suivent logiquement la tendance opposée. A l'inverse, les spectres calculés sont très dépendants de la méthode, MP2 étant la plus apte à décrire une large distribution de fréquences. Le spectre B3LYP/b1 est assez similaire au spectre MP2/b1 alors que le spectre M11/b1 est similaire de celui calculé au niveau MP2/b2, dans les deux cas avec des distributions en fréquences plus faibles. La fonctionnelle M06 [16], de conception pourtant proche de M11 [17], et qui propose des géométries très comparables, échoue à décrire l'amplitude du spectre expérimental. Notons que les gammes de fréquences sont comparées avant application d'un 'scaling factor', mais qui ne change pas les allures générales des spectres. Au vu des conclusions du benchmark de Head-Gordon et de ces quelques tests, la fonctionnelle M11 est validée pour l'optimisation de géométrie et la spectroscopie des sulfates hydratés dans de petits agrégats, confortée, lorsque la taille le permet, par des calculs au niveau MP2.

6.1.3 Détermination des populations relatives d'isomères

Notons $Me1/B1$ le niveau de théorie utilisé pour les optimisations de géométrie et le calcul de fréquences, et $Me2/B2$ un plus haut niveau, utilisée dans un calcul d'énergie à la géométrie $Me1/B1$.

Une distribution de Maxwell-Boltzmann à la température T est habituellement basée sur l'enthalpie libre ΔG_T , qui inclut donc l'entropie ΔS_T . Néanmoins, devant les faibles énergies relatives des systèmes étudiés ici et les difficultés de la chimie quantique à calculer ΔS_T avec précision, nous proposons deux évaluations de population, une basée sur l'énergie et l'autre sur l'énergie libre de Gibbs. Le spectre multiconfigurationnel est construit comme une combinaison de spectres de structures de basse énergie, dont les intensités sont pondérées par leur population de Boltzmann.

Les calculs DFT ont été réalisés avec la suite logicielle GAUSSIAN et les calculs RI-MP2 avec la suite TURBOMOLE. Les corrections thermodynamiques sont calculées à 298 K par défaut mais peuvent être obtenues pour différentes températures (par exemple à 130 K, la température du spectre de Williams), à partir du hessien. Nous avons observé que l'entropie a une dépendance quasi-linéaire à la température. Dans cette approximation, les énergies et les enthalpies libres relatives à la température T sont données par :

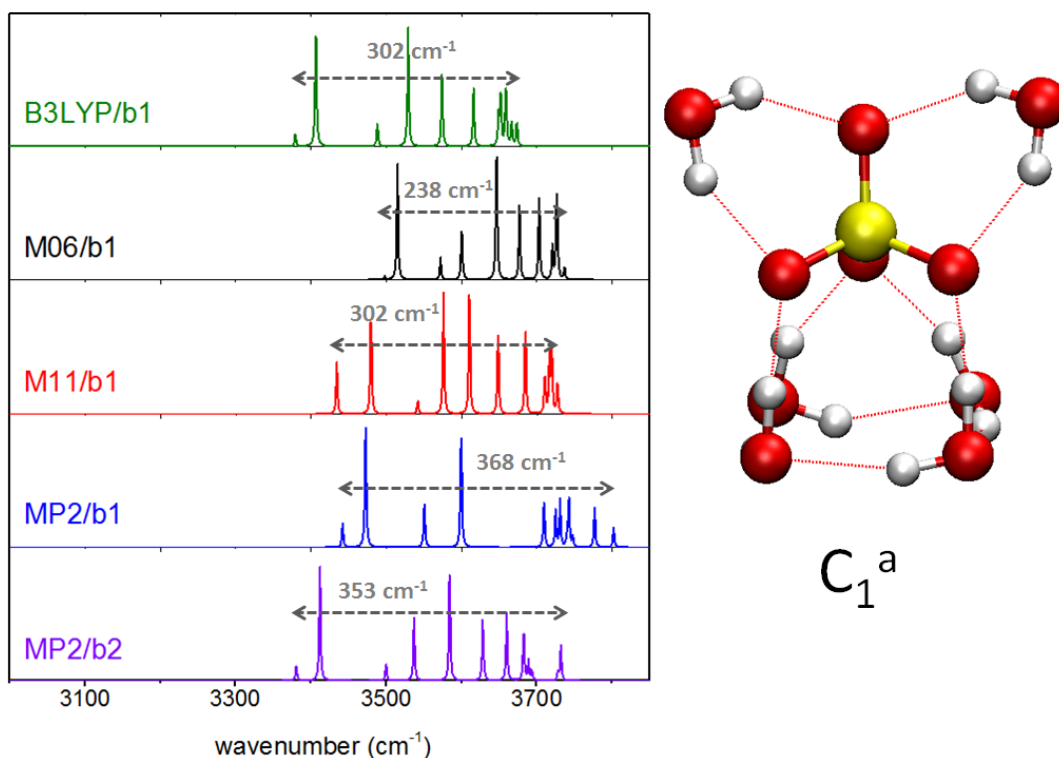


FIGURE 6.7 – Spectre statique de $[SO_4(H_2O)_6]^{2-}$ dans sa géométrie C_1^a calculé à différents niveaux de théorie. b1 = 6-311++G(d,p) et b2 = cc-pVTZ. Les optimisations ont été réalisées aux mêmes niveaux de calcul que les spectres. Aucun 'scaling factor' n'a été appliqué aux fréquences.

$$E(T) = E_{elec}^{Me2/B2} - E_{elec}^{Me1/B1} + (E_{elec} + ZPE)^{Me1/B1} + (E_{therm,298K}^{Me1/B1} - (E_{elec} + ZPE)^{Me1/B1}) \times \frac{T}{298} \quad (6.2)$$

$$\text{soit } E(T) = E_{elec}^{Me2/B2} + \text{correct. } ZPE^{Me1/B1} + \text{correct. } E(T)^{Me1/B1} \quad (6.3)$$

$$G(T) = E_{elec}^{Me2/B2} - E_{elec}^{Me1/B1} + (E_{elec} + ZPE)^{Me1/B1} + (G_{therm,298K}^{Me1/B1} - (E_{elec} + ZPE)^{Me1/B1}) \times \frac{T}{298} \quad (6.4)$$

$$\text{soit } G(T) = E_{elec}^{Me2/B2} + \text{correct. } ZPE^{Me1/B1} + \text{correct. } G(T)^{Me1/B1} \quad (6.5)$$

6.1.4 Spectres statiques multiconfigurationnels

Trois séries de spectres ont été réalisées (Fig 6.8), aux niveaux M11/b3//M11/b1, MP2/b3//MP2/b1, et MP2/b3//MP2/b2, où 'b3' signifie 'aug-cc-pVTZ'. Pour chaque niveau de calcul, les fréquences harmoniques des 5 structures ont été corrigées de facteurs d'échelles. En DFT, pour ne pas les rendre dépendants d'une bande en particulier, la calibration a été réalisée sur la bande de O-H libres d'agrégats $[SO_4(H_2O)_{25}]^{2-}$. L'objectif est de conserver ces facteurs d'échelle pour toutes les tailles traitées. A l'inverse, des résultats MP2 ne sont présentés que pour $n = 6$ et 9 et le facteur d'échelle est choisi afin de positionner au mieux les bandes calculées sous le spectre expérimental. Les convolutions utilisent des fonctions lorentziennes de 50 cm^{-1} de largeur à mi-hauteur. Dans chaque série, 4 spectres supplémentaires sont représentés, combinaisons linéaires des spectres des structures en appliquant les critères de Boltzmann. Dans chaque cas, les populations associées sont indiquées. Comme observé précédemment, les spectres monoconformationnels de M11/b1 et MP2/b2 sont très similaires et en bon accord avec

MP2/b1.

Les spectres multiconfigurationnels calculés sur la base des énergies relatives présentent la même allure en double bande à 130 K, les deux contributions majeures étant apportées par les structures C_2 ou C_3 , dont les spectres sont similaires. A 298 K, les populations MP2 incluent également une contribution importante de la structure C_1^b . Les spectres basés sur les enthalpies libres sont plus différents, M11/b1 et MP2/b2 prévoyant un mélange basé principalement sur C_3 et T_d , tandis qu'en MP2/b1, seule la structure T_d est compétitive, pour les deux températures.

On voit ici la difficulté de bien calculer les corrections thermodynamiques. Le tableau **Tab 6.2** regroupe les énergies relatives des différentes structures. Chaque niveau de calcul identifie C_2 comme la structure la plus stable sur la seule base des énergies relatives. Head-Gordon [8] a montré que C_3 devient la structure de référence lorsqu'on inclut l'énergie vibrationnelle, ce que MP2 reproduit bien.

Les attributions de bandes sur $[SO_4(H_2O)_6]^{2-}$ sont aisées puisque toutes les molécules sont en 1^{ère} sphère et que seuls trois types de vibrateurs sont présents. Des basses vers les hautes fréquences, on trouve les elongations O-H vers le sulfate des molécules monodentates, puis celles vers une autre molécule d'eau, et enfin celles des molécules bidentates.

TABLE 6.2 – Energies relatives (ΔE) avec et sans correction de l'énergie vibrationnelle du point zéro (ZPE) pour les 5 structures de $[SO_4(H_2O)_6]^{2-}$. Toutes les énergies sont données en $\text{kJ}\cdot\text{mol}^{-1}$.

Structures	M11/b3//M11/b1		MP2/b3//MP2/b1		MP2/b3//MP2/b2	
	(ΔE)	($\Delta E + \text{ZPE}$)	(ΔE)	($\Delta E + \text{ZPE}$)	(ΔE)	($\Delta E + \text{ZPE}$)
C_1^a	3.9	3.6	3.0	2.4	4.3	-1.1
C_1^b	4.9	6.5	0.7	1.7	1.3	-0.8
T_d	17.5	4.8	17.2	-0.2	20.3	0.9
C_3	4.1	0.4	3.5	-1.2	5.0	-4.1
C_2	0.0	0.0	0.0	0.0	0.0	0.0

Inversions de cycles de liaisons H

En **chapitre 3**, nous avons introduit un outil pour engendrer, à partir d'une structure donnée, les structures nées de l'inversion de cycles de liaisons hydrogènes. Pour $n = 6$, apparaissent les premiers cycles (**Tab 6.3**), pour 4 des 5 structures proposées. La structure C_2 possède 2 cycles de 3 liaisons hydrogènes, et donc la possibilité d'une combinaison des deux inversions simultanées, mais par symétrie, 1 seule inversion conduit à une structure non superposable. Pour chaque structure engendrée, les énergies relatives au niveau MP2/b3//MP2/b1 ont été calculées (colonne de droite du **Tab 6.3**), et le terme de correction de ZPE est ajouté (**eq 1.3**). Dans la majorité des cas, l'inversion d'un cycle n'apporte aucune différence ($\leq 0.03 \text{ kJ}\cdot\text{mol}^{-1}$), la seule exception étant la structure C_2 , destabilisée lorsqu'elle perd formellement la symétrie C_{2v} . Avec une inversion en moyenne par structure, inclure les nouveaux isomères dans le calcul de population de Boltzmann n'a d'autre effet que de diminuer marginalement la population relative de la structure T_d .

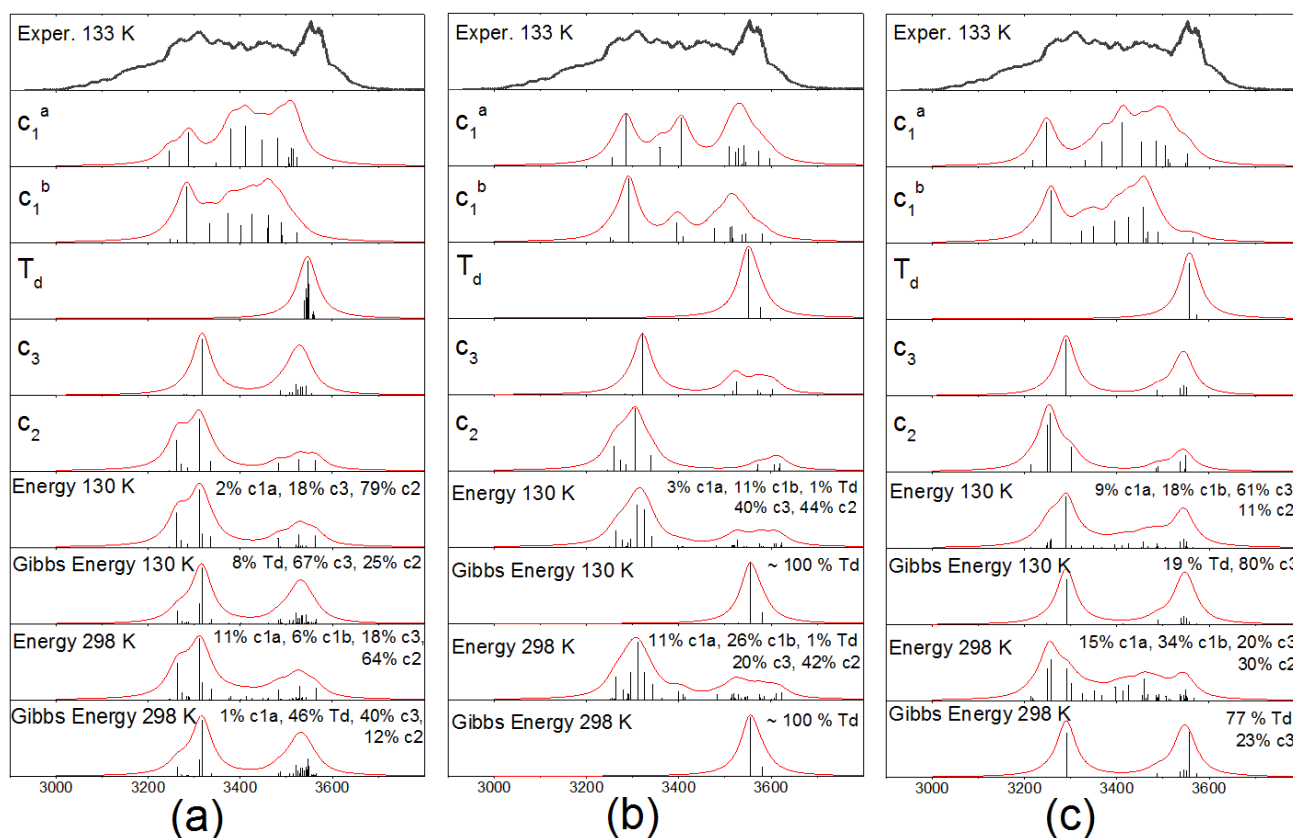


FIGURE 6.8 – Spectres statistiques de 5 structures de $[SO_4(H_2O)_6]^{2-}$ et les spectres multiconfigurationnels associés. Trois niveaux de calculs ont été choisis, associés à des 'scaling factor' (SF) sur les fréquences : (a) M11/b3//M11/b1 + SF 0.945, (b) MP2/b3//MP2/b1 + SF 0.947 et (c) MP2/b3//MP2/b2 + SF 0.952 (avec b1 = 6-311++G(d,p), b2 = cc-pVTZ et b3 = aug-cc-pVTZ).

TABLE 6.3 – Nombre et taille des cycles indépendants dans les structures de basse énergie de $[SO_4(H_2O)_6]^{2-}$. La colonne 'Combinaisons' comptabilise le nombre de structures engendrées à partir d'inversions multiples (jusqu'à 4 inversions de cycles simultanées). La colonne 'Identiques par symétrie' décompte les structures équivalentes. La colonne ' E_{rel} ' affiche l'écart d'énergie, au niveau MP2/b3//MP2/b1 incluant la correction de ZPE, entre les structures engendrées par inversion de cycles et la structure qui a permis de les engendrer.

Structures	Cycles / structure	Atomes / cycles	Combinaisons	Identiques par symétrie	E_{rel} (kJ.mol ⁻¹)
C_1^a	1	[4]			0.01
C_1^b	1	[4]			0.01
T_d	0				
C_3	1	[3]			0.03
C_2	2	[3;3]	1	-2	0.66

6.1.5 Spectres dynamiques AMOEBA

L'approche statique de mécanique quantique est la plus à même de décrire avec précision les géométries les plus stables et leurs énergies relatives. Si les meilleures méthodes sont employées, elle peut permettre d'identifier les contributions relatives de structures au sein d'un mélange, en fonction de la température. Cependant, deux limites apparaissent : d'une part, la précision sur le calcul des propriétés thermodynamiques est très dépendante du niveau théorique utilisé, qui diminue rapidement lorsque la taille augmente ; d'autre part, l'approche statique ignore la cinétique d'échange et la possibilité que des structures assez éloignées de leurs géométries d'équilibre puissent contribuer significativement à la signature expérimentale. Dans ces conditions, la dynamique moléculaire à l'aide

d'un champ de forces polarisable devient compétitive, capable de rendre compte, dans une moindre mesure que la mécanique quantique, des énergies relatives des structures, mais également des effets dynamiques incluant les écarts aux géométries d'équilibre et les interactions spécifiques au mélange de structures.

La gamme de précision d'AMOEBAA03 sur les énergies relatives des isomères les plus stables de $[SO_4(H_2O)_3]^{2-}$ a été évaluée à $\pm 2-3$ kcal.mol⁻¹ par Ponder [18], mais Head-Gordon, sur la gamme $n = 3-7$, identifie que les écarts avec MP2 peuvent atteindre plus de 4.5 kcal.mol⁻¹ [14] sur des structures plus hautes en énergie. Pour des explorations de surface loin des minima, une gamme de confiance de 10 kcal.mol⁻¹ est suggérée. Dans le processus de validation d'AMOEBAA14 [19], Ponder a analysé 42000 agrégats d'eau pure (~ 2400 agrégats de n molécules, sur la gamme $n = 2-22$) et conclut à un RMSD de ~ 10 % sur les énergies relatives comparées à des données *ab initio*.

Nous avons évalué, sur $[SO_4(H_2O)_6]^{2-}$, l'influence d'un changement de version du champ de forces et des jeux de multipôles (**Tab 6.4**). AMOEBA03 reproduit l'ordre des énergies relatives MP2, avec un RMSD de 3.5 kJ.mol⁻¹ et une faible dépendance au jeu de multipôles. En AMOEBA14, l'ordre des stabilités est toujours respecté, mais les écarts entre structures s'intensifient, surtout avec le changement de multipôles. Le meilleur accord de la version AMOEBA03 était attendu, l'ion ayant été calibré avec ce champ de forces. Un objectif de la procédure de paramétrisation de l'ion est la reproduction des énergies relatives des isomères [14]. Dans la philosophie de paramétrisation d'AMOEBAA14, les données théoriques sur les petits systèmes perdent de leur importance au profit des données expérimentales de phase condensée et gazeuse, ce qui peut expliquer une partie des difficultés rencontrées sur les plus petits systèmes. Un dernier point, symptomatique des petites tailles, est la destabilisation associée à une grande ouverture de l'angle de la molécule d'eau dans le champ de forces. Dans $[SO_4(H_2O)_6]^{2-}$, tous les isomères à l'exception de C_2 possèdent une ou plusieurs molécules d'eau doublement liées au sulfate, et cette interaction est systématiquement destabilisante pour l'agrégat. Dans les géométries optimisées AMOEBA14, la symétrie T_d est même localement rompue, les deux liaisons hydrogènes d'une molécule d'eau vers le sulfate n'étant plus équivalentes. Toutefois, ce problème est confiné aux petits sulfates ($n \leq 15-20$), les structures avec molécules d'eau bidentates n'étant rapidement plus compétitives lorsque la taille de l'agrégat augmente.

TABLE 6.4 – Energies relatives AMOEBA (en kJ.mol⁻¹) des 5 structures de basse énergie de $[SO_4(H_2O)_6]^{2-}$, en fonction de la version du champ de forces de l'eau et du jeu de multipôles du sulfate. Les énergies MP2/b3//MP2/b2 sont données pour comparaison.

Structures	AMOEBAA03		AMOEBAA14		MP2*
	Multipôles de Ponder	Nouveaux Multipôles**	Multipôles de Ponder	Nouveaux Multipôles**	
C_1^a	7.3	7.1	13.5	15.2	4.3
C_1^b	5.1	2.6	10.5	8.6	1.3
T_d	27.0	27.8	33.4	39.8	20.3
C_3	12.8	14.0	15.4	19.9	5.0
C_2	0.0	0.0	0.0	0.0	0.0
RMSD	3.5	3.5	4.4	7.0	

MP2* : E_{elec} MP2/b3//MP2/b2

** : Procédure d'extraction décrite au **chapitre 5**

Gamme des élongations O-H

Bien qu'AMOEBAA03 reproduise mieux les énergies relatives des isomères, la version du champ de forces est incapable de décrire le massif des élongations O-H, et notamment la gamme 3200-3400 cm⁻¹ (détails en **chapitre 5**). Pour la spectroscopie, nous utilisons le champ de forces AMOEBA14 et toutes les modifications validées dans le chapitre précédent (jeu de multipôles du sulfate, constante de pliage de l'eau, d'élongation S-O du sulfate, et modèle de charges fluctuantes pour l'eau). A 1 K, (**Fig 6.9**), l'accord avec les méthodes quantiques est globalement bon, à l'exception de la structure T_d , dont la symétrie est rompue.

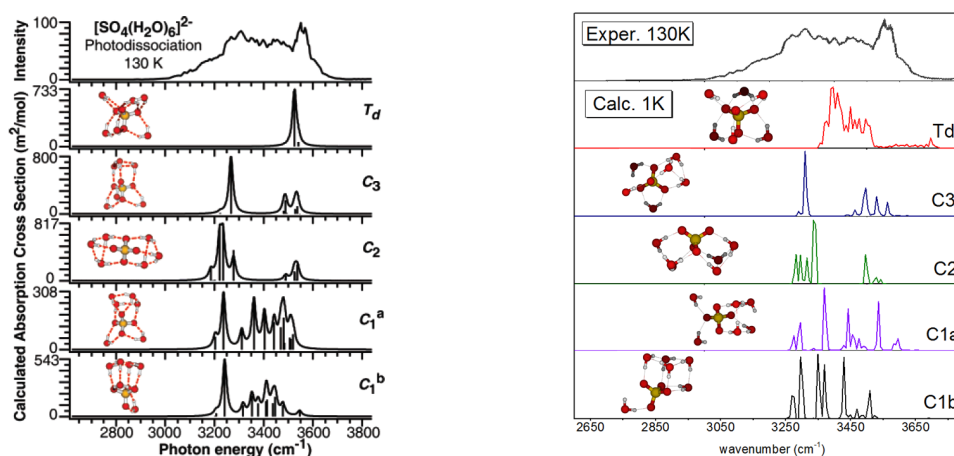


FIGURE 6.9 – Comparaison entre (à gauche) les spectres calculés au niveau B3LYP/aug-cc-pVDZ, dans l’approximation harmonique et avec un facteur d’échelle de 0.962 sur les fréquences, par Williams et coll. [12], et (à droite) des spectres DACF de dynamique moléculaire à 1 K.

En augmentant la température (**Fig 6.10**), des réorganisations ont lieu, sur la structure T_d dès 50 K, puis les structures C_1^a et C_3 à 200 K. Les réorganisations tendent vers la structure C_2 , clairement la plus stable en AMOEBA14. La structure C_1^b , très stable en MP2/b3//MP2/b2 et la seconde plus stable en AMOEBA, est toujours présente au terme de 200 ps de dynamique à 200 K. Avec la température et les écarts aux géométries d’équilibre qu’elle occasionne, les spectres DACF-AMOEBA s’élargissent et s’uniformisent autour de deux larges bandes, l’une entre $3175\text{-}3400\text{ cm}^{-1}$ et l’autre entre $3400\text{-}3650\text{ cm}^{-1}$, avec des allures très similaires aux spectres quantiques multiconfigurationnels pondérés par les énergies. Le massif dans les plus basses fréquences est composé des élongations vers le sulfate et le second des élongations entre molécules d’eau. Dans le mélange de structures, les élongations vers le sulfate sont plus nombreuses que les élongations entre molécules d’eau et sont également plus polarisantes, ce qui justifie les intensités relatives des deux massifs.

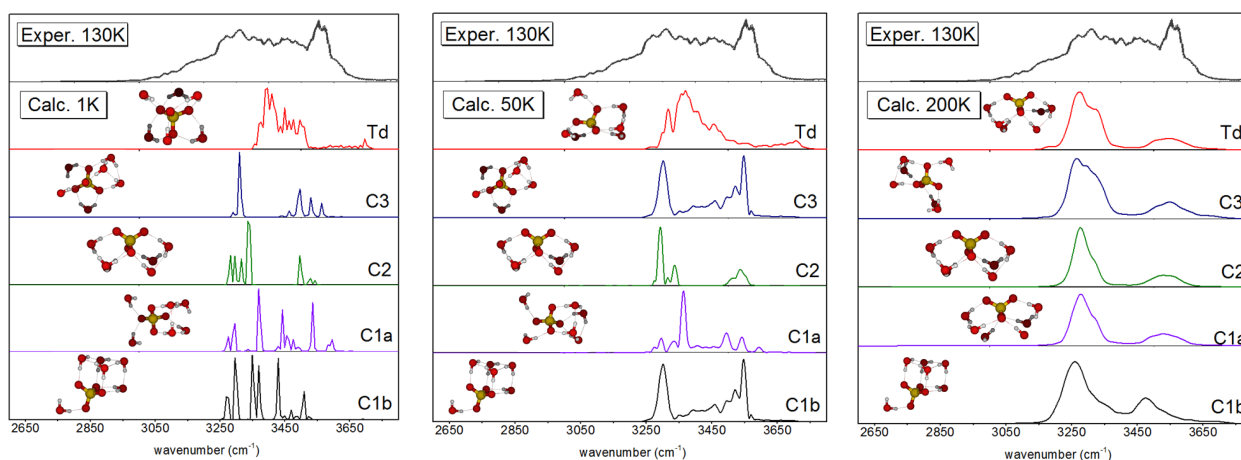


FIGURE 6.10 – Spectres DACF-AMOEBA calculés à 3 températures, 1 K, 50 K et 200 K, et comparés au spectre expérimental de Williams et coll. [12]. Dès 50 K, on observe une dynamique conformationnelle des structures les plus hautes en énergie vers les plus stables.

Gamme des empreintes digitales

La gamme de fréquences $500\text{-}1800\text{ cm}^{-1}$ a été étudiée par Zhou et al. [15] par une approche mixte expérience-théorie. Des spectres IRMPD ont été enregistrés à FELIX sur $[SO_4(H_2O)_n]^{2-}$, $n = 3\text{-}14, 18$ et 24 , à 17 K . Pour quelques tailles, des spectres MP2 ou B3LYP de structures de la littérature sont comparés aux signatures expérimentales. Pour $n = 6$, le spectre expérimental est attribué à la structure T_d , sur la base de spectres calculés MP2/TZVP+ et B3LYP/TZVP+, la structure étant la seule à ne pas présenter de bande dans la gamme $850\text{-}1000\text{ cm}^{-1}$. Nos propres calculs, aux niveaux MP2/b2 et M11/b1, conduisent à la même conclusion, sur la seule base des comparaisons entre spectres, alors que cette structure est la plus haute en énergie. Cette bande est attribuée au pliage hors du plan du vibreur O-H lié au sulfate, pour une molécule monodentate, qui en effet, n'existe pas dans la structure T_d , où toutes les molécules d'eau sont bidentates.

En dynamique AMOEBA, la conclusion est moins catégorique. Deux températures de simulations sont proposées (**Fig 6.11**), 1 K et 50 K . Pour des spectres expérimentaux à basse température, Gaigeot et coll [20] recommandent des températures de simulation plus hautes, afin d'accélérer la dynamique. Ainsi, les spectres calculés à 50 K rendent mieux compte des intensités relatives des bandes expérimentales à 17 K . Expérimentalement, le massif des bandes d'élongation antisymétrique des liaisons S=O est large. Chacune des géométries est éligible pour cette bande. Le pliage de l'eau est trop déplacé vers le bleu, malgré la calibration de la constante de pliage sur les agrégats d'eau pure. C'est la structure C_2 qui rend le mieux compte de cette bande. Les deux petits massifs autour de 600 et 800 cm^{-1} sont bien reproduits par plusieurs géométries, dont les deux géométries les plus stables en AMOEBA, C_3 et C_2 . La structure T_d ne montre pas de bande à 600 cm^{-1} . Aucune structure ne présente de bandes dans la gamme $850\text{-}1000\text{ cm}^{-1}$. En AMOEBA, les spectres des empreintes digitales ne permettent pas d'identifier une structure comme seule candidate au spectre expérimental, mais les plus stables sont capables de reproduire la gamme $500\text{-}1200\text{ cm}^{-1}$.

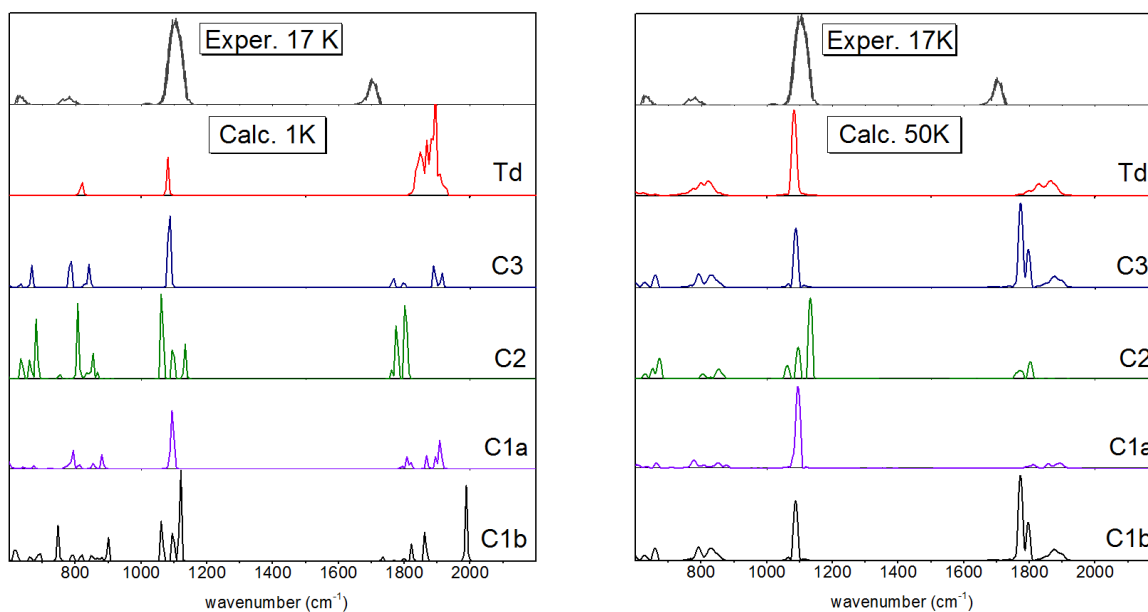


FIGURE 6.11 – Spectres DACF à 1 K et 50 K pour les 5 structures de basse énergie de $[SO_4(H_2O)_6]^{2-}$, comparés au spectre expérimental à 17 K [15].

6.2 $[\text{SO}_4(\text{H}_2\text{O})_9]^{2-}$

La procédure appliquée pour $n = 9$ est celle précédemment décrite. La première étape a consisté à rechercher des structures candidates via plusieurs longues simulations de dynamique AMOEBA (4 ns à 200 K) avec des structures de départ très différentes ou des générations successives de REMD, suivant la procédure décrite en **chapitre 3**. Quelques dizaines de structures ont ensuite été optimisées au niveau M11/b1. Plusieurs sont redondantes par inversion de cycles de liaisons hydrogènes. L'outil 'Traqueur de cycles' a été employé pour engendrer toutes les structures issues de l'inversion de cycles et leurs énergies ont été comparées au niveau M11/b3//M11/b1, incluant la correction de l'énergie vibrationnelle à 0 K (**Tab 6.5**). Au sein d'une famille de structures, peu de variations d'énergie sont observées, et la structure la plus stable de chaque famille est conservée.

TABLE 6.5 – Présence de cycles indépendants dans les structures de basse énergie de $[\text{SO}_4(\text{H}_2\text{O})_9]^{2-}$. La colonne 'Combinaisons' comptabilise le nombre de structures engendrées à partir d'inversions multiples (jusqu'à 4 inversions de cycles simultanées). La colonne 'Identiques par symétrie' décompte les structures équivalentes. La colonne ' E_{rel} ' affiche l'écart d'énergie, au niveau MP2/b3//MP2/b1 incluant la correction de ZPE, entre les structures engendrées par inversion de cycles et la structure qui a permis de les engendrer.

Structures	Cycles / structure	Atomes / cycles	Combinaisons	Identiques par symétrie	E_{rel} (kJ.mol ⁻¹)
W9-1	1	[9]			0.9
W9-2	2	[4;4]	1		0.0, 0.9, 5.4
W9-3	2	[4;4]	1		0.1, 0.3, 0.3
W9-4	1	[4]			1.8
W9-5	2	[3;4]	1		0.9, 0.9, 2.2
W9-6	2	[4;4]	1	-2	0.7
W9-7	3	[3;3;3]	4	-6	2.8

Enfin, 7 structures (**Fig 6.12**) sont identifiées comme étant les plus stables selon leurs énergies relatives au niveau DLPNO-CCSD(T)/b3//M11/b1, incluant les énergies ZPE M11/b1. Les énergies au niveau MP2/b3//MP2/b2 ont également été calculées (**Tab 6.6**).

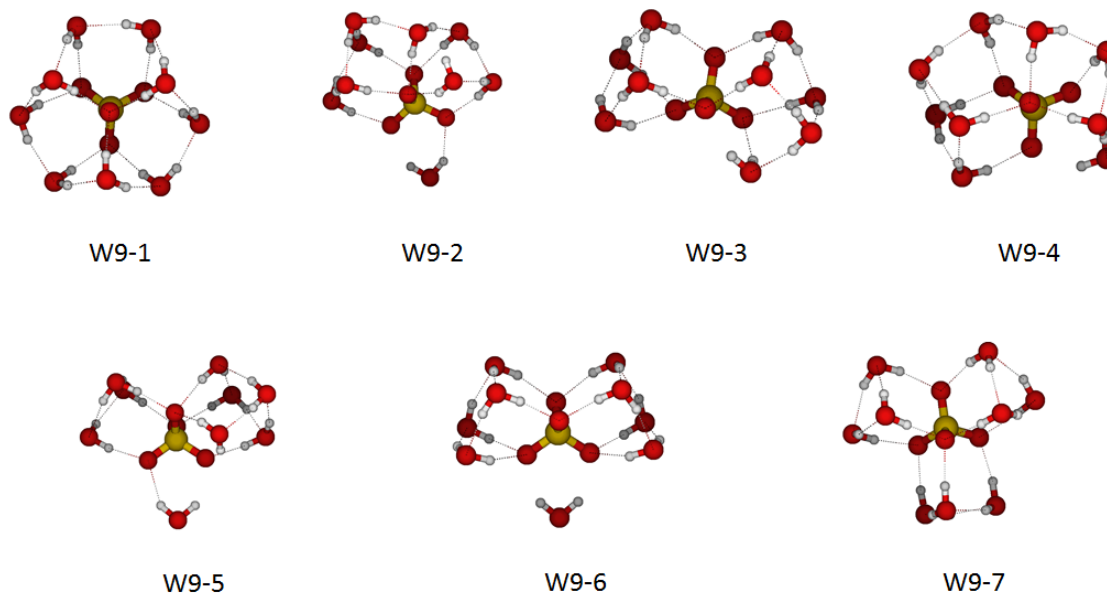


FIGURE 6.12 – Structures de basse énergie de $[\text{SO}_4(\text{H}_2\text{O})_9]^{2-}$. La nomenclature 'W9-x' est basée sur les énergies relatives DLPNO-CCSD(T)/b3//M11/b1, incluant les énergies ZPE M11/b1.

TABLE 6.6 – Energies relatives des structures de $[SO_4(H_2O)_9]^{2-}$ en $\text{kJ}\cdot\text{mol}^{-1}$.

Structures	DLPNO-CCSD(T)/b3//M11/b1			MP2/b3//MP2/b2			AMOEBA
	E_{0K}	E_{298K}	G_{298K}	E_{0K}	E_{298K}	G_{298K}	
W9-1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
W9-2	2.1	3.2	-4.6	6.0	6.6	0.7	14.8
W9-3	4.6	5.5	-1.1	8.5	8.5	5.6	12.2
W9-4	4.6	5.1	1.8	6.6	6.8	4.9	14.1
W9-5	6.0	9.0	-4.2	9.6	12.0	1.8	15.0
W9-6	6.2	9.1	-4.9	13.3	15.4	6.5	12.2
W9-7	8.4	11.6	-5.1	14.0	15.4	5.6	13.2

b1 : 6-311++G(d,p), b2 : cc-pVTZ, b3 : aug-cc-pVTZ

6.2.1 Spectres statiques multiconfigurationnels

A 0 K, la structure W9-1, de haute symétrie, est la plus stable aux deux niveaux de théorie mais l'écart avec les autres structures est en moyenne 2 fois plus important en MP2 qu'en DLPNO-CCSD(T). A 298 K, les stabilités relatives sont conservées, pour les deux méthodes, en appliquant le critère énergétique (E_{298K}). Les corrections MP2 et M11 incluant l'entropie (G_{298K}) sont comparables (un peu plus importantes en M11) mais le biais sur les énergies relatives conduit à des populations très différentes, puisque W9-1 est toujours la structure la plus stable en MP2 mais est absente du mélange en DLPNO-CCSD(T), notre niveau de référence.

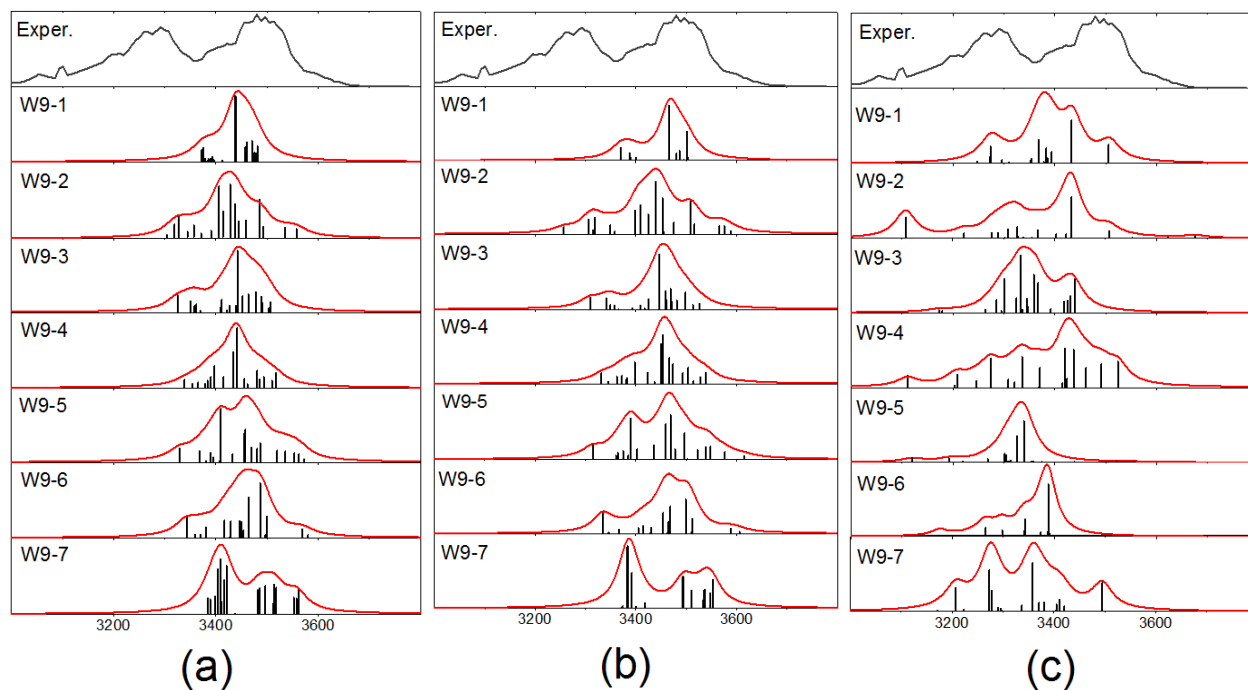


FIGURE 6.13 – Spectres expérimental et calculés statiques de 7 structures de $[SO_4(H_2O)_9]^{2-}$, calculés aux niveaux (a) M11/b1 + SF 0.945 et (b) MP2/b2 + 0.95, dans la double approximation harmonique, et au niveau (c) M11/b1 anharmonique (GVPT2).

Les spectres monoconformationnels sont qualitativement très similaires pour les deux méthodes (Fig 6.13, à gauche et au centre) dans l'approximation harmonique. La distribution en fréquences des spectres MP2 est un peu plus large, notamment vers les basses fréquences, ce que nous avons déjà observé pour $n = 6$. Les structures W9-2, W9-3, W9-5 et W9-6 ont chacune une bande dans la région de basse fréquence du spectre, attribuée à l'élongation

vers le sulfate d'une molécule deux fois donneuse et deux fois acceptrice de liaisons hydrogènes (AADD). Les autres bandes sous 3400 cm^{-1} correspondent à des élongations vers le sulfate de molécules ADD. Dans W9-2, W9-5 ou W9-6, on retrouve une molécule bidentate seulement deux fois donneuse, dont les modes de vibration correspondent aux plus hautes fréquences. Pour des structures très symétriques, pour lesquelles toutes les molécules sont en 1^{ère} sphère (telles que W9-1 ou W9-7), les spectres présentent deux massifs bien séparés qui conduisent aux mêmes attributions que pour $n = 6$, i.e les élongations O-H vers le sulfate aux basses fréquences ($\leq 3400\text{ cm}^{-1}$) et celles vers des molécules d'eau au-delà. Selon ces énergies, à basse température, la structure W9-1 est la plus stable, mais sa gamme de fréquences ne permet pas du tout de reproduire le spectre expérimental. A 130 K, en introduisant l'entropie, le niveau DLPNO-CCSD(T) montre que le spectre multiconformationnel est constitué à 54 % par la structure W9-2, celle décrivant le mieux le massif dans le rouge. A 298 K, les deux méthodes calculent des poids significatifs sur certaines des structures W9-2, W9-3, W9-5 ou W9-6 (**Tab 6.7**). Ces populations ne tiennent pas compte de la préselection des isomères par élimination des structures équivalentes par inversion de cycles. Pour les inclure dans la distribution de Boltzmann, on pondère chacune des 7 structures sélectionnées par le nombre de combinaisons équivalentes détectées par le 'Traqueur de cycles'. Par exemple, en appliquant ce nouveau critère à la population basée sur G_{130K} , au niveau DLPNO-CCSD(T)/b3//M11/b1, le mélange se compose de : 15 % W9-1 + 66 % W9-2 + 7 % W9-3 + 4 % W9-4 + 4 % W9-6 (**Fig 6.14, courbe verte**). Cependant, aucune structure n'est capable de reproduire la gamme de fréquences en dessous de 3300 cm^{-1} , ni de fournir une identification à la résonance $\sim 3100\text{ cm}^{-1}$. Pour espérer reproduire les déplacements les plus intenses dûs à une forte anharmonicité, une approche harmonique et l'application d'un scaling factor unique ne sont pas appropriés. Nous avons recours à un calcul anharmonique de type GVPT2, au niveau M11/6-311++G(d,p). Les spectres obtenus (**Fig 6.13, à droite**) affichent une distribution en fréquences beaucoup plus large que précédemment, et notamment, certaines structures, telles que W9-2 ou W9-4, sont capables de reproduire toute la gamme dans le rouge, y compris la résonance $\sim 3100\text{ cm}^{-1}$, qui est de nouveau attribuée à l'élongation vers le sulfate de molécules deux fois donneuses et deux fois acceptrices de liaisons hydrogènes (AADD). En appliquant la même population de mélange basée sur G_{130K} , le spectre anharmonique (**Fig 6.14, courbe rouge**) présente une distribution de bandes sur toute la gamme expérimentale, et notamment dans la gamme des basses fréquences, même si l'accord général, et notamment les intensités vers les hautes fréquences, n'est toujours pas satisfaisant. Nous avons mentionné en **chapitre 3** les difficultés et les efforts actuels pour bien reproduire les intensités anharmoniques et notamment celles des bandes de combinaison [21–23]. La reproduction simultanée des basses et des hautes fréquences reste un problème non résolu.

TABLE 6.7 – Population de Boltzmann de $[SO_4(H_2O)_9]^{2-}$ en fonction du niveau de théorie, de la propriété thermodynamique, et de la température. Seules les contributions $\geq 3\%$ de la population totale sont données.

Structures	DLPNO					MP2				
	ΔE			ΔG		ΔE			ΔG	
	1 K	130 K	298 K	130 K	298 K	1 K	130 K	298 K	130 K	298 K
W9-1	100	90	64	25	3	100	99	83	95	38
W9-2		8	17	54	21			6	3	28
W9-3			7	4	5					
W9-4			8					6		6
W9-5				6	18					18
W9-6				6	24					
W9-7					26					

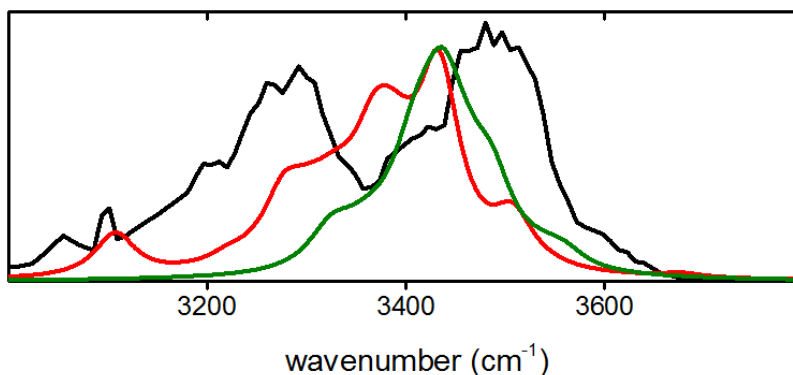


FIGURE 6.14 – Spectres multiconfigurationnels, basés sur le critère ΔG_{130K} en incluant les pondérations des inversions de cycles. Au niveau DLPNO-CCSD(T)/b3//M11/b1, le spectre multiconformationnel est composé de 15 % W9-1 + 66 % W9-2 + 7 % W9-3 + 4 % W9-4 + 4 % W9-6. Les spectres sont calculés dans l’approximation harmonique + SF 0.945 (vert) ou au niveau GVPT2 (rouge), et comparés au spectre de Williams et coll. (noir).

6.2.2 Spectres dynamiques AMOEBA

L’approche quantique statique semble avoir apporté une image cohérente des populations de structures à moyenne température. La petite taille de l’agrégat alliée à sa charge importante induisent certains déplacements de bandes très importants, dont la description échappe à l’approche harmonique et sa prise en compte moyennée de l’anharmonicité par un scaling factor. Si le modèle anharmonique améliore cette représentation, certaines contributions, telles que les plus hautes fréquences, sont assez mal reproduites et on peut se demander si la prise en compte d’effets dynamiques pourrait améliorer cette description. Comme pour $[SO_4(H_2O)_6]^{2-}$, le champ de forces est mis à mal pour décrire les interactions de molécules d’eau doublement liées au sulfate. Les énergies AMOEBA de ces interactions sont systématiquement surestimées (**Tab 6.6**), alors que 5 des 7 isomères sélectionnés possèdent au minimum une molécule de ce type. En conséquence, la structure W9-1 est calculée beaucoup plus stable que toutes les autres, qui se regroupent ensuite dans une gamme de 3 kJ/mol.

A 130 K, la majorité des structures sont cinétiquement stables en dynamique, bien que de nombreux écarts aux géométries d’équilibres soient observés (rupture temporaire de liaison hydrogène, par exemple). A 200 K, quelques inversions de géométries apparaissent dans les simulations avec pour point de départ les structures les plus hautes en énergie. Sans un bon accord sur les énergies relatives des conformères, le champ de forces ne peut prétendre proposer une population de Boltzmann représentative de la température. En revanche, il est possible de simuler un spectre de mélange par une moyenne des spectres dynamiques obtenus de simulations avec des structures de départ différentes. Ces signatures incluent, par la dynamique, l’anharmonicité des bandes due à la température et aux écarts aux géométries d’équilibre. Le spectre de mélange est moyenné à partir des spectres dynamiques des 7 structures de départ, à 200 K (**Fig 6.15**). Comme le massif expérimental, le massif calculé présente une double bande, mais dont le minimum est déplacé vers le bleu de 50 cm^{-1} . Cet écart diminue avec la température. Dans le rouge, le maximum est également déplacé de 50 cm^{-1} , mais l’augmentation de température n’améliore pas cette situation. La bande dans le bleu est bien reproduite (position du maximum et intensité nulle au-dessus de 3650 cm^{-1}). Cette bande est majoritairement attribuée aux interactions entre molécules d’eau alors que les contributions aux plus basses fréquences (i.e $\leq 3400\text{ cm}^{-1}$) mettent principalement en jeu les interactions avec le sulfate, qui sont très fortes dans ces petits agrégats. Les intensités relatives des deux massifs sont proportionnelles aux nombres d’interactions de type eau-sulfate et eau-eau mises en jeu. On note que le modèle à charge fluctuantes n’améliore que modérément la description des bandes de basse fréquence ($\leq 3250\text{ cm}^{-1}$), du fait que sa calibration a été réalisée sur des agrégats de plus grande taille, pour lesquels le sulfate obtient une coordinence moyenne de 12 et chacune de ses interactions est plus faible que celles observées dans les plus petits systèmes. La flexibilité des charges avec la

géométrie pourrait être améliorée par une calibration du modèle sur les petites tailles mais nécessite d'être incluse dans une procédure plus globale de validation des énergies relatives des petits agrégats.

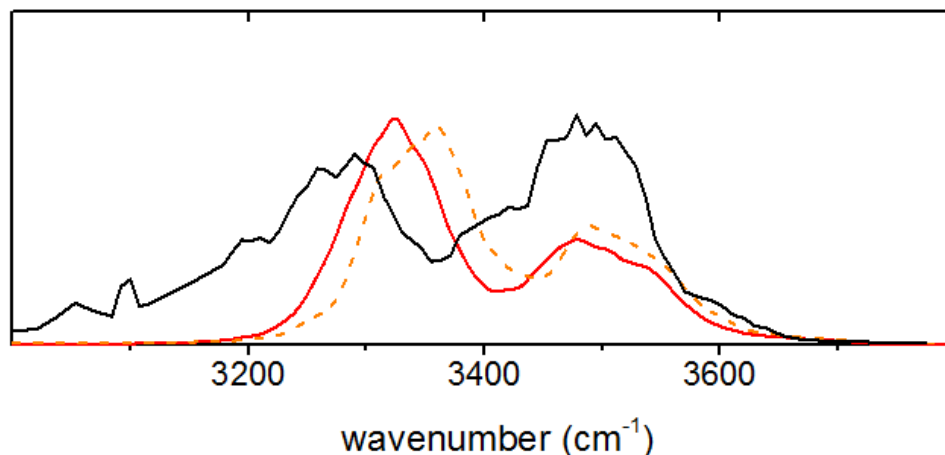


FIGURE 6.15 – Moyenne de spectres AMOEBA à partir des 7 structures stables à 200 K, à partir d'un modèle à charges atomiques fixes (pointillés oranges) ou dépendantes de la géométrie (courbe rouge). Les spectres calculés sont comparés au spectre de Williams et coll. (noir).

Gamme des empreintes digitales :

Dans la gamme 600-1200 cm^{-1} (**Fig 6.16**), toutes les structures présentent des profils assez similaires. La constante d'élongation S=O a été calibrée pour reproduire la bande vers 1100 cm^{-1} de petits agrégats $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$, mais est peu dépendante de l'environnement. A l'inverse, la bande du pliage $\widehat{\text{HOH}}$ a été calibrée à 1630 cm^{-1} sur des agrégats d'eau pure, mais est de nouveau de fréquence trop élevée, s'agissant de décrire le pliage de molécules en 1^{ère} sphère de solvation. La forme et la position de la bande sont donc très dépendantes de la structure de départ, et du nombre d'interactions eau-eau ou eau-sulfate. Le massif à 600-800 cm^{-1} regroupe le pliage $\widehat{\text{OSO}}$ et des modes de rotation frustrée (libration) de l'eau. Les intensités relatives des 3 massifs à 50 K sont en accord très satisfaisant avec le spectre IRMPD à 17 K [15].

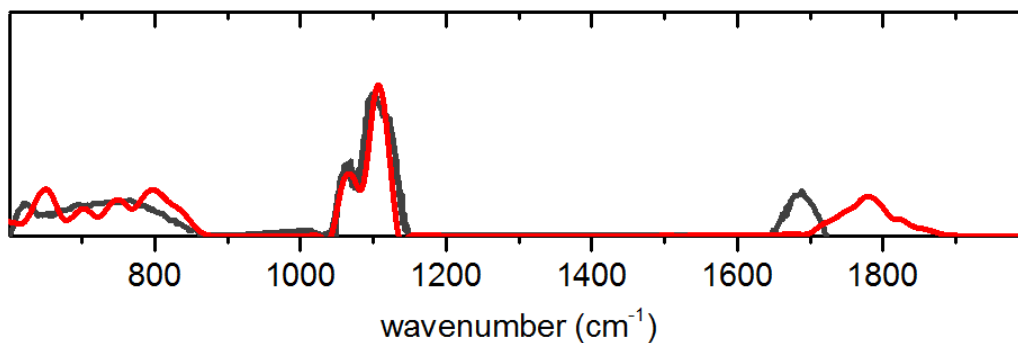


FIGURE 6.16 – Spectre DACF-AMOEBE (rouge) à 50 K moyenné sur les 7 structures de basse énergie de $[\text{SO}_4(\text{H}_2\text{O})_9]^{2-}$ et comparé au spectre expérimental (noir) à 17 K [15].

6.3 $[\text{SO}_4(\text{H}_2\text{O})_{12}]^{2-}$

6.3.1 Spectres statiques multiconfigurationnels

Les structures les plus stables de $[\text{SO}_4(\text{H}_2\text{O})_{12}]^{2-}$ ont été publiées dans [3] (**annexe D**). Sur les 15 structures présentées, les 4 plus basses en énergie (**Tab 6.8**), sont identifiées W12-1a à W12-1d (**Fig 6.17**), car elles ne diffèrent que par l'inversion de quelques liaisons hydrogènes. Ces structures présentent 3 cycles de 4 liaisons hydrogènes. Entre W12-1a et W12-1b, le sens de chaque cycle est inversé (en vue extérieure, tous dans le sens anti-horaire pour W12-1a, et horaire pour W12-1b). W12-1c est un cas intermédiaire, avec 2 cycles dans un sens et un dans l'autre. La géométrie W12-1d est un peu différente, puisque les cycles de 4 liaisons hydrogènes sont identiques à ceux de W12-1a, mais pour chaque groupe de 4 molécules, une molécule d'eau est en 2^{nde} sphère, et ce n'est pas la même dans ces deux géométries. Au niveau de calcul de référence DLPNO-CCSD(T)/b3//M11/b1, les 3 premières sont clairement plus stables que toutes les autres, même en incluant les énergies thermiques. Seule W12-5 devient compétitive sur le critère de l'énergie libre de Gibbs G_{298K} .

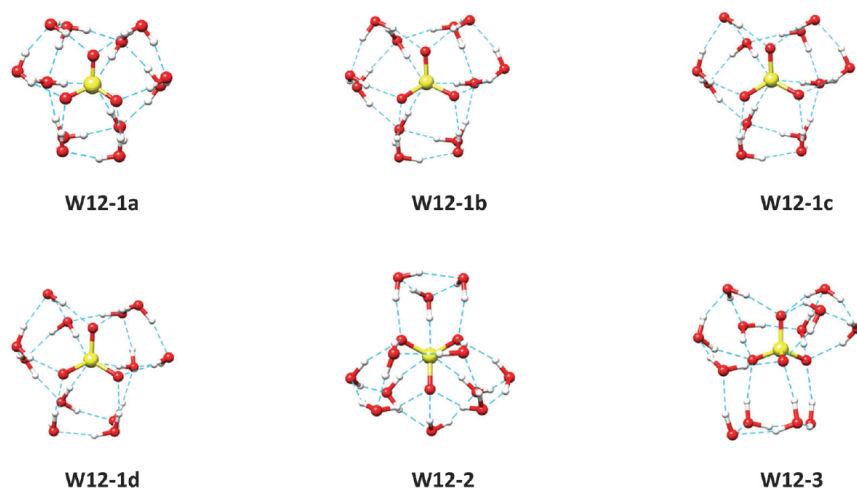


FIGURE 6.17 – 6 isomères de plus basse énergie de $[\text{SO}_4(\text{H}_2\text{O})_{12}]^{2-}$. Figure 1 de l'article [3].

TABLE 6.8 – Energies relatives des structures de $[\text{SO}_4(\text{H}_2\text{O})_{12}]^{2-}$, au niveau DLPNO-CCSD(T)/b3//M11/b1 (kJ/mol).

Structures	ΔE_{0K}	ΔE_{298K}	ΔG_{298K}	AMOEBA
W12-1a	0.0	0.0	0.1	0.9
W12-1b	0.0	0.0	0.0	0.0
W12-1c	2.1	2.3	2.8	2.4
W12-1d	6.2	7.1	8.3	7.5
W12-2	8.2	17.9	6.9	26.3
W12-3	8.5	15.6	6.4	29.7
W12-4	8.9	11.8	8.3	12.0
W12-5	10.0	14.9	1.7	25.5
W12-6	10.2	13.1	4.6	21.7
W12-7	10.8	11.8	12.7	19.8
W12-8	12.9	21.5	14.1	32.8
W12-9	13.1	18.7	12.1	29.7
W12-10	13.3	14.4	15.8	21.9
W12-11	17.7	27.5	19.1	39.6
W12-GL	19.0	26.7	12.6	31.4

La **Fig 6.18** regroupe les spectres statiques harmoniques des structures les plus stables. Les structures W12-1'x' présentent toutes des bandes autour de 3200 cm^{-1} et peuvent reproduire tout le spectre à 130 K. Les bandes à basses fréquences sont attribuées aux élongations vers le sulfate de molécules (AADD) dans trois combinaisons, une symétrique et deux antisymétriques (pour W12-1a, respectivement à 3207 , 3220 et 3282 cm^{-1} , après scaling factor). Pour W12-1a, un calcul anharmonique GVPT2 au niveau M11/b1 calcule une des élongations antisymétrique à 3065 cm^{-1} mais est incapable de décrire le reste du massif OH. Les bandes entre 3300 et 3400 cm^{-1} sont attribuées aux vibreurs vers une molécule d'eau des molécules (AADD) et à d'autres vibreurs (ADD) de 1^{ère} sphère, dans des combinaisons symétriques. Ces structures possèdent trois molécules de 2^{ème} sphère, qui vibrent sous le massif principal dans une combinaison symétrique ($\geq 3400\text{ cm}^{-1}$) et antisymétrique ($\geq 3500\text{ cm}^{-1}$). Sous ce massif apparaissent également les combinaisons antisymétriques des vibreurs ADD de 1^{ère} sphère.

Ce très bon accord valide l'hypothèse énergétique que cette petite série d'isomères est représentative de la signature expérimentale à 130 K. Les calculs d'énergies et enthalpies libres à 130 K et 298 K confirment que seule la structure W12-5 pourrait être compétitive à 298 K. Toutefois, la présence de nombreux cycles dans les structures W12-1'x' (**Tab 6.9**) leur assure une représentation quasi exclusive dans un calcul de population tel que réalisé précédemment. Certaines structures doivent logiquement être décomptées (W12-1b est une image de W12-1a par l'inversion simultanée de 3 cycles, par exemple), mais il est facilement observé que le nombre de structures issues de W12-1'x' associé à leurs énergies, rend marginale toute autre contribution.

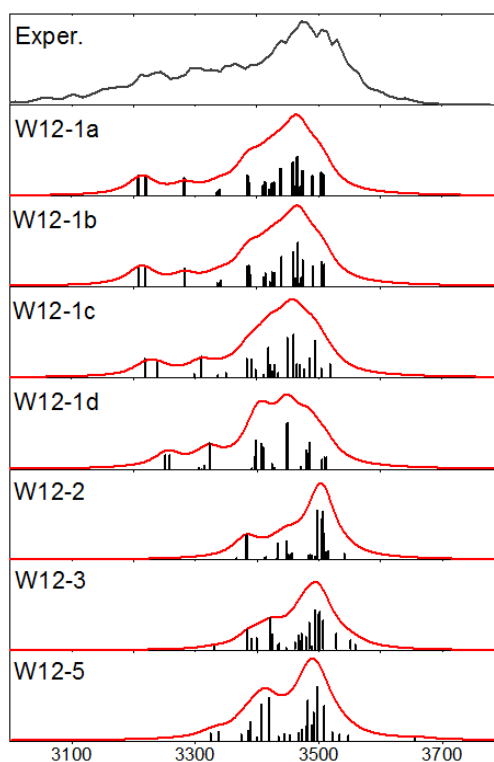


FIGURE 6.18 – Spectres statiques au niveau M11/6-311++G(d,p) + SF 0.945, pour les structures les plus stables de $[SO_4(H_2O)_{12}]^{2-}$, et comparés au spectre expérimental de Williams $\sim 130\text{ K}$ (gris) [2].

TABLE 6.9 – Présence de cycles indépendants dans les structures de basse énergie de $[\text{SO}_4(\text{H}_2\text{O})_{12}]^{2-}$. La colonne 'Combinaisons' comptabilise le nombre de structures engendrées à partir d'inversions multiples (jusqu'à 4 inversions de cycles simultanées).

Structures	Cycles / structure	Atomes / cycles	Combinaisons
W12-1a	4	[4;4;4;12]	4
W12-1b	4	[4;4;4;12]	4
W12-1c	4	[4;4;4;10]	4
W12-1d	4	[4;4;4;6]	4
W12-2	2	[3;9]	1
W12-3	1	[4]	
W12-4	5	[4;5;7;8;10]	2
W12-5	2	[4;4]	1
W12-6	3	[4;4;4]	4
W12-7	4	[5;9;10;12]	
W12-8	1	[9]	
W12-9	1	[4]	
W12-10	2	[3;9]	1
W12-11	1	[11]	
W12-GL	2	[4;4]	1

6.3.2 Spectres dynamiques AMOEBA

Les structures W12-1'x' ne présentant pas de molécules d'eau doublement liées au sulfate, leurs énergies relatives AMOEBA (**Tab 6.8**) sont, de fait, en excellent accord avec les énergies DLPNO-CCSD(T) ($\leq 1 \text{ kJ.mol}^{-1}$). C'est également le cas de W12-4, pour les mêmes raisons. Les autres structures, dans lesquelles ces interactions sont présentes, sont très largement défavorisées en AMOEBA.

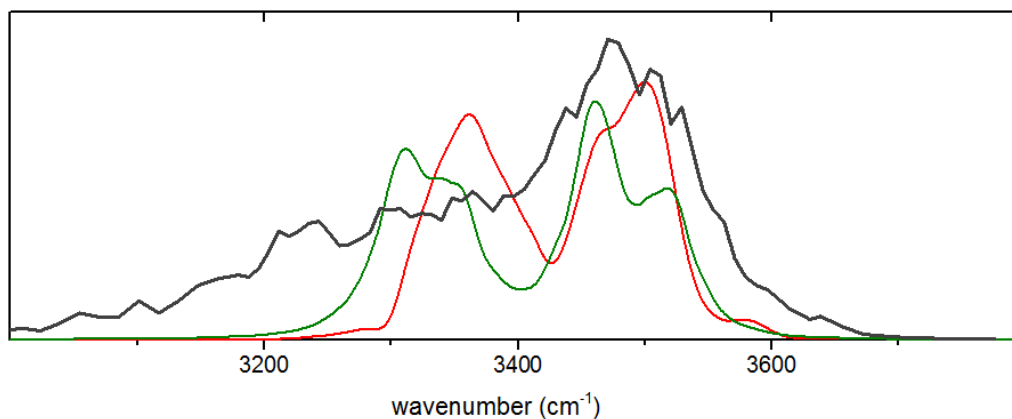


FIGURE 6.19 – Spectres DACF-AMOEBAs à 200 K, à partir d'une moyenne des spectres monoconformationnels des 15 structures (rouge), ou uniquement des structures W12-1'x' (vert), et comparés au spectre expérimental de Williams et coll. $\sim 130 \text{ K}$ (noir) pour $[\text{SO}_4(\text{H}_2\text{O})_{12}]^{2-}$.

Dans ce cas précis, de longues simulations AMOEBA pourraient conduire à une représentation pertinente du mélange de structures à basse température, mais il est beaucoup plus rapide de moyenner des spectres avec des structures de départ différentes. Deux spectres de mélange à 200 K sont proposés, l'un à partir des spectres dynamiques initiés par chacune des 15 structures précédentes, et l'autre uniquement à partir des structures W12-1'x' (**Fig 6.19**), beaucoup plus représentatif des populations à cette température. Ces spectres s'organisent autour de deux massifs, assez similaires à ceux de $n = 9$, à l'exception des intensités relatives puisque la bande vers $3250\text{-}3400 \text{ cm}^{-1}$ perd en intensité. La perte de cette contribution, si elle est moins marquée qu'expérimentalement, témoigne

de la présence de molécules d'eau en 2^{nde} sphère de solvation et de l'augmentation de la proportion d'interactions eau-eau. A cela s'ajoute l'augmentation de la coordination du sulfate, qui affaiblit chacune de ces interactions. Ce massif est large de 125 cm⁻¹, et 50 cm⁻¹ plus déplacé vers les basses fréquences pour les structures W12-1'x' que pour les structures ne présentant pas de molécule (AADD). Ce spectre en double contribution montre que les différences d'environnement (eau-sulfate et eau-eau) sont très marquées dans le champ de forces (électrostatique et vdW), mais qu'au sein de ces environnements, trop peu de flexibilité est apportée, le modèle de polarisation, assisté du modèle de charges fluctuantes, ne suffisant pas à décrire la distribution en fréquences induite par ces environnements proches mais non identiques.

Gamme des empreintes digitales :

Dans cette gamme, peu de différences sont observées. Toutes les structures reproduisent bien la double bande d'élongation du sulfate. La fréquence du pliage de l'eau est encore surestimée, bien que de façon plus raisonnable pour les structures stables W12-1'x'.

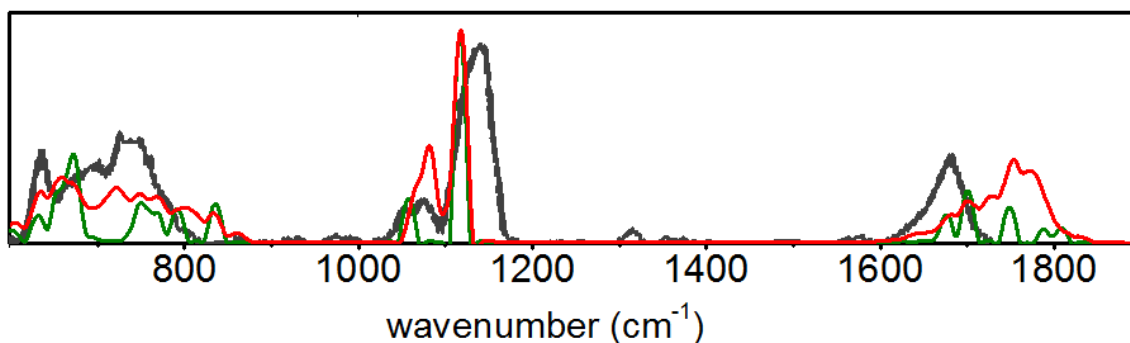


FIGURE 6.20 – Spectres DACF-AMOEBA à 50 K à partir d'une moyenne des spectres monoconformationnels des 15 structures (rouge) ou des 4 structures W12-1'x' (vert), comparés au spectre expérimental à 17 K (noir) publié dans [15].

6.4 [SO₄(H₂O)₁₃]²⁻

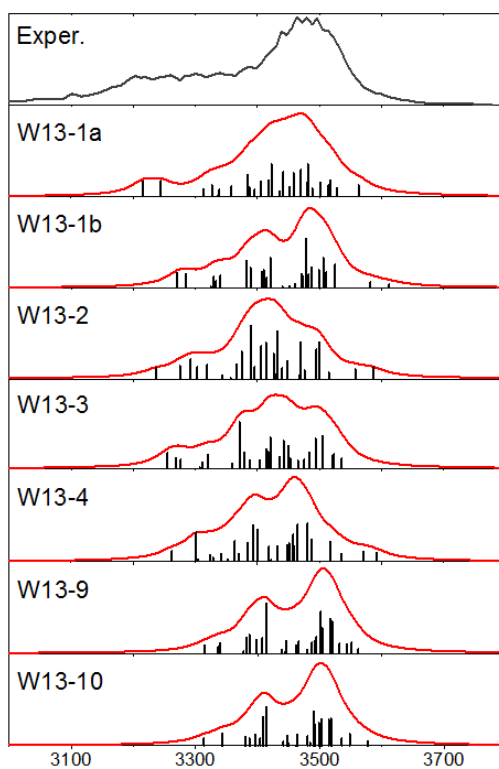
6.4.1 Spectres statiques multiconfigurationnels

Les structures pour $n = 13$ ont également été publiées dans [3] (**annexe D**). La nomenclature diffère quelque peu de $n = 12$ puisque la série W13-1'xy' désigne à présent des agrégats présentant un noyau W12-1, et une 13^{ème} molécule d'eau en divers sites à la surface de l'agrégat. Il a été observé que certaines de ces structures sont très stables (**Tab 6.10**), mais $n = 13$ contraste nettement avec $n = 12$ par la large distribution de structures énergétiquement compétitives (14 structures dans une gamme de 5 kJ.mol⁻¹ sur le critère G_{298K}). La structure W13-1a est la plus stable à 0 K et 298 K (mais participe alors à un fort mélange). Son spectre harmonique au niveau M11/6-311++G(d,p) est capable de décrire l'ensemble du massif des élongations O-H (**Fig 6.21**), de la bande la plus bleue vers 3560 cm⁻¹ aux contributions de molécules AADD vers 3200 cm⁻¹. Cependant, plusieurs structures semblent énergétiquement compétitives, et présentent des allures telles qu'une combinaison est également en bon accord avec le spectre expérimental.

La présence de cycles de plus en plus nombreux au sein de chacune des structures (**Tab 6.11**) diminue l'importance de ce critère, qui peut raisonnablement être négligé pour toute taille supérieure. La population au sein du mélange évolue en incluant la température ou l'entropie (**Tab 6.12**), mais le spectre, lui, est très peu modifié, les

TABLE 6.10 – Energies relatives des structures de $[\text{SO}_4(\text{H}_2\text{O})_{13}]^{2-}$ en kJ/mol. Les énergies quantiques sont calculées au niveau DLPNO-CCSD(T)/b3//M11/b1.

Structures	ΔE_{0K}	ΔE_{298K}	ΔG_{298K}	AMOEBBA
W13-1a	0.0	0.0	0.0	5.1
W13-2	1.2	0.7	2.0	5.5
W13-3	2.9	1.6	5.9	0.0
W13-1b	3.1	4.5	1.1	18.4
W13-1c	3.1	4.5	2.6	3.0
W13-4	3.1	2.9	2.6	5.0
W13-5	3.2	3.9	3.2	16.7
W13-1b2	3.5	4.6	2.0	14.6
W13-6	3.6	4.3	1.7	13.2
W13-2b	4.0	3.4	5.6	5.8
W13-1b3	4.3	5.8	2.3	19.2
W13-1b4	4.4	5.3	3.7	20.5
W13-1d	5.1	4.9	4.3	3.0
W13-7	5.3	6.5	4.3	22.3
W13-8	5.7	5.9	6.9	3.3
W13-9	6.1	8.6	0.3	24.1
W13-1d2	6.3	5.8	6.8	25.6
W13-10	6.3	9.1	0.1	4.8
W13-1e	6.5	7.0	6.4	17.7
W13-11	8.3	7.8	9.8	13.6
W13-12	9.9	10.3	10.2	12.8
W13-13	9.9	9.7	11.2	20.3
W13-AN	19.5	24.5	7.7	27.5

FIGURE 6.21 – Spectres statiques harmoniques calculés au niveau M11/b1 + SF 0.945 pour les structures les plus stables de $[\text{SO}_4(\text{H}_2\text{O})_{13}]^{2-}$ et comparés au spectre expérimental de Williams et coll. (~ 130 K).

environnements devenant de plus en plus équivalents à mesure que la taille de l'agrégat augmente, et les interactions moins spécifiques à une géométrie particulière (**Fig 6.22**).

TABLE 6.11 – Présence de cycles indépendants dans les structures de basse énergie de $[\text{SO}_4(\text{H}_2\text{O})_{13}]^{2-}$

Structures	Cycles / structure	Atomes / cycle	Combinaisons
W13-1a	4	[4;4;4;10]	4
W13-2	6	[4;4;10;10;10;10]	1
W13-3	5	[4;4;4;5;11]	5
W13-1b	4	[4;4;4;10]	4
W13-1c	4	[4;4;4;12]	4
W13-4	4	[4;4;9;9]	1
W13-5	2	[4;9]	1
W13-1b2	4	[4;4;4;10]	4
W13-6	4	[4;4;4;5]	5
W13-2b	5	[4;4;4;10;12]	2
W12-1b3	4	[4;4;4;8]	4
W13-1b4	4	[4;4;4;12]	4
W13-1d	4	[4;4;4;12]	
W13-7	3	[4;4;4]	4
W13-8	4	[4;4;5;5]	3
W13-9	3	[3;3;3]	4
W13-1d2	4	[4;4;4;10]	4
W13-10	3	[4;4;4]	4
W13-1e	4	[4;4;4;10]	4
W13-11	5	[4;4;6;10;10]	2
W13-12	4	[4;4;5;8]	4
W13-13	4	[4;4;4;11]	4
W13-AN	3	[3;3;4]	4

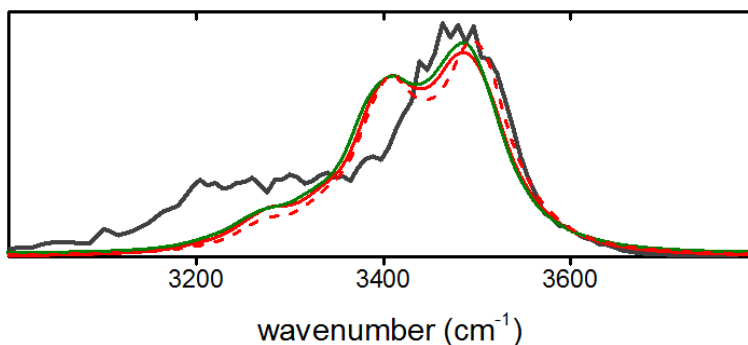


FIGURE 6.22 – Spectres multiconfigurationnels au niveau DLPNO-CCSD(T)/b3/M11/b1 + SF 0.945, basés sur les critères d'énergie à 130 K (pointillés rouges) ou 298 K (traits pleins rouges), ou d'enthalpie libre à 298 K (vert). Les spectres calculés sont comparés au spectre expérimental de Williams et coll. ~ 130 K (noir).

6.4.2 Spectres dynamiques AMOEBA

Les énergies relatives AMOEBA sont en accord correct ($\sim 5 \text{ kJ}\cdot\text{mol}^{-1}$) avec les énergies DLPNO/b3//M11/b1 à 0 K, pour les structures ne présentant pas de molécules d'eau doublement pontées au sulfate (**Tab 6.10**). Les spectres dynamiques sont qualitativement très similaires à ceux obtenus pour $n = 12$ et n'apportent aucune information complémentaire.

TABLE 6.12 – Population de Boltzmann de $[\text{SO}_4(\text{H}_2\text{O})_{13}]^{2-}$ au niveau DLPNO-CCSD(T)/b3/M11/b1, en fonction de la propriété thermodynamique et de la température. Seules les contributions $\geq 4\%$ de la population totale sont données.

Structures	ΔE			ΔG	
	1 K	130 K	298 K	130 K	298 K
W13-1a	100	55	22	52	14
W13-2		22	17	12	6
W13-3		6	12		
W13-1b			4	7	9
W13-1c		3	7	4	5
W13-4			4	4	5
W13-5			4		4
W13-1b2					6
W13-6			4	6	7
W13-2b			6		
W12-1b3					5
W13-9					12
W13-10					13

6.5 Tailles intermédiaires : $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$, $n = 23-25$

Nous rappelons ici les hypothèses formulées à partir des indices fournis par les expériences de Williams et al. (spectres IRPD, abondances ESI et constantes de vitesse de dissociation BIRD). La taille $n = 24$ présente une abondance ESI importante, en lien avec une grande constante de vitesse de dissociation BIRD de $n = 25$, témoignant d'une fragmentation aisée. Certaines des structures les plus stables pour $n = 24$ et 25 pourraient donc être très similaires. En revanche, $n = 23$ devrait présenter une stabilité comparable à $n = 24$ et rien n'indique un schéma de fragmentation simple entre ces deux tailles. De plus, dans cette série, le peuplement de la 2^{ème} sphère de solvation s'accompagne d'une nouvelle bande vers 3530 cm^{-1} **Fig 6.2**, dont on souhaite définir l'environnement.

Nous utilisons la procédure décrite pour les plus petites tailles, en commençant par une recherche de structures stables à partir des outils de la dynamique moléculaire. Pour $n = 24$ et 25 , des simulations classiques AMOEBA à 200 K ont été réalisées, à partir de plusieurs points de départ très différents, tandis que pour $n = 23$, nous avons privilégié des générations successives de REMD. La chimie quantique est ensuite utilisée pour obtenir des énergies fiables de ces structures et calculer les spectres statiques de celles de plus basse énergie. En parallèle, l'augmentation de l'espace des conformations justifie un emploi plus important du DACF pour calculer des spectres de mélange à température finie, qui peuvent être attribués par DMD.

6.5.1 $[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$

Spectres statiques

Pour cette série de tailles, les méthodes post-HF deviennent trop coûteuses, et la DFT a été utilisée pour tous les calculs quantiques (**Tab 6.13**). Le niveau de calcul M11/b1 a été validé sur les petits agrégats à la suite des travaux de calibration de méthodes par Head-Gordon et al. [7,8]. Appliqué à $n = 24$, on observe de grandes disparités dans les énergies relatives lorsqu'on effectue un calcul d'énergie avec une base plus étendue, à ces géométries. Le problème est illustré par les énergies M11/b3//M11/b1, présentées dans le **Tab 6.13**, mais ces calculs ont également été réalisés aux niveaux M11/b2//M11/b1 et M11/jun-cc-pVTZ//M11/b1, pour des résultats similaires à l'utilisation de la base b3. Truhlar et al. ont identifié que la fonctionnelle $\omega\text{B97X-D}$ [11,16] traite mieux les interactions entre molécules d'eau. Afin d'éviter un éventuel effet de la base sur les géométries, les optimisations sont réalisées au

niveau ω B97XD/b2, puis les calculs d'énergies au niveau ω B97XD/b3// ω B97XD/b2.

TABLE 6.13 – Energies relatives des structures de $[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$ (en $\text{kJ}\cdot\text{mol}^{-1}$). Les énergies quantiques incluent la correction de ZPE. La colonne 'Combinaisons' comptabilise le nombre de structures engendrées à partir d'inversions multiples (jusqu'à 4 inversions de cycles simultanées).

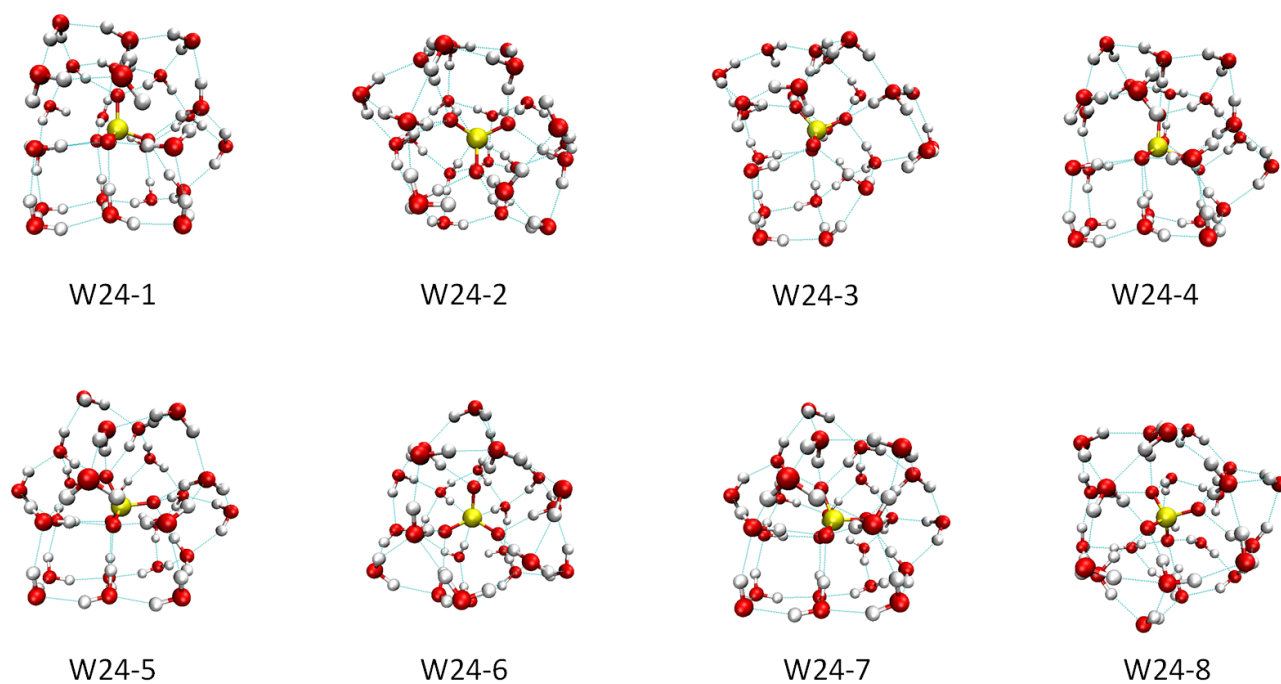
Structures	M11/b1	M11/b3// M11/b1	ω B97XD/b2	ω B97XD/b3// ω B97XD/b2	AMOEBa	Cycles / structure	Combinaisons
W24-1	0.0	0.0	0.0	0.0	0.0	24	60
W24-2	0.8	5.4	6.9	7.9	14.7	40	74
W24-3	7.0	5.8	7.6	5.9	13.7	18	47
W24-4	6.5	8.1	11.0	10.6	-2.9	22	54
W24-5	4.1	9.0	1.6	0.7	-0.2	51	43
W24-6	-7.4	9.4	3.5	2.2	6.9	79	88
W24-7	5.6	10.2	4.1	2.7	-16.5	41	48
W24-8	6.7	18.2	3.0	2.4	15.3	59	108

b1 : 6-311++G(d,p), b2 : cc-pVTZ, b3 : aug-cc-pVTZ

8 structures ont été sélectionnées, identifiées selon leurs énergies relatives au niveau M11/b3//M11/b1 et présentées en **Fig 6.23**. La structure W24-1 est la structure calculée la plus stable au niveau M11/b3//M11/b2 et aux deux niveaux incluant la fonctionnelle ω B97XD. Si certaines structures présentent des motifs de liaisons hydrogènes particuliers ou des symétries locales assez reconnaissables, discriminer des structures ou familles de structures peut être assez difficile sur la seule base des énergies et d'une comparaison visuelle. L'outil 'Traqueur de cycles' facilite l'identification des structures, ou l'élimination de doublons. On notera que le nombre de cycles ne s'homogénéise pas avec l'augmentation de la taille (**Tab 6.13**), puisque certaines structures en possèdent 4 fois plus que d'autres. Le nombre très important de cycles par structure rend impossible d'échantillonner toutes les structures nées de l'inversion de cycles au cours de la recherche par dynamique moléculaire. La génération de l'ensemble de ces structures par le 'Traqueur de cycles' puis leur optimisation avec AMOEBA est une tâche tout à fait réalisable, mais la fiabilité des énergies AMOEBA ne permet pas de discriminer des effets qui sont le plus souvent assez faibles.

Les spectres des 8 structures ont été calculés (**Fig 6.24**) aux niveaux M11/b1 (à gauche) et ω B97XD/b2 (à droite). Qualitativement, les allures des spectres monoconformationnels sont peu dépendantes du niveau de théorie. Une majorité de structures présente des bandes assez intenses en dessous de 3300 cm^{-1} , qui n'existent pas dans le spectre expérimental. Récemment (mais après cette série de spectres), Williams et al. ont remplacé une lampe de la pompe laser de leur FT-ICR. Dans la série $n = 9-13$ (pour laquelle seuls les nouveaux spectres ont été présentés ici), il en a résulté une augmentation très nette des intensités des massifs en dessous de 3300 cm^{-1} , ce qui conduit à être prudent sur l'absence de toute contribution pour $n = 24$. Toutefois, au niveau ω B97XD/b2, certaines structures (W24-5, W24-6, W24-7 et W24-8) ont des intensités plus faibles à basses fréquences. Ces structures sont également celles qui contribuent le plus au mélange à 130 et 298 K (**Tab 6.14**), sur la base d'un calcul de population de Boltzmann incluant l'enthalpie libre ($\sim 75\%$ de la population totale à 130 K). En revanche, au niveau M11/b1, la structure W24-1 reste beaucoup plus stable que les autres, même en incluant les corrections thermodynamiques.

On peut observer que toutes les structures présentent des bandes dans la gamme $3500-3600\text{ cm}^{-1}$, certaines assez nombreuses ou intenses pour conduire à un épaulement, à corrélérer à la bande expérimentale vers 3530 cm^{-1} . A la différence des plus petites tailles pour lesquelles certaines structures présentaient des interactions très spécifiques, les types d'environnements en 1^{ère} et 2^{ème} sphères sont très similaires d'une structure à l'autre. Ainsi, une structure (par exemple W24-1) permet d'attribuer la plupart des bandes observées dans les autres structures. Afin de définir systématiquement chaque type de molécule d'eau et l'environnement auquel elle est attachée, nous utilisons une notation dérivée de celle de Williams [24] : A = acceptrice de liaison H, D = donneuse de liaison H. Ainsi, les

FIGURE 6.23 – Structures de basse énergie de $[SO_4(H_2O)_{24}]^{2-}$.

molécules de 1^{ère} sphère sont généralement (AADD) (exceptionnellement (AAADD)) tandis que celles de 2^{ème} sphère sont toutes (ADD). Cette nomenclature peut être précisée afin d'inclure le type de chaque liaison hydrogène. Par exemple, une molécule de 2^{ème} sphère (ADD) est (211) si elle est acceptrice d'une liaison hydrogène d'une molécule de 2^{ème} sphère et donne à deux molécules de 1^{ère} sphère. Une molécule de 1^{ère} sphère (AADD) est également (22S2) si elle accepte de 2 molécules de 2^{ème} sphère, et donne vers le sulfate et vers une molécule de 2^{ème} sphère. Des exemples d'identifications de molécules sont proposés en **Fig 6.25**.

Nous nous limitons aux attributions des bandes de hautes ($\geq 3500\text{ cm}^{-1}$) et basses ($\leq 3360\text{ cm}^{-1}$) fréquences, les fréquences intermédiaires mettant en jeu des mouvements concertés d'un grand nombre de vibreurs les rendant difficile à décrire, et qui ne permettent pas d'attribuer un spectre à une structure. Les autres bandes de W24-1 sont regroupées dans le **Tab 6.15**. W24-1 possède 12 molécules en 1^{ère} sphère, dont $1 \times (12S1)$, $4 \times (22S1)$, $5 \times (12S2)$ et $2 \times (22S2)$, et 12 molécules en 2^{ème} sphère, $5 \times (111)$, $2 \times (112)$ et $5 \times (211)$.

La majorité des contributions au dessus de 3520 cm^{-1} sont attribuées aux molécules de 2^{ème} sphère (211), bien que d'autres contributions, de molécules de 2^{ème} sphère (111) ou (112), ou encore des mouvements concertés de vibreurs de 1^{ère} sphère, puissent être relevées. Le nombre de molécules (211) croit avec la taille de l'agrégat et le peuplement de la 2^{ème} sphère, ce qui permet d'expliquer l'apparition de la bande expérimentale vers 3530 cm^{-1} (**Fig 6.2**). Vers les plus basses fréquences ($\leq 3340\text{ cm}^{-1}$), on trouve exclusivement des élongations de molécules de 1^{ère} sphère vers la 2^{ème}.

Dynamique moléculaire

Cette gamme de tailles rentre plus dans le domaine d'application du champ de forces AMOEBA que les plus petites. Certaines interactions mal décrites et propres aux petites tailles n'existent plus (molécules d'eau de 1^{ère} sphère doublement liées au sulfate). La calibration du modèle de charges fluctuantes a été réalisée sur les sulfates de grande taille ($n = 36-80$). Enfin, l'intérêt de la dynamique est renforcé par la taille de l'agrégat et la flexibilité de la 2^{ème} sphère, qui contribuent à un plus grand mélange de structures.

En dynamique AMOEBA, quelques réorganisations ont lieu au cours des simulations de 200 ps à 100 K, et un

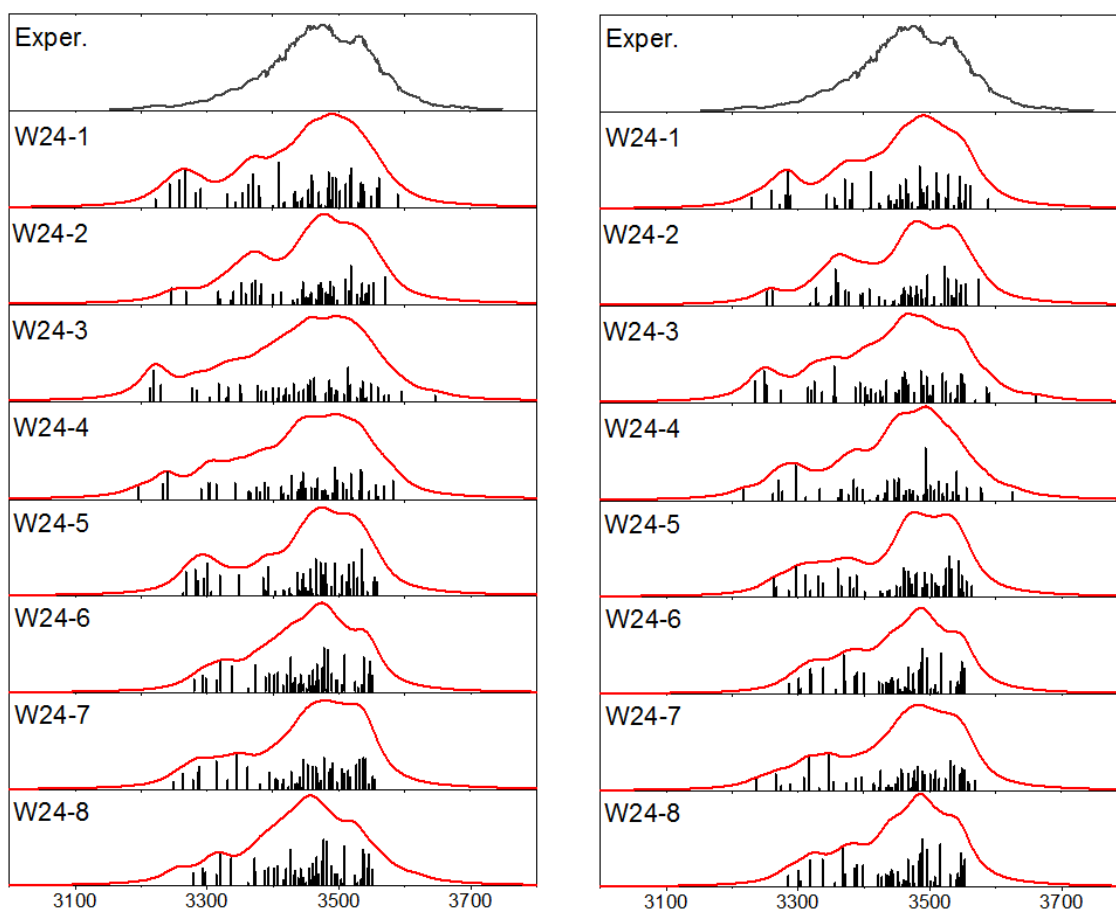


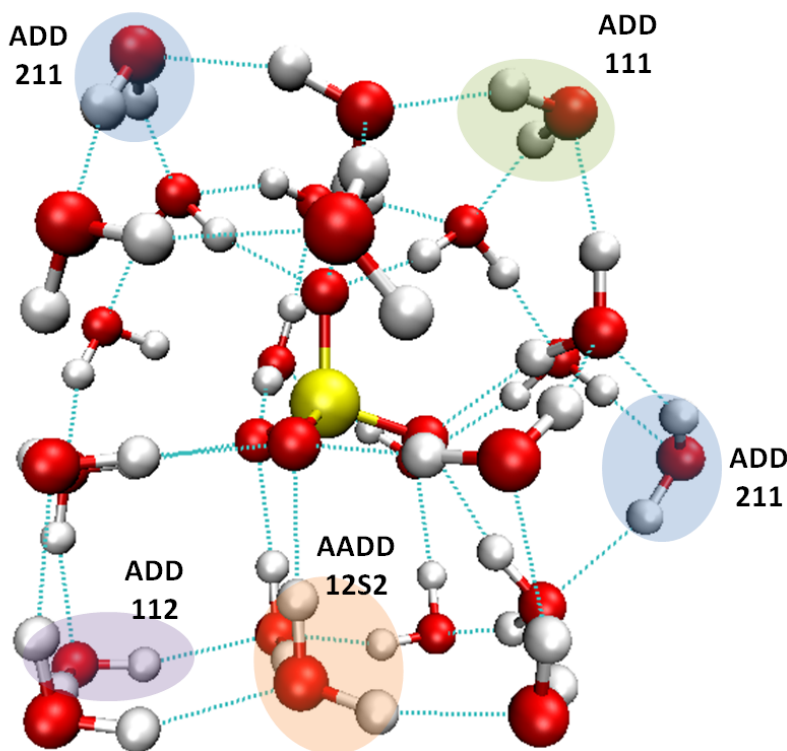
FIGURE 6.24 – Spectres statiques aux niveaux M11/b1 + SF 0.945 (à gauche) et ω B97XD/b2 + SF 0.947 (à droite), pour les structures les plus stables de $[SO_4(H_2O)_{24}]^{2-}$, et comparés au spectre expérimental de Williams ~ 130 K (gris) [2].

peu plus à 200 K. Cette tendance est observée sur les spectres DACF (**Fig 6.26**) à 100 et 200 K, dont chacun est une moyenne de 3 simulations de 200 ps. Les temps de simulations ne permettent pas de profondes réorganisations et l'élargissement avec la température est majoritairement dû aux écarts à la géométrie optimisée, bien plus qu'à un mélange de structures. A 100 K, les massifs sont découpés en 2, 3, parfois 4 massifs, qui décrivent des localisations assez systématiques des fréquences des vibreurs de 1^{ère} ou 2^{ème} sphère. Avec la température, les différences d'environnement se réduisent, et les massifs s'élargissent un peu vers les basses fréquences.

Les spectres des deux structures de plus basse énergie W24-1 et W24-2 ont été attribués par des simulations de DMD (**Fig 6.27**), dans les conditions validées pour $n = 36$ et 80. Pour les 2 structures, les bandes de plus basses fréquences (≤ 3325 cm^{-1}) sont attribuées à des élongations de molécules de 1^{ère} sphère vers la 2^{ème} sphère. Comme en DFT, les deux types de vibreurs (22S2) et (12S2) résonnent dans la même gamme de fréquences (**Tab 6.16**). Les bandes de plus hautes fréquences (≥ 3500 cm^{-1}) sont attribuées majoritairement à des molécules de 2^{ème} sphère. On y retrouve plus de molécules (211) que (111), même si les gammes de résonances s'homogénéisent, en partie par une évaluation moins fine des interactions par le champ de forces que par la DFT, et en partie par les phénomènes de dynamique, qui éloignent les vibreurs de leur longueur de référence et diminuent les différences d'environnements. Pour W24-2, une résonance à 3593 cm^{-1} est relevée, attribuée à des élongations de molécules de 1^{ère} sphère vers le sulfate, et qui peut être corrélée avec les mouvements concertés observés en DFT.

TABLE 6.14 – Population de Boltzmann de $[SO_4(H_2O)_{24}]^{2-}$ en fonction du niveau de théorie, de la propriété thermodynamique, et de la température. Seules les contributions $\geq 2\%$ de la population totale sont données.

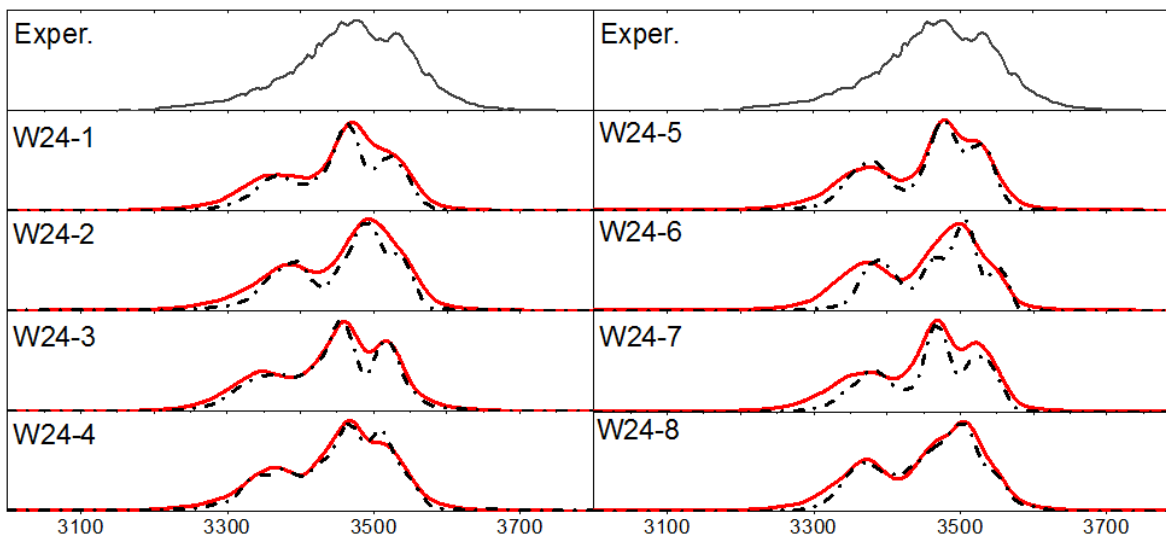
Structures	M11					ω B97XD				
	ΔE			ΔG		ΔE			ΔG	
	1 K	130 K	298 K	130 K	298 K	1 K	130 K	298 K	130 K	298 K
W24-1	100	99	80	99	87	100	69	41	25	20
W24-2			7		2					
W24-3			6		8					7
W24-4			4							
W24-5							24	26	24	20
W24-6								10	19	18
W24-7								9	19	18
W24-8								10	10	13

FIGURE 6.25 – Identification de molécules d'eau sur la structure W24-1, en 1^{ère} ((12S2), orange) ou 2^{ème} sphère ((111), vert, (211), bleu, et (112), violet).

Gamme des empreintes digitales Des spectres DACF ont été réalisés à 50 K à partir des 8 structures de basse énergie. Avec l'augmentation de la taille de la structure et l'élargissement naturel des bandes, les différences entre les spectres des 8 structures deviennent minimales, et seul l'élargissement du massif du pliage \widehat{HOH} témoigne des différences d'environnement. En **Fig 6.28**, une moyenne de ces spectres est comparée au spectre expérimental de Zhou et al. [15]. Le massif entre $600-800\text{ cm}^{-1}$ et l'élongation antisymétrique S=O sont bien reproduits. La surestimation de la fréquence du pliage de l'eau pour les molécules de 1^{ère} sphère perd en importance avec l'augmentation de la taille de l'agrégat. Les 3 petits massifs autour de 1400 cm^{-1} n'apparaissent pas expérimentalement pour les autres tailles d'agrégats [15], et ne sont observés ni par la mécanique quantique, ni par la dynamique moléculaire.

TABLE 6.15 – Attributions des bandes de $[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$ au niveau $\omega\text{B97XD}/\text{b2}$ sur la base des fréquences de W24-1 (en cm^{-1}).

Type de vibrateur	Fréquences $\omega\text{B97XD}/\text{b2}$	Type de vibrateur	Fréquences $\omega\text{B97XD}/\text{b2}$
(211)	3589, 3553, 3546	(112)	3358, 3355, 3344
(211)	3544, 3524	(22S2)	3358, 3355, 3344
(111)	3562	(22S2)	3260, 3229
concerté vers sulfate	3555, 3547	(12S2)	3289, 3284
(112)	3537	(12S2)	3280, 3229

FIGURE 6.26 – Spectres DACF à 100 K (noir) et 200 K (rouge) des structures les plus stables de $[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$, comparés au spectre expérimental de Williams ~ 130 K (gris) [2].

6.5.2 $[\text{SO}_4(\text{H}_2\text{O})_{23}]^{2-}$ et $[\text{SO}_4(\text{H}_2\text{O})_{25}]^{2-}$

Pour $n = 23$ et 25 , seuls les résultats utilisant la fonctionnelle ωB97XD sont présentés. La recherche de structures stables pour $n = 23$ s'est appuyée sur 5 générations successives de REMD, qui ont toutes fait apparaître des structures compétitives après optimisation DFT, ce qui justifiera au minimum une génération REMD supplémentaire. Les 8 structures de plus basse énergie sont présentées en **Fig 6.29** et identifiées selon leurs énergies relatives au niveau $\omega\text{B97XD}/\text{b2}$, incluant la correction de ZPE (**Tab 6.17**). Au niveau $\omega\text{B97XD}/\text{b3}/\omega\text{B97XD}/\text{b2}$, les structures W23-1 et W23-2 sont isoénergétiques; elles ne diffèrent que de quelques réorganisations de liaisons hydrogènes. Dans ces deux structures, de même que dans W23-5 et W23-8, on relève un motif à deux cycles, respectivement de 4 et 5 liaisons hydrogènes (en vert, **Fig 6.29**), également observé sur plusieurs structures de $[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$, dont W24-1. Le sens des liaisons hydrogènes au sein de ce motif varie d'une structure à l'autre, mais semble un élément général de stabilité.

Les spectres monoconformationnels des 8 structures sont présentés en **Fig 6.30**, à gauche. Les structures W23-5, W23-6 et W23-7 présentent des bandes très déplacées vers les basses fréquences. Toutes sont attribuées à des élongations vers la 2^{nde} sphère de molécules de 1^{ère} sphère. La structure W23-5 a une bande à 3069 cm^{-1} , de sa seule molécule (AAADD), W23-6 a deux bandes, à 3185 (AADD) et 3211 cm^{-1} (AAADD), et W23-7 deux bandes également, à 3161 (AADD) et 3212 cm^{-1} (AAADD). Pour les structures W23-1 et W23-2, les bandes intenses calculées dans la gamme $3300\text{-}3350 \text{ cm}^{-1}$ sont en bon accord avec la petite bande expérimentale (bande orange sur la **Fig 6.30**) et correspondent à des combinaisons d'élongations de molécules de 1^{ère} sphère vers la 2^{nde} sphère, et d'élongations de molécules de 2^{ème} sphère. Seules les structures W23-1, W23-2 et W23-3 (et dans une

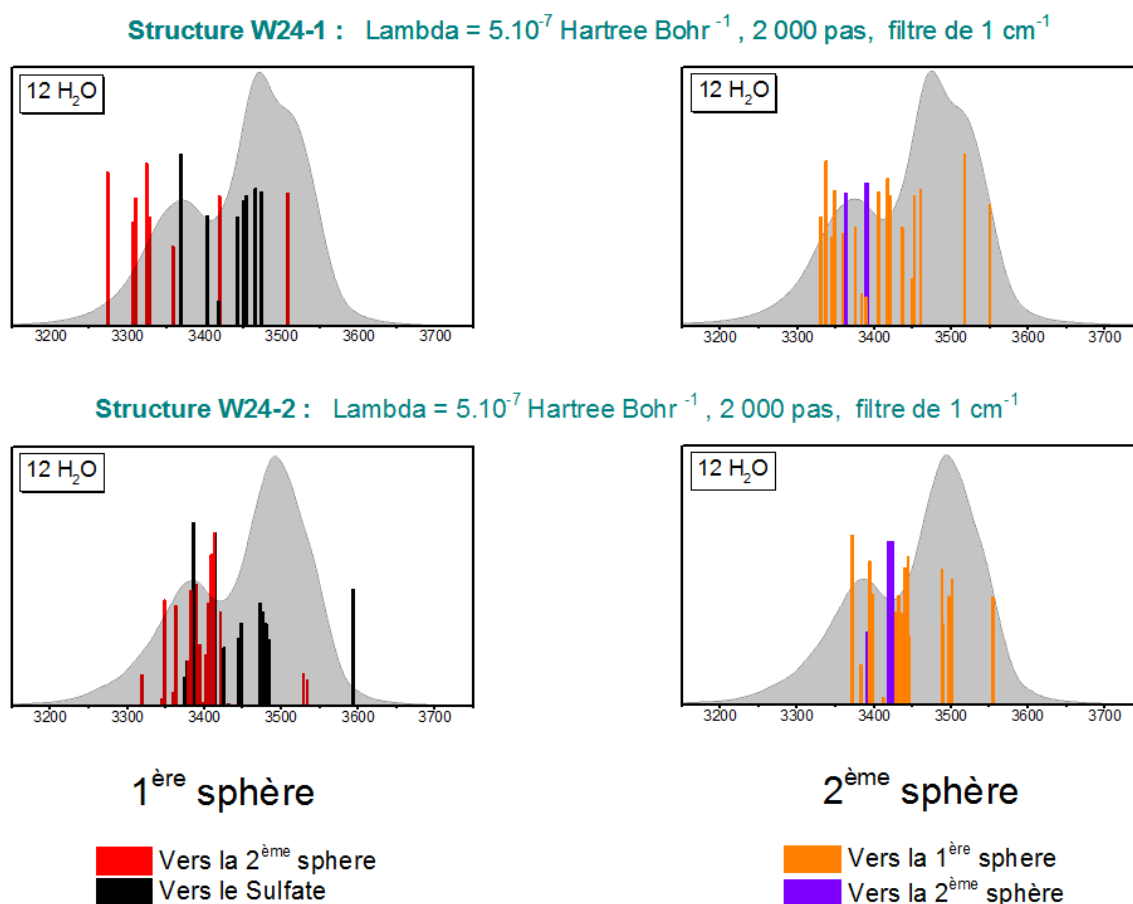


FIGURE 6.27 – Attribution dynamique de modes normaux pour $[SO_4(H_2O)_{24}]^{2-}$ par DMD à partir des 2 structures W24-1 et W24-2. Les fréquences de résonance DMD sont superposées au spectre DACF à 200 K (gris). Les bandes identifiées d'un * sont attribuées dans le **Tab 6.16**.

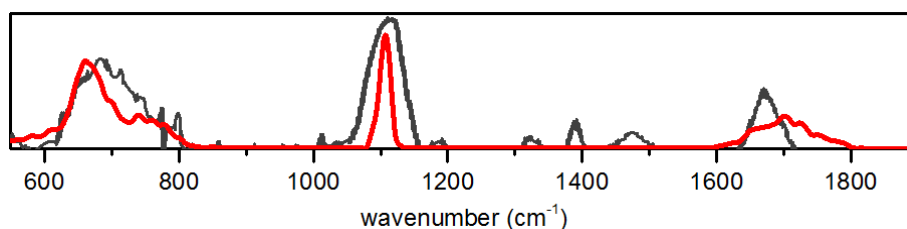


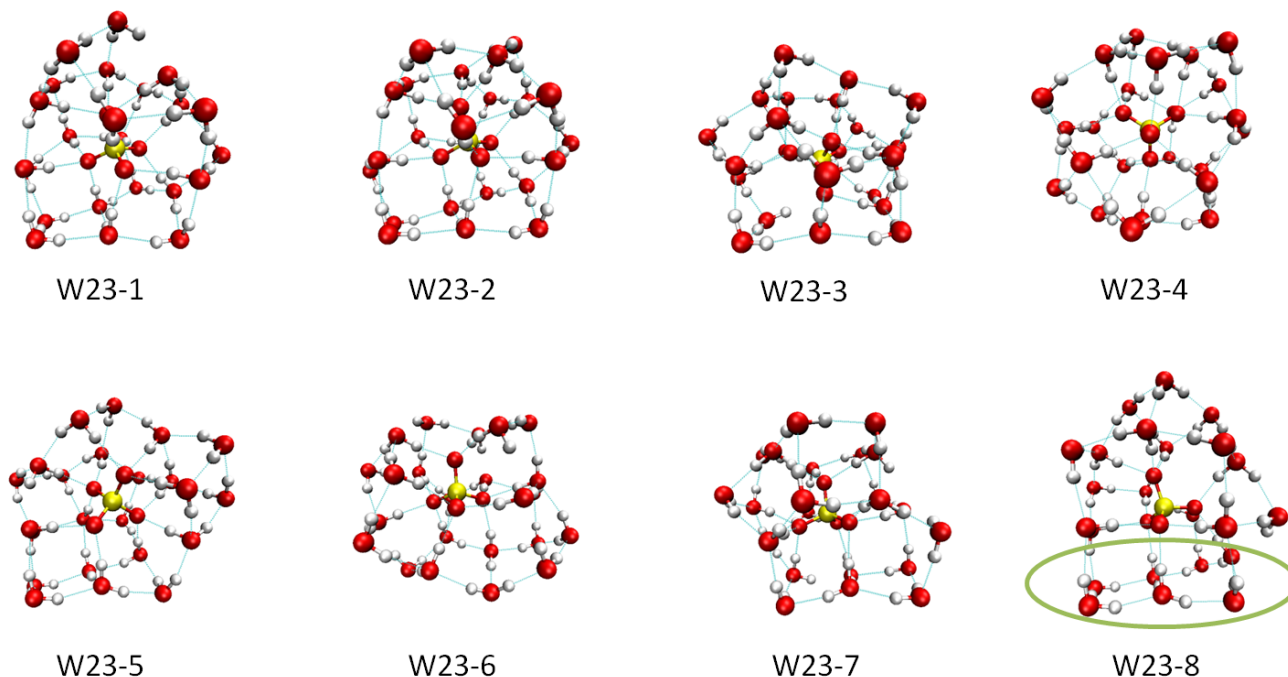
FIGURE 6.28 – Spectre DACF-AMOEBa de $[SO_4(H_2O)_{24}]^{2-}$ à 50 K à partir d'une moyenne des spectres monomériques des 8 structures (rouge), comparé au spectre expérimental à 17 K (noir) publié dans [15].

moindre mesure W23-4) sont capables de reproduire les 3 bandes expérimentales caractéristiques (bleue, verte et orange) tout en ne proposant pas de bande supplémentaire en dessous de 3250 cm $^{-1}$. Sur la base des spectres, cette famille de structures est la meilleure candidate pour attribuer le spectre expérimental. En incluant les corrections thermodynamiques aux énergies relatives (**Tab 6.18**), les structures W23-1 et W23-2 participent pour plus de 50 % de la population totale, pour toutes les températures, et cette participation monte à 80 % en incluant les structures W23-3 et W23-4.

Pour $[SO_4(H_2O)_{25}]^{2-}$, la recherche de structures a été réalisée par des simulations de dynamique AMOEBa, à partir de points de départ très différents. La structure de plus basse énergie W25-1 (**Fig 6.31**) présente des similarités avec la structure W24-1, en particulier la présence du motif de 2 cycles à 4 et 5 liaisons hydrogènes,

TABLE 6.16 – Attributions des spectres de $[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$ en DMD sur la base des fréquences de W24-1 et 24-2 (en cm^{-1}).

Nomenclature	Fréquences W24-1	Fréquences W24-2
(211)	3550, 3517, 3516	3501, 3497, 3489
(111)	3460	3554, 3488, 3445
...
(22S2)	3275, 3326	3378
(12S2)	3308, 3311, 3330	3344, 3348, 3359, 3363
(22S1)		3319

FIGURE 6.29 – Structures de basse énergie de $[\text{SO}_4(\text{H}_2\text{O})_{23}]^{2-}$.TABLE 6.17 – Energies relatives des structures de $[\text{SO}_4(\text{H}_2\text{O})_{23}]^{2-}$ (en $\text{kJ}\cdot\text{mol}^{-1}$), incluant la correction de ZPE. La colonne 'Combinaisons' comptabilise le nombre de structures engendrées à partir d'inversions multiples (jusqu'à 4 inversions de cycles simultanées).

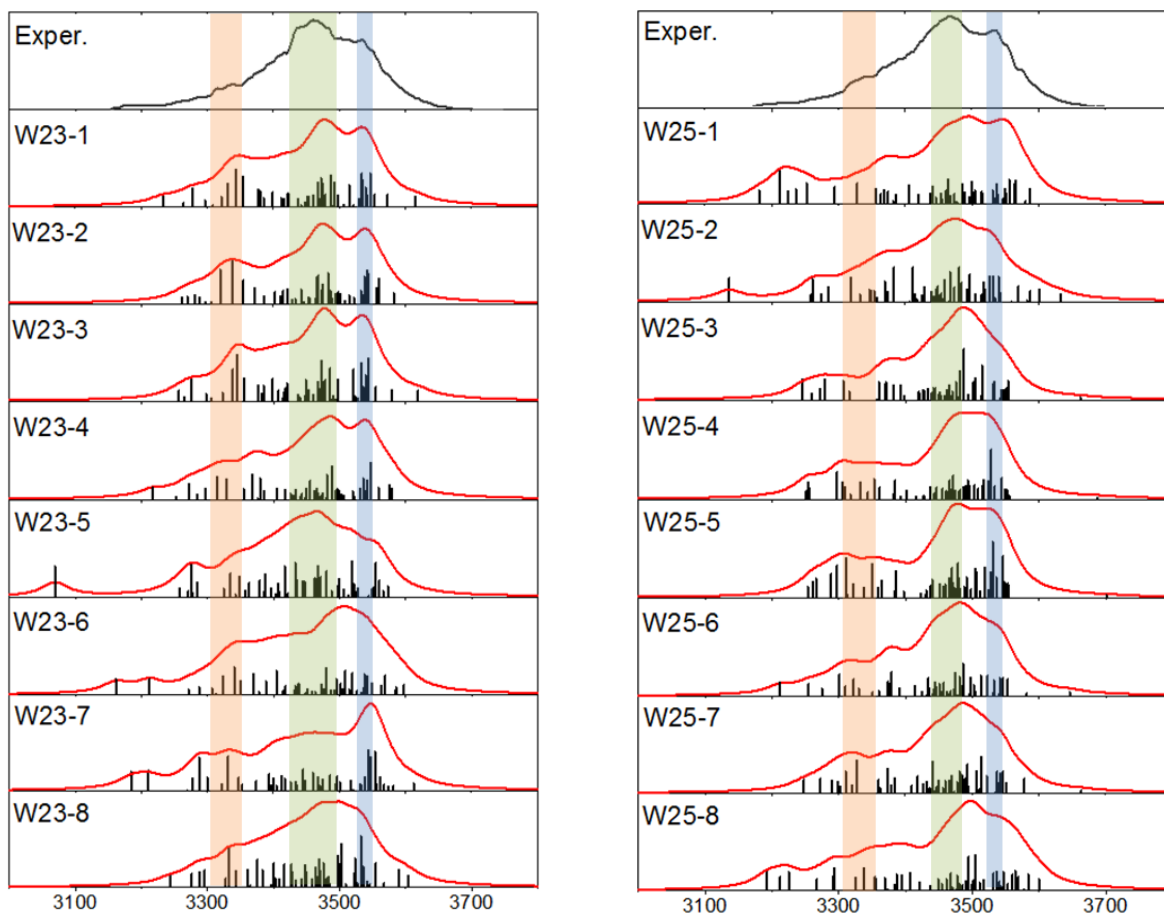
Structures	$\omega\text{B97XD/b2}$	$\omega\text{B97XD/b3//}$ $\omega\text{B97XD/b2}$	Cycles / structure	Combinaisons
W23-1	0.0	0.0	18	56
W23-2	1.2	0.0	18	65
W23-3	2.1	2.3	18	56
W23-4	2.4	1.4	22	63
W23-5	5.4	4.1	41	57
W23-6	6.8	4.3	16	53
W23-7	8.8	6.1	17	21
W23-8	8.8	6.7	13	41

b2 : cc-pVTZ, b3 : aug-cc-pVTZ

bien que la plupart des cycles soient inversés entre les deux structures. Cette observation renforce notre hypothèse que certaines des structures les plus stables de $n = 25$ possèdent une molécule facile à arracher, et conduisent à

TABLE 6.18 – Population de Boltzmann de $[SO_4(H_2O)_{23}]^{2-}$ en fonction de la propriété thermodynamique, et de la température. Seules les contributions $\geq 1\%$ de la population totale sont données.

Structures	ΔE			ΔG	
	1 K	130 K	298 K	130 K	298 K
W23-1	99	43	29	29	24
W23-2		37	27	55	31
W23-3		4	11	4	10
W23-4		13	17	4	10
W23-5		3	9		3
W23-6			3	3	9
W23-7			1	4	9
W23-8			1		3

FIGURE 6.30 – Spectres statiques au niveau $\omega B97XD/b2 + SF 0.947$, pour les structures les plus stables de $[SO_4(H_2O)_{23}]^{2-}$ (à gauche) et $[SO_4(H_2O)_{25}]^{2-}$ (à droite), et comparés au spectre expérimental de Williams ~ 130 K (gris) [2].

des structures de type W24-1, à partir d'un minimum de réorganisations. Nous avons donc cherché à construire une telle structure, à partir de simulations de dynamique AMOEBA, avec pour point de départ une structure construite à partir de W24-1 à laquelle nous avons ajouté une molécule d'eau en surface de l'agrégat, loin de ce motif caractéristique. Des simulations à 100 et 200 K ont permis à la molécule d'eau de s'insérer dans le réseau de molécules d'eau. Nous avons réalisé une optimisation systématique avec AMOEBA des structures échantillonnées pendant la dynamique, puis au niveau $\omega B97XD/b2$ des quelques structures les plus stables. La plus stable en DFT est la structure W25-2, qui s'avère très compétitive, à moins de $1 \text{ kJ}\cdot\text{mol}^{-1}$ de la structure W25-1 (Tab 6.19). 6

autres structures sont identifiées et ont des profils très différents, notamment par leur nombre de cycles (**Tab 6.19**), mais des énergies dans une gamme de 4 kJ.mol⁻¹.

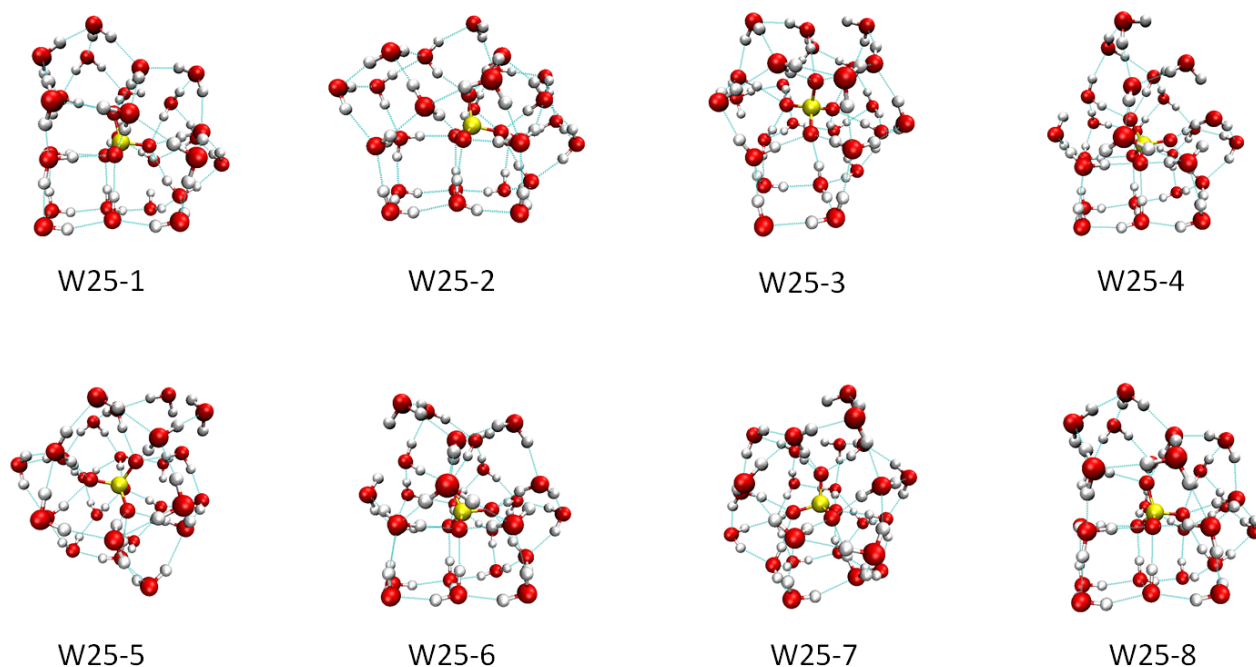


FIGURE 6.31 – Structures de basse énergie de $[SO_4(H_2O)_{25}]^{2-}$.

TABLE 6.19 – Energies relatives des structures de $[SO_4(H_2O)_{25}]^{2-}$ en kJ.mol⁻¹. Les énergies quantiques incluent la correction de ZPE. La colonne 'Combinaisons' comptabilise le nombre de structures engendrées à partir d'inversions multiples (jusqu'à 4 inversions de cycles simultanées).

Structures	ω B97XD/b2	ω B97XD/b3// ω B97XD/b2	Cycles / structure	Combinaisons
W25-1	0.0	0.0	53	34
W25-2	0.4	0.8	27	69
W25-3	1.5	1.1	59	110
W25-4	2.3	1.3	54	75
W25-5	2.5	4.0	59	110
W25-6	3.2	2.4	54	68
W25-7	3.9	3.5	59	108
W25-8	4.6	4.0	33	38

b2 : cc-pVTZ, b3 : aug-cc-pVTZ

Les spectres statiques au niveau ω B97XD/b2 sont présentés en **Fig 6.30**, à droite. Le spectre IRPD de $[SO_4(H_2O)_{25}]^{2-}$ est très similaire à celui de $[SO_4(H_2O)_{23}]^{2-}$ et les structures sont jugées sur leur capacité à reproduire les 3 bandes caractéristiques. Si aucune ne présente un accord parfait, beaucoup présentent des profils intéressants. Le maximum sous la bande bleue de W24-1 est déplacé de 25 cm⁻¹ vers les hautes fréquences, mais est très bien reproduit par W25-2 ou W25-5. La bande vers 3450-3500 cm⁻¹ (bande verte) est bien reproduite par W25-1, W25-2 et W25-5. Dans la gamme 3300-3350 cm⁻¹, W25-2 et W25-4 possèdent une grande densité de bandes, à même de reproduire la bande expérimentale (bande orange). Aux basses fréquences, l'accord est très bon pour W25-6, tandis que les bandes autour de 3200 cm⁻¹ de W25-1 sont nombreuses et intenses. La bande calculée

à 3150 cm^{-1} pour W25-2 est en dehors de la zone d'enregistrement du spectre expérimental et peut donc tout à fait exister.

En intégrant les corrections thermodynamiques (**Tab 6.20**), le spectre multiconformationnel est un mélange assez homogène des structures présentées, en accord avec la nécessaire description multiconformationnelle pour les attributions.

TABLE 6.20 – Population de Boltzmann de $[SO_4(H_2O)_{25}]^{2-}$ en fonction de la propriété thermodynamique, et de la température. Seules les contributions $\geq 1\%$ de la population totale sont données.

Structures	ΔE			ΔG	
	1 K	130 K	298 K	130 K	298 K
W25-1	100	47	26	11	13
W25-2		23	19	23	18
W25-3		10	14	14	14
W25-4		9	13	34	21
W25-5		4	9	8	11
W25-6		5	10	4	8
W25-7			5	4	8
W25-8			4	1	5

Nous avons calculé les énergies associées à la perte d'une molécule d'eau au sein des agrégats $n = 24$ et 25 , à partir des structures les plus stables de $n = 23, 24$ et 25 . Au niveau $\omega B97XD/b3//\omega B97XD/b2$, cette énergie de liaison est de $-41\text{ kJ}\cdot\text{mol}^{-1}$ pour $n = 25$ et $-43\text{ kJ}\cdot\text{mol}^{-1}$ pour $n = 24$, ce qui ne permet pas de conclure sur les constantes de vitesse de dissociation BIRD. Lorsque toutes les explorations seront complètes, ce calcul devra être effectué à partir des structures les plus stables de chaque taille (et représentatives du mélange), et des agrégats formés par la perte d'une molécule d'eau et après peu de réorganisation.

6.6 Conclusion

L'étude de l'ion sulfate hydraté dans de petits agrégats d'eau a été réalisée sur les séries $n = 9, 12$ et 13 , puis $n = 23, 24$ et 25 , à travers leurs propriétés énergétiques et spectroscopiques.

Plusieurs approches ont été mises en oeuvre, à partir de calculs quantiques statiques en simulant un spectre de mélange sur un critère de Maxwell-Boltzmann, ou par dynamique moléculaire à l'aide du champ de forces polarisable. Notre niveau de calcul quantique de référence DLPNO-CCSD(T)/b3//M11/b1 nous a permis d'identifier, pour les petites tailles, des isomères de basse énergie. Pour $n = 12$ et 13 , ces structures sont capables de décrire la gamme complète des élongations O-H du spectre expérimental avec un très bon accord. Pour $n = 9$, nos efforts pour engendrer des structures de basse énergie par dynamique moléculaire nous rendent confiants dans le fait d'avoir identifié les structures les plus stables. Toutefois, du fait des interactions plus fortes avec le sulfate, la description de son spectre échappe au rayon d'action du modèle harmonique, même s'il a été observé qu'un niveau de théorie plus élevé (RI-MP2/cc-pVTZ) pouvait sensiblement améliorer l'allure des spectres monoconformationnels. L'approche statique anharmonique permet de reproduire plus qualitativement certaines bandes du spectre, même si la question du niveau de calcul et de l'évaluation des intensités reste posée. L'objectif de proposer un spectre de mélange à partir de spectres monoconformationnels est atteint. La dépendance au niveau de calcul est évidente, celle à la propriété visée (énergie ou enthalpie libre), est de première importance pour $n = 9$, mais beaucoup moins pour $n = 12$ et 13 . Ce constat de dépendance au niveau de calcul est encore plus marqué pour les plus grandes tailles, $n = 23-25$, pour lesquelles le choix de la DFT influe beaucoup sur les ordres énergétiques des conformères. Pour ces tailles, la fonctionnelle M11 a été abandonnée au profit de la fonctionnelle $\omega B97XD$, et les bases d'optimisation et de calcul d'énergie ont été homogénéisées. Face à ces contraintes, il devient difficile de se fier à une population

de Boltzmann basée sur des énergies DFT, même si la prise en compte des corrections thermodynamiques reste essentielle pour discuter des propriétés de mélange. La prise en compte des possibilités d'inversions de cycles en tant que contribution d'entropie de mélange est limitée aux plus petites tailles, non pas parce qu'à plus grande taille, toutes les structures ont un nombre de cycles équivalents, mais parce que l'incertitude sur les énergies DFT rend inutile ces tentatives d'ajustement fin des populations de Boltzmann. Toutefois, cette information reste pertinente, même à ces tailles, représentant un des seuls moyens d'identifier les similitudes entre structures.

Les attributions ont confirmé que la bande observée sur les spectres de Williams et coll. pour $[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$, $n = 9-13$ (**Fig 6.1**) à $\sim 3100 \text{ cm}^{-1}$ est une contribution de molécule(s) (AADD). Cette attribution confirme que W9-2 est une structure importante pour décrire $[\text{SO}_4(\text{H}_2\text{O})_9]^{2-}$, et qu'il est nécessaire d'inclure les corrections thermodynamiques aux énergies à 0 K pour décrire un spectre de mélange à température finie. Nous avons pu observer que les molécules de 2^{ème} sphère de $[\text{SO}_4(\text{H}_2\text{O})_{12}]^{2-}$ vibrent sous le massif principal, entre 3400 et 3600 cm^{-1} et que l'augmentation de l'intensité de ce massif avec la taille de l'agrégat peut être attribuée au peuplement progressif de la 2^{ème} sphère de solvation. Des calculs sur $n = 11$ sont en cours pour confirmer les attributions dans la gamme 3200-3400 cm^{-1} (bandes orange et rouge du spectre IRPD).

L'exploration de la SEP par REMD, et les calculs DFT ont permis d'identifier une famille de structures stables pour $[\text{SO}_4(\text{H}_2\text{O})_{23}]^{2-}$, dont les spectres sont en très bon accord avec le spectre IRPD de Williams. Plusieurs structures de basse énergie ont également été identifiées pour $[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$ et $[\text{SO}_4(\text{H}_2\text{O})_{25}]^{2-}$. Parmi elles, la seconde structure la plus stable de $[\text{SO}_4(\text{H}_2\text{O})_{25}]^{2-}$ est une très bonne candidate pour engendrer la structure la plus stable de $[\text{SO}_4(\text{H}_2\text{O})_{24}]^{2-}$. Toutefois, si ces observations confortent les hypothèses basées sur les spectres ESI de Williams et al., d'autres structures sont énergétiquement très compétitives et le spectre expérimental doit être vu comme un spectre multiconformationnel. De plus, si la structure W25-2 a un accord très correct avec le spectre expérimental, le spectre de W24-1 est dans un accord bien moindre.

L'approche dynamique par champ de forces polarisable a montré des limites, majoritairement dues à la paramétrisation des systèmes et modèles. La calibration des paramètres du sulfate pour le nouveau champ de forces de l'eau sera une étape nécessaire. L'étude de petits systèmes contenant des interactions très spécifiques est généralement en dehors du domaine de compétences des champs de force. AMOEBA s'y emploie raisonnablement bien, en partie du fait qu'une grande part de sa paramétrisation est issue de calculs *ab-initio*, ce qui doit être surveillé au vu des nouvelles procédures de paramétrisation [25] plus orientées vers la reproduction de données macroscopiques. Si l'intérêt d'étudier de petits agrégats par dynamique est confirmé, le modèle de charges fluctuantes devra être calibré pour cette application dans le futur. Pour les agrégats de plus grandes tailles, les spectres DACF sont en accord bien meilleur avec l'expérience, en partie grâce au modèle de charges fluctuantes dont la gamme de calibration ($[\text{SO}_4(\text{H}_2\text{O})_n]^{2-}$, $n = 36-80$) est plus proche des tailles intermédiaires que des petites. A ces tailles, l'intérêt de la DMD est éclipsé par la performance de la DFT, mais le bon accord entre les deux méthodes d'attribution valide la méthode d'attribution dynamique pour les tailles supérieures.

Bibliographie

- [1] Prell, J. S. ; O'Brien, J. T. ; Williams, E. R. *J. Am. Soc. Mass. Spectrom.* **2010**, *21*, 800-809.
- [2] Cooper, R. J. ; Williams, E. R. "personal communication", **2016**.
- [3] Thaunay, F. ; Clavaguéra, C. ; Ohanessian, G. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25935-25945.
- [4] Wong, R. L. ; Williams, E. R. *J. Phys. Chem. A* **2003**, *107*, 10976–10983.
- [5] Zhan, C.-G. ; Zheng, F. ; Dixon, D. A. *J. Chem. Phys.* **2003**, *119*, 781-793.
- [6] Blades, A. T. ; ; Kebarle, P. *J. Phys. Chem. A* **2005**, *109*, 8293-8298.
- [7] Mardirossian, N. ; Lambrecht, D. S. ; McCaslin, L. ; Xantheas, S. S. ; Head-Gordon, M. *J. Chem. Theory Comput.* **2013**, *9*, 1368-1380.
- [8] Lambrecht, D. S. ; McCaslin, L. ; Xantheas, S. S. ; Epifanovsky, E. ; Head-Gordon, M. *Mol. Phys.* **2012**, *110*, 2513–2521.
- [9] Riplinger, C. ; Sandhoefer, B. ; Hansen, A. ; Neese, F. *J. Chem. Phys.* **2013**, *139*, 134101.
- [10] Sparta, M. ; Neese, F. *Chem. Soc. Rev.* **2014**, *43*, 5032-5041.
- [11] Leverentz, H. R. ; Qi, H. W. ; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 995-1006.
- [12] Bush, M. F. ; Saykally, R. J. ; Williams, E. R. *J. Am. Chem. Soc.* **2007**, *129*, 2220–2221.
- [13] Wang, X.-B. ; Sergeeva, A. P. ; Yang, J. ; Xing, X.-P. ; Boldyrev, A. I. ; Wang, L.-S. *J. Phys. Chem. A* **2009**, *113*, 5567–5576.
- [14] Lambrecht, D. S. ; Clark, G. N. I. ; Head-Gordon, T. ; Head-Gordon, M. *J. Phys. Chem. A* **2011**, *115*, 11438–11454.
- [15] Zhou, J. ; Santambrogio, G. ; Brümmer, M. ; Moore, D. T. ; Woste, L. ; Meijer, G. ; Neumark, D. M. ; Asmis, K. R. *J. Chem. Phys.* **2006**, *125*, 111102.
- [16] Zhao, Y. ; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *120*, 215–241.
- [17] Peverati, R. ; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2012**, *14*, 11363-11370.
- [18] Ponder, J. W. ; Wu, C. ; Ren, P. ; Pande, V. S. ; Chodera, J. D. ; Schnieders, M. J. ; Haque, I. ; Mobley, D. L. ; Lambrecht, D. S. ; DiStasio, R. A. ; Head-Gordon, M. ; Clark, G. N. I. ; Johnson, M. E. ; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- [19] Laury, M. L. ; Wang, L.-P. ; Pande, V. S. ; Head-Gordon, T. ; Ponder, J. W. *J. Phys. Chem. B* **2015**, *119*, 9423-9437.
- [20] Jaque, S. ; Oomens, J. ; Cimas, A. ; Gageot, M.-P. ; Rijs, A. M. *Angew. Chem. Int. Ed.* **2014**, *53*, 3663–3666.
- [21] Bloino, J. ; Biczyński, M. ; Barone, V. *J. Phys. Chem. A* **2015**, *119*, 11862-11874.
- [22] Gohaud, N. ; Begue, D. ; Darrigan, C. ; Pouchan, C. *J. Comput. Chem.* **2005**, *26*, 743–754.
- [23] Bégulé, D. ; Baraille, I. ; Garrain, P. A. ; Dargelos, A. ; Tassaing, T. *J. Chem. Phys.* **2010**, *133*, 034102.

- [24] Bush, M. F.; Saykally, R. J.; Williams, E. R. *J. Am. Chem. Soc.* **2008**, *130*, 15482–15489.
- [25] Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. *J. Phys. Chem. B* **2013**, *117*, 9956–9972.

Annexe D

Article

Hydration of the sulfate dianion in cold nanodroplets : $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$ and $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$



Cite this: *Phys. Chem. Chem. Phys.*,
2015, 17, 25935

Hydration of the sulfate dianion in cold nanodroplets: $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$ and $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}^\dagger$

Florian Thauay, Carine Clavaguéra and Gilles Ohanessian*

The structures, energetics and infrared spectra of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$ and $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$ have been investigated by a combination of classical polarizable molecular dynamics and static quantum chemical calculations. Snapshots extracted from MD trajectories were used as inputs for local DFT optimization. Energies of the most stable structures were further refined at the *ab initio* level. A number of new low energy structures have thus been identified. The most stable structures of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$ have the sulfate on the surface of the water cluster, while it may be slightly more buried in $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$, however still with an incomplete first hydration shell. Differences in the infrared spectra arise in part from mixing of sulfate stretching and water librational modes in the 900–1100 cm^{-1} region, leading to some sensitivity of the IR spectrum to the structure. Second shell water molecules however do not generate signatures that are specific enough to relate spectra to structures straightforwardly, at least in this frequency range. Thus the emergence of a new band at 970 cm^{-1} in the $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$ spectrum cannot be taken as a clue as to the number of water molecules which is necessary for a cluster to close the first hydration shell of sulfate. This number is at least 14 and possibly larger. However the density of low energy isomers is large enough that individual structures may lose meaning at all but the lowest temperatures.

Received 1st May 2015,
Accepted 15th June 2015

DOI: 10.1039/c5cp02557e

www.rsc.org/pccp

1. Introduction

Hydrates of the sulfate dianion SO_4^{2-} are found to be key components in the environment as well as in several biological contexts. They are present in the atmosphere as aerosol particles.^{1,2} Field measurements show that nucleation rates correlate with the concentration of sulfuric acid, eventually leading to the formation of cloud condensation nuclei.³ Aerosols of hydrated sulfate also play a key role in the homogeneous nucleation of ice particles in the upper troposphere.⁴ In biological context, SO_4^{2-} is an important cell nutrient and the fourth most abundant anion in mammalian plasma. However as a hydrophilic anion, it requires a transporter inward and outward of cells. Such a mechanism is regulated by the competition between ion hydration and capture by a macromolecule.⁵ Thus a detailed understanding of sulfate hydration and its competition with other media is of high importance in a number of contexts.

A first step towards this understanding is the knowledge of the structure and dynamics of bare sulfate in water. To this end, sulfate–water cluster $\text{SO}_4^{2-}(\text{H}_2\text{O})_n$ in the gas phase are appropriate prototype systems. Thus it is not surprising that these clusters have been the subject of intense research for more than a decade. While most early studies focused on small clusters in

the $n = 3$ –6 range, data have now grown to encompass n ranging from 3 to 80⁶ and even *ca.* 250.⁷

The gaseous sulfate dianion is unstable against electron detachment.^{8–10} It has been shown that stabilizing the dianion requires three water molecules, *i.e.* $\text{SO}_4^{2-}(\text{H}_2\text{O})_3$ is the smallest stable sulfate hydrate in the gas phase.^{10,11} Small clusters in the $n = 3$ –7 range have been extensively characterized.^{12–21} In particular computational investigations by Head-Gordon *et al.* have provided both a comprehensive picture of the low energy structure manifold¹⁹ and a thorough calibration of the performances of quantum chemical methods.²⁰ Many density functionals were compared to *ab initio* results at the current accuracy limit.²¹

There have been several reports of quantum chemical computations for $n = 12$.^{13,14,23,24,27} They were largely limited to the same shell-closing 12+0 structure (named W12-AN below) although Wong and Williams²⁴ found that many structures with varying numbers of molecules in the second shell are expected to lie in a small energy window, and Gao and Liu¹⁴ did report a 9+3 structure (named W12-GL below) which they found to be significantly more stable than 12+0. A recent work reported by Galli *et al.*²⁷ appears to be the first to address the issue of whether or not the first hydration shell of sulfate is closed at $n = 12$. In a Born–Oppenheimer molecular dynamics (MD) study carried out using the PBE and PBE0 functionals, these authors found structures with the sulfate on the cluster surface which were slightly more stable than those invoked previously, with the sulfate in the

Laboratoire de chimie moléculaire, Ecole polytechnique, CNRS,
91128 Palaiseau Cedex, France. E-mail: gilles.ohanessian@polytechnique.edu

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5cp02557e

interior of the water cluster. Comparison between computed and experimental IR spectra showed that both types of structures are compatible with experiment (both infrared (IR) and photoelectron spectra). The authors concluded that both may be formed as a mixture under the experimental conditions used.

A number of experimental techniques have been used to characterize the properties of $\text{SO}_4^{2-}(\text{H}_2\text{O})_n$ in a wide size range. Photoelectron spectra at room temperature of $n = 3\text{--}40$ ²² displayed several bands corresponding to rather different binding energies, and gradual changes with increasing size. In particular, the relative intensity of the low binding energy feature arising from SO_4^{2-} decreases and almost disappears in the large clusters, while a very intense peak emerges at the high binding energy side. This was interpreted as indicating that electron detachment occurs from the sulfate in the small size regime, while it occurs from the water cluster for larger sizes. While this strongly suggests that the sulfate ion is deeply buried inside the water droplet for the larger clusters, there appeared to be no breaking point which would provide a clear indication of a structural transition in a specific size range.

IR spectra were recorded at low temperature in the $n = 3\text{--}24$ range.²³ Direct comparison with the spectrum in bulk water suggested easy band assignments to either sulfate or water bands. Splittings observed for specific sizes between the sulfate antisymmetric stretching bands were used to infer structure information. In addition, the appearance of a new band in the $n = 13$ spectrum was taken as an indication that the first hydration shell may be complete at $n = 12$, with the 13th molecule bound in a different manner, most probably in the second shell. Earlier BIRD experiments,²⁴ whereby the rate of water evaporation can be determined for each cluster size, did indicate a break at $n = 12$, with a similar interpretation. Finally, high pressure mass spectrometric experiments^{25,26} yielded size-selective binding enthalpies and entropies. These results showed a relatively uniform trend with size, with no specific properties for $n = 12$. However the authors noted that the populations of structures undergoing dissociation may be different if BIRD experiments involve non-Boltzmann distributions. The temperature dependence of isomer populations has also been pointed out for smaller clusters in the discussion of photoelectron spectra.¹⁸ Thus the question of the existence of a structural transition between $n = 12$ and 13 remains partly unsettled from the experimental point of view.

In the present paper we reconsider the structures of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$ and $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$ from a computational viewpoint and attempt to provide answers to the following questions: (i) is there a shell-closing transition for these clusters, and (ii) how structure-telling are IR spectra in the fingerprint region in this size range?

2. Computational details

The potential energy surfaces for $n = 12$ and 13 were explored using a combination of molecular dynamics (MD) using a classical force field and of quantum chemical calculations. MD simulations were carried out using the polarizable AMOEBA

force field (using the AMOEBA03 parameters for water as available in the Tinker software³¹ and the sulfate parameters from ref. 19). They were propagated from several starting points chosen to ensure both structural and energetic diversity. Snapshots extracted from these trajectories at regular intervals were then used as starting points for local geometry optimization using density functional theory (DFT), followed by vibrational spectrum calculations in the harmonic approximation. Final energetics were refined at the *ab initio* level. The details of these procedures are detailed below.

Simulations for $n = 12$ started from 3 structures: the most often invoked in the literature^{13,14,23,24,27} has all water molecules, arranged in four $(\text{H}_2\text{O})_3$ clusters, interacting with the sulfate oxygens thereby forming a closed first hydration shell. With no H_2O in the second shell, this structure is thus described as 12+0. A second structure from the literature,¹⁴ described as more stable than the previous, in which three $(\text{H}_2\text{O})_4$ clusters surround the sulfate ion. Each has one H_2O in the second shell, making this a 9+3 type structure. (iii) Another 12+0 structure manually generated to be less symmetrical than the previous, in which the sulfate ion is closer to the cluster surface, although the total numbers of sulfate–water (12) and water–water (12) hydrogen bonds are the same. From each of these 3 structures, MD trajectories were propagated using the Beeman integrator, and a time step of 1 fs. The temperature was first raised from 0 to 200 K in the *NVT* ensemble for 2 ps, using the Nosé-Hoover thermostat. The trajectory was then propagated for 3 ns at 200 K in the *NVE* ensemble. Snapshots were extracted at 100 ps time intervals. This time step is much larger than that for any relevant vibrational mode. We expect that exchange between conformations may require a number of vibration periods, thus selecting snapshots with a much smaller time interval may generate a large fraction of redundancies. In order to scan the energy surface extensively, a large time interval may be more efficient. On the other hand, the value taken here is probably too large to ensure a thorough identification of all low energy structures. Rather, our aim here is to identify all low energy *structural families*. Snapshots at 200 ps intervals were then subjected to geometry optimization using the M11 density functional and the 6-311++G(d,p) atomic basis set; M11 has been shown to be accurate on relative energies for $n = 3\text{--}7$ by Head-Gordon *et al.*²¹ Whenever two consecutive optimized structures were within 10 kJ mol⁻¹ of the lowest in energy so far, an additional calculation was run using the intermediate snapshot from MD, thus reducing the time interval from 200 to 100 ps. A total of 71 DFT calculations were run in this manner.

Simulations for $n = 13$ followed essentially the same lines as for $n = 12$. Starting structures were manually constructed by adding one H_2O to previously obtained structures of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$ from four different families: three structures were used as starting points for $n = 12$ and the most stable 9+3 type structure found at the *ab initio* level. This ensures structural diversity as well as some very low energy starting points. The exact same procedure was used throughout as described above for $n = 12$, except for using 4 starting structures instead of 3 and propagating each trajectory for 4 instead of 3 ns. This generated a total of 99 DFT calculations.

Some considerations of the efficiency of the exploration procedure are provided in the ESI,† see Section S-1.

As described previously for clusters with $n = 3-7$,¹⁹ the energy ordering is modified by inclusion of zero-point vibrational energy (ZPE). Thus all results below include ZPE computed at the M11 level. These are however not corrected for entropy effects since experiments in ref. 23 were carried out at 17 K, where they are negligible. Computations indicate that entropy effects would be significant at room temperature.

Vibrational spectra were computed at the M11/6-311++G(d,p) level. Sulfate frequencies, especially those of the antisymmetric group of bands of central interest in this work, turned out not to require scaling. This has been observed before with other density functionals²⁸ and also for the phosphate vibrational frequencies.²⁹ It would be conceivable to use a standard scaling for water bands and none for sulfate bands, however the 900–1100 cm^{-1} region was found to involve significant mixing between some water libration and sulfate stretching motions. Therefore it appeared more appropriate to use no scaling for any band.

Considering the rather complex manifolds of low energy structures, it was deemed important to obtain final energetics at more accurate levels than M11/6-311++G(d,p). To this end, energetics at the M11 geometries were obtained at a number of levels, including M11/aug-cc-pVTZ, MP2, CC2 and the recently developed Domain-based Local Pair Natural Orbital approximation to CCSD(T) [DLPNO-CCSD(T)]³⁰ levels. Each of these post-Hartree-Fock levels was combined with both the TZVPP and aug-cc-pVTZ basis sets. All the results are gathered in Tables S1 for $n = 12$ and S2 for $n = 13$ (ESI†). Only those obtained at the highest level, DLPNO-CCSD(T)/aug-cc-pVTZ//M11/6-311++G(d,p), are reported in Tables 1 and 2 and in the text below.

MD calculations were run using the Tinker 6 package.³¹ DFT calculations were done using the Gaussian09 package.³² MP2 and CC2 wavefunctions were computed using the Turbomole 6.6 package.³³ DLPNO-CCSD(T) wavefunctions were computed using the Orca 3.0.3 package.³⁴

Table 1 Relative energies (kJ mol^{-1}) of low energy structures of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$, computed at the DLPNO-CCSD(T)/aug-cc-pVTZ//M11/6-311++G(d,p) level with M11-computed 0 K vibrational energies. Structures with names in bold are displayed in Fig. 1, with IR spectra in Fig. 2

Structure name	Coordination of sulfate oxygens	Relative energy
W12-1a	[3,2,2,2]	0.0
W12-1b	[3,2,2,2]	0.0
W12-1c	[3,2,2,2]	2.1
W12-1d	[3,2,2,2]	6.2
W12-2	[3,3,3,3]	8.2
W12-3	[3,"3","3",2]	8.5
W12-4	[3,2,2,2]	8.9
W12-5	[3,"3","3",2]	10.0
W12-6	[3,3,2,2]	10.2
W12-7	[3,2,2,2]	10.8
W12-8	[3,"3","3",2]	12.9
W12-9	[3,"3","3",2]	13.1
W12-10	[3,2,2,2]	13.3
W12-11	[3,3,3,2]	17.7
W12-GL	[4,2,2,2]	19.0
W12-AN	[3,3,3,3]	27.0

3. Results

A. Structures and energetics of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$

A total of 71 structure optimizations were carried out at the DFT level for $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$. The low-energy results are summarized in Fig. 1 and 2 and in Table 1. A recurring motif, well known in pure water clusters, is the cyclic tetramer. It bears one dangling OH bond per water molecule, in alternating orientations relative to the oxygen network in the most stable structure of $(\text{H}_2\text{O})_4$. Only slightly higher in energy lies the cyclic tetramer with all dangling OH bonds pointing in roughly the same direction. Among the most stable hydrated sulfate clusters studied herein, this $(\text{H}_2\text{O})_4$ structure is found as a common building block, as it is well adapted to binding to three sulfate oxygens, while the fourth dangling OH can be oriented towards an oxygen of another water cluster (*e.g.* another tetramer). This fourth OH belongs to a water molecule lying in the second hydration shell of the sulfate ion, while the other three are in the first shell. The cluster can also receive a hydrogen bond from a third water cluster on its other side. Putting side to side three such clusters makes the most stable structure of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$, W12-1a as shown in Fig. 1. It is of 9+3 type, indicating that 9 water molecules are in the first and 3 are in the second hydration shell. In addition, each tetramer can accommodate degenerate reversal of its cyclic hydrogen bonding orientation. Interactions with sulfate and with other water clusters make this reversal no longer degenerate in the present case; however energy differences were found to be only a few kJ mol^{-1} , see W12-1b to W12-d. Since hydrogen bonding orientation reversal can occur in a single tetramer or in two or three simultaneously, seven very similar structures can thus be generated for this particular motif.

With a heavy atom skeleton of C_3 symmetry, three of the sulfate oxygens in the most stable structures W12-1a to W12-1d are 2-coordinated and lie at the cluster surface while the fourth oxygen is 3-coordinated in the cluster interior. Thus the coordination shell scheme is 9+3 and the sulfate oxygen coordination pattern is [3,2,2,2]. W12-b to W12-d are all deduced from W12-a by reversal of some of the hydrogen bonds. Two views of W12-1b are shown in Scheme 1; the left view highlights the position of the sulfate ion near the water cluster surface. The right view is a top view along an S–O axis with the interior O in the back. The latter perspective is chosen for all other W12-1 isomers shown in Fig. 1 to highlight the hydrogen bond network.

The first structure with a complete hydration shell around the ion is W12-2. It is made of 3 water trimers, each making a 4-membered ring with one of the sulfate oxygens. This motif is quite stable, probably one of the low energy structures of $\text{SO}_4^{2-}(\text{H}_2\text{O})_9$. W12-2 is completed with a cyclic water trimer donating 3 hydrogen bonds to 3 sulfate oxygens (shown at the top of W12-2 in Fig. 1). This structure is less stable than W12-1 by 8.2 kJ mol^{-1} . It is quite different from, and much more stable than, the shell-completing isomer W12-AN invoked previously as the most stable of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$: the latter is computed to lie 27.0 kJ mol^{-1} higher in energy than W12-1 and 18.8 kJ mol^{-1} higher than W12-2. Another isomer, W12-3, although it is quite different than W12-2, has a nearly complete hydration shell,

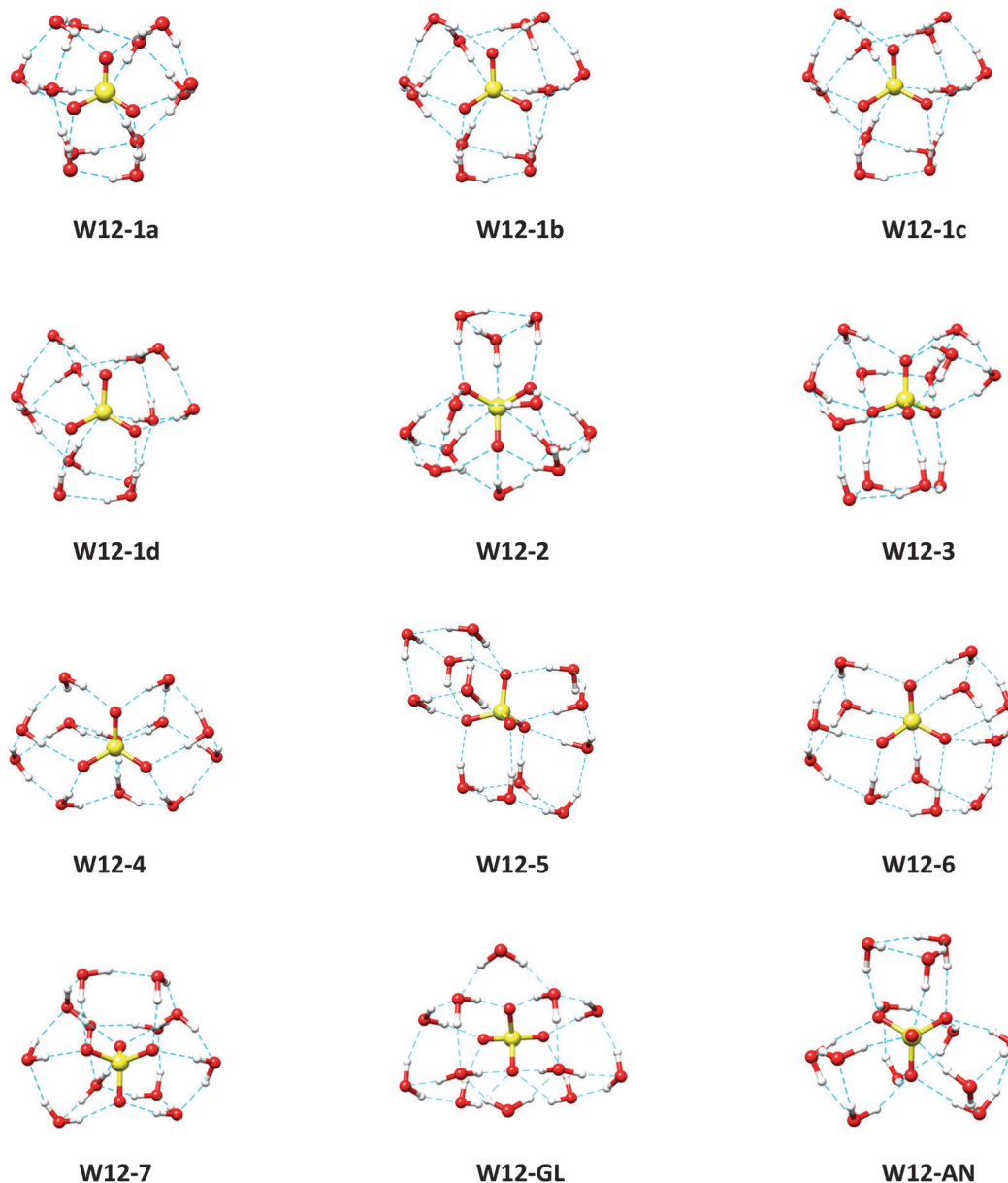


Fig. 1 Structures of low energy isomers of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$. The ordering in “W12-*x*” is based on DLPNO-CCSD(T)/aug-cc-pVTZ//M11/6-311+G(d,p) relative energies, including M11 zero-point vibrational energies. “W12-AN” is from ref. 23.

although a single water molecule bridges between two sulfate oxygens with larger hydrogen bonding distances.

The next most stable structure W12-4 is distorted from W12-1 with only two water tetramers connected in a fused

manner by a $(\text{H}_2\text{O})_5$ cluster. It is also of 9+3 and [3,2,2,2] types. W12-5 has two distorted “cube-like” structures involving five H_2O and the sulfate. The next structures in the energy list (W12-6 and W12-7) have rather similar global shapes, somewhat “2-dimensional” with the sulfate again near the cluster surface. Finally, the 9+3 isomer W12-GL, the most stable found in a previous study,¹⁴ is also included for comparison. It is confirmed to be significantly more stable than W12-AN, however the large difference with W12-1 makes it very unlikely to be observable experimentally. The present search thus leads to a complete reappraisal of the structures and energies of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$, since most of the above structures had not been identified before. The notable exception is the work by Galli *et al.*²⁷ who did find surface-bound sulfate among the low energy structures.



Scheme 1

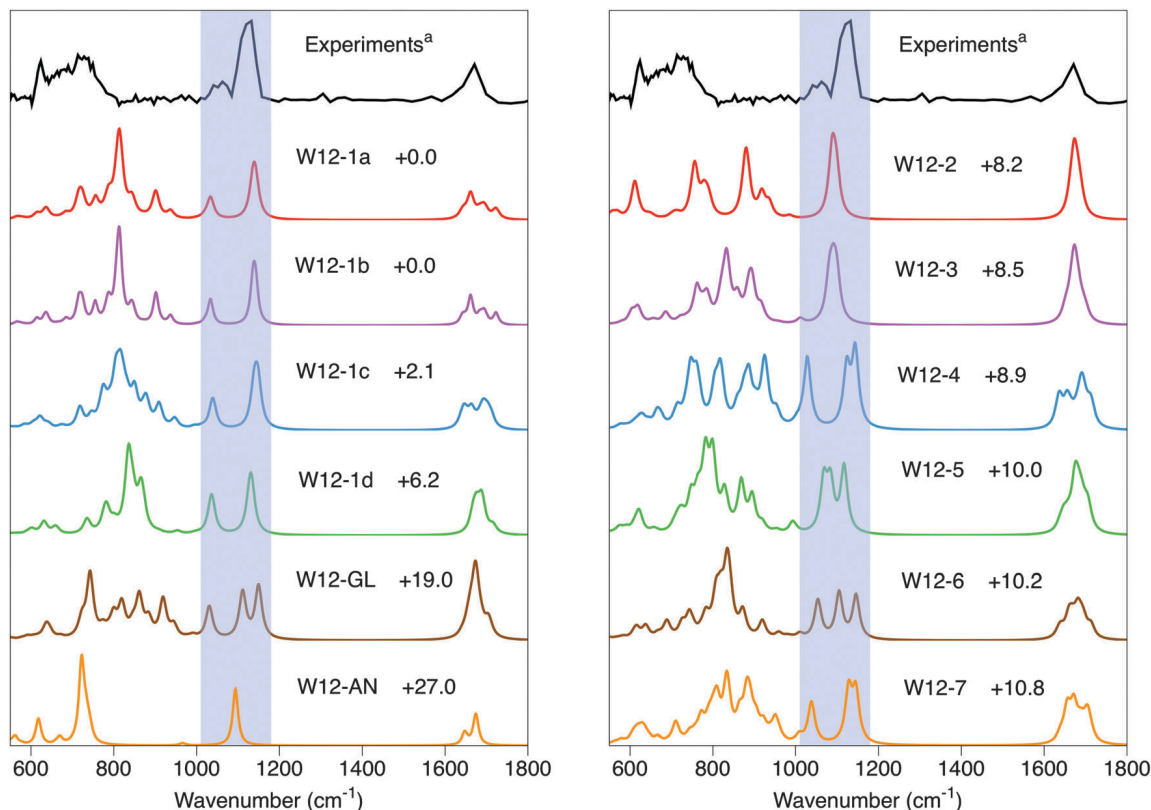


Fig. 2 Experimental IRPD spectrum (ref. 23, top) and computed absorption spectra of several low energy isomers of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$. Structures and spectra are at the M11/6-311++G(d,p) level. Relative energies, in kJ mol^{-1} , are at the DLPNO-CCSD(T)/aug-cc-pVTZ level using the same geometries.

In the most stable structures W12-1a to W12-1d, the water–water hydrogen bond lengths range from 1.86 to 1.95 Å, the shortest are made by water molecules interacting with the sulfate; globally expanded relative to those in the equivalent free water tetramer (1.83 and 1.91 Å in a puckered conformation). The inter-cluster bonds are all near 1.91 Å. The water–sulfate hydrogen bonds are all very similar, ranging from 1.86–1.90 Å. While these values are typical of many other structures, there is some variation associated with the types of water clusters formed around the ion: in the 3-membered rings of W12-AN, they are significantly longer at 1.98 Å because of larger cyclic strain, while the water–sulfate distances of 1.92 Å are only slightly longer than for 4-membered rings.

Very recently, a sophisticated basin-hopping Monte Carlo algorithm for potential energy surface exploration was used to survey the low energy structures of $\text{SO}_4^{2-}(\text{H}_2\text{O})_n$, $n = 3\text{--}50$.³⁵ A rigid-body potential was used to calculate energies, TIP4P for water and fixed atomic charge electrostatic plus 12-6 Lennard-Jones interaction terms for the sulfate ion.³⁶ The low energy structures identified are W12-AN and (similar to) W13-AN for $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$ and $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$, respectively. As described above, these are high energy structures, at least 20 kJ mol^{-1} above the lowest energy structure found herein. This large energy discrepancy is reflected by rather different mean hydrogen bond lengths computed with the force field and with M11/6-311++G(d,p), for the 12+0 structure of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$: ~ 1.87 vs. 1.98 Å for water–water and ~ 1.79 vs. 1.92 Å for sulfate–water hydrogen bonds,

respectively. This force field favors 3-membered water rings interacting with 2 sulfate O's, while we find that 4-membered rings are distinctly more stable, at least in this size range. Thus the use of a rigid-body, non-polarizable potential with simple electrostatics to calculate energies does not appear to provide enough accuracy to identify low energy structures reliably.

B. IR spectra of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$

As discussed in ref. 23, the IR spectra of $\text{SO}_4^{2-}(\text{H}_2\text{O})_n$ are expected to be easy to interpret, based on the known IR spectrum of sulfate in bulk water. The sulfate bending and antisymmetric stretching bands are expected to arise near 600 and 1100 cm^{-1} , respectively, while water libration and bends should appear as wide features centered at around 800 and 1600 cm^{-1} , respectively. This is what is observed in most experimental spectra for $n = 3\text{--}14$, 18 and 24,²³ with water bands rising in width and intensity relative to those of sulfate, for clusters of growing size. Some of the details in band shapes were used as structural fingerprints, especially for structures involving different levels of symmetry breaking in small clusters. As shown below, this simple scenario does not appear to hold in the 900–1100 cm^{-1} range with extensive water–sulfate mixing in some of the normal modes for $n = 12$ and $n = 13$.

Computed spectra are displayed in Fig. 2. The lowest energy structures W12-1a-d have rather similar spectra as expected, except in the water libration massif for W12-1c and W12-1d. Band assignments will therefore be detailed for W12-1a only.

Three intense bands appear in the 1000–1200 cm^{-1} range. The most intense at 1136 and 1143 cm^{-1} correspond to sulfate stretching motions involving primarily the three S–O bonds on the cluster surface. The band at 1034 cm^{-1} is a combination dominated by the libration of several water molecules, with a minor contribution of sulfate stretching arising mainly from the S–O bond pointing towards the water cluster. These three bands account well for the experimental doublet in the 1000–1200 cm^{-1} area with the component to the blue being much more intense. A second combination with the “inside” sulfate stretching as the major component and water libration as the minor component is found at 938 cm^{-1} , to the blue of two more intense, delocalized water libration bands at 901 and 902 cm^{-1} . There are several weak bands in the 600–620 cm^{-1} range with significant components of sulfate bending modes, in addition to water. All of the features above, plus the water bending feature in the 1630–1710 cm^{-1} region, agree rather well with the experimental bands. It is less true for the large feature from 600 to 800 cm^{-1} , for which the computed frequencies appear to be systematically overestimated.

The computed spectra of W12-4 and W12-7 are similar to those of the W12-1 series in the 1000–1200 cm^{-1} range, while the higher IR activity in 900–1000 cm^{-1} range appears to be in less good agreement with experiment. The spectra of W12-2, W12-3 and W12-AN all have a single feature between 1000 and 1200 cm^{-1} , in disagreement with experiment. These 3 structures have complete or nearly complete hydration shells forming identical or similar bonding environments for all sulfate oxygens, leading to three very similar sulfate stretching frequencies. This highlights the inconsistency of the $n = 12$ spectrum with a saturated first shell. The reverse situation occurs for W12-6 in which although the sulfate is on the cluster surface, it has an unsymmetrical coordination motif with one of the surface oxygen bound to three H_2O . This lifts band degeneracy to yield three well separated bands at 1054, 1105 and 1146 cm^{-1} , clearly at variance with the experimental spectrum. For all of these structures the sulfate and water bending frequencies are rather insensitive to structural details, while the water libration group of bands is much more sensitive, however always to the blue of the experimental feature except in the case of W12-AN for which it is well positioned, albeit much too narrow because of the high symmetry.

The results described above do not fit with the shell-closing picture at $n = 12$, both from computed energetics and from comparison of computed and experimental IR spectra. Three S–O bonds of the sulfate are found to lie at the cluster surface in the several most stable structures, with the first 12+0 isomer lying 8.2 kJ mol^{-1} higher in energy than the most stable one. The computed IR spectra also challenge the view that band assignments can be safely made on the basis of those established for the bulk spectrum in the 900–1200 cm^{-1} region.

C. Structures and energetics of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$

A total of 99 structure optimizations were carried out at the DFT level for $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$. The low-energy results are summarized in Fig. 3 and 4 and in Table 2. All structures lying in an energy

Table 2 Relative energies (kJ mol^{-1}) of low energy structures of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$, computed at the DLPNO-CCSD(T)/aug-cc-pVTZ//M11/6-311++G(d,p) level with M11-computed 0 K vibrational energies. Structures with names in bold are displayed in Fig. 3, with IR spectra in Fig. 4

Structure name	Coordination of sulfate oxygens	Relative energy
W13-1a	[3,2,2,2]	0.0
W13-2	[3,2,2,2]	1.2
W13-3	[3,2,2,2]	2.9
W13-4	[3,2,2,2]	3.1
W13-1b	[3,2,2,2]	3.1
W13-1c	[3,2,2,2]	3.1
W13-5	[3,3,3,2]	3.2
W13-1b2	[3,2,2,2]	3.5
W13-6	[3,3,2,2]	3.6
W13-2b	[3,2,2,2]	4.0
W13-1b3	[3,2,2,2]	4.3
W13-1b4	[3,2,2,2]	4.4
W13-1d	[3,2,2,2]	5.1
W13-7	[3,“3”,“3”,2]	5.3
W13-8	[3,2,2,2]	5.7
W13-9	[3,3,3,“3”]	6.1
W13-1d2	[3,2,2,2]	6.3
W13-10	[3,3,3,“3”]	6.3
W13-1e	[3,3,2,2]	6.5
W13-11	[3,2,2,2]	8.3
W13-12	[3,3,2,2]	9.9
W13-13	[3,3,2,2]	9.9
W13-AN	[3,3,3,3]	19.5

window of 10 kJ mol^{-1} are listed in Table 2, and a selection is shown in Fig. 3 and 4.

As for $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$, a number of isomers of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$ were found to be similar in the sense that their oxygen network and connectivity are identical and they differ only by the partial reversal of the hydrogen bonds. Another type of similarity is when the 13th water molecule is added to one of the several sites of a single $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$ core. A particularly clear case is that of W12-1. To this structure, the 13th water molecule can be added either to a “vertical” or a “horizontal” edge of one of the $(\text{H}_2\text{O})_4$ clusters (in the orientation shown in Fig. 3, see W13-1d and W13-1c), or bridging between two such clusters (W13-1a), or bridging between one cluster and a sulfate oxygen (W13-1e), or bridging between two sulfate oxygens (W13-1b). This W13-1 structural family is found to span an energy range of at least 6.5 kJ mol^{-1} (see Table 2).

While W13-1a is the most stable structure found for $n = 13$, there are several other structure types which appear in the lower energy manifold. They are based on the same building blocks described above for $n = 12$. Cubic-like fragments can be seen in most structures among W13-2 to W13-7. Additional water cluster fragments are also found, including a cyclic pentamer. As for the tetramer described above, the pentamer can donate OH's to three sulfate oxygens, however its larger size allows it to donate two and receive one hydrogen bonds with other water clusters. The most stable structures with nearly complete first hydrations shells are W13-9 and W13-10 (not shown in Fig. 3 and 4), 6.1 and 6.5 kJ mol^{-1} higher in energy than W13-1a. Both have a water molecule bridging between two sulfate O's, with bond distances larger than those for a truly saturated case such as W13-AN, the structure previously invoked in ref. 23 whose

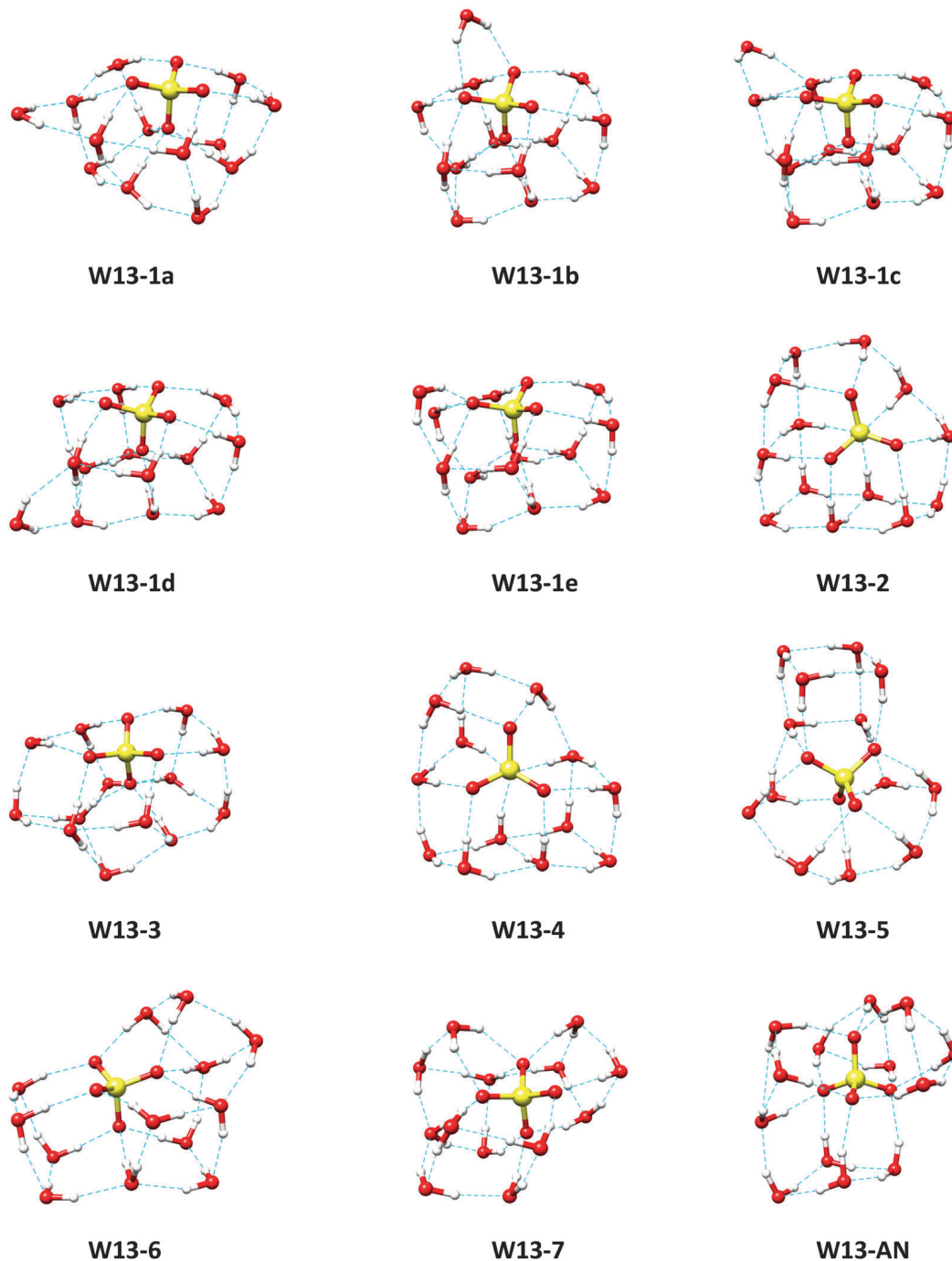


Fig. 3 Structures of low energy isomers of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$. W13-AN is from ref. 23. The ordering in "W13-x" is based on DLPNO-CCSD(T)/aug-cc-pVTZ//M11/6-311++G(d,p) relative energies, including M11 zero-point vibrational energies.

energy is 19.5 kJ mol^{-1} higher than that of W13-1a. Thus this 12+1 structure cannot contribute to the IR spectrum.

Finally, comparison of Tables 1 and 2 shows that the low energy density of isomers is significantly larger for $n = 13$ than it is for $n = 12$.

D. IR spectra of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$

As discussed above for $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$, most of the bands in the IR spectrum of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$ can be readily assigned to specific sulfate or water motions, except in the $900\text{--}1100 \text{ cm}^{-1}$ range

where coupling occurs between the stretching of the inside S–O bond and some water librations.

The lowest energy structure, W13-1a, has a $(\text{H}_2\text{O})_{12}$ core identical to W12-1. Most of their vibrational bands being very similar, only the most relevant features and the main differences will be mentioned below. For instance the pure sulfate antisymmetric stretching bands appear at 1130 and 1153 instead of 1136 and 1143 cm^{-1} , slightly more different because of the symmetry breaking effect of the 13th water molecule. Again the band at 1037 cm^{-1} is a combination

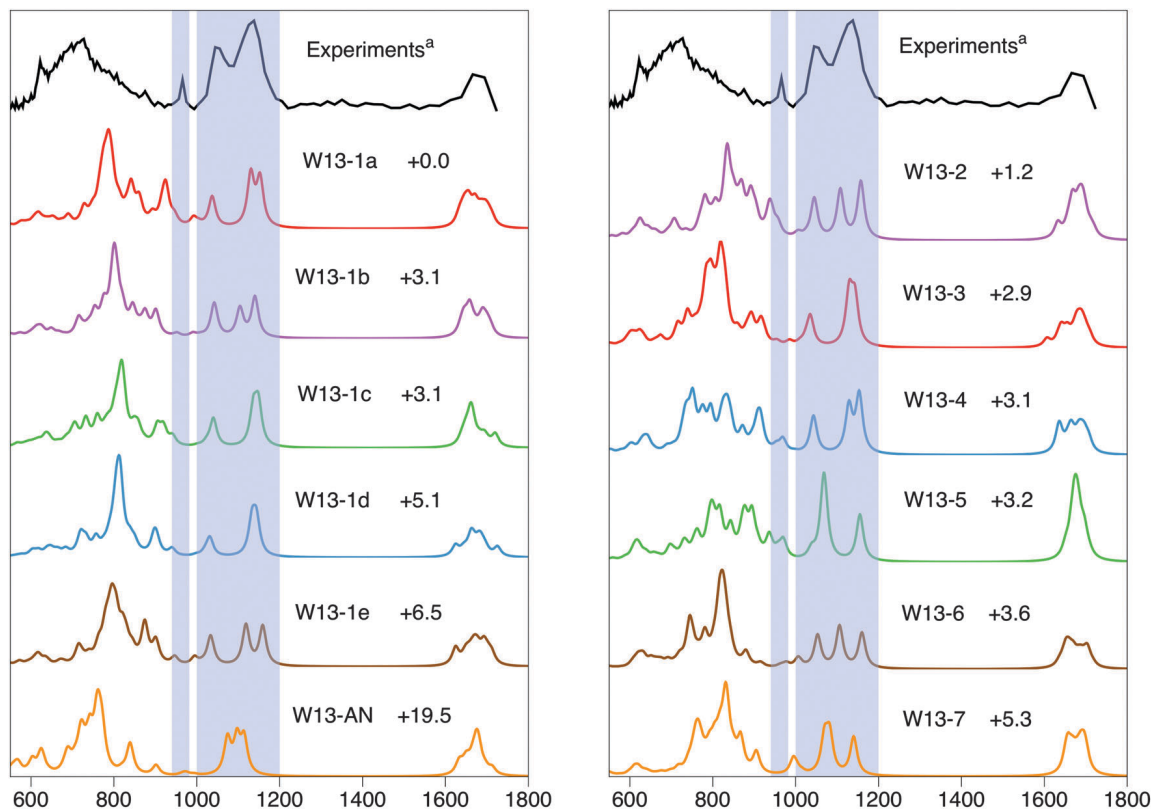


Fig. 4 Experimental IRPD spectrum (ref. 23, top) and computed absorption spectra of several low energy isomers of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$. Structures and spectra are at the M11/6-311++G(d,p) level. Relative energies, in kJ mol^{-1} , are at the DLPNO-CCSD(T)/aug-cc-pVTZ level using the same geometries.

dominated by the libration of several water molecules, with a minor contribution of sulfate stretching arising mainly from the S–O bond pointing towards the water cluster. These three bands account well for the intense experimental features in the $1000\text{--}1200\text{ cm}^{-1}$ area. A second combination with the “inside” sulfate stretching being the major component and water libration minor is found at 946 cm^{-1} . It is our best candidate to account for the experimental band at 970 cm^{-1} . To the blue of the previous band lie two more intense water libration bands at 917 and 925 cm^{-1} , blue-shifted relative to the analogous bands in the spectrum of W12-1. None of the above libration motions involve a significant component from the “13th” water molecule. The most intense is the coupled motion of the two water molecules which are bound to the sulfate and receive hydrogen bonds for the outer water, the latter making their motion more constrained and thus blue-shifted. Finally a less intense band at 992 cm^{-1} arises from coupled librations in one of the water tetramers, to which the outer water is bound. This is a second possible assignment for the experimental band at 992 cm^{-1} . As for $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$, there are several weak bands in the $600\text{--}620\text{ cm}^{-1}$ range with significant components of sulfate bending modes. All of the features above, plus the water bending bands in the $1630\text{--}1710\text{ cm}^{-1}$ region, agree rather well with the experimental bands. It is less true for the large feature from 600 to 900 cm^{-1} , for which the computed frequencies appear to be systematically overestimated.

Structures W13-1b to W13-1e, in which the 13th water molecule occupies different binding sites of a common $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$

core which is essentially the same as W12-1, have similar computed spectra as expected. Some easily discernible differences arise from the local interactions in the various binding sites. Near degeneracy of the surface S–O bond stretch frequencies occurs for W13-1c and W13-1d in which the additional water molecule bind to an edge of a $(\text{H}_2\text{O})_4$ cluster, either horizontal in 1c or vertical in 1d, far away from the sulfate ion. When it bridges between a sulfate and a water O, or between two of the sulfate Os, degeneracy is lifted again (W13-1e and W13-1b). In all cases a sulfate stretch motion combined to some water libration gives rise to a relatively weak band in the $940\text{--}950\text{ cm}^{-1}$ range.

Inspection of Fig. 4 indicates that the spectra computed for the other low energy structures, W13-2 to W13-6, are all relatively similar to those described above for W13-1b to W13-1e. Again the varying degree of degeneracy of the sulfate stretching bands can be traced to the more or less symmetrical environment of the sulfate S–O bonds, both on the surface and in the interior of the water cluster since the structures are more different. The spectra of W13-2, W13-4 and W13-5 all display a sulfate stretching band in the $940\text{--}970\text{ cm}^{-1}$ range.

The spectrum of W13-4 may have the best overall agreement with the experimental spectrum, especially since it displays a band at 969 cm^{-1} and a weaker one at 953 cm^{-1} (not resolved in Fig. 4). Both arise from libration of water molecules in a four-membered ring, coupled to stretching of the interior S–O band. The $(\text{H}_2\text{O})_4$ cluster has a different H bonding pattern to the rest of the water cluster than has W13-1, thus the band shifting.

A discussion on the computed spectra of W13-5 and W13-7 may be found as Section S62 in the ESI.†

Comparison of experimental and computed spectra makes it clear that the shell-closing 12+1 structure W13-AN described in the literature does not fit. A clear indication is provided by the sulfate antisymmetric stretch bands. While the H₂O molecule in the second shell breaks the quasi-*C*₃ symmetry of the SO₄²⁻(H₂O)₁₂ core, the degeneracy lift of the 3 frequencies is clearly insufficient to account for the experimental width of *ca.* 200 cm⁻¹, contrary to those of the structures described above. In addition, the band at 902 cm⁻¹ is in significantly worse agreement with the 970 cm⁻¹ experimental band than those of most structures described above, which lie either in the 940–960 or 990 cm⁻¹ range. Thus the inescapable conclusion is that the first hydration shell is not closed at *n* = 13. Interestingly, the wide water libration massif is somewhat different for W13-AN than for the others displayed in Fig. 4, better matching the experimental features on the red side.

4. Discussion

A. Comparison with previous computations

The present results indicate that nearly all structures published previously for SO₄²⁻(H₂O)₁₂ are of high energy. The single exception is from the work of Galli *et al.*²⁷ with which the present results are qualitatively consistent. There are however considerable quantitative differences, as the more extensive structural search performed herein has allowed identification of a number of new low energy structures. The present results, obtained at the DLPNO-CCSD(T)/aug-cc-pVTZ level, are also expected to be more reliable than the PBE0 results obtained previously. As a consequence, the interpretation²⁷ of IR spectra in terms of a mixture of surface and interior structures being formed experimentally is unlikely to hold, both in terms of energy and of IR spectra. Kinetic trapping of high energy structures such as W12-AN and W13-AN during cryogenic cooling is not expected to occur for non-covalent complexes which can interconvert through small energy barriers. Our results do support a mixture of several structures of very similar stability, especially for *n* = 13, however at very low temperature all have the sulfate near the cluster surface.

It remains that the agreement between the present computations and experimental spectra is unsatisfactory in the 600–900 cm⁻¹ region. BOMD computations by Galli *et al.*²⁷ have shown significant improvement over static calculations in this frequency range. It would be very interesting to run BOMD calculations with the lowest energy structures found herein as starting point. In addition, our picture for *n* = 13, in which the 13th water molecule can move around a compact SO₄²⁻(H₂O)₁₂ core, may be associated with large configurational entropy, which our current approach does not treat. BOMD calculations on such structures may be revealing in this respect.

B. Comparison of IR spectra of SO₄²⁻(H₂O)₁₂ and SO₄²⁻(H₂O)₁₃

The most stable structure for *n* = 13 has lowered symmetry compared to that for *n* = 12 because of the additional water

molecule. This leads to a larger degeneracy breaking of the sulfate antisymmetric stretching frequencies at 1130–1150 cm⁻¹, accounting for the larger experimental feature for *n* = 13, which extends toward 1200 cm⁻¹. Coupling of the sulfate inner S–O stretch with water librations, dominated by those waters interacting with the inner S–O bond, leads to two bands of very similar positions for *n* = 12 and 13. The sulfate-dominated motion has frequencies of 937 and 946 cm⁻¹, respectively, while the water-dominated motion has frequencies of 1034 and 1037 cm⁻¹, respectively. As these motions are remote from the outer water molecule, their similarity is expected. The computed water libration bands for both *n* = 12 and 13 do not reproduce the experimental features very well, since the latter extend down to 600 cm⁻¹ while calculated intensities are rather small below 700 cm⁻¹. This may be due to the static, harmonic approximation used. They do reproduce, however, the overall blue shift of the blue part of the massif for *n* = 13 as compared to *n* = 12, with significantly more numerous bands in the 800–950 cm⁻¹ area. This extends as far as a band at 992 cm⁻¹, corresponding to coupled librations in one of the water tetramers, to which the outer water is bound. Since this band appears in a completely silent portion of the *n* = 12 spectrum, it is one of the two possible assignments for the experimental band at 970 cm⁻¹. In such a scenario, the experimental differences observed in the spectra of *n* = 12 and 13 do arise from a 13th water molecule lying in the second solvation shell, even though it is unrelated to closing the first. In addition the new band at 970 cm⁻¹ is not related to this outer molecule's motion. It is the perturbation of several of the inner water libration modes by this additional molecule which leads to the most visible spectral changes, let aside the degeneracy lift of the sulfate stretches discussed above. The outer water molecule donates two hydrogen bonds, however it receives none. This weaker participation to the hydrogen bond network makes its libration less constrained thus the bands for which its contribution is major appear much farther to the red, at around 700 cm⁻¹.

The computed IR spectra also challenge the view that band assignments can be safely made on the basis of those established for the bulk solution spectrum in the 900–1200 cm⁻¹ region. It should be noted that the series of experimental spectra,^{2,3} encompassing *n* = 4–14, 18 and 24 does show progressive overlap between the water libration and sulfate antisymmetric stretch features from *n* = 10 to 14, while this no longer exists for *n* = 18 and 24. There appears to exist a region of intermediate size in which the first and second hydration shells are perturbed enough, compared to the bulk, that some parts of their spectra are not directly comparable.

C. Closure of the first hydration shell

The appearance of a band at 970 cm⁻¹ at *n* = 13 appears not to be related to hydration shell closure at *n* = 12. The single low-energy saturated isomer found for *n* = 12 (W12-2) does not appear to generate a low-energy isomer for *n* = 13 by simple addition of H₂O. This is likely due to the smaller binding energy at any site that has to be remote from the ion, compared to W12-1 for which direct access to the charge is possible.

The computed energetics and spectra show that shell closure is still not effected at $n = 13$. In addition, the experimental spectra of $n = 13$ and 14 are strongly similar,²³ including the 970 cm^{-1} band which is common to both. Finally, the shapes of the lowest energy structures for $n = 13$ shown in Fig. 3 suggest that at least one more strongly bound water molecule could easily be accommodated in the second shell. All of these points suggest that closure of the first hydration shell might occur even beyond $n = 14$. At least the experimental IRPD spectra do not provide any indication of shell closure at $n = 14$.

5. Conclusions

Extensive exploration of the potential energy surfaces of $\text{SO}_4^{2-}(\text{H}_2\text{O})_{12}$ and $\text{SO}_4^{2-}(\text{H}_2\text{O})_{13}$ and computation of the structures, energetics and IR spectra have been carried out near the current computational accuracy limit. The results provide a new picture of these clusters, as most of the low energy structures identified herein have not been described previously. While the difference in spectra between $n = 12$ and $n = 13$ does provide a structural fingerprint, the latter turns out not to correspond to a qualitative transition in the cluster structure. In particular, it appears not to be related to hydration shell closure at $n = 12$. The computed energetics and spectra show that this closure is still not effected at $n = 13$. In addition, the experimental IRPD spectra of $n = 12$ –14 do not provide any indication in favor (nor against) shell closure at $n = 14$.

Comparison of experimental and computed IR spectra also demonstrates that the IR “fingerprint” region does not provide a clear identification of a single or a few structures being populated, even at very low temperature. There exists a number of structures, computed to lie in a few kJ mol^{-1} window, whose spectra are compatible with the experimental spectrum. It would be particularly interesting to investigate the O–H stretching portion of IRPD spectra in the 3200 – 3800 cm^{-1} range to see if more discriminating features can be obtained there.

Finally, it is apparent from the results given in Tables 1 and 2 that even for the small-sized clusters discussed herein, the few most stable isomers may not be the most appropriate tool for structural characterization. Their energies are so similar that except at the very lowest temperatures, a collection of structure families may be required to describe these objects. This will become even truer for larger clusters. To this end, resorting to molecular dynamics (MD) appears to be the method of choice. The results of the present structural search indicate that AMOEBA is a good candidate for running classical MD simulations on hydrated sulfate clusters. IR spectra may then be computed directly from the simulations, as described recently.³⁷ Such studies are currently underway.

Acknowledgements

We gratefully thank Prof K. R. Asmis for sharing his experimental data of ref. 23, as reproduced in Fig. 2 and 4 of this paper. This work was granted access to the HPC resources of

[CCRT/CINES/IDRIS] under the allocation c2015077128 made by GENCI (Grand Equipement National de Calcul Intensif).

References

- 1 B. J. Finlayson-Pitts and J. N. Pitts Jr., *Chemistry of the upper and lower atmosphere – theory, experiments and applications*, Academic Press, San Diego, 2000.
- 2 S. Twomey, *Atmospheric aerosols*, Elsevier, New York, 1977.
- 3 A. Laaksonen, M. Kulmala, T. Berndt, F. Stratmann, S. Mikkonen, A. Ruuskanen, K. E. J. Lehtinen, M. Dal Maso, P. Aalto and T. Petäjä, *et al.*, *Atmos. Chem. Phys.*, 2008, **8**, 7255.
- 4 V. Ramanathan, P. J. Crutzen, J. T. Kiehl and D. Rosenfeld, *Science*, 2001, **294**, 2119–2124.
- 5 D. Markovich, *Physiol. Rev.*, 2001, **81**, 1499–1533.
- 6 J. T. O'Brien, J. S. Prell, M. F. Bush and E. R. Williams, *J. Am. Chem. Soc.*, 2010, **132**, 8248–8249.
- 7 J. T. O'Brien and E. R. Williams, *J. Am. Chem. Soc.*, 2012, **134**, 10228–10236.
- 8 A. I. Boldyrev and J. Simons, *J. Phys. Chem.*, 1994, **88**, 2298–2300.
- 9 A. Whitehead, R. Barrios and J. Simons, *J. Chem. Phys.*, 2002, **116**, 2848–2851.
- 10 X. B. Wang, J. B. Nicholas and L. S. Wang, *J. Chem. Phys.*, 2000, **113**, 10837–10840.
- 11 A. T. Blades and P. Kebarle, *J. Am. Chem. Soc.*, 1994, **116**, 10761–10766.
- 12 C. C. Pye and W. W. Rudolph, *J. Phys. Chem. A*, 2001, **105**, 905–912.
- 13 C. G. Zhan, F. Zheng and D. A. Dixon, *J. Chem. Phys.*, 2003, **119**, 781–793.
- 14 B. Gao and Z. F. Liu, *J. Chem. Phys.*, 2004, **121**, 8299–8306.
- 15 B. Gao and Z. Liu, *J. Chem. Phys.*, 2005, **123**, 224302.
- 16 Y. Miller, G. M. Chaban, J. Zhou, K. R. Asmis, D. M. Neumark and R. B. Gerber, *J. Chem. Phys.*, 2007, **127**, 094305.
- 17 M. F. Bush, R. J. Saykally and E. R. Williams, *J. Am. Chem. Soc.*, 2007, **129**, 2220–2221.
- 18 X. B. Wang, A. P. Sergeeva, J. Yang, X. P. Xing, A. I. Boldyrev and L. S. Wang, *J. Phys. Chem. A*, 2009, **113**, 5567–5576.
- 19 D. S. Lambrecht, G. N. I. Clark, T. Head-Gordon and M. Head-Gordon, *J. Phys. Chem. A*, 2011, **115**, 11438–11454.
- 20 D. S. Lambrecht, L. McCaslin, S. S. Xantheas, E. Epifanovsky and M. Head-Gordon, *Mol. Phys.*, 2012, **110**, 2513–2521.
- 21 N. Mardirossian, D. S. Lambrecht, L. McCaslin, S. S. Xantheas and M. Head-Gordon, *J. Chem. Theory Comput.*, 2013, **9**, 1368–1380.
- 22 X. B. Wang, X. Yang, J. B. Nicholas and L. S. Wang, *Science*, 2001, **294**, 1322–1325.
- 23 J. Zhou, G. Santambrogio, M. Brümmer, D. T. Moore, L. Wöste, G. Meijer, D. M. Neumark and K. R. Asmis, *J. Chem. Phys.*, 2006, **125**, 111102.
- 24 R. L. Wong and E. R. Williams, *J. Phys. Chem. A*, 2003, **107**, 10976–10983.

- 25 A. T. Blades, J. S. Klassen and P. Kebarle, *J. Am. Chem. Soc.*, 1995, **117**, 10563–10571.
- 26 A. T. Blades and P. Kebarle, *J. Phys. Chem. A*, 2005, **109**, 8293–8298.
- 27 Q. Wan, L. Spanu and G. Galli, *J. Phys. Chem. B*, 2012, **116**, 9460–9466.
- 28 S. Katsyuba and E. Vandyukova, *Chem. Phys. Lett.*, 2003, **377**, 658–662.
- 29 A. Sharma, G. Ohanessian and C. Clavaguéra, *J. Mol. Model.*, 2014, **20**, 2426–2434.
- 30 (a) C. Riplinger, B. Sandhoefer, A. Hansen and F. Neese, *J. Chem. Phys.*, 2013, **139**, 134101; (b) M. Sparta and F. Neese, *Chem. Soc. Rev.*, 2014, **43**, 5032–5041.
- 31 J. W. Ponder, *Tinker 6 – Software Tools for Molecular Design*, 2010, available at <http://dasher.wustl.edu/tinker>.
- 32 M. J. Frisch, *et al.*, *Gaussian09, rev. D.01*, Gaussian, Inc., Wallingford CT, 2013.
- 33 Turbomole V6.6 2013, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007, available at <http://www.turbomole.com>.
- 34 F. Neese, Orca version 3.0.3., *WIREs Comput. Mol. Sci.*, 2012, **2**, 73–78, available at <http://www.thch.uni-bonn.de/tc/orca/>.
- 35 L. C. Smeeton, J. D. Farrell, M. T. Oakley, D. J. Wales and R. L. Johnston, *J. Chem. Theory Comput.*, 2015, **11**, 2377–2384.
- 36 W. R. Cannon, B. M. Pettitt and J. A. McCammon, *J. Phys. Chem.*, 1994, **98**, 6225–6230.
- 37 D. Semrouni, A. Sharma, J.-P. Dognon, G. Ohanessian and C. Clavaguéra, *J. Chem. Theory Comput.*, 2014, **10**, 3190–3199.

Conclusions et perspectives

L'objectif de ce travail était de fournir des développements à la méthodologie nécessaire à la modélisation et l'attribution de spectres IR(M)PD. Les systèmes étudiés, de petits peptides ou des ions hydratés dans des agrégats d'eau en phase gazeuse, ont comme point commun qu'un modèle basé sur une description statique et monoconformationnelle est en général incapable de reproduire le spectre expérimental.

La première étape a consisté à explorer les surfaces d'énergie potentielle par dynamique moléculaire classique avec le champ de forces polarisable AMOEBA, afin d'identifier les structures les plus stables. Pour les surfaces les plus complexes, AMOEBA a été couplé à la dynamique par échange de répliques (REMD). Pour les ions solvatés, nous avons développé un outil 'Traqueur de cycles', qui permet d'identifier les réseaux de liaisons hydrogènes au sein des agrégats et d'engendrer des inversions de cycles. Ceci fournit une base solide pour des calculs basés sur la mécanique quantique, permettant d'obtenir des énergies fiables. Une étape de validation des calculs quantiques, WFT ou DFT, reste néanmoins nécessaire; elle est gouvernée à la fois par la taille des systèmes étudiés et par la propriété calculée.

Généralement, le spectre expérimental à température finie ne peut pas être décrit par une seule structure. Pour les ions solvatés dans de petits agrégats, nous avons simulé des spectres multiconfigurationnels à partir de spectres statiques calculés en chimie quantique puis pondérés par un critère de Boltzmann, dans lequel nous avons ajouté une contribution liée aux possibilités de réorganisation des cycles de liaisons hydrogènes. Pour les systèmes de grande taille, le recours à la dynamique moléculaire et à la méthode DACF (Dipole Autocorrelation Function) est le moyen de combiner les effets de dynamique et de température, tels que l'anharmonicité des modes de vibration, la largeur des bandes ou le mélange de structures, pour un coût raisonnable. Le DACF a notamment été employé pour modéliser le spectre dynamique à température finie du dipeptide Ace-Phe-Ala-NH₂. A cette occasion, nous avons pu constater la forte dépendance des spectres calculés à la procédure d'extraction des multipôles pour AMOEBA.

La modélisation de la spectroscopie des ions solvatés de grande taille a nécessité des étapes de validation de paramètres puis de calibration. Le prérequis à l'utilisation du champ de forces dans une problématique d'attribution de spectres IR à température finie, était d'établir sa capacité à reproduire les comportements dynamiques des ions solvatés, c'est à dire leur capacité à organiser le réseau de molécules d'eau autour d'eux sous la forme de sphères de solvation plus ou moins rigides. Nous avons réalisé une étude de dynamique des ions dans des nanogouttes de 100 molécules d'eau. Différents paramètres ont été évalués, tels que le nombre et la mobilité des molécules d'eau en 1^{ère} ou en 2^{ème} sphères, mais également le comportement privilégié de l'ion, fortement solvaté au sein de l'agrégat, ou au contraire migrant à sa surface. Le champ de forces de l'eau AMOEBA03 a permis de reproduire les comportements dynamiques pour tous les ions de charge -II à +III. A l'inverse, il a montré que, dans une version non modifiée, il était incapable de reproduire les spectres d'élongations OH d'ions hydratés. Ponder et al. ont proposé de nouveaux paramètres de l'eau, qui apportent de vraies améliorations du modèle en terme de spectroscopie. La version AMOEBA14 a passé le même processus de validation des comportements dynamiques, qui restent globalement bons, mais avec un accord un peu plus faible sur les coordinences et les temps de résidence par rapport aux données expérimentales, dû à l'inhomogénéité des versions des paramètres pour l'eau et les ions. Le

modèle a ensuite été calibré spécifiquement pour la spectroscopie des ions solvatés de grande taille. Les nouveaux paramètres de l'eau ont été validés sur leur capacité à reproduire les bandes de petits agrégats d'eau, ce qui a nécessité de modifier la constante de pliage \widehat{HOH} . Afin de mieux reproduire les élargissements des massifs des élongations O-H, nous avons augmenté la flexibilité de réponse des vibreurs à leur environnement, soit par une augmentation de la flexibilité intramoléculaire, soit par un modèle de charges fluctuantes. Ces étapes de calibration font apparaître plusieurs voies d'amélioration, telles que l'homogénéisation des versions de paramètres pour l'eau et les ions, ou une augmentation de la flexibilité du modèle de polarisation, par exemple en incluant les effets de transfert de charge. Toutefois, notre version modifiée du champ de forces a été capable de décrire certaines observations expérimentales, telles que les déplacements des massifs de O-H liés et des bandes de O-H libres dans des agrégats $M(H_2O)_n$, $n = 36$ et 100 , en fonction de la charge de l'ion, ou l'apparition d'une bande de O-H libre dans des agrégats de sulfates hydratés de taille croissante. Sur des tailles intermédiaires ($n = 24-36$), le modèle reproduit avec une bonne précision l'ensemble du massif des O-H liés.

La méthode DACF n'offre pas d'accès direct aux modes normaux de vibration, qui sont généralement traités *a posteriori*. Nous avons implémenté une attribution dynamique des spectres DACF par des outils basés sur la Driven molecular Dynamics (DMD), dont la première application a été l'attribution de spectres anharmoniques à température finie de peptides. En choisissant les conditions d'entraînement au cours de courtes simulations DMD, la méthode peut soit calculer des modes à basse température, comparables à ceux obtenus par un calcul du Hessien, soit permettre d'observer l'apparition de couplages avec la température. Appliquée à la gamme des élongations O-H de sulfates hydratés, elle offre une vue de l'évolution de l'environnement des vibreurs avec l'augmentation de la taille de l'agrégat, et des déplacements en fréquences associés. Toutefois, identifier des contributions parmi de fortes densités de vibreurs peut être assez délicat, surtout dans les agrégats de grande taille pour lesquels les molécules d'eau des sphères extérieures ont une dynamique importante. Les conditions d'entraînement ont dû être drastiquement contrôlées pour limiter les redistributions d'énergie et les résonances multiples. Il est possible de moduler les conditions d'entraînement sur une partie des vibreurs. Cette voie pourrait être suivie afin d'observer des résonances particulières, même au sein de systèmes de grande taille fortement couplés.

Pour les sulfates hydratés $SO_4^{2-}(H_2O)_n$, nous avons identifié une famille de structures permettant de reproduire, par calculs quantiques, l'ensemble des spectres expérimentaux pour $n = 12$ et 13 , alors qu'un mélange est nécessaire pour décrire celui de $n = 9$. Pour les tailles intermédiaires, les spectres DACF produisent une image fidèle des spectres IRPD et des attributions des composantes principales des bandes peuvent être proposées par calculs DMD. D'autre part, Williams et al. identifient plusieurs agrégats de stabilité particulière, dont $n = 12$ et 24 , par des mesures d'abondance et de vitesse d'évaporation d'eau en fonction de la taille de l'agrégat, en spectrométrie de masse. Pour les paires $n = 12, 13$ d'une part et $n = 24, 25$ d'autre part, les structures les plus stables présentent des similitudes permettant d'en proposer une première interprétation. Des calculs en cours permettront d'évaluer les énergies d'évaporation à ces tailles, qui pourraient avoir un effet primordial sur les abondances observées.

Enfin, afin de reproduire les signatures caractéristiques du partage de proton au sein de petits peptides, nous avons implémenté un modèle EVB dans TINKER. La principale force de ce modèle est le calcul des états diabatiques, qui est réalisé par le champ de forces polarisable AMOEBA, garantissant notamment une bonne directionnalité de la liaison hydrogène. Le potentiel de couplage a été voulu le simple possible, afin d'en tester la transférabilité. Ce modèle a permis de décrire le comportement dynamique du transfert de proton au sein d'acides aminés ou de diacides déprotonés, et ainsi de reproduire la signature spectroscopique observée expérimentalement. L'objectif de transférabilité est partiellement atteint, puisqu'un même potentiel de couplage a pu être appliqué aux diacides déprotonés $HO_2C-(CH_2)_n-CO_2^-$, $n = 3, 4$ et 6 , dont les contraintes autour du proton sont similaires. Cette transférabilité pourrait être facilitée lorsque le champ de forces inclura directement les effets de 'pénétration de charge'. Ces premiers résultats ouvrent de nombreuses perspectives au modèle EVB-AMOEBA, qui pourrait être étendu à un

modèle général à deux états, dans lequel les deux sites de protonation sont différents, voire à un modèle multi-états. Dans ce cas, bien qu'il n'y ait aucune restriction dans l'utilisation du champ de forces polarisable comme d'un bloc de construction d'un modèle EVB multi-états, le coût calculatoire associé au calcul de l'énergie électrostatique et de polarisation demandera des algorithmes de parallélisation efficaces, afin de conserver la capacité du modèle à reproduire les mouvements de basse fréquence et les réarrangements lents.

Titre : Développement de champs de forces polarisables et applications à la spectroscopie vibrationnelle

Mots clés : Spectroscopie IR, champ de forces polarisables, hydratation des ions, transfert de proton, modes normaux

Résumé : La spectroscopie IRPD permet d'obtenir les signatures vibrationnelles d'espèces chargées en phase gazeuse. L'attribution des modes de vibration pour établir une relation entre le spectre expérimental et une structure moléculaire est une tâche délicate et nécessite le recours à la modélisation moléculaire. Ce manuscrit présente un ensemble d'outils théoriques pour le calcul et l'attribution de spectres vibrationnels, basé principalement sur la dynamique moléculaire (DM) classique et le champ de forces polarisable AMOEBA, ainsi que son application à des ions gazeux de tailles diverses. Les ions hydratés dans des agrégats d'eau $M(H_2O)_n$ (n allant de 6 à 100) sont caractérisés par une dynamique importante, et leur spectre expérimental ne peut pas être décrit par une seule structure. La signature des peptides évolue avec la température et les effets d'anharmonicité dynamique. Ils peuvent également être le siège de mécanismes de transfert de proton, présentant une signature vibrationnelle très caractéristique. La surface d'énergie potentielle de ces systèmes est explorée par DM classique en trajectoires individuelles ou avec échange de répliques, afin d'engendrer des structures énergétiquement stables. Pour les plus petits systèmes, les méthodes quantiques DFT et post-HF sont utilisées pour confirmer les structures de plus basse énergie, calculer leurs spectres IR statiques et proposer des attributions des modes de vibration. Pour les systèmes de plus grandes tailles, c'est-à-dire les ions dans

des gouttes d'eau de plusieurs dizaines de molécules, la simulation des spectres IR à température finie est basée sur la transformée de Fourier de la fonction d'autocorrélation du moment dipolaire (DACF), calculée pour une trajectoire de dynamique moléculaire classique. Cette méthode n'offrant pas d'accès direct aux modes normaux de vibration, nous avons implémenté une méthode d'attribution dynamique, basée sur la Driven Molecular Dynamics (DMD) et couplée au DACF. La combinaison AMOEBA/DACF/DMD a été utilisée pour reproduire et attribuer le spectre du dipeptide Ace-Phe-Ala-NH₂, et ceux d'ions hydratés dans des agrégats d'eau. Enfin, la signature vibrationnelle d'un transfert de proton ne peut être décrite, ni par des méthodes statiques quantiques, ni par la dynamique classique. Sa modélisation a nécessité le développement d'un modèle Empirical Valence Bond (EVB) à deux états, couplé au champ de forces polarisables AMOEBA. Le modèle EVB a été implémenté dans TINKER. Il permet de reproduire le comportement dynamique du transfert de proton au sein de petits peptides et de diacides déprotonés, ainsi que la signature spectroscopique observée expérimentalement. Une partie importante des applications de ces développements concerne des ions simples hydratés dans des nano-gouttes, et en particulier l'ion sulfate de grande importance environnementale. Nous avons pu reproduire de façon satisfaisante, pour la première fois, les spectres d'agrégats contenant jusqu'à 100 molécules d'eau.

Title : Polarizable force fields development and applications to vibrational spectroscopy

Keywords : infrared Spectroscopy, polarizable force fields, ion hydration, proton transfer, normal modes

Spectroscopy IRPD provides vibrational signatures of charged species in the gas phase. The vibrational normal modes assignment to establish a relationship between the experimental spectrum and molecular structure is a delicate task and requires the use of molecular modeling. This manuscript presents a set of theoretical tools for calculation and assignment of vibrational spectra, based mainly on classical molecular dynamics (DM) and polarizable AMOEBA force field, and its application to gaseous ions of various sizes. Hydrated ions in water clusters $M(H_2O)_n$ (n in 6-100 range) are characterized by a dynamic behavior, and their experimental spectrum can not be described by a single structure. The signature of peptides changes with temperature and dynamic anharmonicity effects. They can also be the site of proton transfer mechanisms, with a very characteristic vibrational signature. The potential energy surface of these systems is explored by classical DM in individual trajectories or replica exchange to generate energetically stable structures. For smaller systems, quantum methods, as DFT and post-HF, are used to confirm the lowest energy structures, calculate their static IR spectrum and propose normal modes assignments. For larger systems, i.e ions in

water drops of several tens of molecules, the simulation of IR spectra at finite temperature is based on the DACF, calculated during a classical molecular dynamics trajectory. As this method does not allow direct access to the vibrational normal modes, we implemented a method of dynamic assignments, based on the Driven Molecular Dynamics (DMD) and coupled to the DACF. The combination AMOEBA/DACF/DMD was used to reproduce and assign the spectrum of the dipeptide Ace-Phe-Ala-NH₂, and those of hydrated ions in water clusters. Finally, the vibrational signature of a proton transfer can not be described by quantum static methods or by classical dynamics. Its modeling required the development of a two states Empirical Valence Bond Model (EVB), coupled with AMOEBA polarizable force field. The two states EVB model was implemented in the software TINKER. It can reproduce the dynamic behavior of proton transfer in small peptides and deprotonated acids, as well as the spectroscopic signatures observed experimentally. An important part of the applications of these developments relates simple hydrated ions in nano-droplets, and in particular the sulfate ion of great environmental importance. We were able to reproduce satisfactorily, for the first time, the spectra of clusters containing up to 100 water molecules.

