



HAL
open science

Énergétique et spectroscopie de polypeptides par dynamique moléculaire : champ de force de seconde génération et chimie quantique

David Semrouni

► **To cite this version:**

David Semrouni. Énergétique et spectroscopie de polypeptides par dynamique moléculaire : champ de force de seconde génération et chimie quantique. Chimie théorique et/ou physique. Ecole Polytechnique X, 2010. Français. NNT : . pastel-00528244

HAL Id: pastel-00528244

<https://pastel.hal.science/pastel-00528244v1>

Submitted on 21 Oct 2010

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.



École doctorale de l'École Polytechnique

THÈSE

pour obtenir le grade de

DOCTEUR DE L'ÉCOLE POLYTECHNIQUE

Spécialité : Chimie

par

David SEMROUNI

Énergétique et spectroscopie de peptides attachés à un cation métallique, champ de force de seconde génération et chimie quantique

| | | |
|--------------------------|----------------------|---|
| Rapporteurs : | Terry B. MCMAHON | Professeur – Université de Waterloo (Canada) |
| | Jean-Philip PIQUEMAL | Maître de Conférences – Université de Paris 6 |
| Examineurs : | Philippe MAÎTRE | Directeur de recherche au CNRS – Université de Paris 11 |
| | Gérald MONARD | Professeur – Nancy-Université |
| | Thomas SIMONSON | Directeur de recherche au CNRS – École polytechnique |
| Co-directrice de thèse : | Carine CLAVAGUÉRA | Chargé de recherche au CNRS – École polytechnique |
| Directeur de thèse : | Gilles OHANESSIAN | Directeur de recherche au CNRS – École polytechnique |

Laboratoire des Mécanismes Réactionnels (DCMR)

UMR N° 7651 CNRS-École Polytechnique

91128, Palaiseau Cedex, France

Je remercie Gilles Ohanessian pour m'avoir accueilli au sein du Laboratoire des Mécanismes Réactionnels et ainsi permis d'effectuer sous sa direction ces trois années de travail de recherche qui s'achèvent. Les discussions scientifiques, liées ou non à mon sujet, m'auront beaucoup apporté, tout comme ses encouragements.

Je remercie, bien évidemment Carine Clavaguéra pour son encadrement de tout jour et à toute heure, son soutien de toujours et la patience dont elle a su faire preuve à maintes occasions durant ces trois ans.

Je souhaite remercier MM Therry McMahon et Jean-Philip Piquemal qui ont accepté d'être rapporteurs de ma thèse. Ma gratitude va aussi aux examinateurs : MM Philippe Maître, Gérald Monard et Thomas Simonson. L'intérêt porté à mon travail par ce jury a été un honneur pour moi.

J'ai particulièrement apprécié la collaboration que nous avons eue avec Florent Calvo, principalement autour du REMD. Je tiens à le remercier pour notre agréable séjour à Lyon qui fut très enrichissant et utile pour le travail effectué.

Je remercie aussi Edith pour les discussions que nous avons pu avoir sur l'IRMPD.

Je suis très reconnaissant aux personnes qui m'ont fait découvrir et donné goût à la chimie théorique et calculatoire. Je remercie donc Xavier Assfeld, une seconde fois Gérald Monard et également les autres membres de l'équipe de Chimie et Biochimie Théorique de Nancy. Je remercie également les (autres) amis dans le monde de la chimie théorique : Amy, Aude, Gabriella, Hassoun, Éric Perpète et l'UCPTS de Namur...

Passer trois années en thèse, c'est aussi des rencontres, notamment sur le campus avec X'Doc et bien sûr au labo. Je remercie Magalie pour les marionnettes, Moussa pour les conversations des plus intéressantes, Saïd pour l'aide avec la GC-MS, Renjie, Joe, Yasmine, Ophélie, toute la grande famille du DCMR... les stagiaires et post-docs bien trop nombreux pour être énumérés mais qui ont participé à rendre ces trois années de thèse encore plus intéressantes.

Je tiens à exprimer ma gratitude envers les autres personnes m'ayant soutenu de près ou de loin : HH, les anciens collègues de Nancy et d'ailleurs, ainsi que, bien sûr, ma famille.

Sommaire

| | |
|---|-----------|
| Introduction | 3 |
| I Méthodes et outils | 11 |
| 1 Généralités | 13 |
| Méthodes quantiques | 13 |
| Méthodes classiques | 16 |
| 2 AMOEBA, un champ de force polarisable flexible | 31 |
| 2.1 Généralités sur AMOEBA | 31 |
| 2.1.1 Formalisme et philosophie d'AMOEBA | 32 |
| 2.1.2 Classes et types d'atomes | 35 |
| 2.2 Adaptation de paramètres d'AMOEBA | 36 |
| 2.2.1 Exemple d'optimisation de paramètres de torsion | 37 |
| 2.2.2 Optimisation des multipôles atomiques | 38 |
| 3 Exploration de l'espace des phases | 51 |
| 3.1 Quelques méthodes d'exploration | 51 |

| | | |
|----------|--|-----------|
| 3.2 | Dynamique moléculaire | 54 |
| 3.2.1 | Propagation de la trajectoire | 54 |
| 3.2.2 | Thermostats | 57 |
| 3.2.3 | Discussion sur la dynamique moléculaire avec le champ de force AMOEBA | 58 |
| 3.3 | Dynamique moléculaire par échange de répliques (REMD) | 59 |
| 3.3.1 | Principe | 59 |
| 3.3.2 | Tirex : implémentation du REMD avec Tinker | 62 |
| 3.3.3 | Traitement des données issues d'une simulation REMD avec Tirex . | 63 |
| 3.4 | Comparaison et bilan des méthodes d'exploration | 69 |
| 4 | Sélection, identification, différenciation de structures | 77 |
| 4.1 | Discernement de structures | 78 |
| 4.1.1 | Propriétés géométriques | 78 |
| 4.1.2 | Outils mathématiques | 79 |
| 4.1.3 | Discussion | 81 |
| 4.2 | Tri et groupement de structures | 87 |
| 5 | Simulation de spectre infrarouge | 91 |
| 5.1 | Approximation harmonique : exemple du NMA | 92 |
| 5.2 | Traitement statique de l'anharmonicité | 94 |
| 5.3 | Calcul de spectre par DACF | 95 |
| 5.3.1 | Bases théoriques | 95 |
| 5.3.2 | Mise en œuvre et exploitation | 98 |

| | | |
|-----------|--|------------|
| 5.3.3 | Spectres DACF et transitions structurales | 106 |
| 5.4 | Application du DACF | 108 |
| II | Applications | 123 |
| 6 | Structure, énergétique et champ de force polarisable | 127 |
| 7 | Structure de l'octa-glycine sodée | 141 |
| 8 | Évaluation de fonctionnelles pour le spectre IR de G₈-Na⁺ | 153 |
| 9 | Spectroscopie IRMPD et modélisation de polyglycines sodées | 183 |
| | Conclusions et perspectives | 213 |
| | Liste des figures | 218 |
| | Liste des tableaux | 219 |

Introduction

L'importance de la structure des biomolécules sur leur fonction ou leur dysfonction est bien établie. Les caractéristiques des biopolymères sont loin de ne dépendre que de la séquence (structure primaire) des unités (résidus) qui les composent. La structure tridimensionnelle contrôle par exemple les phénomènes de reconnaissance non covalente : anticorps-antigène, drogue-récepteur, protéine-protéine, peptide-membrane, etc. La conformation spatiale des systèmes protéiques est responsable des maladies neuro-dégénératives à prion¹. La conception de médicament ou *drug design* repose en grande partie sur les relations entre structure et réactivité (SAR) et d'homologie.

Pour caractériser les structures moléculaires, toutes les méthodes spectroscopiques sont utiles *a priori*. Pour les biomolécules, ce sont surtout la diffraction des rayons X ou des neutrons et la RMN, mais aussi les spectroscopies infrarouge et Raman² qui donnent des informations sur le contenu de macromolécules en structures secondaires : motifs périodiques de plus ou moins grande longueur (hélices 3_{10} , α ou π ; feuillet β parallèles ou anti-parallèles) ou éléments locaux de structuration (coudes γ ou β).

L'analyse de biomolécules par spectrométrie de masse a explosé depuis le début des années 90³, du fait du développement de sources d'ions préservant l'intégrité des molécules à liaisons labiles, y compris non covalentes, lors du passage de la phase condensée à la phase gazeuse. Pour autant, la question de la préservation de la structure tridimensionnelle n'est pas résolue de manière générale. Malgré cela, la possibilité d'étudier des biomolécules en routine par spectrométrie de masse confère un grand intérêt pratique à cette technique. En particulier, leur fragmentation par activation, le plus souvent collisionnelle, a ouvert la voie au séquençage très efficace de peptides, protéines et acides nucléiques, et ainsi donné lieu à une révolution en protéomique et en génomique.

Est-il possible d'aller au-delà du séquençage (étude de la structure primaire) dans l'analyse structurale de biomolécules après qu'elles ont été passées en phase gazeuse ? Il faut pour cela coupler une technique spectroscopique à la spectrométrie de masse. Après de nombreux essais depuis la fin des années 80, ce couplage a été réalisé au début des années 2000 par la technique de dissociation induite par absorption de multiples photons infrarouge (IRMPD)^{4,5}.

Le principe de la spectroscopie IRMPD consiste à former des ions gazeux et à les piéger

pendant des temps longs (jusqu'à la seconde ou plus), à les sélectionner en masse pour étudier les espèces une à une, et à les irradier avec des photons de longueur d'onde variable. Comme le nombre d'ions piégés est bien trop faible pour réaliser une spectroscopie d'absorption ou de transmission, on utilise la fragmentation comme sonde de l'absorption des photons. Il s'agit donc d'une spectroscopie d'action, facilitée par la capacité de la spectrométrie de masse à détecter des ions en très faible quantité.

La spectroscopie d'action peut être réalisée en principe dans des gammes de longueur d'onde différentes. Celle qui a connu le plus de succès est l'infrarouge moyen, mais on peut noter que ce type de couplage existe aussi pour la spectroscopie UV/Visible^{6,7}, même s'il est moins diagnostique sur les structures. En spectroscopie IR, l'absorption d'un seul photon ne suffisait généralement pas à déclencher la fragmentation d'un ion moléculaire, deux possibilités existent : faire en sorte qu'un seul photon suffise en accrochant aux ions un messager (par exemple un gaz rare) qui les perturbe peu et se dissocie facilement⁸, ou utiliser des conditions d'irradiation de forte puissance. Dans le premier cas, on doit utiliser des espèces refroidies, par détente supersonique ou par refroidissement cryogénique⁹. Dans le second, on utilise des lasers IR accordables puissants. Pour cette raison, ce sont les lasers à électrons libres qui ont été les premiers à donner accès à cette spectroscopie, avec le désavantage qu'il s'agit d'un appareillage très lourd. Depuis quelques années le développement des lasers de table OPO/OPA a rendu cette technique accessible, bien qu'à nettement plus faible puissance, dans la gamme des « hautes fréquences »¹⁰ de 3 000 à 4 000 cm^{-1} puis maintenant aux énergies plus basses.

Puisque cette technique nécessite le plus souvent l'absorption de plusieurs photons pour observer une fragmentation, il s'agit d'un processus complexe dont la première étape est l'absorption d'un photon si son énergie est résonante avec un mode vibrationnel de la molécule ou du système chimique considéré. Puis, pour que l'absorption d'un second photon de même énergie puisse avoir lieu, le système doit retourner de l'état excité formé initialement vers l'état fondamental avec redistribution de l'énergie interne dans le bain thermique constitué par l'ensemble des autres modes. Le processus peut alors être répété jusqu'à ce que l'énergie interne accumulée soit suffisante pour déclencher la fragmentation et donc l'apparition de nouveaux pics dans le spectre de masse. Les intensités relatives des ions fragments et leur parent permettent d'attribuer une intensité à l'absorption. Répéter cette opération

pour une série de longueurs d'onde des photons permet d'engendrer un spectre vibrationnel. Il n'est pas équivalent au spectre d'absorption IR traditionnel monophotonique, mais l'expérience montre que les spectres à un photon calculés sont généralement proches des spectres IRMPD ainsi obtenus, même si les intensités relatives peuvent être sensiblement différentes.

Il existe actuellement deux centres laser dans le monde permettant l'acquisition de spectres IRMPD avec des lasers à électrons libres. Il s'agit du FOM à Rijnhuizen aux Pays-Bas et du CLIO (Centre Laser Infrarouge d'Orsay) en France. Ils ont permis un développement très rapide de cette technique et son application à la détermination de structure d'ions très variés. On peut citer les structures de dimères protonés de molécules organiques¹¹, celles de complexes organométalliques¹² et d'agrégats métalliques¹³, ou encore la structure des couches successives de micro-hydratation de cations métalliques¹⁴. L'élucidation de la structure de peptides forme une part importante de ce développement foisonnant. Les sites de protonation d'oligoglycines ont réservé des surprises puisque pour certaines tailles, un isomère dans lequel un oxygène peptidique est protoné a été observé¹⁵. Les interactions de petits peptides avec des cations métalliques ont été étudiées, en particulier pour déterminer si la complexation peut conduire à la formation d'un pont salin¹⁶. Les spectres sont assez congestionnés, ce qui permet en général la distinction entre isomères mais pas toujours la distinction entre conformères. Pour dépasser cette limite, le refroidissement cryogénique utilisé dans un montage développé à Lausanne, rend aujourd'hui possible l'enregistrement de spectres IR très bien résolus, conformère par conformère¹⁷. Une des tendances actuelles est de repousser les limites de cette spectroscopie en explorant ses possibilités pour des assemblages non covalents¹⁸ et des protéines^{19,20}. L'interprétation des spectres vibrationnels de biomolécules n'est pas chose facile. D'une part, il s'agit d'ions en phase gazeuse et les bases de données accumulées pour des molécules neutres en solution ne sont pas directement utilisables. D'autre part, l'attribution des bandes une à une n'est possible que pour des molécules assez simples, dont les spectres sont peu congestionnés. En effet, les expériences réalisées à la température ambiante fournissent des bandes élargies de contours rotationnels, qui se recouvrent rapidement dès que la taille des molécules augmente, d'autant que la résolution intrinsèque du laser à électrons libres à CLIO est d'environ 0,5 %, soit 5 à 10 cm^{-1} dans la région des empreintes digitales (de 1 000 cm^{-1} à 2 000 cm^{-1}). Attribuer les bandes observées pour établir une relation entre spectre expérimental et structure moléculaire n'est

donc pas une chose aisée. En pratique, le recours massif à la modélisation moléculaire est nécessaire. Lorsque ce travail a commencé au Laboratoire des Mécanismes Réactionnels, l'étude par IRMPD de polyglycines sodées avait déjà fourni les spectres des peptides comportant 2 à 8 résidus dans la gamme spectrale $1\ 000\text{-}2\ 000\text{ cm}^{-1}$. Pour les plus petits peptides, une recherche manuelle des structures les plus stables, suivie de calculs de spectres IR avec des méthodes de chimie quantique dans l'approximation harmonique standard, avait donné accès à des conclusions structurales partielles²¹. À partir de 5 résidus, toutes ces méthodes commençaient à poser problème, aussi bien pour l'exploration des surfaces d'énergie potentielle que pour les calculs de spectres IR. Le développement de nouvelles approches s'avérait incontournable. Simultanément, des spectres IRMPD commençaient à être obtenus dans la gamme spectrale $3\ 000\text{ cm}^{-1}$ pour plusieurs polyglycines sodées, faisant de cette série l'objet idéal d'application de nouvelles approches théoriques. Notre travail a donc eu pour objectif la mise au point d'une démarche qui permette l'attribution d'un spectre expérimental à une structure tridimensionnelle pour des peptides sodés de taille croissante. Pour cela, il a fallu commencer par une recherche théorique efficace des structures les plus probables en utilisant un champ de force apte à traiter l'interaction peptide-cation. La combinaison de dynamique moléculaire avec échange de répliques et du champ de forces polarisable AMOEBA, étendu en fonction de nos besoins successifs, s'est avérée être efficace. La mécanique quantique est ensuite utilisée pour obtenir une énergétique fiable de ces structures et le calcul de leurs spectres. Cette étape demande une calibration soignée des méthodes utilisées, en particulier des fonctionnelles de la densité. Les spectres infrarouge calculés traditionnellement à 0 K ne prennent pas en compte les effets dus à la température expérimentale qui est le plus souvent l'ambiante. Des spectres obtenus par dynamique moléculaire à température finie incluent des effets comme l'anharmonicité des modes de vibration, leurs couplages anharmoniques, les largeurs de bande, les combinaisons. . . . Ce type de méthodologie associée à un champ de force polarisable est encore très peu développé pour des molécules biologiques. Nous l'avons mis en œuvre avec AMOEBA et obtenu ainsi les spectres IR de systèmes modèles qui démontrent la pertinence de cette approche.

Bibliographie

- [1] S. B. PRUSINER, M. R. SCOTT, S. J. DEARMOND et F. E. COHEN : Prion Protein Biology. *Cell*, 93:337–348, 1998.
- [2] F. SIEBERT et P. HILDEBRANDT : *Vibrational Spectroscopy in Life Science*. Wiley-VCH, 2008.
- [3] R. M. CAPRIOLI et M. L. GROSS, éditeurs. *The Encyclopedia of Mass Spectrometry*, volume volume 2 : Biological Applications, chapitre Part A : Peptides and Proteins. Elsevier, 2005.
- [4] J. OOMENS, van ROIJ A. J. A., G. MEIJER et G. von HELDEN : Gas-phase infrared photodissociation spectroscopy of cationic polyaromatic hydrocarbons. *Astrophys. J.*, 542:404–410, 2000.
- [5] J. LEMAIRE, P. BOISSEL, M. HENINGER, G. MAUCLAIRE, G. BELLEC, H. MESTDAGH, A. SIMON, S. LE CAËR, J. M. ORTEGA, F. GLOTIN et P. MAÎTRE : Gas Phase Infrared Spectroscopy of Selectively Prepared Ions. *Phys. Rev. Lett.*, 89:273002, 2002.
- [6] J. P. REILLY : Ultraviolet photofragmentation of biomolecular ions. *Mass Spectrom. Rev.*, 28:425–447, 2009.
- [7] A. RACAUD, R. ANTOINE, L. JOLY, N. MESPLET, P. DUGOURD et J. LEMOINE : Wavelength-Tunable Ultraviolet Photodissociation (UVPD) of Heparin-Derived Disaccharides in a Linear Ion Trap. *J. Am. Soc. Mass Spectrom.*, 16:2045–2051, 2009.
- [8] J. M. LISY : Infrared studies of ionic clusters : the influence of Yuan T. Lee. *J. Chem. Phys.*, 125:132302, 2006.
- [9] T. R. RIZZO, J. A. STEARNS et O. V. BOYARKIN : Spectroscopic studies of cold, gas-phase biomolecular ions. *Intern. Rev. Phys. Chem.*, 28:481–515, 2009.

- [10] J. M. BAKKER, T. BESSON, J. LEMAIRE, D. SCUDERI et P. MAÎTRE : Gas-Phase Structure of a π -Allyl-Palladium Complex : Efficient Infrared Spectroscopy in a 7 T Fourier Transform Mass Spectrometer. *J. Phys. Chem. A*, 111:13415–13424, 2007.
- [11] T. D. FRIDGEN : Infrared consequence spectroscopy of gaseous protonated and metal ion cationized complexes. *Mass Spectrom. Rev.*, 28:586–607, 2009.
- [12] L. MACALEESE et P. MAÎTRE : Infrared spectroscopy of organometallic ions in the gas phase : from model to real world complexes. *Mass Spectrom. Rev.*, 26:583–605, 2007.
- [13] K. R. ASMIS et J. SAUER : Mass-selective vibrational spectroscopy of vanadium oxide cluster ions. *Mass Spectrom. Rev.*, 26:542–562, 2007.
- [14] M. F. BUSH, J. T. O'BRIEN, J. S. PRELL, C. C. WU, R. J. SAYKALLY et E. R. WILLIAMS : Hydration of Alkaline Earth Metal Dications : Effects of Metal Ion Size Determined Using Infrared Action Spectroscopy. *J. Am. Chem. Soc.*, 131:13270–13277, 2009.
- [15] R. H. WU et T. B. MCMAHON : Protonation Sites and Conformations of Peptides of Glycine ($\text{Gly}_{1-5}\text{H}^+$) by IRMPD Spectroscopy. *J. Phys. Chem.*, 113:8767–8775, 2009.
- [16] N. C. POLFER et J. OOMENS : Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass Spectrom. Rev.*, 28:468–494, 2009.
- [17] N. S. NAGORNOVA, T. R. RIZZO et O. V. BOYARKIN : Highly Resolved Spectra of Gas-Phase Gramicidin S : A Benchmark for Peptide Structure Calculations. *J. Am. Chem. Soc.*, 132:4040–4041, 2010.
- [18] J. C. POULLY, F. LECOMTE, N. NIEUWJAER, B. MANIL, J. P. SCHERMANN, C. DESFRANÇOIS, F. CALVO et G. GRÉGOIRE : Probing the specific interactions and structures of gas-phase vancomycin antibiotics with cell-wall precursor through IRMPD spectroscopy. *Phys. Chem. Chem. Phys.*, 12:3606–3615, 2010.
- [19] J. OOMENS, N. C. POLFER, D. T. MOORE, L. van der MEER, A. G. MARSHALL, J. R. EYLER, G. MEIJER et G. von HELDEN : Charge-state resolved mid-infrared spectroscopy of a gas-phase protein. *Phys. Chem. Chem. Phys.*, 7:1345–1348, 2005.
- [20] Y. M. E. FUNG, T. BESSON, J. LEMAIRE et R. A. MAÎTRE, P. Zubarev : Room-Temperature Infrared Spectroscopy Combined with Mass Spectrometry Distinguishes Gas-Phase Protein Isomers. *Angew. Chem. Int.*, 48:8340–8342, 2009.

- [21] O. P. BALAJ, C. KAPOTA, J. LEMAIRE et G. OHANESSIAN : Vibrational signatures of sodiated oligopeptides (GG-Na⁺, GGG-Na⁺, AA-Na⁺ and AAA-Na⁺) in the gas phase. *Int. J. Mass Spectrom.*, 269:196–209, 2008.

Première partie

Méthodes et outils

Chapitre 1

Généralités

Notre objectif était la mise en œuvre de méthodes de modélisation pour l'identification structurale de peptides attachés à un cation métallique en phase gazeuse, à partir de spectres expérimentaux IRMPD. Nous avons pour cela utilisé des méthodes diverses, basées sur la mécanique quantique pour certaines, sur la mécanique moléculaire pour d'autres. Dans cette première partie, nous les plaçons brièvement dans leur contexte avant d'entrer plus en détail dans la description des méthodes effectivement utilisées ou développées.

Méthodes quantiques

Depuis Schrödinger, la chimie quantique a connu le développement continu de méthodes de modélisation de plus en plus précises ou utilisables pour des systèmes de tailles croissantes. Elles rendent possible le calcul numérique presque exact des noyaux, voire des électrons, pour des systèmes constitués de quelques atomes. 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) pour des molécules contenant quelques dizaines d'atomes. Le travail décrit dans ce mémoire n'a pas pour but d'atteindre une précision ultime, mais utilisera à l'occasion de telles méthodes quantiques sophistiquées pour obtenir des valeurs de référence.

Les méthodes quantiques de calcul des propriétés électroniques des molécules se déclinent

aujourd'hui en deux grandes catégories : celles qui ont pour objet le calcul de la fonction d'onde (WFT pour *Wave Function Theory*) et celles qui visent la densité (DFT pour *Density Functional Theory*)¹. Les méthodes WFT utilisent presque toutes l'approximation de Hartree-Fock comme première étape, dans laquelle la répulsion coulombienne entre les électrons est moyennée. Certaines traitent les autres termes (la « corrélation électronique ») par perturbation de la solution Hartree-Fock. Le niveau le plus utilisé est le second ordre de perturbation sur l'énergie (MP2), 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. Nous utiliserons très largement la méthode MP2 dans ce travail. Les méthodes variationnelles permettent de traiter exactement les couplages entre configurations, et d'inclure des degrés d'excitation aussi développés que souhaité pour s'approcher de la solution exacte, mais le coût de leur mise en œuvre devient rapidement explosif. La formulation qui s'est imposée depuis une dizaine d'années pour inclure des excitations d'ordre élevé est dite « Coupled Cluster », dans laquelle une expression exponentielle de la fonction d'onde est utilisée. La fonction d'onde de l'état fondamental pour cette méthode s'écrit selon l'équation :

$$|\Psi_{CC}\rangle = e^{\hat{T}} |\Phi_o\rangle \quad (1.1)$$

avec Φ_o la fonction d'onde Hartree-Fock et \hat{T} l'opérateur qui permet d'inclure toutes les excitations des électrons des orbitales occupées vers les orbitales virtuelles. L'opérateur \hat{T} s'écrit selon l'équation :

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_n \quad (1.2)$$

où n est le nombre d'électrons du système. En pratique, l'approximation appelée CCSD, $\hat{T} = \hat{T}_1 + \hat{T}_2$, est la plus courante avec \hat{T}_1 l'opérateur des mono-excitations et \hat{T}_2 l'opérateur des di-excitations. Afin de traiter des systèmes de plus grande taille, la méthode CC2 a été implémentée comme une approximation du CCSD en tronquant le terme des di-excitations. C'est le niveau CC2 que nous utiliserons dans ce qui suit pour obtenir des énergies de référence. Il est aujourd'hui praticable sans difficulté pour les systèmes tels que $G_{10}\text{-Na}^+$, parmi les plus grands de ceux que nous avons considérés.

Les méthodes DFT ont été développées pour contourner le mur auquel les méthodes WFT

semblaient faire face il y a une vingtaine d'années pour le traitement de molécules de grande taille. La recherche de l'énergie sous la forme d'une fonctionnelle de la densité électronique, dépendant de seulement 3 variables spatiales quelle que soit la taille du système considéré alors que la fonction d'onde dépend de $3N$ variables d'espace pour un système à N atomes, est particulièrement séduisante même si l'expression et la nature de cette fonctionnelle sont toujours inconnues après plus de 40 ans de travaux. Les méthodes DFT sont aujourd'hui très utilisées parce qu'elles fournissent un compromis attrayant entre coût et précision. Elles ont ainsi permis d'étendre les domaines d'application de la chimie quantique de manière spectaculaire. Pour autant, il est apparu depuis une dizaine d'années que les fonctionnelles les plus utilisées, en particulier B3LYP, rencontrent des problèmes systématiques². Il s'agit entre autres des interactions non covalentes qui sont cruciales pour les biomolécules qui nous intéressent. Ceci a conduit à un foisonnement de nouvelles fonctionnelles, dont les objectifs principaux sont d'éliminer le problème de la « self-interaction » des électrons, d'optimiser l'introduction d'échange exact et d'introduire des termes modélisant les forces de dispersion absents des premières générations de fonctionnelles. Dans ce travail, nous avons utilisé principalement la fonctionnelle B3LYP complétée par un terme de dispersion empirique en $1/r^6$ (noté B3LYP-D) après l'avoir calibrée par comparaison aux méthodes WFT, post-Hartree-Fock. Nous avons également réalisé une évaluation des capacités de nombreuses fonctionnelles récentes à reproduire la structure, l'énergétique et le spectre infrarouge d'un peptide sodé, notre objectif principal.

L'utilisation effective de toutes les méthodes quantiques requiert le choix de bases de fonctions atomiques. Pour les méthodes WFT et DFT, nous avons presque toujours utilisé des bases de qualité double- ζ avec polarisation pour les couches de valence, pour l'optimisation des géométries et le calcul de spectres IR statiques. Pour obtenir des résultats énergétiques fiables, des bases plus étendues de qualité triple- ζ avec double polarisation pour les couches de valence ont été le plus souvent choisies. Il peut s'agir des bases cc-pVTZ ou TZVPP. Nous avons utilisé aussi souvent que possible l'approximation de résolution de l'identité (RI)³⁻⁷ pour les calculs d'énergie et les optimisations de géométrie. Cette méthode permet de réduire grandement les temps de calcul grâce à une transformation des intégrales biélectroniques à 4 centres en intégrales à 3 centres, au moyen d'un changement de base et de l'utilisation de bases spécifiques dites auxiliaires.

En général, 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 suivants : Gaussian⁸, Turbomole⁹ et GAMESS^{10,11}.

Méthodes classiques

En contraste avec la mécanique quantique, la mécanique moléculaire utilise des modèles de physique classique, plus simples pour calculer l'énergie potentielle comme la somme de termes décrivant les interactions considérées entre les atomes du système (équ. 1.3). La plupart du temps, cette somme est effectuée sur des potentiels de liaison covalente, E_{str} , d'angle de valence pour décrire l'énergétique de déformation angulaire, E_{bnd} , de van der Waals, E_{vdW} , électrostatiques, E_{elec} et de torsion, $E_{torsion}$ pour compléter l'énergie des interactions 1-4 entre deux atomes séparés par 3 liaisons covalentes quand les potentiels E_{vdW} et E_{elec} ne suffisent pas.

$$E_{totale} = \sum E_{str} + \sum E_{bnd} + \sum E_{torsion} + \sum E_{vdW} + \sum E_{elec} \quad (1.3)$$

On distingue ces potentiels ou interactions en deux catégories. E_{str} , E_{bnd} et $E_{torsion}$ n'existent qu'entre des paires d'atomes reliés par une, deux ou trois liaisons covalentes. De ce fait, on parle d'interactions covalentes ou de termes liés. D'un autre côté, les potentiels E_{vdW} et E_{elec} existent, certes, entre des atomes d'une même molécule mais aussi entre des atomes qui ne sont pas reliés, même indirectement par des liaisons covalentes. Ces composantes de l'énergie totale du système sont donc considérées comme des termes non-liés ou non-covalents. Alors que les interactions liées sont exclusivement intramoléculaires, les termes non-liés participent au calcul de l'énergétique intra- et intermoléculaire.

Ces modèles paramétrés sont plus simples et plus rapides à appliquer que la résolution des équations de Kohn-Sham ou de Schrödinger, même approchée. Les valeurs des différents paramètres utilisés sont obtenues sur la base de données expérimentales ou de résultats de calculs de modélisation moléculaire de niveau élevé, généralement de mécanique quantique

ab initio. On parle de champ de force. Ce terme définit le formalisme du calcul de l'énergie potentielle à lui seul ou en combinaison avec un jeu de paramètres.

Il existe un nombre important de champs de force plus ou moins spécialisés dans un ou des domaines d'application, qu'il s'agisse du type de systèmes étudiés, des propriétés à déterminer ou du genre de simulation effectuée.

En ce qui concerne la modélisation des biomolécules et en particulier des protéines, parmi les plus utilisés, on trouve :

- AMBER¹²⁻¹⁶ (Assisted Model Building with Energy Refinement) dont les charges atomiques ont été optimisées pour reproduire le potentiel électrostatique et qui possède un terme énergétique pour traiter explicitement les liaisons hydrogène ;
- OPLS-AA¹⁷ (Optimized Potential For Liquid Simulations - All Atoms), utilisant une partie du formalisme d'AMBER pour les termes liés, originellement développé pour reproduire les densités des liquides de molécules organiques et étendu aux peptides ;
- CHARMM^{18,19} (Chemistry at HARvard using Molecular Mechanics) dont le formalisme est très proche de celui d'AMBER. CHARMM en diffère notamment par la présence d'un terme de couplage Urey-Bradley et par la paramétrisation des charges atomiques basée sur des charges de Mulliken.

Il existe d'autres champs de force non-spécialisés pour les systèmes biologiques. Citons UFF²⁰ (United Force Field), un champ de force très général dont la précision est limitée mais qui dispose de paramètres pour l'ensemble de la classification périodique des éléments, ainsi que de règles pour les combiner entre eux.

Allinger et coll. ont développé une série de champs de force principalement destinés aux molécules organiques. Les champs de force MM1²¹, MM2²² et MM3^{23,24} les potentiels covalents de l'énergie incluent des termes supplémentaires d'anharmonicité et certains d'entre ces potentiels sont couplés entre eux. Nous reverrons plus en détail le contenu de MM3 à l'occasion de la description du champ de force AMOEBA.

Les champs de force cités jusqu'à ce point sont parfois qualifiés de classiques ou encore de champs de force de première génération. Bien qu'elle soit importante pour des systèmes contenant des ions ou des macromolécules, la polarisation n'est pas explicitement traitée par ces champs de force. Une autre génération de champs de force a été développée pour prendre

en compte cet effet de déformation d'une densité électronique par une ou des charges.

Comme les champs de force sont dans leur grande majorité développés pour effectuer des simulations en phase condensée, les effets de polarisation du solvant font l'objet de nombreuses études. L'effet polarisant du solvant peut être vu en première approximation comme celui d'un « continuum » diélectrique^{25,26}. Ces modèles ne tiennent pas compte d'effets de polarisation causés par des molécules discrètes du solvant. Étant donné les caractéristiques de l'eau et son importance en chimie, comme en biologie, il est intéressant d'en décrire le pouvoir polarisant et la polarisabilité. Ainsi, d'autres modèles, polarisables, ont été développés²⁷⁻³² dont nous ne discuterons pas s'ils ne sont pas étendus à des systèmes d'intérêt biologique.

L'énergie d'interaction entre deux entités (molécules) A et B peut être perçue soit avec une approche supermoléculaire dans laquelle $E_{int} = E_{AB} - E_A - E_B$, soit par la théorie de perturbation. L'énergie intermoléculaire, au premier ordre de la théorie de perturbation est purement électrostatique. C'est la somme des interactions des multipôles du système et des intégrales de pénétration. Au deuxième ordre, il faut ajouter des termes de polarisation et de dispersion. La polarisation n'est pas un terme additif, par opposition aux interactions de paires (forces coulombiennes, de vdW, ...). Jeziorski et coll. ont introduit un opérateur d'antisymétrisation intermoléculaire pour satisfaire au principe de Pauli dans leur théorie de perturbation adaptée à la symétrie (SAPT)³³. Cela permet de considérer des termes d'échange électronique à chaque ordre de perturbation. En mécanique moléculaire, l'échange et la répulsion interatomique sont assurés par les termes de van der Waals.

Lorsque l'on prend en compte les états à transfert de charge du système, typiquement les états A^+B^- et A^-B^+ , une énergie de transfert de charge vient s'ajouter à l'énergie d'interaction comme un terme à n corps, fortement anisotrope.

La composante électrostatique de l'énergie est calculée entre des charges ponctuelles dans la plupart des champs de force dits classiques. Certains champs de force font intervenir des multipôles d'ordre supérieur comme des dipôles ou des quadripôles. C'est le cas de MM3 dans lequel seuls quelques atomes portent des charges, en général entières ou semi-entières. Les interactions intermoléculaires ont été étudiées par Claverie et coll. en approximant la distribution de charge par un ensemble de charges et dipôles ponctuels et polarisables³⁴. Le

groupe de Claverie a présenté une méthode utilisant des charges et dipôles ponctuels pour modéliser la topologie électrostatique au voisinage d'une molécule³⁵. Parmi les méthodes et outils développés pour extraire multipôles et polarisabilités à partir de données quantiques, on peut citer ceux du groupe de Stone^{36,37}, d'Ángyán et coll.³⁸ ou encore de Sokalski et coll.³⁹.

Les modèles de dipôles ponctuels induits d'Applequist⁴⁰ présentent le risque de « polarisation catastrophe », c'est-à-dire la non-convergence de l'énergie de polarisation entre deux atomes à courte distance l'un de l'autre. Pour éviter cette divergence, Tholé a introduit un facteur d'amortissement⁴¹. Le modèle de Tholé connaît des variantes⁴². Le facteur d'amortissement de Tholé est utilisé dans différents champs de force comme celui développé par Soteras et coll.⁴³.

À l'exception de NEMO⁴⁴, les multipôles utilisés dans la plupart des champs de forces polarisables sont limités à la charge, au dipôle et au quadrupôle. On peut noter le développement d'« AMBER polarizable ff02 »⁴⁵ qui ajoute à AMBER des dipôles polarisables et utilise des charges ajustées sur le potentiel électrostatique. Le champ de force polarisable d'Allinger, MM4^{46,47} exploite des dipôles polarisables sur les liaisons. Le champ de force TCPEp (*Topological and Classical Polarization Effects for proteins*) de Masella et coll.⁴⁸⁻⁵⁰ est utilisé pour des systèmes de grande taille. Dans ce champ de force, l'expansion multipolaire est limitée, en dehors des charges atomiques, à des dipôles polarisables attribuée à chaque atome « lourd », c'est-à-dire autre qu'un hydrogène.

L'oscillateur de Drude⁵¹, ou *Shell model*, constitue un autre modèle de polarisation dans lequel chaque atome est divisé en un *core*, centré sur le noyau et un *shell* qui peut s'en éloigner. Chacune de ces parties est caractérisée par une charge. À l'image de la déformation de densités électroniques en mécanique quantique, le *shell* oscille autour du *core* en fonction des influences dues à l'environnement. Cette méthode permet de prendre en compte de manière relativement simple les effets de polarisation. Elle a en outre le bénéfice d'éviter le phénomène de « polarization catastrophe ». Elle a par exemple été utilisée dans une version polarisable de CHARMM^{52,53}.

Le champ de force SIBFA (*Sum of Interactions Between Fragments Ab initio computed*) a été développé par Gresh et coll.⁵⁴ incluant un traitement complet des effets à n corps,

y compris le transfert de charge. Le point fondamental des champs de force polarisables est la découpe de l'énergie de référence, *ab initio* Hartree-Fock pour SIBFA, en termes transférables. Les termes ont tous une signification physique, ce qui aide à la compréhension des succès et à la limitation des problèmes. Les sites des multipôles, jusqu'aux quadrupôles, sont disposés aux noyaux, aux milieux des liaisons et aux paires libres. De plus, les polarisabilités dipolaires sont anisotropes dans SIBFA. Ce champ de force permettrait vraisemblablement de traiter de manière très satisfaisante les peptides attachés à un cation métallique auxquels nous nous intéressons. Cependant, l'utilisation de fragments rigides et le fait que la dynamique moléculaire avec SIBFA est en cours de développement sont très limitants pour les applications envisagées à la spectroscopie vibrationnelle. En 2009, Sakharov et Lim ont publié un autre champ de force polarisable avec transfert de charge CTPOL⁵⁵.

Le modèle de champ de force GEM⁵⁶ (« Gaussian Electrostatic Model ») comme GMM⁵⁷ (« Gaussian Multipole Model ») calcule l'électrostatique, la polarisation et le transfert de charge dans l'esprit de SIBFA mais en utilisant des fonctions gaussiennes de type s pour représenter la densité électronique. Ces gaussiennes permettent de prendre en compte des effets quantiques à courte distance comme la répulsion d'échange. En 2007, X-Pol développé par Gao et coll.⁵⁸ permet aussi d'obtenir le champ de force directement à partir de la structure électronique, comme par exemple un calcul DFT.

Le modèle de charges fluctuantes⁵⁹ est basé sur l'ajustement des charges en fonction de leur environnement chimique et permet de prendre en compte des effets de transfert de charge. Ce modèle a notamment été utilisé pour étendre CHARMM en CHARMM-FQ^{60,61}. 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 charge⁶².

Dans les champs de force non polarisables, l'absence de polarisation est en partie compensée par une sous-estimation de la répulsion⁶³ et une surestimation des charges atomiques afin d'exagérer les moments dipolaires locaux et globaux des molécules. Ainsi, dans les champs de force dits classiques, certains résultats satisfaisants sont dus à des compensations d'erreur et on peut redouter que le manque de sens physique causé par ces compensations conduise dans certaines situations à des cas pathologiques. La polarisation électronique peut s'avérer capitale pour traiter les interactions biomoléculaires⁶⁴ ou pour traiter différents états

de la matière avec un même champ de force⁶⁵. Pour les besoins de nos études nous avons opté pour le champ de force AMOEBA qui fait l'objet d'une description détaillée dans le chapitre suivant.

Bibliographie

- [1] R. J. HINDE : *Quantum Chemistry, 5th Edition (by Ira N. Levine)*, volume 77 de *J. Chem. Educ.* Prentice Hall, Department of Chemistry, University of Tennessee, Knoxville, TN 37996, 2000.
- [2] Y. ZHAO et D. G. TRUHLAR : Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.*, 41:157–167, 2008.
- [3] K. EICKORN, O. TREUTLER, H. ÖHM, M. HÄSER et R. AHLRICH : Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.*, 240:283–290, 1995.
- [4] K. EICKORN, O. TREUTLER, H. ÖHM, M. HÄSER et R. AHLRICH : Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.*, 242:652–660, 1995.
- [5] K. EICKORN, F. WEIGEND, O. TREUTLER et R. AHLRICH : Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theo. Chem. Acc.*, 97:119–124, 1997.
- [6] F. WEIGEND et M. HÄSER : RI-MP2 : first derivatives and global consistency. *Theor. Chem. Acc.*, 97:331, 1997.
- [7] O. CHRISTIANSEN, H. KOCH et P. JØRGENSEN : For the approximate coupled-cluster singles-and-doubles model CC2. *Chem. Phys. Lett.*, 243:409–418, 1995.
- [8] M. J. FRISCH, G. W. TRUCKS, H. B. SCHLEGEL, G. E. SCUSERIA, M. A. ROBB, J. R. CHEESEMAN, J. A. MONTGOMERY, Jr., T. VREVEN, K. N. KUDIN, J. C. BURANT, J. M. MILLAM, S. S. IYENGAR, J. TOMASI, V. BARONE, B. MENNUCCI, M. COSSI, G. SCALMANI, N. REGA, G. A. PETERSSON, H. NAKATSUJI, M. HADA, M. EHARA, K. TOYOTA, R. FUKUDA, J. HASEGAWA, M. ISHIDA, T. NAKAJIMA, Y. HONDA, O. KITAO, H. NAKAI, M. KLENE, X. LI, J. E. KNOX, H. P. HRATCHIAN, J. B. CROSS, V. BAKKEN, C. ADAMO, J. JARAMILLO, R. GOMPERTS, R. E. STRAT-

- MANN, O. YAZYEV, A. J. AUSTIN, R. CAMMI, C. POMELLI, J. W. OCHTERSKI, P. Y. AYALA, K. MOROKUMA, G. A. VOTH, P. SALVADOR, J. J. DANNENBERG, V. G. ZAKRZEWSKI, S. DAPPRICH, A. D. DANIELS, M. C. STRAIN, O. FARKAS, D. K. MALICK, A. D. RABUCK, K. RAGHAVACHARI, J. B. FORESMAN, J. V. ORTIZ, Q. CUI, A. G. BABOUL, S. CLIFFORD, J. CIOSLOWSKI, B. B. STEFANOV, G. LIU, A. LIASHENKO, P. PISKORZ, I. KOMAROMI, R. L. MARTIN, D. J. FOX, T. KEITH, M. A. AL-LAHAM, C. Y. PENG, A. NANAYAKKARA, M. CHALLACOMBE, P. M. W. GILL, B. JOHNSON, W. CHEN, M. W. WONG, C. GONZALEZ et J. A. POPLE : Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.
- [9] TURBOMOLE V6.1 2009, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007 ; <http://www.turbomole.com>.
- [10] M. W. SCHMIDT, K. K. BALDRIDGE, J. A. BOATZ, S. T. ELBERT, M. S. GORDON, J. H. JENSEN, S. KOSEKI, N. MATSUNAGA, K. A. NGUYEN, S. SU, T. L. WINDUS, M. DUPUIS et J. A. MONTGOMERY : General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.*, 14:1347–1363, 1993.
- [11] M. S. GORDON et M. W. SCHMIDT : *Advances in electronic structure theory : GAMES a decade later*, pages 1167–1189. Elsevier, Amsterdam, 2005.
- [12] W. S. ROSS et C. C. HARDIN : Ion-Induced Stabilization of the G-DNA Quadruplex : Free Energy Perturbation Studies. *J. Am. Chem. Soc.*, 116:6070–6080, 1994.
- [13] W. D. CORNELL, P. CIEPLAK, C. I. BAYLY, I. R. GOULD, Jr MERZ, K. M., D. M. FERGUSON, D. C. SPELLMEYER, T. FOX, J. W. CALDWELL et P. A. KOLLMAN : A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.
- [14] P. KOLLMAN, R. DIXON, W. CORNELL, T. FOX, C. CHIPOT et A. POHORILLE : *The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab Initio calculations and experimental data*, volume 3, pages 83–96. Springer, 1997.
- [15] T. E. CHEATHAM, P. CIEPLAK et P. A. KOLLMAN : A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, 16:845–862, 1999.

- [16] J. WANG, P. CIEPLAK et P. A. KOLLMAN : How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules ? *J. Comput. Chem.*, 21:1049–1074, 2000.
- [17] W. L. JORGENSEN, D. S. MAXWELL et J. TIRADO-RIVES : Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.
- [18] A. D. MACKERELL, D. BASHFORD, BELLOTT, R. L. DUNBRACK, J. D. EVANSECK, M. J. FIELD, S. FISCHER, J. GAO, H. GUO, S. HA, D. JOSEPH-MCCARTHY, L. KUCHNIR, K. KUCZERA, F. T. K. LAU, C. MATTOS, S. MICHNICK, T. NGO, D. T. NGUYEN, B. PRODHOM, W. E. REIHER, B. ROUX, M. SCHLENKRICH, J. C. SMITH, R. STOTE, J. STRAUB, M. WATANABE, J. WIORKIEWICZ-KUCZERA, D. YIN et M. KARPLUS : All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.
- [19] N. FOLOPPE et Jr MACKERELL, A. D. : All-atom empirical force field for nucleic acids : I. parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.*, 21:86–104, 2000.
- [20] A. K. RAPPÉ, C. J. CASEWIT, K. S. COLWELL, III GODDARD, W. A. et W. M. SKIFF : UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.*, 114:10024–10035, 1992.
- [21] N. L. ALLINGER, M. T. TRIBBLE, M. A. MILLER et D. H. WERTZ : Conformational analysis. LXIX. An improved force field for the calculation of the structures and energies of hydrocarbons. *J. Amer. Chem. Soc.*, 93:1637–48, 1971.
- [22] N. L. ALLINGER : Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.*, 99:8127–8134, 1977.
- [23] N. L. ALLINGER, Y. H. YUH et J. H. LII : Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.*, 111:8551–8566, 1989.
- [24] J. H. LII et N. L. ALLINGER : Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics. *J. Am. Chem. Soc.*, 111:8556–8575, 1989.

- [25] M. K. GILSON et B. HONIG : Calculation of the total electrostatic energy of a macromolecular system : Solvation energies, binding energies, and conformational analysis. *Proteins Struct. Funct. Bioinf.*, 4:7–18, 1988.
- [26] A. ONUFRIEV, D. BASHFORD et D. A. CASE : Effective Born radii in the generalized Born approximation : The importance of being perfect. *J. Comp. Chem.*, 23:1297–1304, 2002.
- [27] M. MASELLA et J.-P. FLAMENT : Hydrogen bond network fluctuations in small $(\text{H}_2\text{O})_n$ clusters ($n=8-12-24$). *J. Chem. Phys.*, 111:5081–5090, 1999.
- [28] C. J. BURNHAM, J. LI, S. S. XANTHEAS et M. LESLIE : The parametrization of a Thole-type all-atom polarizable water model from first principles and its application to the study of water clusters ($n=2-21$) and the phonon spectrum of ice I_h . *J. Chem. Phys.*, 110:4566–4581, 1999.
- [29] S. O. YESYLEVSKYY, L. V. SCHÄFER, D. SENGUPTA et S. J. MARRINK : Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.*, 2010. à paraître.
- [30] E. HARDER, B. KIM et B. J. FRIESNER, R. A. and Berne : Efficient Simulation Method for Polarizable Protein Force Fields : Application to the Simulation of BPTI in Liquid Water. *J. Chem. Theory Comput.*, 1:169–180, 2005.
- [31] F. ARCHAMBAULT, C. CHIPOT, I. SOTERAS, F. J. LUQUE, K. SCHULTEN et F. DEHEZ : Polarizable Intermolecular Potentials for Water and Benzene Interacting with Halide and Metal Ions. *J. Chem. Theory Comput.*, 5:3022–3031, 2009.
- [32] W. XIE, J. PU, A. D. MACKERELL et J. GAO : Development of a Polarizable Intermolecular Potential Function (PIPF) for Liquid Amides and Alkanes. *J. Chem. Theory Comput.*, 3:1878–1889, 2007.
- [33] B. JEZIORSKI, R. MOSZYNSKI et K. SZALEWICZ : Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.*, 94:1887–1930, 1994.
- [34] P. CLAVERIE et R. REIN : Theory of intermolecular interactions : The long range terms in the dipole-dipole, monopoles-dipole, and monopoles-bond polarizabilities approximations. *I. J. Quantum Chem.*, 3:537–551, 2004.

- [35] F. VIGNÉ-MAEDER et P. CLAVERIE : The exact multicenter multipolar part of a molecular charge distribution and its simplified representations. *J. Chem. Phys.*, 88:4934–4948, 1988.
- [36] A. J. STONE : Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.*, 83:233–239, 1981.
- [37] A. J. STONE et M. ALDERTONA : Distributed multipole analysis Methods and applications. *Mol. Phys.*, 56:1047–1064, 1985.
- [38] J. G. ÁNGYÁN, C. CHIPOT, F. DEHEZ, C. HÄTTIG, G. JANSEN et C. MILLOT : OPEP : A tool for the optimal partitioning of electric properties. *J. Comput. Chem.*, 24:997–1008, 2003.
- [39] W. A. SOKALSKI et R. A. POIRIER : Cumulative atomic multipole representation of the molecular charge distribution and its basis set dependence. *Chem. Phys. Lett.*, 98:86–92, 1983.
- [40] J. APPLEQUIST, J. R. CARL et K.-K. FUNG : Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.*, 94:2952–2960, 1972.
- [41] B. T. THOLÉ : Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.*, 59:341 – 350, 1981.
- [42] P. Th. van DUIJNEN et M. SWART : Molecular and Atomic Polarizabilities : Thole's Model Revisited. *J. Phys. Chem. A*, 102:2399–2407, 1998.
- [43] I. SOTERAS, C. CURUTCHET, A. BIDON-CHANAL, F. DEHEZ, J. G. ÁNGYÁN, M. OROZCO, C. CHIPOT et F.J. LUQUE : Derivation of Distributed Models of Atomic Polarizability for Molecular Simulations. *J. Chem. Theory Comput.*, 3:1901–1913, 2007.
- [44] J. M. HERMIDA-RAMON, S. BRDARSKI, G. KARLSTROM et U. BERG : Inter- and intramolecular potential for the N -formylglycinamide-water system. A comparison between theoretical modeling and empirical force fields. *J. Comput. Chem.*, 24:161–76, 2003.
- [45] P. CIEPLAK, J. W. CALDWELL et P. A. KOLLMAN : Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation : aqueous solution free energies of methanol and N-methyl acetamide,

- nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comput. Chem.*, 22:1048–1057, 2001.
- [46] K. CHEN, J. LII, Y. FAN et N. L. ALLINGER : Molecular mechanics (MM4) study of amines. *J. Comput. Chem.*, 28:2391–2412, 2007.
- [47] N. L. ALLINGER, K. CHEN, J. LII et K. A. DURKIN : Alcohols, ethers, carbohydrates, and related compounds. I. The MM4 force field for simple compounds. *J. Comput. Chem.*, 24:1447–72, 2003.
- [48] M. MASELLA et P. CUNIASSE : A many-body model to study proteins. I. Applications to ML_n^{m+} complexes, $M^{m+} = Li^+, Na^+, K^+, Mg^{2+}, Ca^{2+},$ and Zn^{2+} , $L = H_2O, CH_3OH, HCONH_2$, $n = 1-6$, and to small hydrogen bonded systems. *J. Chem. Phys.*, 119:1866–1873, 2003.
- [49] P. CUNIASSE et M. MASELLA : A many-body model to study proteins. II. Incidence of many-body polarization effects on the interaction of the calmodulin protein with four Ca^{2+} dications and with a target enzyme peptide. *J. Chem. Phys.*, 119:1874–1878, 2003.
- [50] M. MASELLA, D. BORGIS et P. CUNIASSE : Combining a polarizable force-field and a coarse-grained polarizable solvent model : application to long dynamics simulations of bovine pancreatic trypsin inhibitor. *J. Comput. Chem.*, 29:1707–24, 2008.
- [51] S. W. RICK et S. J. STUART : *Potentials and Algorithms for Incorporating Polarizability in Computer Simulations*, volume 18, pages 89–146. Wiley, 2002.
- [52] G. LAMOUREUX et B. ROUX : Modeling induced polarization with classical Drude oscillators : Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.*, 119:3025–3039, 2003.
- [53] G. LAMOUREUX, E. HARDER, I.V. VOROBYOV, B. ROUX et A. D. MACKERELL : A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.*, 418:245–9, 2006.
- [54] N. GRESH, G. A. CISNEROS, T. A. DARDEN et J.-P. PIQUEMAL : Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.*, 3:1960–1986, 2007.

- [55] D. V. SAKHAROV et C. LIM : Force fields including charge transfer and local polarization effects : Application to proteins containing multi/heavy metal ions. *J. Comput. Chem.*, 30:191–202, 2009.
- [56] M. MASIA, M. PROBST et R. REY : On the performance of molecular polarization methods. II. Water and carbon tetrachloride close to a cation. *J. Chem. Phys.*, 123:164505, 2005.
- [57] J.-P. PIQUEMAL, G. A. CISNEROS, P. REINHARDT, N. GRESH et T. A. DARDEN : Towards a force field based on density fitting. *J. Chem. Phys.*, 124:104101, 2006.
- [58] W. XIE et J. GAO : Design of a Next Generation Force Field : The X-POL Potential. *J. Chem. Theory Comput.*, 3:1890–1900, 2007.
- [59] A. K. RAPPÉ et III GODDARD, W. A. : Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.*, 95:3358–3363, 1991.
- [60] S. PATEL et C. L. BROOKS : CHARMM fluctuating charge force field for proteins : I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.*, 25:1–16, 2004.
- [61] S. PATEL, A. D. MACKERELL et C. L. BROOKS : CHARMM fluctuating charge force field for proteins : II Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.*, 25:1504–1514, 2004.
- [62] H. A. STERN, G. A. KAMINSKI, J. L. BANKS, R. ZHOU, B. J. BERNE et R. A. FRIESNER : Fluctuating Charge, Polarizable Dipole, and Combined Models : Parameterization from ab Initio Quantum Chemistry. *J. Phys. Chem. B*, 103:4730–4737, 1999.
- [63] F. DEHEZ, J. G. ÁNGYÁN, I. SOTERAS GUTIÉRREZ, F. J. LUQUE, K. SCHULTEN et C. CHIPOT : Modeling Induction Phenomena in Intermolecular Interactions with an Ab Initio Force Field. *J. Chem. Theory Comput.*, 3:1914–1926, 2007.
- [64] D. JIAO, P. A. GOLUBKOV, T. A. DARDEN et P. REN : Calculation of protein-ligand binding free energy by using a polarizable potential. *PNAS*, 105:6290–6295, 2008.
- [65] P. CIEPLAK, P. A. KOLLMAN et T. LYBRAND : A new water potential including polarization : Application to gas-phase, liquid, and crystal properties of water. *J. Chem. Phys.*, 92:6755–6760, 1990.

Chapitre 2

AMOEBA, un champ de force polarisable flexible

2.1 Généralités sur AMOEBA

Le champ de force polarisable AMOEBA (*Atomic Multipole iOptimized Energetics for Biomolecular Applications*) est développé par Ponder et coll.^{1,2}. Nous l'avons choisi pour ses capacités à reproduire diverses propriétés physico-chimiques en phase gazeuse et en phase condensée et aussi pour sa complète flexibilité intramoléculaire qui est nécessaire pour le calcul de spectres vibrationnels. Une de ses principales caractéristiques est que la construction du champ de forces repose sur un ensemble de termes décrivant chacun des phénomènes physiques impliqués dans les interactions, en particulier les effets à n corps. La transférabilité de ses paramètres a permis notamment de reproduire les énergies libres de solvation de cations³⁻⁵ tels que K^+ , Na^+ , Mg^{2+} , Ca^{2+} et Zn^{2+} . AMOEBA a aussi été comparé à d'autres champs de forces par Jensen et coll.^{6,7} pour reproduire des énergies conformationnelles d'acides aminés calculées en chimie quantique. Il a été montré que les champs de force 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%, en particulier grâce à la prise en compte des effets de polarisation.

Nous avons vu qu'il existe d'autres champs de force polarisables plus élaborés dans leur formalisme qu'AMOEBa. Nous avons cependant choisi d'utiliser AMOEBA car il a été développé pour des applications biologiques et il inclut une flexibilité structurale complète, nécessaire au calcul de spectres vibrationnels de ces molécules. De plus, il est largement distribué avec la suite logicielle Tinker⁸ qui permet de l'utiliser pour des simulations, entre autres, de dynamique moléculaire. En outre, le code source de Tinker, comme les paramètres d'AMOEBa, est accessible et modifiable selon les besoins.

2.1.1 Formalisme et philosophie d'AMOEBa

La philosophie générale du champ de force AMOEBA est de décrire analytiquement les forces intermoléculaires à partir de la théorie des perturbations au deuxième ordre⁹. Ainsi l'énergie d'interaction est écrite comme un ensemble de termes ayant chacun un sens physique.

$$E_{int} = E_{elec} + E_{rep-ech} + E_{pol} + E_{pol-ech} + E_{disp} + E_{disp-ech}$$

avec E_{elec} l'énergie électrostatique, $E_{rep-ech}$ l'énergie de répulsion d'échange, E_{pol} l'énergie de polarisation, $E_{pol-ech}$ l'énergie de polarisation-échange, E_{disp} l'énergie de dispersion, $E_{disp-ech}$ l'énergie de dispersion-échange.

Dans le but de traiter efficacement des systèmes biologiques en interaction avec leur environnement, AMOEBA se limite à l'inclusion des termes d'énergie électrostatique, de polarisation, de dispersion et de répulsion d'échange dans le traitement des interactions non liées. Il est à noter que les phénomènes de transfert de charge ne sont pas explicitement pris en compte dans le champ de force. Afin de modéliser le transfert de charge du gadolinium micro-solvaté¹⁰, Dognon et coll. ont utilisé AMOEBA en ajoutant un modèle à charges fluctuantes pour le transfert de charge. Des études sont en cours afin de dériver les expressions analytiques et d'implémenter ce terme dans AMOEBA¹¹.

L'énergie électrostatique

Pour calculer le terme d'énergie électrostatique, AMOEBA utilise des multipôles pour chaque atome qui n'est donc plus traité uniquement comme une charge ponctuelle (comme

dans Amber¹² ou CHARMM¹³) mais comme une charge, un dipôle et un quadrupôle permanents. Il s'agit de multipôles atomiques par opposition à la définition de dipôles de liaison comme dans MM2 ou MM3. Ces paramètres sont extraits de fonctions d'onde *ab initio*. AMOEBA peut ainsi mieux représenter les potentiels électrostatiques issus de la mécanique quantique¹⁴. Une description détaillée des termes électrostatiques est donné au paragraphe 2.1.2.

Le terme explicite de polarisation

AMOEBA traite explicitement les effets de polarisation à n corps. Le dipôle induit sur chaque site i est :

$$\mu_{i,\alpha}^{ind} = \alpha_i E_{i,\alpha} \quad (\alpha \in \{x,y,z\}) \quad (2.1)$$

α_i étant la polarisabilité atomique et $E_{i,\alpha}$ la composante selon l'axe α du champ électrique sur l'atome i . Le terme de polarisation est écrit comme une somme de termes de réponse des multipôles de l'atome au champ électrique créé par les atomes non liés et de termes d'interaction entre les dipôles atomiques induits :

$$\mu_{i,\alpha}^{ind} = \alpha_i \sum_{\{j\}} T_{\alpha}^{i,j} M_j + \alpha_i \sum_{\{j'\}} T_{\alpha\beta}^{i,j'} \mu_{j',\beta'}^{ind} \quad (2.2)$$

T étant la matrice d'interaction et M_j contient les composantes des moments multipolaires permanents. Le terme de polarisation doit donc être calculé de manière auto-cohérente. Pour remédier au risque de *polarization catastrophe*, à courte distance, un facteur d'amortissement de Tholé¹⁵ permet d'éviter la divergence de l'énergie de polarisation.

Les termes de répulsion d'échange et de dispersion

Les termes de répulsion d'échange et de dispersion représentent généralement dans un champ de forces les interactions de van der Waals (vdW). AMOEBA utilise un formalisme « buffered 14-7 » proposé par Thomas A. Halgren¹⁶. Ce potentiel s'est montré efficace pour les gaz rares. Il a aussi l'avantage de mieux reproduire à la fois des résultats de calculs *ab initio* en phase gazeuse et des grandeurs expérimentales en phase liquide qu'un potentiel standard 12-6 de Lennard-Jones que l'on peut trouver dans AMBER, par exemple.

L'équation 2.3 donne la contribution énergétique d'une interaction vdW dans ce formalisme pour des atomes i et j dont la distance relative à la séparation idéale est nommée ρ_{ij} . Le puits de potentiel ainsi décrit a une profondeur maximale ε_{ij} à la distance idéale R_{ij_0} .

Les règles de combinaison permettent de déterminer la profondeur d'un puits (Équ. 2.4)) ε_{ij} pour toute paire d'atomes i et j en fonction des paramètres définis pour chaque paire homoatomique ε_{ii} . L'équation 2.5 est utilisée pour calculer la séparation idéale R_{ij_0} en fonction des distances de minimum d'énergie R_{ii_0} déterminées pour chaque paire homoatomique.

$$E_{vdW}(ij) = \varepsilon_{ij} \left(\frac{1,07}{\rho_{ij} + 0,07} \right)^7 \left(\frac{1,12}{\rho_{ij}^7 + 0,12} - 2 \right) \quad (2.3)$$

$$\text{avec } \varepsilon_{ij} = \frac{4\varepsilon_{ii}\varepsilon_{jj}}{(\sqrt{\varepsilon_{ii}} + \sqrt{\varepsilon_{jj}})^2} \quad (2.4)$$

$$\text{et } \rho_{ij} = \frac{R_{ij}}{R_{ij_0}} \text{ tel que } R_{ij_0} = \frac{(R_{ii_0})^3 + (R_{jj_0})^3}{(R_{ii_0})^2 + (R_{jj_0})^2} \quad (2.5)$$

Dans AMOEBA, les paramètres de vdW sont définis par classe d'atome (définie au paragraphe 1.1.2).

Les interactions liées

AMOEBA dérive du champ de force MM3¹⁷⁻¹⁹ d'Allinger pour le traitement des interactions liées. Il en conserve le formalisme, notamment les composantes d'élongation (Équ. 2.6) celles de déformation angulaire (Équ. 2.7) et le terme de couplage *stretch-bend* (Équ. 2.8).

Si on considère la distance de liaison b et sa valeur idéale b_0 ainsi que la constante de force K_b , alors la composante d'élongation vaut :

$$E_{str} = K_b(b - b_0)^2 \left[1 - 2,55(b - b_0) + \frac{7}{12} 2,55^2 (b - b_0)^2 \right] \quad (2.6)$$

L'énergie provenant d'un angle déformé à la valeur θ par rapport à sa valeur idéale θ_0 avec la constante de force K_θ est donné par :

$$E_{bnd} = K_\theta(\theta - \theta_0)^2 \left[1 - 0,014(\theta - \theta_0) + 5,6 \times 10^{-5}(\theta - \theta_0)^2 - 7 \times 10^{-7}(\theta - \theta_0)^3 + 2,2 \times 10^{-8}(\theta - \theta_0)^4 \right] \quad (2.7)$$

Le couplage entre élongation et déformation angulaire dépend d'une constante de force $K_{str-bnd}$, de la déformation angulaire $(\theta - \theta_0)$ et des élongations $(b - b_0)$ et $(b' - b'_0)$ pour les deux liaisons formant l'angle considéré.

$$E_{str-bnd} = K_{str-bnd}[(b - b_0) + (b' - b'_0)](\theta - \theta_0) \quad (2.8)$$

Pour les atomes trivalents et « plans » (hybridation sp^2), le terme de déformation hors du plan, *out-of-plane bend*, provient lui aussi de MM3. Ce potentiel de Wilson-Decius-Cross est défini par le produit entre la constante de force K_χ et le carré de χ , l'angle dièdre formé par les 4 atomes considérés.

$$E_{oop} = K_\chi \chi^2 \quad (2.9)$$

Dans le formalisme du champ de force AMOEBA, la composante énergétique de torsion est définie comme le développement en série de Fourier d'une fonction périodique. Ce formalisme est fréquent dans les champs de force. Dans le cas d'AMOEBA, il se limite à 6 termes (Équ. 2.10). Seuls les 3 premiers sont utilisés en pratique.

$$E_{torsion}(\alpha) = \frac{1}{2} \sum_{n=1}^6 A_n [1 + \cos(n\alpha - \Phi_n)] \quad (2.10)$$

Dans ce développement, un coefficient A_n est défini pour chaque périodicité n à laquelle est aussi associée une phase Φ_n de valeur 0 ou 180°.

On peut noter que du fait des différences de traitement des interactions non liées, la composante de torsion en AMOEBA ne doit pas être équivalente à celle de MM3. C'est pourquoi les paramètres de torsion de ces deux champs de force sont différents bien que l'expression de $E_{torsion}$ leur soit commune.

2.1.2 Classes et types d'atomes

Comme pour d'autres champs de force, le formalisme d'AMOEBA^{1-3,20} propose des classes et des types d'atomes.

Le calcul des interactions liées (élongation, déformation angulaire, torsion...) et des interactions de van der Waals utilise des paramètres définis pour différentes classes d'atomes.

Bien qu'étant le même élément, deux atomes peuvent avoir des propriétés si différentes à cause de leur environnement qu'ils nécessitent des paramètres distincts : valeur de référence, constante de force, rayon de van der Waals, profondeur du puits de potentiel... Pour toute la partie électrostatique au sens large des interactions (coulombiennes, polarisation), les paramètres sont définis par type d'atomes afin de rendre mieux compte de l'environnement chimique. Les types d'atomes sont des sous-divisions des classes d'atomes : des atomes de types différents peuvent être de la même classe, alors que l'inverse n'est pas vrai. Par exemple, les carbones C_α des acides aminés alanine, sérine et cystéine appartiennent à la même classe mais ont chacun un type d'atome propre afin que la distribution multipolaire soit spécifique à l'environnement.

2.2 Utilisation d'AMOEBA pour des cas non standard, adaptation de paramètres

Un des inconvénients de tout champ de force est l'incomplétude de ses paramètres. Nous avons ainsi été amenés à étendre le jeu de paramètres d'AMOEBA en diverses occasions. Les champs de force biologiques comportent toujours des terminaisons de zwitterion : ammonium et carboxylate, qui ne sont pas adaptés aux acides aminés et peptides en phase gazeuse dont les terminaisons sont neutres : amine et acide carboxylique. Si AMOEBA comporte des paramètres pour les acides carboxyliques des chaînes latérales des acides aspartique et glutamique, il n'en a pas pour une amine primaire. Des charges atomiques ont d'abord été déterminées pour $-NH_2$ selon la procédure décrite dans la partie II, chapitre 6. Une meilleure description des terminaisons a ensuite été obtenue par ajustement sur leur densité électronique de la distribution de multipôles atomiques sur leur densité électronique. La procédure utilisée est décrite dans le paragraphe 2.1.2 de ce chapitre. Après ces modifications, l'étude des conformations de G_2-Na^+ et de leurs énergies relatives a montré que des problèmes subsistaient. Il s'est avéré qu'ils étaient dus à certaines torsions de la chaîne principale des peptides, pour lesquelles les paramètres étaient tous nuls : les $N-C_\alpha-C-N$ ainsi que le cas particulier de l'extrémité C-terminale $N-C_\alpha-C_{ter}-OH$. Nous avons donc optimisé ces paramètres selon la procédure décrite ci-dessous. Les résultats sont décrits dans la partie II, chapitre 6.

2.2.1 Exemple d'optimisation de paramètres de torsion

Le but de l'optimisation d'un paramètre est de rapprocher la surface d'énergie potentielle AMOEBA d'une surface d'énergie potentielle de référence. Le choix de cette dernière est donc crucial. Les paramètres initiaux d'AMOEBEA ayant été ajustés sur des calculs *ab initio* MP2, nous avons fait le même choix. Il s'agit en général de calculs MP2/aug-cc-pVTZ. La base TZVPP a été utilisée lorsque la construction de la surface de référence exigeait un nombre important de calculs ou des calculs difficilement réalisables avec la base aug-cc-pVTZ.

L'optimisation d'une fonction dépendant de paramètres peut être effectuée soit analytiquement, soit numériquement. L'optimisation correspond à la minimisation d'une fonction de score reflétant la différence entre la surface de potentiel à optimiser et la surface de référence. Parmi les algorithmes numériques explorant l'espace des paramètres point par point, on peut citer le Simplexe²¹ ou sa variante de Nelder-Mead²². Des algorithmes plus sophistiqués sont souvent utilisés en modélisation moléculaire, tels que des algorithmes génétiques pour l'optimisation de paramètres semi-empiriques²³ ou de champs de force classiques comme MM3^{24,25}. On peut aussi s'appuyer sur une exploration systématique de l'espace des paramètres, en engendrant une grille de points dans une région délimitée à des pas fixés. La fonction de score associée à chaque jeu de paramètres permet ensuite de sélectionner le ou les meilleurs d'entre eux.

Lorsqu'il est possible de calculer facilement le gradient de la fonction à optimiser par rapport aux paramètres, la minimisation analytique de la fonction de score peut devenir une alternative intéressante. C'est la procédure que nous avons choisie pour l'optimisation de torsions du squelette peptidique, en utilisant une régression non-linéaire via la méthode des moindres carrés en raison de la définition de la fonction de score choisie.

L'hyper-surface d'énergie potentielle de référence a été construite de manière à représenter uniquement la composante énergétique de torsion pour laquelle on cherche des paramètres. Pour engendrer une telle courbe d'énergie potentielle, nous faisons l'approximation qu'en dehors de cette torsion, l'énergie AMOEBA reproduit exactement l'énergie MP2. Dans cette approximation, la différence entre l'énergie MP2 et l'énergie AMOEBA amputée du terme

de torsion en question est due exclusivement à la torsion MP2. En la nommant E_{mixte} , on a :

$$E_{mixte} = E_{MP2} - (E_{AMOEBA} - E_{torsion})$$

Il faut donc reproduire au mieux la courbe de E_{mixte} en fonction de l'angle de torsion, par une fonction définie par le champ de force. Pour une torsion, l'optimisation analytique est effectuée en utilisant l'expression AMOEBA (Équ. 2.10), c'est-à-dire en déterminant les paramètres A_1 , A_2 et A_3 qui permettent à $E_{torsion}$ d'être la plus proche possible de E_{mixte} .

Des algorithmes de régression non-linéaire sont particulièrement adaptés à ces problèmes. Nous avons opté pour l'utilisation d'un algorithme de Levenberg-Marquardt modifié²⁶. Les courbes d'énergie ont été construites par simple variation d'un angle dièdre du peptide G₃ isolé sans relaxation des autres paramètres : si la structure est relaxée en chaque point, la courbe ne représente plus une pure contribution de torsion. Une autre limitation de la relaxation dans ce contexte est l'obtention de surfaces discontinues. La contrepartie de la non-relaxation est que l'on risque de parcourir des zones du potentiel peu significatives puisqu'au cours d'une torsion, des répulsions importantes entre atomes non liés peuvent apparaître. Il convient alors de choisir parmi les solutions celles qui représentent un bon fit dans les gammes de torsions les plus représentatives. Les résultats de ces calculs sont présentés dans le chapitre 6 de la partie II.

2.2.2 Optimisation des multipôles atomiques

Nous avons vu qu'au moyen de multipôles définis pour chaque atome, AMOEBA effectue un traitement relativement sophistiqué des composantes électrostatique et de polarisation de l'énergie, qui permet de reproduire le champ électrique d'un système ou d'une partie d'un système de manière plus précise que les champs de forces disposant uniquement de charges atomiques ponctuelles.

Le champ de force MM3 comprend un terme électrostatique bâti sur des dipôles de liaisons covalentes et un terme énergétique empirique spécifique pour le traitement des liaisons hydrogène. L'énergie d'une liaison hydrogène calculée dépend donc des interactions dipôle-dipôle prises en compte, de la distance des atomes impliqués et de l'angle qu'ils forment.

Dans le formalisme de MM3, l'alignement des 3 atomes d'une liaison hydrogène est favorisé. Ce terme n'existe pas dans AMOEBA. La force et la directionnalité de la liaison hydrogène sont assurées par l'attribution d'un dipôle et d'un tenseur de quadrupôle à chaque atome, ou plus généralement à chaque site si on considère des paires libres d'électrons.

La distribution multipolaire a ainsi une importance considérable pour le calcul des énergies d'interaction et la structuration des systèmes par liaisons hydrogène ou plus largement, électrostatiques et de polarisation. La modulation de la surface d'énergie potentielle par le traitement des interactions entre multipôles a, en outre, d'importantes conséquences sur les spectres infrarouge calculés.

Définition des multipôles atomiques dans AMOEBA

Dans AMOEBA, les paramètres des multipôles atomiques sont déclarés dans des repères locaux pour chaque site (atome ou paire libre). Les vecteurs de ces repères sont définis par l'environnement des atomes considérés. Pour un centre de repère O avec les atomes A, B et C comme voisins, on peut définir la base locale associée. L'axe z est déterminé par la position d'un atome voisin : \vec{k} est colinéaire à \vec{OA} . L'axe x est ensuite défini orthogonalement à \vec{k} dans le semi-plan $(\vec{k}; \vec{OB})$. Dans ce cas, on parle de convention « z-then-x ». Dans certains cas, on applique une autre convention, « z-then-bisector » qui définit l'axe x non plus par rapport au semi-plan $(\vec{k}; \vec{OB})$ mais par rapport à la bisectrice de l'angle \widehat{BOC} . Finalement, l'axe y est obtenu de manière à former le trièdre direct $(\vec{i}; \vec{j}; \vec{k})$.

À titre d'exemple, considérons le N-méthylacétamide (NMA) (Fig. 5.1). Cette molécule

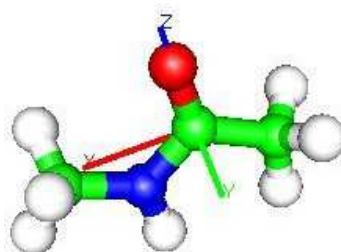


FIG. 2.1: N-méthylacétamide, repère local associé à l'atome C du groupement carbonyle

est un modèle de la liaison peptidique, potentiellement accepteur ou donneur de liaison hydrogène au même titre que le sont les acides aminés et leurs résidus. Dans cette molécule, la

base des multipôles associés à l'atome C du groupement carbonyle est définie en AMOEBA par rapport aux atomes O et N. L'axe z , en bleu dans la figure 2.1 est donc défini selon \vec{CO} alors que l'axe x , en rouge, se situe dans le plan de l'amide (CNO). L'axe y , en vert, est orthogonal à ce plan. Dans le cas du NMA, les atomes lourds sont tous dans le même plan et seuls les hydrogène des groupements méthyle sont hors du plan. En conséquence, on peut s'attendre à une composante en y nulle des multipôles de l'atome C considéré ici.

Analyse de la distribution des multipôles

Les multipôles atomiques d'AMOEBA proviennent de données basées sur des résultats *ab initio*. L'analyse de la distribution des multipôles (DMA) consiste en la répartition de la densité de charge du système entre les sites afin d'en déterminer les moments multipolaires²⁷. Pour cela, le programme GDMA de Stone²⁸ utilise la fonction d'onde issue d'un calcul effectué avec Gaussian²⁹. La méthode DMA est similaire à une généralisation de la méthode de calcul des charges atomiques de Mulliken pour déterminer les moments multipolaires. La flexibilité dans le choix des sites et dans le rang des multipôles est un avantage du logiciel de Stone.

Considérant cela et afin de conserver au mieux la cohérence avec AMOEBA tel qu'il est distribué, les fonctions d'onde sur lesquelles ont été effectuées les DMA proviennent de calculs MP2/aug-cc-pVTZ. Maintenir ce niveau de calcul limite la taille des systèmes pour lesquels on peut extraire des multipôles. L'analyse des multipôles est volontairement limitée aux seuls ordres définis dans les paramètres AMOEBA : monopoles (charges), dipôles et quadrupôles.

La version 2.2 de GDMA laisse à l'utilisateur le choix parmi plusieurs algorithmes pour extraire des multipôles à partir de fonctions gaussiennes. On peut utiliser l'algorithme original (un algorithme de Gauss-Hermite), DMA qui est basé sur une expansion des gaussiennes. Les multipôles atomiques issus de DMA ont une moins bonne convergence que ceux optimisés sur le potentiel électrostatique³⁰. L'algorithme original peut parfois présenter des problèmes dans la reproduction du potentiel électrostatique au niveau des liaisons hydrogène. Cet algorithme est connu pour être sensible à l'utilisation de fonctions diffuses. Les autres algorithmes disponibles dans GDMA sont basés sur des grilles de points représentant

définissant la topologie du potentiel électrostatique et sont supposés être moins sensibles aux effets de recouvrement de fonctions diffuses mais moins précis que l'algorithme original. L'utilisation de fonctions de base plus localisées limite les recouvrements orbitaux et par conséquent les avantages des algorithmes de grilles sur l'algorithme original. Il est possible dans GDMA de fixer un seuil, pour la somme des exposants des gaussiennes ζ de deux fonctions, à partir duquel un algorithme de grille est utilisé alors qu'en deçà, c'est l'algorithme original qui est utilisé.

GDMA utilise des rayons atomiques pour pondérer les sites correspondants. L'utilisation d'un rayon unique pour la plupart des atomes semble améliorer la convergence de l'expansion multipolaire du potentiel électrostatique. Stone conseille un rayon de 0,65 Å par défaut pour la plupart des atomes et 0,325 Å pour les atomes d'hydrogène, ce qui permet d'obtenir des charges atomiques plus en accord avec l'intuition chimique. La forte dépendance de l'hydrogène vis-à-vis de son environnement en terme d'expansion multipolaire est amoindrie avec un rayon petit. Les charges atomiques issues du jeu de paramètres original d'AMOEBEA que nous nommons AMOEBEA0 sont comparées avec celles obtenues par GDMA en utilisant différents rayons pour les atomes d'hydrogène. Les charges présentées dans le tableau 2.1 résultent de DMA, l'algorithme original de GDMA. Le tableau 2.2 présente les valeurs obtenues avec un algorithme basé sur une grille, l'algorithme de Lebedev. Ces deux tableaux montrent que les charges obtenues pour des atomes supposés équivalents diffèrent entre elles car ces atomes « équivalents » ont des environnements tridimensionnels différents. En l'occurrence, les hydrogènes H^7 , H^8 , H^9 et H^{10} , H^{11} , H^{12} des méthyles présentent des variations de charge approchant le dixième d'électron avec DMA. En particulier, H^7 qui est en vis-à-vis de l'hydrogène d'amide a une charge plus faible que ses équivalents H^8 et H^9 . Par opposition à celui-ci, H^{11} qui se situe en position anti par rapport à l'oxygène d'amide a une charge plus élevée que ses équivalents H^{10} et H^{12} .

Avec l'algorithme de Lebedev, ces différences sont de moindre importance et si la charge de H^7 est toujours sous-estimée par rapport aux charges des atomes équivalents, H^{11} a une charge cette fois-ci plus faible que celles de H^{10} et H^{12} .

Pour les deux algorithmes, les charges de C et O du carbonyle sont peu sensibles au choix du rayon des hydrogènes. Il en va autrement des atomes N et H de l'amide pour lesquels les variations de charges peuvent atteindre 0,2 électron. Dans le cas de l'algorithme de Lebedev,

les charges des atomes C et H des méthyles sont elles-aussi sensibles au rayon choisi pour les atomes d'hydrogène. Les variations des charges atomiques dans les méthyles sont peu dépendantes de ce choix lorsque DMA est utilisé.

Le choix de l'algorithme et du rayon utilisé fait appel à l'intuition chimique. Les charges portées par les atomes d'hydrogène dans AMOEBA0 sont toutes inférieures à $0,1 |e^-|$. L'algorithme DMA se démarque d'AMOEBAA0 et de l'algorithme de Lebedev en attribuant des charges atomiques de l'ordre de $0,2 |e^-|$ à ces hydrogènes. En conséquence, les carbones des méthyles voient eux-même leur charge augmenter en valeur absolue à des valeurs proches de $0,6$ voire $0,7 |e^-|$, ce qui semble plus en accord avec l'intuition chimique. Les atomes C et O du carbonyle sont eux-aussi plus chargés quand ils sont calculés par DMA que lorsqu'ils le sont par l'algorithme de Lebedev. Les charges de N et H d'amide semblent plus intuitives lorsqu'un rayon relativement petit est utilisé pour les hydrogène. Un rayon de $0,31$ conduit à des charges plus importantes dans ce groupement. Cette valeur de rayon est choisie pour effectuer l'analyse de la distribution multipolaire avec l'algorithme original, DMA.

| rayon | 0,31 | 0,325 | 0,35 | AMOEBAA0 |
|-----------------|---------|---------|---------|----------|
| C ¹ | -0,6039 | -0,6009 | -0,5587 | -0,0362 |
| N ² | -0,2456 | -0,1642 | -0,0558 | -0,1875 |
| C ³ | 0,6411 | 0,6411 | 0,6411 | 0,6422 |
| C ⁴ | -0,6892 | -0,6881 | -0,6598 | -0,2439 |
| O ⁵ | -0,7988 | -0,7988 | -0,7988 | -0,6652 |
| H ⁶ | 0,2388 | 0,1574 | 0,0490 | 0,0773 |
| H ⁷ | 0,1940 | 0,1931 | 0,1937 | 0,0609 |
| H ⁸ | 0,2520 | 0,2514 | 0,2286 | 0,0609 |
| H ⁹ | 0,2545 | 0,2531 | 0,2330 | 0,0609 |
| H ¹⁰ | 0,2241 | 0,2231 | 0,2092 | 0,0769 |
| H ¹¹ | 0,2993 | 0,2980 | 0,2986 | 0,0769 |
| H ¹² | 0,2338 | 0,2349 | 0,2200 | 0,0769 |

TAB. 2.1: Influence du rayon (en Å) des atomes d'hydrogène sur les charges atomiques du N-méthyle acétamide obtenues par DMA, l'algorithme original de GDMA

De GDMA à AMOEBA

Outre les éventuels changements d'unités, le passage de résultats issus de GDMA à un jeu de paramètres pour AMOEBA nécessite quelques étapes de transformation. Les multipôles

| rayon | 0,31 | 0,325 | 0,35 | AMOEBAA0 |
|-----------------|---------|---------|---------|----------|
| C ¹ | 0,1818 | 0,2359 | 0,3308 | -0,0362 |
| N ² | -0,4345 | -0,3280 | -0,2874 | -0,1875 |
| C ³ | 0,5209 | 0,5256 | 0,5341 | 0,6422 |
| C ⁴ | 0,0472 | 0,1010 | 0,1969 | -0,2439 |
| O ⁵ | -0,7446 | -0,7421 | -0,7378 | -0,6652 |
| H ⁶ | 0,2212 | 0,1171 | 0,0805 | 0,0773 |
| H ⁷ | 0,0181 | -0,0010 | -0,0347 | 0,0609 |
| H ⁸ | 0,0348 | 0,0150 | -0,0195 | 0,0609 |
| H ⁹ | 0,0483 | 0,0282 | -0,0069 | 0,0609 |
| H ¹⁰ | 0,0448 | 0,0253 | -0,0094 | 0,0769 |
| H ¹¹ | 0,0183 | -0,0014 | -0,0362 | 0,0769 |
| H ¹² | 0,0438 | 0,0243 | -0,0104 | 0,0769 |

TAB. 2.2: Influence du rayon (en Å) des atomes d'hydrogène sur les charges atomiques du N-méthyle acétamide obtenues par l'algorithme de Lebedev

obtenus *via* GDMA sont exprimés en harmoniques sphériques, alors que les paramètres mutipolaires AMOEBA sont définis dans des repères atomiques, *a fortiori* locaux.

Connaissant les coordonnées atomiques dans le repère cartésien global, on peut convertir les harmoniques sphériques en coordonnées cartésiennes du repère global³¹ $\{O; \vec{i}; \vec{j}; \vec{k}\}$. Considérons le multipôle d'un atome O_n . Le repère atomique y correspondant peut s'écrire : $\{O_n; \vec{i}_n; \vec{j}_n; \vec{k}_n\}$. Une projection des composantes du multipôle atomique selon \vec{i} , \vec{j} et \vec{k} sur les vecteurs \vec{i}_n , \vec{j}_n et \vec{k}_n permet le passage du repère global au repère local.

Les changements de repère effectués dans nos travaux ont été effectués dans un premier temps *via* le programme Orient³². Dans la suite Tinker, le programme *poledit* a été développé au fil des versions et permet de réaliser les mêmes opérations, présentant l'avantage pour nous d'être intégré à Tinker.

Reparamétrisation des multipôles du N-méthylacétamide

Nous avons constaté, dans le contexte d'AMOEBEA, une mauvaise réponse des fréquences de vibration aux liaisons hydrogène, et plus généralement aux interactions électrostatiques. Afin de comprendre si ces limites sont intrinsèques au formalisme d'AMOEBEA ou aux paramètres utilisés, les paramètres des multipôles atomiques ont été réoptimisés. Ceux du jeu de paramètres AMOEBA mod. résultent d'une analyse effectuée avec DMA, l'algorithme original de GDMA et des rayons de 0,31 Å pour les hydrogènes. La topologie électrostatique

à reproduire provient d'un calcul MP2/aug-cc-pVTZ sur une molécule de NMA isolée.

Dans le cas du NMA, la présence des méthyles entraîne une dissymétrie par rapport au plan de l'amide. Elle est responsable, entre autres, des différences de charges entre atomes d'hydrogène équivalents. La présence des méthyles entraîne aussi l'existence, bien que de faible importance, de composantes selon l'axe y non-nulles pour les moments dipolaires et quadripolaires des atomes C, O et N qui sont dans le même plan, ainsi que l'hydrogène d'amide (Fig. 5.1).

La transférabilité de ces paramètres est étudiée sur 3 systèmes (Fig. 2.2.2) : NMA-Na⁺ permet de percevoir l'influence des interactions avec le sodium (cation d'intérêt biologique important) sur les structures, l'énergétique et les propriétés spectroscopiques du NMA. Dans NMA-H₂O, l'hydrogène de l'amide est lié à la molécule d'eau mais le carbonyle est libre. Dans le dimère du N-méthylacétamide, NMA-NMA, un monomère est accepteur pendant que l'autre est donneur de liaison hydrogène. Ce cas est d'autant plus intéressant pour étudier la transférabilité des paramètres que tous les multipôles atomiques considérés sont les mêmes pour les deux molécules. Dans la géométrie MP2/aug-cc-pVTZ, les deux monomères sont inscrits dans des plans orthogonaux.

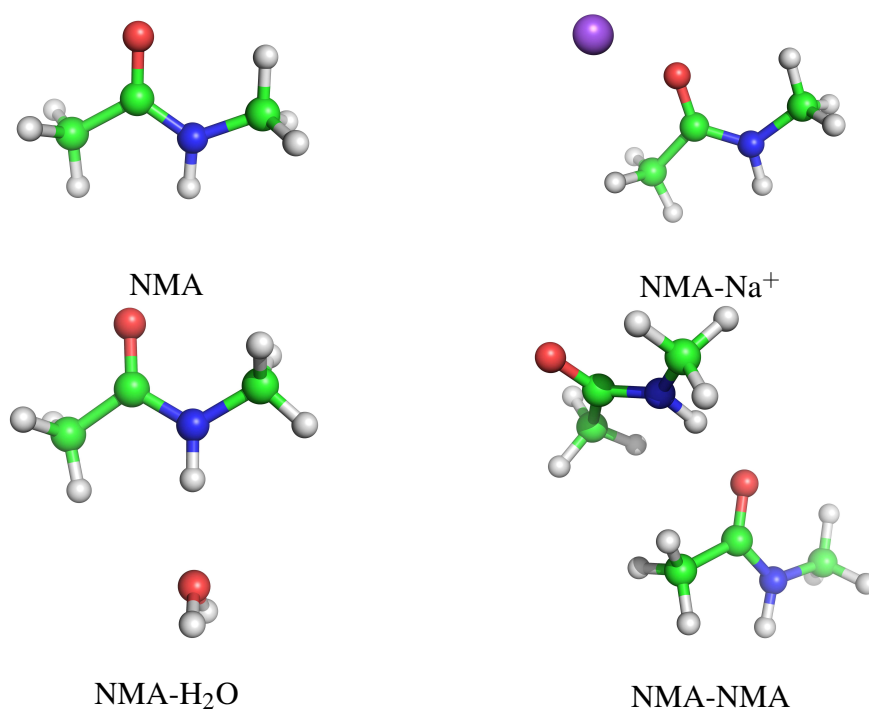


FIG. 2.2: Structures du N-méthylacétamide isolé et en interaction

| système | MP2/aug-cc-pVTZ | AMOEBAA0 | AMOEBEA mod. |
|----------------------|-----------------|----------|--------------|
| NMA-Na ⁺ | -44,3 | -37,7 | -45,6 |
| NMA-H ₂ O | -6,7 | -5,2 | -5,9 |
| NMA-NMA | -11,7 | -9,4 | -12,4 |

TAB. 2.3: Énergies d'interaction intermoléculaire (en kcal.mol⁻¹) de systèmes à base de N-méthylacétamide

Les énergies d'interaction intermoléculaire pour ces systèmes sont présentées dans le tableau 2.3. Pour chacun de ces systèmes, les énergies calculées avec AMOEBEA mod. sont plus proches de la référence *ab initio* que ne le sont les énergies calculées avec AMOEBAA0. Par rapport à la valeur MP2 de référence, l'interaction avec le sodium est très nettement améliorée, passant de -37,7 avec AMOEBAA0 à -45,6 kcal.mol⁻¹ après reparamétrisation. L'énergie d'interaction de NMA-H₂O est améliorée de 0,7 kcal.mol⁻¹. Elle est alors de -5,9 contre -6,7 kcal.mol⁻¹ en MP2. Par rapport à la valeur MP2 de -11,7 kcal.mol⁻¹, l'interaction du dimère NMA-NMA est sous-estimée de +2,3 kcal.mol⁻¹ par AMOEBAA0, alors qu'elle est surestimée de -0,6 kcal.mol⁻¹ avec les paramètres modifiés. Sur ces 3 exemples, AMOEBAA0 sous-estime systématiquement l'interaction intermoléculaire. AMOEBEA mod. a tendance à surestimer de manière moindre ces interactions, sauf dans le cas de NMA-H₂O. Dans NMA-Na⁺, la distance d'attachement du carbonyle au cation est de 2,02 Å à la géométrie MP2/aug-cc-pVTZ et en AMOEBEA, elle est ramenée de 2,09 à 2,01 Å par la modification des multipôles. Ce raccourcissement est à mettre en relation avec l'amélioration de l'énergie d'interaction entre le peptide et le cation. Dans le cas du NMA interagissant avec une molécule d'eau, la distance de la liaison hydrogène est détérioré par le nouveau jeu de paramètres. Elle passe de 2,10 à 2,16 Å contre 1,99 Å en MP2/aug-cc-pVTZ. Les distances de la liaison hydrogène dans le cas du dimère de NMA sont respectivement de 1,94 et 1,90 Å en AMOEBEA avant et après modification des multipôles, alors qu'elle vaut 1,88 Å en MP2/aug-cc-pVTZ.

L'amélioration par les nouveaux multipôles de la directionnalité des liaisons hydrogène et des interactions électrostatiques est importante. Bien que l'orientation relative des deux molécules de N-méthylacétamide diffère en MP2 et en AMOEBEA mod., les angles d'attachement ou d'interaction des carbonyles sont rapprochés des valeurs de référence. Cette amélioration est à mettre en contraste avec le potentiel d'AMOEBAA0 qui présente une forte tendance à restreindre ces angles à des valeurs proches de 120°.

Dans le cas du dimère, la différence de l'angle C-O-HN entre AMOEBA0 (122°) et sa version reparamétrée AMOEBA mod. (125°) est faible, alors que la valeur MP2/aug-cc-pVTZ est de 114°. La directionnalité des autres interactions électrostatiques est par contre améliorée. L'angle C-O-Na⁺ dans NMA-Na⁺ vaut 171° en MP2/aug-cc-pVTZ et le jeu de paramètres incluant les nouveaux multipôles ramène la valeur de l'angle en AMOEBA de 133° à 160°.

Bien que non-uniforme, l'amélioration des propriétés obtenues avec les nouveaux multipôles atomiques est remarquable en termes de géométrie. Elle l'est au moins autant en terme de fréquences vibrationnelles du NMA et de leurs modifications dues aux interactions intermoléculaires. Ces améliorations dans le domaine spectroscopique seront discutées dans la section 5.4.

Limites, difficultés et perspectives possibles

Les atomes de même type AMOEBA ont par définition les mêmes paramètres. Cependant, leur environnement dans un système chimique n'est *a priori* pas le même. Cela entraîne des différences, parfois non négligeables, sur les multipôles extraits par GDMA. Pour les valeurs de multipôles obtenus par GDMA pour des atomes équivalents, on peut avoir recours au moyennage des valeurs. À partir de la version de Tinker disponible depuis avril 2010, le programme *poledit* permet le moyennage direct des paramètres des multipôles pour les atomes équivalents.

Le moyennage des multipôles est une façon de traiter le problème des atomes équivalents. Pourtant, il n'y a pas de raison que la topologie électrostatique résultant de la distribution multipolaire après moyennage soit la meilleure possible. Pour parvenir à la meilleure distribution possible en considérant les contraintes d'équivalence de multipôles pour des atomes de même type, il faudrait être capable d'imposer des contraintes à l'analyse des multipôles effectuée par GDMA. On aimerait pouvoir imposer des contraintes supplémentaires telles que fixer certains multipôles lors de la détermination d'autres par GDMA. Ce point est capital pour extraire des multipôles de systèmes de taille plus ou moins grande mais surtout si on souhaite conserver des paramètres existants et les transférer.

Bibliographie

- [1] P. REN et J. W. PONDER : Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.*, 23:1497–506, 2002.
- [2] P. REN et J. W. PONDER : Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B*, 107:5933–47, 2003.
- [3] A. GROSSFIELD, P. REN et J. W. PONDER : Ion solvation thermodynamics from simulation with a polarizable force field. *J. Am. Chem. Soc.*, 125:15671–15682, 2003.
- [4] J.-P. PIQUEMAL, L. PERERA, G. A. CISNEROS, P. REN, L. G. PEDERSEN et 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.*, 125:054511, 2006.
- [5] J. C. WU, J.-P. PIQUEMAL, P. REINHARDT, R. CHAUDRET et P. REN : Polarizable molecular dynamics simulation of Zn(II) in water using the AMOEBA force field. *J. Chem. Theory Comput.*, 2010. sous presse.
- [6] T. D. RASMUSSEN, P. REN, J. W. PONDER et F. JENSEN : Force field modeling of conformational energies : Importance of multipole moments and intramolecular polarization. *I. J. Quantum Chem.*, 107:1390–1395, 2007.
- [7] J. KAMINSKÝ et F. JENSEN : Force Field Modeling of Amino Acid Conformational Energies. *J. Chem. Theory Comput.*, 3:1774–1788, 2007.
- [8] J. W. PONDER : TINKER, Software Tools for Molecular Design, 2004. Version 4.2, Washington University School of Medicine, Saint Louis, MO.
- [9] A. J. STONE : *The Theory of Intermolecular Forces*. Oxford University Press, 1996.

- [10] C. CLAVAGUÉRA, F. CALVO et J.-P. DOGNON : Theoretical study of the hydrated Gd^{3+} ion : Structure, dynamics, and charge transfer. *J. Chem. Phys.*, 124:074505, 2006.
- [11] R. CHAUDRET, N. GRESH, P. REN et J.-P. PIQUEMAL : communication personnelle.
- [12] P. KOLLMAN, R. DIXON, W. CORNELL, T. FOX, C. CHIPOT et A. POHORILLE : *The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab Initio calculations and experimental data*, volume 3, pages 83–96. Springer, 1997.
- [13] B. R. BROOKS, III BROOKS, C. L., A. D. MACKERELL, Jr., L. NILSSON, R. J. PETRELLA, B. ROUX, Y. WON, G. ARCHONTIS, C. BARTELS, S. BORESCH, A. CAFLISCH, L. CAVES, Q. CUI, A. R. DINNER, M. FEIG, S. FISCHER, J. GAO, M. HODOSCEK, W. IM, K. KUCZERA, T. LAZARIDIS, J. MA, V. OVCHINNIKOV, E. PACI, R. W. PASTOR, C. B. POST, J. Z. PU, M. SCHAEFER, B. TIDOR, R. M. VENABLE, H. L. WOODCOCK, X. WU, W. YANG, D. M. YORK et M. KARPLUS : CHARMM : The Biomolecular Simulation Program. *J. Comput. Chem.*, 30:1545–1614, 2009.
- [14] A. J. STONE : Intermolecular Potentials. *Science*, 321:787–789, 2008.
- [15] B. T. THOLÉ : Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.*, 59:341 – 350, 1981.
- [16] T. A. HALGREN : The representation of van der Waals (vdW) interactions in molecular mechanics force fields : potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.*, 114:7827–7843, 1992.
- [17] N. L. ALLINGER, Y. H. YUH et J. H. LII : Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.*, 111:8551–8566, 1989.
- [18] J. H. LII et N. L. ALLINGER : Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics. *J. Am. Chem. Soc.*, 111:8556–8575, 1989.
- [19] J. H. LII et N. L. ALLINGER : Molecular mechanics. The MM3 force field for hydrocarbons. 3. The van der Waals' potentials and crystal data for aliphatic and aromatic hydrocarbons. *J. Am. Chem. Soc.*, 111:8576–8582, 1989.
- [20] 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 et T. HEAD-GORDON : Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B*, 114:2549–2564, 2010.
- [21] J. C. LAGARIAS, J. A. REEDS, M. H. WRIGHT et P. E. WRIGHT : Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions. *SIAM J. Optim.*, 9:112–147, 1998.
- [22] J. A. NELDER et R. MEAD : A Simplex Method for Function Minimization. *Comput. J.*, 7:308–313, 1965.
- [23] J. P. MCNAMARA, M. SUNDARARAJAN et I. H. HILLIER : Development of parameter sets for semi-empirical MO calculations of transition metal systems : Iron parameters for iron-sulfur proteins. *J. Mol. Graphics Modell.*, 24:128–137, 2005. Biomolecular Recognition and Reactivity Special Issue in honour of the contributions of Professor Ian Hillier.
- [24] T. STRASSNER, M. BUSOLD et W. A. HERRMANN : MM3 parametrization of four- and five-coordinated rhenium complexes by a genetic algorithm - Which factors influence the optimization performance? *J. Comput. Chem.*, 23:282–290, 2002.
- [25] Peter WOLOHAN et David E. REICHERT : Molecular modeling of hexakis(areneisonitrile)technetium(I), tricarbonyl η^5 cyclopentadienyl technetium and technetium(V)-oxo complexes : MM3 parameter development and prediction of biological properties. *J. Mol. Graphics Modell.*, 25:616–632, 2007.
- [26] W. PRESS, S. TEUKOLSKY, W. VETTERLING et B. FLANNERY : *Numerical Recipes in C*, pages 683–688. Cambridge University Press, 1992.
- [27] A. J. STONE : Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.*, 83:233–239, 1981.
- [28] A. J. STONE : Distributed Multipole Analysis for Gaussian Wavefunctions, version 2.2.03, 2005-2007.
- [29] M. J. FRISCH, G. W. TRUCKS, H. B. SCHLEGEL, G. E. SCUSERIA, M. A. ROBB, J. R. CHEESEMAN, J. A. MONTGOMERY, Jr., T. VREVEN, K. N. KUDIN, J. C. BURANT, J. M. MILLAM, S. S. IYENGAR, J. TOMASI, V. BARONE, B. MENNUCCI, M. COSSI, G. SCALMANI, N. REGA, G. A. PETERSSON, H. NAKATSUJI, M. HADA, M. EHARA, K. TOYOTA, R. FUKUDA, J. HASEGAWA, M. ISHIDA, T. NAKAJIMA, Y. HONDA, O. KITAO, H. NAKAI, M. KLENE, X. LI, J. E. KNOX, H. P. HRATCHIAN,

- J. B. CROSS, V. BAKKEN, C. ADAMO, J. JARAMILLO, R. GOMPERTS, R. E. STRATMANN, O. YAZYEV, A. J. AUSTIN, R. CAMMI, C. POMELLI, J. W. OCHTERSKI, P. Y. AYALA, K. MOROKUMA, G. A. VOTH, P. SALVADOR, J. J. DANNENBERG, V. G. ZAKRZEWSKI, S. DAPPRICH, A. D. DANIELS, M. C. STRAIN, O. FARKAS, D. K. MALICK, A. D. RABUCK, K. RAGHAVACHARI, J. B. FORESMAN, J. V. ORTIZ, Q. CUI, A. G. BABOUL, S. CLIFFORD, J. CIOSLOWSKI, B. B. STEFANOV, G. LIU, A. LIASHENKO, P. PISKORZ, I. KOMAROMI, R. L. MARTIN, D. J. FOX, T. KEITH, M. A. AL-LAHAM, C. Y. PENG, A. NANAYAKKARA, M. CHALLACOMBE, P. M. W. GILL, B. JOHNSON, W. CHEN, M. W. WONG, C. GONZALEZ et J. A. POPLE : Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.
- [30] W. T. M. MOOIJ, F. B. van DUIJNEVELDT, J. G. C. M. VAN DUIJNEVELDT-VAN DE RIJDT et B. P. VAN EIJCK : Transferable ab Initio Intermolecular Potentials. 1. Derivation from Methanol Dimer and Trimer Calculations. *J. Phys. Chem. A*, 103: 9872–9882, 1999.
- [31] H. B. SCHLEGEL et M. J. FRISCH : Conversion from normalised spherical form to normalised Cartesian. *I. J. Quantum Chem.*, 54:83–87, 1995.
- [32] A. J. STONE, A. DULLWEBER, O. ENKVIST, E. FRASCHINI, M. P. HODGES, A. W. MEREDITH, D. R. NUTT, P. L. A. POPELIER et D. J. WALES : Orient : a program for studying interactions between molecules, version 4.6, University of Cambridge, Enquiries to A. J. Stone, ajs1@cam.ac.uk, 2002-2010.

Chapitre 3

Exploration de l'espace des phases

Nous avons vu que la structure tridimensionnelle d'un peptide et à plus forte raison d'une protéine a une importance capitale dans sa fonction, ce qui motive les études spectroscopiques donnant accès à des informations structurales, et en particulier la spectroscopie IRMPD pour ce qui concerne la phase gazeuse.

Afin de mieux comprendre les spectres expérimentaux, la modélisation moléculaire des systèmes étudiés est nécessaire. Pour attribuer une ou plusieurs structures à un spectre, il faut d'abord disposer de ces structures.

Pour des peptides de taille croissante, l'espace conformationnel est de plus en plus complexe. L'exploration conformationnelle devient alors non seulement plus longue mais elle nécessite aussi l'utilisation de méthodes plus sophistiquées.

3.1 Quelques méthodes d'exploration

Le développement de l'informatique et de méthodes mathématiques robustes a permis la simulation de systèmes de plus en plus complexes. Devant le grand nombre de dimensions des problèmes posés, les méthodes analytiques basées sur le gradient du potentiel ne sont capables d'apporter de solution qu'à des cas simples. Elles sont très facilement piégées dans des *minima* locaux : elles ne permettent pas de trouver le minimum absolu de la surface de potentiel si la recherche démarre avec une structure appartenant à un puits de potentiel

éloigné de celui du minimum absolu.

Pour rechercher des structures de basses énergies de systèmes complexes, il est nécessaire d'utiliser des méthodes numériques dont la caractéristique principale n'est pas forcément la convergence vers le fond d'un puits de potentiel mais qui puisse explorer des bassins de potentiel éloignés. La littérature fournit de nombreuses méthodologies numériques pour une recherche du minimum global. Parmi les méthodes d'optimisation, on trouve les méthodes de programmation linéaire et les méthodes stochastiques.

Parmi les méthodes de programmation linéaire, on peut citer celle du Simplex¹. Pour explorer la topologie d'un problème de dimension n dans une région limitée, l'algorithme du Simplex évalue un ensemble de $n + 1$ points, un simplexe. Après évaluation de ces points, le Simplex remplace le plus mauvais point et, en se basant sur la position et la valeur des autres points, estime le meilleur endroit pour le placer au pas suivant de l'optimisation. Le passage d'un pas de l'optimisation au suivant ne nécessite donc pas de gradient. Typiquement pour la recherche, la position de chaque point d'un simplexe correspond à une structure et le Simplex estime l'énergie de chacune d'entre elles. Le Simplex est surtout utilisé pour des optimisations lorsque la fonction à optimiser (ici l'énergie) est facile à calculer et le gradient inaccessible ou l'est difficilement. Le résultat d'une telle optimisation est très dépendant de l'ensemble du simplexe (jeu de $n + 1$ structures) initial. Même les variantes du Simplex ou de la méthode de Nelder-Mead² qui en dérive peuvent trop facilement être piégées dans un minimum local.

Les méthodes stochastiques, quant à elles sont à même de trouver un minimum global de façon quasi-certaine après un nombre de pas « suffisant ».

La méthode Monte Carlo³ repose sur la génération, à chaque pas, d'une structure obtenue par variation aléatoire d'une ou plusieurs coordonnées internes ou cartésiennes de la structure précédente. L'algorithme de Metropolis⁴ est de loin le plus utilisé pour les simulations de type Monte Carlo. Dans ce cadre, si la nouvelle structure est plus basse en énergie que la structure précédente, elle est systématiquement acceptée. Si par contre, elle est plus élevée en énergie, alors cette nouvelle structure est aléatoirement acceptée avec une probabilité de Boltzmann pour une température de simulation. Les simulations de type Monte Carlo permettent d'accéder à certaines grandeurs thermodynamiques mais ne permettent

généralement pas de déterminer des propriétés dynamiques du système.

La méthode Monte Carlo disponible dans Tinker utilise l'algorithme de Minimisation Monte Carlo développé par Li et Scheraga⁵. Dans le cadre de cet algorithme, après une variation structurelle induite à chaque pas de Monte Carlo, la nouvelle structure est optimisée. Ainsi, on explore une surface de potentiel modifiée dans laquelle les seuls niveaux énergétiques possibles correspondent aux *minima* locaux de la surface de potentiel « vraie ». Cette méthode d'exploration est plus efficace pour trouver des bassins de potentiel⁶.

La méthode du recuit simulé⁷ est basée sur la diminution de la température, ce qui a pour conséquence la descente en énergie dans un puits de potentiel alors que des phases de montée en température permettent au système de franchir des barrières énergétiques afin d'échapper à un minimum local. Le principe du recuit simulé est utilisé en combinaison avec d'autres méthodes, principalement Monte Carlo⁸ et dynamique moléculaire⁹ pour faciliter l'exploration des surfaces d'énergie potentielle des biomolécules.

Une autre famille d'algorithmes s'inspire de la génétique des populations. D'un point de vue biologique, dans une population fermée et isolée d'événement extérieur, après un nombre suffisant de générations, les allèles les plus favorables à la reproduction des individus deviennent majoritaires. Dans le cadre des minimiseurs basés sur des algorithmes génétiques, les modifications structurales sont engendrées par des mutations aléatoires de paramètres géométriques par analogie aux mutations génétiques en biologie. Plus le taux de mutations et la taille de la population d'individus (structures) par génération sont élevés et plus la diversité de chaque paramètre structural est assurée. Des étapes de reproduction comprenant des *crossing-over* contribuent à la diversité structurale de manière plus globale. Les étapes de sélection conservent systématiquement les structures de plus basse énergie rencontrées. D'autres peuvent tout de même être conservées par des tournois. Cela permet de maintenir de la diversité et contribue à échapper aux *minima* locaux. Ces algorithmes sont utilisés pour la recherche conformationnelle de peptides¹⁰, ou de systèmes protéine-substrat¹¹

De nombreux autres algorithmes existent. Nous nous intéresserons par la suite à deux méthodes d'exploration de l'espace conformationnel : la dynamique moléculaire et la dynamique moléculaire par échange de répliques.

3.2 Dynamique moléculaire

Le principe de la dynamique moléculaire est de simuler l'évolution temporelle d'un système. Elle permet l'étude de propriétés et de processus dynamiques, par exemple biologiques : le transport d'ions, le repliement des protéines. . . .

Nous nous intéressons particulièrement aux caractéristiques vibrationnelles de nos systèmes (cf. chapitre 5.3), ainsi qu'à leurs changements conformationnels et l'exploration de l'espace des phases que cela permet.

3.2.1 Propagation de la trajectoire

Le principe fondamental de la dynamique nous permet, à travers les équations de Newton (Éq. 3.1), d'étudier l'évolution dans le temps des coordonnées spatiales des atomes du système. Pour une particule donnée, la somme des forces \vec{F}_i exercées par chaque particule sur la particule considérée, est égale au produit de sa masse m par son accélération \vec{a} .

$$\sum \vec{F}_i = m\vec{a} \quad (3.1)$$

Un potentiel appliqué au principe fondamental de la dynamique suffit à calculer l'accélération pour chaque particule du système. Pour intégrer les équations de Newton, le temps est discrétisé sur l'intervalle de simulation en une suite d'instantanés régulièrement espacés par un pas de temps δt . Par intégration des équations de Newton, on obtient alors les positions atomiques pour le pas suivant. Le potentiel utilisé peut être de nature quantique ou classique. On parlera alors de dynamique moléculaire classique si un potentiel de mécanique moléculaire est utilisé ou de dynamique moléculaire quantique si les forces sont calculées à un niveau quantique.

En fonction du traitement des électrons lors des dynamiques moléculaires quantiques, on distingue trois types de dynamique : Ehrenfest, Born-Oppenheimer et Carr-Parrinello.

Dans le cadre de la dynamique moléculaire Ehrenfest^{12,13} introduite en 1927, le mouvement des noyaux traités de manière classique est couplé à la propagation des électrons. Le calcul des forces nécessite la résolution de l'équation de Schrödinger dépendante du temps.

Le pas de temps est régi par la propagation du mouvement des électrons.

En dynamique moléculaire de type Born-Oppenheimer¹⁴, on considère que le mouvement des électrons peut être découplé de celui des noyaux. À chaque pas de la trajectoire, les orbitales électroniques sont relaxées avant le calcul des forces. Le pas de temps est ici régi par la propagation du mouvement des noyaux. Par conséquent, il peut être plus grand qu'en dynamique Ehrenfest. La dynamique moléculaire Born-Oppenheimer peut utiliser des méthodes semi-empiriques¹⁵ pour le calcul des forces.

En dynamique moléculaire Carr-Parrinello¹⁶, les mouvement des électrons et des noyaux sont couplés. La résolution de l'hamiltonien électronique n'est pas exactement effectuée à chaque pas de la trajectoire ; les matrices de la fonction d'onde ne sont pas diagonalisées à chaque pas, ce qui réduit considérablement le temps de calcul des forces. En contre-partie, le pas de temps est limité dans ce type de dynamique.

Dans la pratique, la dynamique moléculaire basée sur des méthodes quantiques est limitée en faisabilité par la taille des systèmes et les durées de simulation. Les méthodes post-Hartree-Fock et la DFT sont généralement utilisées pour des simulations ne dépassant pas quelques dizaines de ps. La dynamique moléculaire Born-Oppenheimer Semi-Empirique reste encore peu utilisée. En outre, pour des perspectives de simulations longues et de systèmes de tailles croissantes, ces méthodes sont coûteuses en temps de calcul par rapport à la dynamique moléculaire classique. C'est pourquoi les simulations de dynamique moléculaire présentées dans nos travaux utilisent un potentiel de mécanique moléculaire issu du champ de force polarisable AMOEBA.

Une simulation de dynamique moléculaire s'effectue dans un ensemble thermodynamique défini. L'ensemble thermodynamique détermine ce que le système est autorisé à échanger avec l'extérieur : de l'énergie ou de la matière. Pour étudier des systèmes isolés en phase gazeuse qui par définition n'évoluent pas dans un volume restreint, nous considérons l'ensemble thermodynamique isotherme-isobare conservant la matière. Nos systèmes ne peuvent alors échanger que de l'énergie avec l'extérieur. Nous nommons **NPT** cet ensemble à quantité de matière, pression et température constantes.

L'intégration des équations du mouvement nécessite de définir des conditions initiales. Les positions atomiques initiales sont fixées par la géométrie de la structure initiale. Les vitesses

initiales peuvent quant à elles, soit être définies explicitement par l'utilisateur, soit le plus souvent être engendrées de manière aléatoire.

Il existe plusieurs algorithmes pour l'intégration temporelle. Verlet a proposé l'algorithme qui porte son nom pour l'intégration des équations du mouvement, qui a été décliné sous plusieurs formes. Dans l'algorithme original, les positions r aux instants t et $t + \delta t$ sont définies par les équations 3.2 avec δt comme pas de temps de propagation.

$$\begin{aligned} r(t + \delta t) &= r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \\ r(t - \delta t) &= r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \\ r(t + \delta t) &= 2r(t) - r(t - \delta t) + a(t)\delta t^2 \end{aligned} \quad (3.2)$$

L'algorithme de Verlet utilise l'accélération au temps t et les positions aux temps t et $t - \delta t$ pour calculer les positions à l'instant $t + \delta t$. Cette formulation a l'avantage de ne pas faire intervenir les vitesses de manière explicite dans le calcul de la trajectoire. En contrepartie, son désavantage est que l'énergie cinétique n'est pas directement calculée.

Une variante de l'algorithme de l'algorithme Verlet, Verlet-vitesse (Éq. 3.4) permet de calculer les positions et les vitesses pour chaque instant t de la trajectoire.

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \quad (3.3)$$

$$v(t + \delta t) = v(t) + \frac{1}{2}[a(t) + a(t + \delta t)]\delta t \quad (3.4)$$

Beeman a proposé un autre algorithme¹⁷ (Éq. 3.6). Il permet un calcul simultané des positions et des vitesses. Leurs expressions plus précises permettent une meilleure conservation de l'énergie et des moments. Le coût de cette meilleure description est un stockage plus important de données nécessaires à la propagation de la trajectoire.

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{2}{3}a(t)\delta t^2 - \frac{1}{6}a(t - \delta t)\delta t^2 \quad (3.5)$$

$$v(t + \delta t) = v(t) + \frac{1}{3}a(t + \delta t)\delta t + \frac{5}{6}a(t)\delta t - \frac{1}{6}a(t - \delta t)\delta t \quad (3.6)$$

L'élévation du temps de calcul due au choix de Beeman par rapport à Verlet-vitesse est

petite devant le temps du calcul des forces. Lors des simulations de dynamique moléculaire, nous utilisons un intégrateur de Beeman avec des coefficients modifiés par B. R. Brooks en 1988 (Better Beeman¹⁸) tel qu'il est disponible dans Tinker. Cette modification améliore encore la conservation de l'énergie par l'intégrateur de Beeman¹⁹.

Pour un intégrateur donné, le pas de temps δt détermine le nombre de calculs des forces à effectuer. Cela impose des limites inférieures au pas de temps en fonction de la durée de la simulation. C'est aussi un facteur déterminant pour l'aptitude d'une trajectoire à représenter la propriété physique recherchée. En plus de l'exploration de surfaces de potentiel, nous souhaitons exploiter des trajectoires de dynamique moléculaire pour calculer des spectres infrarouge. Pour étudier les vibrations d'une molécule autour d'une structure d'équilibre, le pas de temps est généralement choisi comme environ un dixième de la période d'oscillation la plus courte du système.

3.2.2 Thermostats

Comme nous l'avons précisé, pour nos simulations nous considérons l'ensemble thermodynamique isotherme-isobare conservant la matière (NPT). Puisque la matière est conservée dans les équations de Newton, l'énergie du système demeure en principe constante au cours de la trajectoire. Des thermostats permettent de passer de l'énergie constante à température fixée. Les thermostats introduisent des biais dans les trajectoires en termes dynamiques mais permettent d'échantillonner plus ou moins correctement l'ensemble thermodynamique de la simulation.

Tinker donne à l'utilisateur le choix entre trois thermostats : Berendsen, Andersen et Nosé-Hoover.

Le thermostat de Berendsen²⁰ échelonne périodiquement les vitesses atomiques pour maintenir le système à la température imposée. Il est efficace pour conserver une température constamment proche de la température choisie dans l'ensemble NPT.

Le thermostat d'Andersen²¹ modifie les vitesses en faisant intervenir une particule fictive subissant des collisions stochastiques avec des atomes du système aléatoirement tirés. La vitesse d'un atome sélectionné pour la collision est remplacée par une valeur tirée d'une

distribution de Maxwell-Boltzmann. Nous avons constaté sur nos systèmes une moins bonne conservation de l'énergie dans les trajectoires utilisant le thermostat d'Andersen par rapport aux trajectoires utilisant le thermostat de Berendsen.

L'échelonnage des vitesses atomiques plutôt que des vitesses internes peut causer ce qui est décrit dans la littérature comme l'effet « flying ice cube »²². Cet effet est principalement visible pour les systèmes dans le vide. Les états vibrationnels des hautes fréquences peuvent être dépeuplés au profit des modes de basses fréquences. Une solution consiste à régulièrement supprimer les rotations et translation du système au cours d'une trajectoire de dynamique moléculaire. C'est ce que fait Tinker. Bien que le thermostat d'Andersen semble être moins responsable de cet effet que celui de Berendsen, c'est le thermostat de Nosé-Hoover qui est recommandé pour l'éviter efficacement.

Le thermostat de Nosé-Hoover^{23,24} corrige les vitesses à chaque pas et à chaque demi-pas. L'échange d'énergie est effectué avec un ou des degrés de liberté additionnel(s). Nosé-Hoover a une meilleure signification physique en terme de fluctuations de température et reproduit bien l'ensemble thermodynamique. Ce thermostat est disponible dans les versions de Tinker mises en ligne à partir de l'année 2009.

Dans les simulations de dynamique moléculaire que nous présenterons, nous préciserons le thermostat utilisé.

3.2.3 Discussion sur la dynamique moléculaire avec le champ de force AMOEBA

Nous nous intéressons à des systèmes en phase gazeuse. Par conséquent, aucune condition aux limites périodiques n'est appliquée. Le système est donc moins contraint que les systèmes en phase condensée. Par conséquent, la dynamique moléculaire des peptides en phase gazeuse met en jeu des temps de thermalisation et d'équilibration moins importants que dans les cas de simulations de phases condensées.

L'utilisation d'un champs de force polarisable incluant les effets à n-corps pour le calcul des forces en dynamique moléculaire n'est pas sans incidence sur le déroulement de celle-ci. La procédure de calcul de la polarisation nécessite un temps de calcul plus important qu'avec

les champs de force dits classiques (p. 16). Par conséquent, l'augmentation du temps CPU demandé par l'intégrateur de Beeman par rapport à celui de Verlet ou au Verlet-vitesse est négligeable.

Un point important concernant le calcul de la polarisation et sa convergence est, nous l'avons vu, le possible phénomène de *polarization catastrophe* (décrit p. 33). Il convient de choisir un pas d'intégration de la dynamique suffisamment petit pour que les structures rencontrées au cours d'une trajectoire ne soient pas responsables d'une polarisation dont l'énergie est trop difficile à converger. Notre expérience a montré qu'un pas de temps de 1 fs permet d'éviter ce problème. Ce pas de temps reste suffisamment élevé pour permettre des simulations de l'ordre de quelques dizaines ou centaines de ns pour les peptides en phase gazeuse. Il convient donc parfaitement pour effectuer des trajectoires de dynamique moléculaire dans le but de l'exploration de l'espace conformationnel des systèmes d'intérêt.

3.3 Dynamique moléculaire par échange de répliques (REMD)

3.3.1 Principe

L'échange de répliques est un algorithme d'ensemble généralisé^{25,26} dans lequel les simulations reposent sur une pondération par un facteur non-boltzmannien de chaque état, ce qui assure la possibilité d'effectuer un pas aléatoire dans l'espace de l'énergie potentielle. Le *replica exchange*^{27,28}, aussi appelé *parallel tempering*²⁹ présente l'avantage d'avoir un facteur de pondération connu.

Le principe du REMD est d'effectuer plusieurs trajectoires à des températures différentes en parallèle, des répliques, et de les échanger entre elles. Lorsqu'une réplique est échangée vers une température de simulation plus élevée, elle est plus apte à franchir des barrières de potentiel et quitter un bassin énergétique. Quand elle est échangée vers une température plus basse, elle explore mieux le fond des bassins. Cela permet de passer des barrières énergétiques élevées et de surmonter la tendance des méthodes d'exploration conventionnelles à rester piégées dans des bassins énergétiques locaux.

L'implémentation originale de l'échange de répliques a été utilisée en couplage avec des

simulations de Monte Carlo. La méthode Monte Carlo par échange de répliques ou REMC permet un échantillonnage efficace de la surface de potentiel pour des polypeptides³⁰. En 1999, Sugita et coll. ont proposé la dynamique moléculaire par échange de répliques³¹. Dès lors, l'échange de répliques n'affecte pas uniquement les coordonnées atomiques, mais aussi les moments cinétiques atomiques qui sont échelonnés.

Un critère souvent utilisé pour définir la probabilité d'échange w entre deux répliques i et j d'énergies E_i et E_j , respectivement aux températures m et n est celui de Metropolis :

$$w = \begin{cases} 1, & \text{si } \Delta \leq 0, \\ e^{-\Delta} & \text{si } \Delta > 0. \end{cases} \quad (3.7)$$

avec

$$\Delta = (\beta_{m(i)} - \beta_{n(j)})(E_i - E_j) \quad (3.8)$$

Seibert et coll.³² ont montré que le REMD constitue une méthode plus efficace que la dynamique moléculaire conventionnelle en terme d'échantillonnage de l'espace des phases. Pour se différencier vraiment d'un ensemble de simulations conventionnelles dans des ensembles NPT à différentes températures et être plus efficace, une simulation de REMD doit mettre en œuvre de nombreux échanges. Pour cela, le choix de l'intervalle de temps entre deux tentatives d'échange, $\delta t_{tentative}$ est important. Il doit être choisi petit, le plus petit possible pour Sindhikara et coll.³³ ou Abraham et Gready³⁴.

Plusieurs auteurs ont étudié l'influence des autres paramètres du REMD sur son efficacité³⁵⁻³⁷. Trois principaux points sont à considérer pour vérifier la validité d'une simulation de REMD :

1. la distribution des températures est-elle bonne ?

Une distribution adéquate des températures doit permettre une répartition uniforme du taux d'échange entre les différentes paires de températures voisines.

2. le nombre de répliques est-il suffisant ?

Si les températures sont bien réparties, leur nombre va conditionner le recouvrement énergétique entre pseudo-trajectoires et ainsi permettre un taux d'échange raisonnable. Un taux correspondant à au moins 10 % d'échange réussi serait satisfaisant.

3. la température la plus élevée est-elle assez haute pour empêcher le système d'être

piégé dans un bassin local d'énergie ?

La pseudo-trajectoire à la plus haute température de simulation doit permettre d'importants changements conformationnels. L'analyse des paramètres structuraux issus de cette pseudo-trajectoire y apporte une réponse.

Il a été montré par Okamoto et coll.³⁸ qu'une distribution exponentielle ou géométrique des températures est préférable à une distribution arithmétique dans laquelle les températures seraient réparties avec une différence constante entre deux températures consécutives. En effet, plus les températures sont élevées, plus larges sont les gammes énergétiques des structures rencontrées lors des simulations, ce qui a pour conséquence d'augmenter la probabilité d'échange. Pour avoir une probabilité d'échange homogène, les températures ont donc besoin d'être plus proches lorsqu'elles sont basses.

Nymeyer³⁹, et Rosta et Hummer⁴⁰ ont étudié l'influence de la distribution et du nombre de répliques, donc de températures, de la simulation lorsque ces valeurs sont déjà assez grandes. Ces auteurs ont démontré qu'augmenter encore le nombre de répliques n'améliore que très peu l'efficacité du REMD en terme d'échantillonnage de l'espace conformationnel si le temps de simulation est suffisamment long.

Par contre, Nymeyer a montré que le choix de la température maximale détermine grandement l'efficacité de la simulation en terme d'exploration et recherche du minimum global. La température maximale doit permettre de franchir facilement les barrières énergétiques pour pouvoir échantillonner les différents puits de potentiel. Si en revanche, la température maximale est déjà suffisamment élevée pour permettre d'explorer facilement les différents bassins, une augmentation de cette température pénaliserait la recherche conformationnelle en sondant un peu plus des zones de la surface de potentiel qui ne sont pas pertinentes.

Nous avons évalué par des calculs MP2/TZVPP et MP2/aug-cc-pVTZ à 12 kcal.mol⁻¹ une barrière de rotation autour du dièdre N-C_α-C-N du système G₃ isolé dans sa forme non-zwitterionique⁴¹ (voir p. 127 pour la référence). Notons qu'il ne s'agit d'une évaluation sur-estimée puisque calculée sans relaxation de la structure pour chaque valeur de l'angle dièdre. Considérons une simulation REMD d'un système de polyalanine sodée que nous décrirons plus en détail (p. 63). Pour échantillonner correctement l'espace des phases de A₁₂-Na⁺, il est nécessaire d'engendrer toutes les combinaisons possibles de rotations autour des 12 dièdres NC_α-C-N. Si la présence du groupement méthyle augmente la barrière de rotation

pour les polyalanines par rapport au polyglycines, 12 kcal.mol^{-1} rest une surestimation. La somme des barrières individuelles de chaque rotation autour de chaque dièdre est une surestimation des barrières à franchir pour explorer tout l'espace conformationnel. En effet, des mécanismes faisant intervenir des déformations concertées, impliquant d'autres modes de vibration, sont probablement bien plus favorables. On peut alors estimer que les barrières les plus hautes nécessaires à franchir pour explorer tous les bassins de potentiel sont toutes inférieures à $144 \text{ kcal.mol}^{-1}$, même en considérant des changements de coordination du cation. Or, à 600 K, l'énergie thermique de la structure utilisée comme point de départ de la dodéca-alanine attachée à un sodium est évaluée à $147,4 \text{ kcal.mol}^{-1}$ en B3LYP-D/SVP. Nous estimons donc qu'une température maximale de 600 K pour la simulation REMD de système $A_{12}\text{-Na}^+$ est suffisante pour explorer aisément son espace conformationnel.

L'aptitude de la dynamique moléculaire par échange de répliques à trouver les conformations de plus basses énergies a été démontrée pour des systèmes biologiques en solution⁴² ou en phase gazeuse⁴³.

3.3.2 Tirez : implémentation du REMD avec Tinker

Penev et coll. ont réalisé le couplage entre REMD et Tinker, *via* le logiciel Tirez⁴⁴. Comme il appelle Tinker pour effectuer les portions de trajectoires de dynamique moléculaire, Tirez nous permet de réaliser des simulations REMD avec le champ de force polarisable AMOEBA.

La parallélisation de Tirez 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 cœur, unité de calcul d'un processeur qui peut en contenir un ou plus. Toutefois, parfois par abus de langage, le terme « multiprocesseur » est encore utilisé. Le REMD peut ainsi être parallélisé sur autant de cœurs que de répliques impliquées. Comme Tirez ne nécessite pas d'importants échanges *via* le réseau, le temps réel de la simulation dépend peu du nombre de répliques demandées, à condition d'y consacrer le même nombre de cœurs. Cette affirmation est vraie si on fait l'approximation que l'énergie de polarisation dans AMOEBA converge en moyenne avec le même nombre de pas pour les portions de toutes les trajectoires d'une même simulation.

Une particularité de l'implémentation de Tirez est que l'échange de répliques est substitué par l'échange de températures, ce qui y est équivalent en principe. Après chaque intervalle de temps d'échange, lorsque le critère de Metropolis est satisfait entre deux répliques de températures consécutives, l'échange des températures a lieu. Comme proposé par Sugita et coll.³¹, Tirez échelonne les vitesses, en conséquence, de la manière suivante :

$$v_{nouvelle} = \sqrt{\frac{T_{nouvelle}}{T_{ancienne}}} v_{ancienne} \quad (3.9)$$

avec $v_{nouvelle}$ et $v_{ancienne}$, respectivement les vitesses correspondant aux températures $T_{nouvelle}$ et $T_{ancienne}$ des deux portions de trajectoire encadrant l'échange. Ainsi, chaque trajectoire continue avec la même structure (réplique) et les positions atomiques au pas suivant sont calculées par intégration des équations du mouvement à partir des coordonnées précédentes du potentiel qui y est associé et des nouvelles vitesses échelonnées.

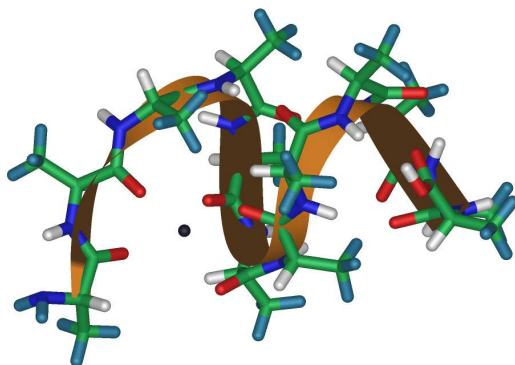
3.3.3 Traitement des données issues d'une simulation REMD avec Tirez

Si Tirez donne une interface simple pour la réalisation de simulations REMD utilisant Tinker, les outils proposés pour le traitement des résultats n'étaient pas directement exploitables. Il nous a donc fallu mettre au point ou adapter les outils qui nous semblaient nécessaires.

Tout d'abord, les trajectoires sont calculées par réplique et non par température. Nous re-concaténons les pseudo-trajectoires par température, afin de disposer d'ensembles cohérents d'un point de vue thermodynamique. Les fichiers y correspondant sont divisés et réassemblés après lecture dans un fichier contenant les index de température par réplique après chaque tentative d'échange.

Une fois ces pseudo-trajectoires reconstruites par température, nous pouvons contrôler l'efficacité du REMD. Comme exemple, utilisons une simulation REMD d'une polyalanine sodée : $A_{12}\text{-Na}^+$ sous sa forme canonique, *i.e.* non-zwitterionique. Les séquences de polyalanines sont connues pour former facilement des hélices en solution. De plus, des études par mobilité ionique menées par le groupe de Jarrold⁴⁵ ont montré que la structure de $A_{12}\text{-Na}^+$ est probablement hélicoïdale et non-zwitterionique. La structure initiale (Fig. 3.1) de la simulation est pentacoordinée au sodium par les carbonyles 1, 2, 5, 6 et 7 (la numérotation

| $N_{répliques}$ | T_{min} | T_{max} | $N_{tentatives}$ | $N_{pas}(tentatives)$ | δt | $\delta t_{sauvegarde}$ | durée |
|-----------------|-----------|-----------|------------------|-----------------------|------------|-------------------------|-------|
| 12 | 200 K | 600 K | 499 | 10000 | 1 fs | 5 ps | 5 ns |

TAB. 3.1: Paramètres des simulations REMD de $A_{12}Na^+$ FIG. 3.1: Simulation REMD de $A_{12}-Na^+$: structure initiale

des résidus commence à l'extrémité N-terminale). Le cation est volontairement éloigné du carboxyle terminal pour rendre la structure très éloignée de l'hélice. On peut considérer le peptide comme étant formé d'une partie globulaire organisée autour du cation et une partie plus éloignée du sodium qui est structurée par un réseau de liaisons hydrogène.

Les paramètres de la simulation REMD de $A_{12}Na^+$ sont regroupés dans le tableau 3.1. Le champ de force utilisé pour calculer les forces est AMOEBA. Chaque simulation REMD de $A_{12}Na^+$ effectue 499 tentatives d'échange séparées par des portions de trajectoire de dynamique moléculaire dans l'ensemble thermodynamique NPT. La simulation met en jeu 12 répliques évoluant à des températures réparties de manière géométrique entre 200 et 600 K. Ces portions de trajectoires sont propagées avec un pas de temps de 1 fs afin d'éviter le phénomène de *polarization catastrophe* durant le calcul des forces AMOEBA. Ces tentatives ont lieu tous les 10 000 pas de dynamique, soit toutes les 10 ps. Chaque trajectoire a donc une durée de 5 ns. Les structures rencontrées sont sauvegardées toutes les 5 ps, ce qui amène à sauvegarder 1 000 structures par trajectoire. Le nombre de structures sauvegardées sur l'ensemble de la simulation, ici 12 000, n'est pas un facteur limitant en terme de stockage mais il est important pour le post-traitement de la simulation. Il doit être suffisant pour échantillonner correctement la trajectoire mais le traitement d'un nombre plus important de structures pourrait être prohibitif. Après cette première simulation REMD, la structure d'énergie la plus basse obtenue a été utilisée comme structure de départ pour

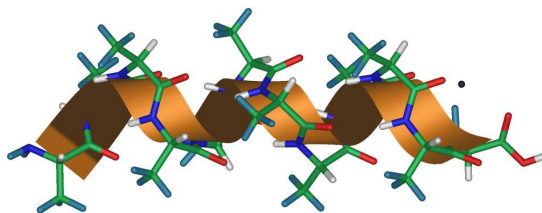


FIG. 3.2: Simulation REMD de $A_{12}\text{-Na}^+$: structure la plus stable

une deuxième simulation ou génération. De la même manière, la structure la plus basse en énergie a été utilisée comme structure initiale pour une troisième simulation qui n'a pas fourni de structure plus stable. Il n'a donc pas été considéré nécessaire d'effectuer une quatrième génération de REMD. Cette structure (Fig. 3.2) la plus basse en énergie est une structure en hélice α dans laquelle le sodium est coordonné aux carbonyles 9, 10 et 12. Ces simulations REMD ont donc permis d'obtenir une structure basse en énergie. Elle est plus basse en énergie que la structure de départ de $24,6 \text{ kcal.mol}^{-1}$ au niveau B3LYP-D/TZVPP aux géométries B3LYP-D/SVP. Cette énergie relative est $25,9 \text{ kcal.mol}^{-1}$ en AMOEBA. De surcroît, la structure obtenue est très différente de la structure initiale en termes de complexation de Na^+ et de liaisons hydrogène, ce qui montre que les simulations ont permis d'explorer de nombreux autres bassins de potentiel. Les illustrations présentées dans ce chapitre proviennent de la deuxième simulation de REMD. Les échanges de températures ayant eu lieu lors de cette simulation sont présentés dans la figure 3.3. Dans cette figure, chaque couleur correspond à une réplique. On observe des échanges quand pour une réplique, la température passe à celle qui lui est immédiatement inférieure ou supérieure. Les taux d'échange sont certes disponibles avant reconcaténation mais après, nous pouvons vérifier les recouvrements énergétiques des différentes pseudo-trajectoires (ex. Fig. 3.4). Cela nous renseigne sur la qualité de répartition des températures et du choix du nombre de celles-ci. Le graphique montre que plus la température est élevée, plus les structures rencontrées dans les simulations sont hautes en énergie et plus la répartition énergétique des structures est étalée. Cet étalement énergétique permet de meilleurs recouvrements énergétiques aux hautes températures. Les pseudo-trajectoires à 543 et 600 K se recouvrent mieux que les pseudo-trajectoires à 200 et 221 K alors que l'écart en température dans le premier cas est plus du double de celui du second. Cela illustre la nécessité d'avoir une distribution plus

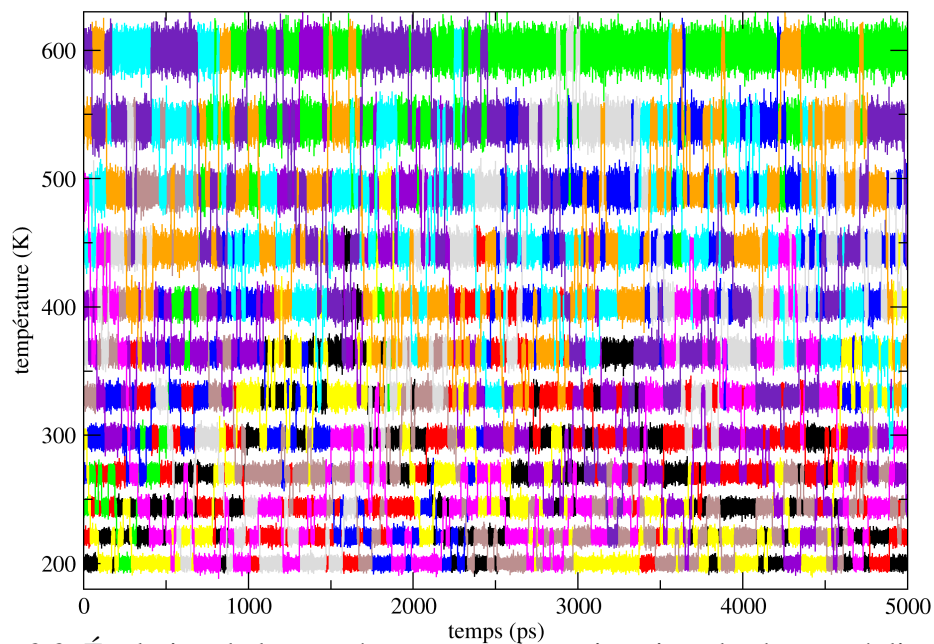


FIG. 3.3: Évolution de la température pour les trajectoires de chaque réplique de la simulation REMD, exemple de $A_{12}\text{-Na}^+$

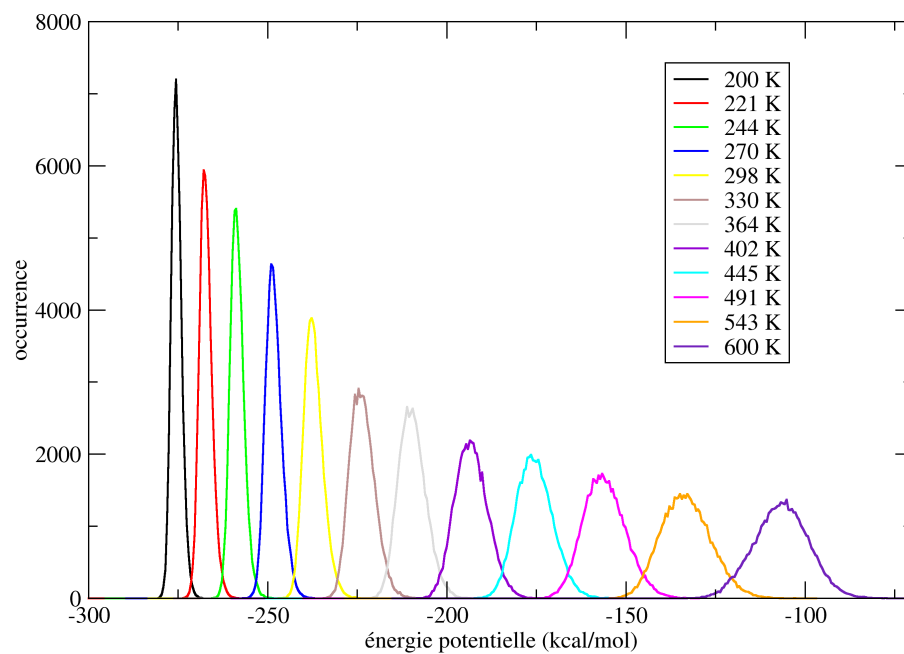


FIG. 3.4: Recouvrement des énergies potentielles pour les différentes température du REMD, exemple de $A_{12}\text{-Na}^+$

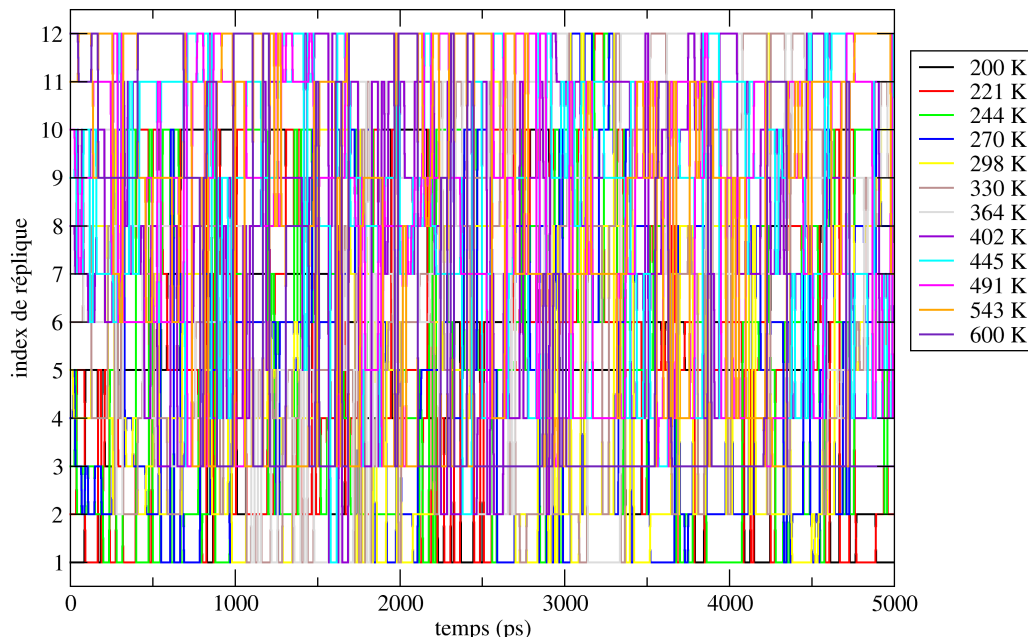


FIG. 3.5: Évolution de l'index de réplique de chaque température, exemple de $A_{12}\text{-Na}^+$

resserrée dans les basses températures. Dans le cas présent, l'utilisation de 12 températures sélectionnées selon une distribution géométrique semble satisfaisante en terme de recouvrement énergétique et permet donc un taux raisonnable d'échanges réussis (26%).

Le suivi des index de réplique par température (ex. Fig. 3.5) nous permet aussi de nous assurer qu'une réplique est passée par toutes les températures de la simulation, ce qui est souhaitable pour pouvoir affirmer qu'une surface de potentiel a été bien échantillonnée.

Enfin, la visualisation ou la caractérisation de la pseudo-trajectoire à la température la plus haute est une étape nécessaire pour vérifier que la simulation n'est pas restée piégée dans un puits. Dans l'exemple de la dodéca-alanine sodée, nous avons choisi de présenter les distances d'attachement au sodium (Fig. 3.6) pour montrer la diversité structurale obtenue au cours de la pseudo-trajectoire de plus haute température (600 K). Pour mieux les discerner, nous n'avons représenté ces données que pour une portion de la pseudo-trajectoire, qui s'étend sur un dixième de la durée totale, soit de 4,5 ns à 5 ns. Durant cette période, chacun des sites coordinables au sodium a été au moins une fois lié et au moins une fois non-lié au cation, la frontière étant fixée à 2,6 ns. Cela montre que la température maximale utilisée

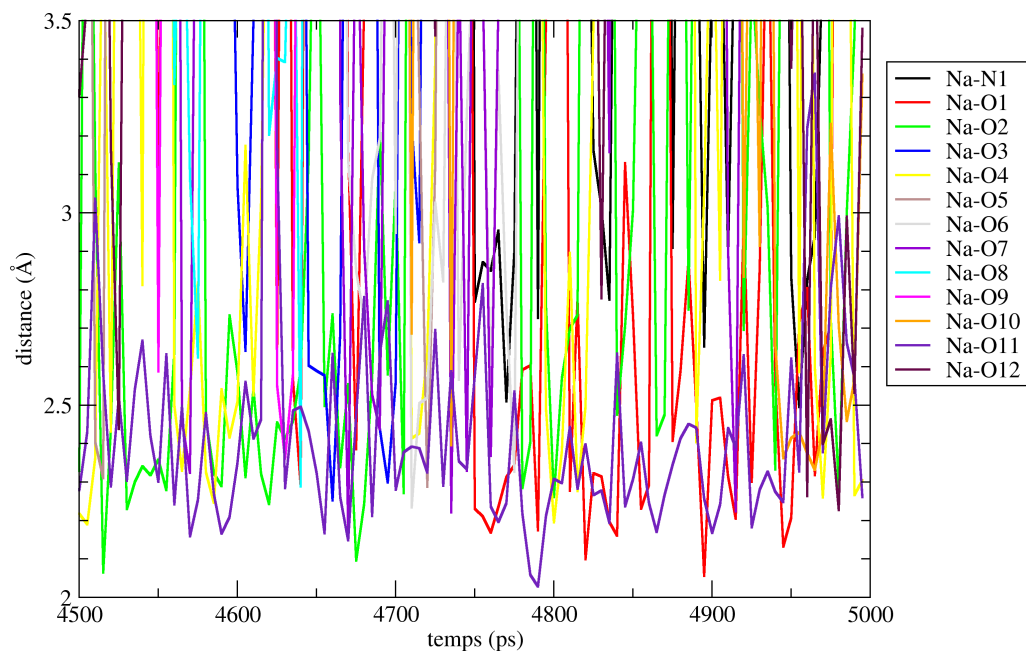


FIG. 3.6: Évolution des distances d'attachement au sodium au cours d'une portion de la pseudo-trajectoire de plus haute température du REMD de $A_{12}\text{-Na}^+$

(600 K) est suffisante pour franchir les barrières énergétiques liées à des changements de coordination et des changements structuraux importants dans $A_{12}\text{-Na}^+$.

Si ces conditions sont satisfaites dans une simulation REMD, on peut considérer la simulation comme bonne et commencer à l'exploiter. L'énergie cinétique à chaque instant des trajectoires est un facteur d'éloignement des structures sauvegardées par rapport au fond des bassins énergétiques rencontrés. Dans l'optique d'une recherche conformationnelle, nous optimisons toutes les structures sauvegardées lors du calcul des trajectoires. Nous pouvons ainsi identifier plus facilement la répartition de population dans les bassins au cours des pseudo-trajectoires par le nombre d'occurrence de chacun des *minima*.

En pratique, dès que la simulation Tirez est terminée, les trajectoires sont reconcaténées, ainsi que les valeurs instantanées de propriétés comme la température, l'énergie cinétique, potentielle... L'optimisation des structures sauvegardées est distribuée sur les cœurs réservés pour la simulation. Comme en principe on a une réplique, soit une trajectoire simulée, par cœur, toutes les structures d'une même pseudo-trajectoire sont optimisées avec Tinker de

manière séquentielle. On a ainsi une distribution des tâches facilement insérée de façon cohérente dans un script de soumission sur une grappe de calcul.

3.4 Comparaison et bilan des méthodes d'exploration

Parmi les nombreux algorithmes de recherche conformationnelle du minimum global, peu sont couplés efficacement avec le champ de force AMOEBA. Ces algorithmes doivent être interfacés avec Tinker et les paramètres utilisés, soigneusement choisis pour éviter une *polarization catastrophe*, ce qui n'est pas toujours trivial.

En dynamique moléculaire, nous avons constaté au cours de toutes nos simulations que le phénomène de *polarization catastrophe* est évité à condition d'utiliser un pas d'intégration suffisamment petit. Pour des pas inférieurs ou jusqu'à 1 fs nous n'avons pas observé ce problème.

La dynamique moléculaire et le REMD qui en dérive présentent l'avantage de ne pas nécessiter la définition de variables géométriques, autres que les coordonnées cartésiennes. De plus, Tirez nous donne une interface pour le REMD avec la dynamique moléculaire disponible dans Tinker qui dispose du champs de force polarisable AMOEBA.

Des variantes du REMD permettent d'améliorer l'efficacité d'échantillonnage et de l'utilisation des ressources informatiques du REMD. On peut citer par exemple le REMD à réservoir qui consiste à utiliser une simulation simple ou multiple à haute température pour permettre une plus grande diversité structurale. Par exemple, on peut utiliser comme réservoir de structures, une trajectoire de dynamique moléculaire conventionnelle longue et à haute température ou un REMD à hautes températures proches entre elles pour maximiser les échanges. Plus de détails sur le REMD à réservoir sont présentés dans un article de T. Head-Gordon⁴⁶ qui y discute aussi la température maximale à utiliser.

Nous avons comparé la variété de structures rencontrées dans le REMD, au cours de la pseudo-trajectoire à 298 K avec celles issues d'une trajectoire de dynamique moléculaire conventionnelle à 300 K. L'évolution temporelle des énergies potentielles des bassins rencontrés, *i.e.* après minimisation des structures (Fig. 3.7), montre une diversité bien plus im-

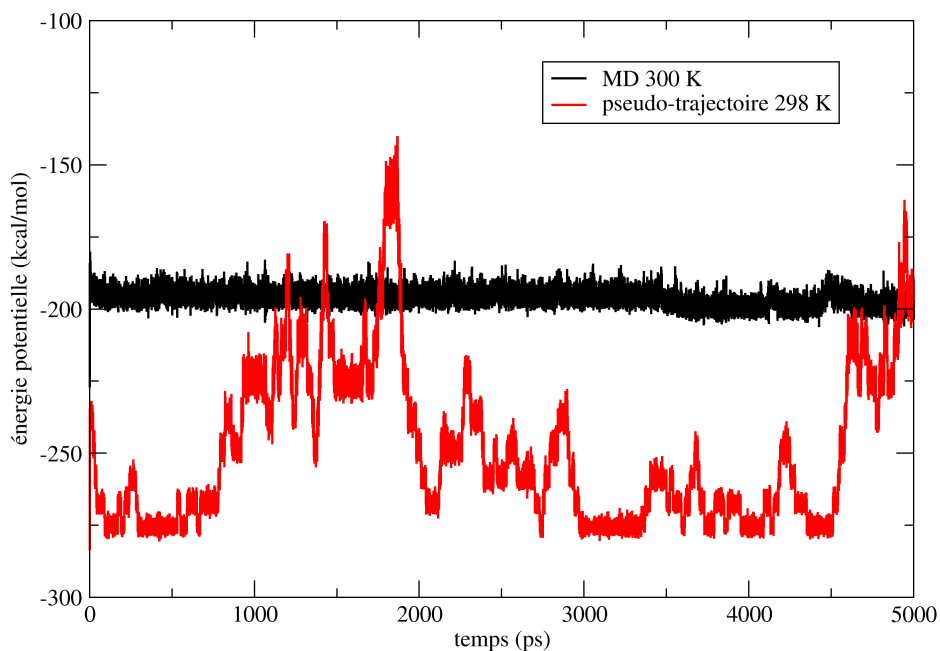


FIG. 3.7: Évolution de l'énergie potentielle des bassins : comparaison sur l'exemple de A_{12} - Na^+ de l'efficacité d'échantillonnage entre une pseudo-trajectoire à 298 K de REMD et une trajectoire de dynamique moléculaire conventionnelle à 300 K

portante pour la pseudo-trajectoire à 298 K que pour la trajectoire conventionnelle à 300 K.

Ces méthodes permettent, en outre, d'accéder aux propriétés dynamiques et thermodynamiques du système et de mieux comprendre les mécanismes des transitions structurales, deux points auxquels nous nous intéresserons particulièrement par la suite.

Bibliographie

- [1] J. C. LAGARIAS, J. A. REEDS, M. H. WRIGHT et P. E. WRIGHT : Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions. *SIAM J. Optim.*, 9:112–147, 1998.
- [2] J. A. NELDER et R. MEAD : A Simplex Method for Function Minimization. *Comput. J.*, 7:308–313, 1965.
- [3] N. METROPOLIS : The beginning of the Monte Carlo method. *Los Alamos Sci.*, 15:125–130, 1987.
- [4] N. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER et E. TELLER : Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21:1087–1092, 1953.
- [5] Z. LI et H. A. SCHERAGA : Monte carlo-minimization approach to the multiple-minima problem in protein folding. *PNAS*, 84:6611–6615, 1987.
- [6] E. O. PURISIMA et H. A. SCHERAGA : An approach to the multiple-minima problem in protein folding by relaxing dimensionality : Tests on enkephalin. *J. Mol. Biol.*, 196:697–709, 1987.
- [7] E. H. L. AARTS et P. J. M. van LAARHOVEN : Statistical cooling : A general approach to combinatorial optimization problems. *Philips J. R.*, 40:193–226, 1985.
- [8] Hongyu ZHANG, Yuting YANG, Luhua LAI et Yougi TANG : Conformational analysis of two glycoproteins : a Monte Carlo simulated annealing approach using a soft-sphere potential. *Carbohydr. Res.*, 284:25–34, 1996.
- [9] V. VILLANI et A. M. TAMBURRO : Improvements to the molecular dynamics simulated annealing procedures : conformational search of the elastin tetrapeptide Boc-Gly-Leu-Gly-Gly-NMe. *J. Mol. Struct. THEOCHEM*, 308:141–157, 1994.

- [10] L. P. B. SCOTT, J. CHAHINE et J. R. RUGGIERO : use of genetic algorithms and solvation potential to study peptides structure. *Appl. Math. and Comput.*, 195:515–522, 2008.
- [11] E. THIRIOT et G. MONARD : Combining a genetic algorithm with a linear scaling semiempirical method for protein-ligand docking. *J. Mol. Struct. THEOCHEM*, 898: 31–41, 2009.
- [12] G. D. BILLING : On the use of Ehrenfest's theorem in molecular scattering. *Chem. Phys. Lett.*, 100:535–539, 1983.
- [13] P. EHRENFEST : Aus der Schrödingerschen Gleichung läßt sich durch eine kurze elementare Rechnung ohne Veruachlässigung die Beziehung. *Z. Phys.*, 45:455, 1927.
- [14] G. BERGERON, P. C. HIBERTY et C. LEFORESTIER : A quantum-mechanical investigation of collinear models for collision-induced dissociation. *Chem. Phys.*, 93:253–264, 1985.
- [15] G. MONARD, M. I. BERNAL-URUCHURTU, A. van der VAART, K. M. Jr MERZ et M. F. RUIZ-LÓPEZ : Simulation of Liquid Water Using Semiempirical Hamiltonians and the Divide and Conquer Approach. *J. Phys. Chem. A*, 109:3425–3432, 2005.
- [16] R. CAR et M. PARRINELLO : Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.*, 55:2471–2474, 1985.
- [17] D. BEEMAN : Some multistep methods for use in molecular dynamics calculations. *J. Comput. Phys.*, 20:130–139, 1976.
- [18] B. R. BROOKS : Algorithms for Molecular Dynamics at Constant Temperature and Pressure. Rapport technique, DCRT Report, NIH, April 1988.
- [19] M. LEVITT, M. HIRSHBERG, R. SHARON et V. DAGGETT : Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Comput. Phys. Commun.*, 91:215–231, 1995.
- [20] H. J. C. BERENDSEN, J. P. M. POSTMA, W. F. van GUNSTEREN, A. DiNOLA et J. R. HAAK : Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
- [21] H. C. ANDERSEN : Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 1980.

- [22] S. C. HARVEY, R. K.-Z. TAN et T. E. III CHEATHAM : The flying ice cube : Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.*, 19:726–740, 1998.
- [23] S. NOSÉ : A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511–519, 1984.
- [24] W. G. HOOVER : Canonical dynamics : Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.
- [25] U. H. E. HANSMANN et Y. OKAMOTO : Numerical comparisons of three recently proposed algorithms in the protein folding problem. *J. Comput. Chem.*, 18:920–933, 1997.
- [26] Ulrich H.E. HANSMANN et Yuko OKAMOTO : New Monte Carlo algorithms for protein folding. *Curr. Opin. Struct. Biol.*, 9:177–183, 1999.
- [27] R. H. SWENDSEN et J.-S. WANG : Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.*, 57:2607–2609, 1986.
- [28] K. HUKUSHIMA et K. NEMOTO : Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phy. Soc. Jpn*, 65:1604–1608, 1996.
- [29] M. MARINARI, G. PARISI et J. J. RUIZ-LORENZO : *A. P. Young, ed., Spin Glasses and Random Fields (Series on Directions in condensed Matter Physics)*, volume 12, page 59. World Scientific, Singapore and New Jersey, 1998.
- [30] A. A. PODTELEZHNIKOV et D. L. WILD : Exhaustive Metropolis Monte Carlo sampling and analysis of polyalanine conformations adopted under the influence of hydrogen bonds. *Proteins*, 61:94–104, 2005.
- [31] Y. SUGITA et Y. OKAMOTO : Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314:141–151, 1999.
- [32] M. M. SEIBERT, A. PATRIKSSON, B. HESS et D. van der SPOEL : Reproducible Polypeptide Folding and Structure Prediction using Molecular Dynamics Simulations. *J. Mol. Biol.*, 354:173–183, 2005.
- [33] D. SINDHIKARA, Y. MENG et A. E. ROITBERG : Exchange frequency in replica exchange molecular dynamics. *J. Chem. Phys.*, 128:024103, 2008.
- [34] M. J. ABRAHAM et J. E. GREASY : Ensuring Mixing Efficiency of Replica-Exchange Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, 4:1119–1128, 2008.

- [35] W. ZHENG, M. ANDREC, E. GALLICCHIO et R. M. LEVY : Simulating replica exchange simulations of protein folding with a kinetic network model. *P. Natl Acad. Sci. USA*, 104:15340–15345, 2007.
- [36] S. TREBST, M. TROYER et U. H. E. HANSMANN : Optimized parallel tempering simulations of proteins. *J. Chem. Phys.*, 124:174903, 2006.
- [37] R. DENSCHLAG, M. LINGENHEIL et P. TAVAN : Efficiency reduction and pseudo-convergence in replica exchange sampling of peptide folding-unfolding equilibria. *Chem. Phys. Lett.*, 458:244–248, 2008.
- [38] Y. OKAMOTO, M. FUKUGITA, T. NAKAZAWA et H. KAWAI : alpha-Helix folding by Monte Carlo simulated annealing in isolated C-peptide of ribonuclease A. *Protein Eng.*, 4:639–647, 1991.
- [39] H. NYMEYER : How Efficient Is Replica Exchange Molecular Dynamics ? An Analytic Approach. *J. Chem. Theory Comput.*, 4:626–636, 2008.
- [40] E. ROSTA et G. HUMMER : Error and efficiency of replica exchange molecular dynamics simulations. *J. Chem. Phys.*, 131:165102, 2009.
- [41] D. SEMROUNI, G. OHANESSIAN et C. CLAVAGUÉRA : Structural, Energetic, and Dynamical Properties of Sodiated Oligoglycines : Relevance of a Polarizable Force Field. *Phys. Chem. Chem. Phys.*, 12:3450–3462, 2010.
- [42] A. BAUMKETNER et J.-E. SHEA : The Structure of the Alzheimer Amyloid [beta] 10-35 Peptide Probed through Replica-Exchange Molecular Dynamics Simulations in Explicit Solvent. *J. Mol. Biol.*, 366:275–285, 2007.
- [43] J.-C. POULLY, F. LECOMTE, N. NIEUWJAER, B. MANIL, J.-P. SCHERMANN, C. DESFRANÇOIS, F. CALVO et G. GRÉGOIRE : Probing the specific interactions and structures of gas-phase vancomycin antibiotics with cell-wall precursor through IRMPD spectroscopy. *Phys. Chem. Chem. Phys.*, 12:3606–3615, 2010.
- [44] E. S. PENEV, S. LAMPOUDI et J.-E. SHEA : TiReX : Replica-exchange molecular dynamics using Tinker. *Comput. Phys. Commun.*, 180:2013–2019, 2009.
- [45] M. KOHTANI, B. S. KINNEAR et M. F. JARROLD : Metal-Ion Enhanced Helicity in the Gas Phase. *J. Am. Chem. Soc.*, 122:12377–12378, 2000.

- [46] J. Z. RUSCIO, N. L. FAWZI et T. HEAD-GORDON : How hot ? Systematic convergence of the replica exchange method using multiple reservoirs. *J. Comput. Chem.*, 31:620–627, 2010.

Chapitre 4

Sélection, identification, différenciation de structures

L'exploration d'une hyper-surface d'énergie potentielle à l'aide d'un champ de force peut facilement engendrer quelques milliers de conformations stables pour un polypeptide. Beaucoup de ces structures ne sont pas tout à fait uniques. Pour un grand nombre des structures obtenues lors de l'exploration, il en existe au moins une autre qui lui soit quasi-identique. Les causes potentielles sont multiples et dépendent des algorithmes utilisés. Afin de limiter le nombre de structures à considérer, il convient de les discerner entre elles et les filtrer. Ce filtrage peut être effectué sur des critères énergétiques ou structuraux.

Après cette étape de discernement, ces structures peuvent être considérées comme uniques. Seules certaines de ces structures sont sélectionnées pour être étudiées avec une méthode quantique, plus fiable (*ab initio*, DFT). Les critères utilisés pour le tri des structures sont très dépendants du système et de ce que l'on attend des structures sélectionnées (sur la base de données expérimentales, de l'intuition chimique...).

L'objet de ce chapitre est de décrire les outils de discernement, de tri et de regroupement de structures que nous avons construits au cours de ce travail.

4.1 Discernement de structures

Cette étape conditionne le nombre de structures à étudier avec des méthodes plus fiables. Le but, ici, est de limiter la redondance dans ces calculs coûteux en temps de processeur. Cette étape est donc d'importance capitale.

Le discernement de quelques structures peut certes être effectué « à l'œil » mais cette démarche est subjective. La subjectivité et la répétition de cette tâche pour la quantité de géométries engendrées lors de chaque recherche conformationnelle justifient la mise en œuvre d'une procédure objective et automatisée.

4.1.1 Propriétés géométriques

L'expérience numérique indique que la probabilité de dégénérescence énergétique à une précision de 10^{-3} kcal.mol⁻¹ est très faible pour des structures vraiment différentes. En guise de première étape de filtrage, on considère donc comme étant identiques deux structures isoénergétiques à 10^{-4} kcal.mol⁻¹ près.

Un critère énergétique n'est pas suffisant pour satisfaire à un tri à la fois efficace qui filtrerait beaucoup de structures, et sûr dans la mesure où il n'écarterait pas des structures différentes à cause d'une proximité énergétique. Il convient alors d'effectuer un filtrage des structures selon des considérations géométriques. L'identité de deux structures équivaut à leur superposabilité mais leur similarité est une notion subjective. Elle repose sur des évaluations de similitude globale ou locale et dépend des seuils fixés pour ces dernières. La question de la similarité structurale est donc une question complexe et ouverte.

Dans un milieu isotrope, le système étudié conserve ses propriétés indépendamment de son orientation dans ce milieu. L'étude de la superposabilité de deux molécules nécessite la réalisation d'une série de rotations d'une molécule vis-à-vis de l'autre. La rotation s'arrête quand les structures sont à leur maximum de superposabilité. On peut parler d'alignement de structures mais cette terminologie peut avoir une signification sensiblement différente.

En bioinformatique, notamment, l'alignement structural de biopolymères ne nécessite pas que les structures comparées correspondent au même système. L'alignement structural, en

bioinformatique, évoque en général un alignement de séquences basé sur des données structurales. L'alignement structural est alors notamment utilisé pour mesurer l'homologie structurale entre deux séquences protéiques dans l'objectif d'étudier la relation entre séquence et fonction protéiques ou une relation liée à l'évolution biologique. Pour éviter les confusions, nous parlerons d'étape de superposition en ce qui concerne les rotations maximisant la superposabilité.

Il est vrai que par définition, la comparaison en coordonnées internes possède l'avantage de ne nécessiter aucune étape de superposition préalable. Toutefois la comparaison des coordonnées cartésiennes est plus facile que celle des coordonnées internes. En effet, celles-ci sont à déterminer pour chaque système et leur définition est subjective. De plus, de petites variations des coordonnées internes comme les angles dièdres peuvent avoir des conséquences importantes sur la structure du système¹.

4.1.2 Outils mathématiques

Les mathématiques fournissent divers outils pour construire des fonctions de similarité ou de dissimilarité.

Moyenne quadratique : RMSD

On appelle RMSD, la moyenne quadratique de la différence entre deux jeux de valeurs x_1 et x_2 . Elle est exprimée dans l'équation 4.1 pour N valeurs dans chaque jeu.

$$RMSD = \frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^N (x_1(i) - x_2(i))^2} \quad (4.1)$$

On peut pondérer le RMSD pour obtenir des valeurs sans unité, renseignant sur la distribution des valeurs.

Le RMSD peut être pondéré par l'amplitude des valeurs $d(i) = x_1(i) - x_2(i)$. On parle alors de moyenne quadratique normalisée (NRMSD, voir équation 4.2). Le NRMSD² est utilisé pour la comparaison de structure(s) théorique(s) à des données expérimentales³. L'avantage

de cette mesure est que son domaine de valeur est borné à 1 et ne dépend pas de la taille du système.

$$NRMSD = \frac{1}{(x_{max} - x_{min})\sqrt{N}} \sqrt{\sum_{i=1}^N (x_1(i) - x_2(i))^2} \quad (4.2)$$

Un RMSD relatif (équation 4.3) est parfois utilisé. Il s'agit d'une pondération du RMSD entre deux structures par le RMSD moyen pour deux structures (protéiques) de rayon de giration équivalent⁴. Cette mesure est sans dimension et présente l'avantage de pouvoir être exprimée comme un pourcentage.

$$RRMSD = \frac{1}{RMSD_{moyen}\sqrt{N}} \sqrt{\sum_{i=1}^N (x_1(i) - x_2(i))^2} \quad (4.3)$$

Pour toutes les variations de RMSD, une pondération atomique par le coefficient w_i est possible (éq. 4.4). Ce coefficient peut simplement être la proportion massique de l'atome dans le système ou avoir une expression plus élaborée comme une gaussienne pour des besoins particuliers d'alignement de domaines protéiques⁵.

$$WRMSD = \frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^N w_i (x_1(i) - x_2(i))^2} \quad (4.4)$$

Erreur moyenne sans signe (UME)

L'erreur moyenne sans signe entre deux jeux de valeurs est la moyenne arithmétique de la valeur absolue des différences entre les paires de données des deux jeux (éq. 4.5).

$$UME = \frac{1}{N} \sum_{i=1}^N \|x_1(i) - x_2(i)\| \quad (4.5)$$

Certaines fonctions de dissimilarité entre deux structures peuvent être faussées si elles ne tiennent pas compte d'une rotation possible du système. Il convient dès lors de pratiquer une superposition de ces structures à comparer. Concrètement cela revient à trouver une matrice de rotation permettant de réduire au maximum la valeur de dissimilarité. Par conséquent, elle peut être calculée après itération du calcul de la fonction de dissimilarité jusqu'à minimisation de sa valeur⁶. Kabsch a développé une méthode^{7,8} pour rechercher la matrice de

rotation minimisant la dissimilarité de deux structures de manière non-itérative, ce qui réduit le temps de calcul du RMSD. Cette méthode est, entre autres, utilisée par le logiciel VMD⁹. La fonction de superposition de Tinker utilise une méthode des quaternions^{10,11}.

4.1.3 Discussion

Le problème de ce type de fonction est celui de la multidimensionnalité : on veut associer une mesure globale à un ensemble de propriétés, de mesures partielles. Ces outils mathématiques ne donnent pas accès à la localisation des différences structurales.

Cette importance est très subjective et il appartient à chacun de définir des critères d'unicité pour les fonctions de dissimilarité. Ce critère sera d'autant plus petit que l'on s'intéresse aux détails des structures.

Pour les structures de taille importante (protéines et autres biopolymères), afin d'accélérer le tri, il arrive que l'on ne compare pas les positions de tous les atomes mais qu'on considère uniquement les atomes nécessaires pour donner une description globale de la structure. Il est par exemple fréquent de comparer des valeurs de RMSD prenant en compte les C_α uniquement dans le cas de protéines^{12,13}, ce qui revient à négliger les chaînes latérales.

En ce qui nous concerne, le RMSD est systématiquement effectué, soit pour tous les atomes, soit pour tous à l'exception des atomes d'hydrogène lorsque ce sera précisé. De même, nous le préciserons lorsque nous utiliserons des RMSD pondérés par la masse des atomes.

Influence de la localisation d'une modification structurale sur la dissimilarité

Nous avons étudié cette question sur des variations de G8_ha3-0 (Fig. 4.1). Il s'agit d'une structure hélicoïdale d'octaglycine zwitterionique isolée, avec $\Phi = -58^\circ$ et $\Psi = -47^\circ$ sur l'ensemble de son squelette peptidique. Les structures de G8_ha3-a à G8_ha3-h correspondent chacune à une modification d'un angle dièdre Φ (de -58 à $+58$ rad) de G8_ha3 en allant du premier résidu (N-terminal) au dernier résidu (C-terminal). Les comparaisons de RMSD entre ces structures (Tab. 4.1) montrent qu'une modification, certes importante,

| structure | Φ_1 | Φ_2 | Φ_3 | Φ_4 | Φ_5 | Φ_6 | Φ_7 |
|-----------|----------|----------|----------|----------|----------|----------|----------|
| G8_ha3-0 | - | - | - | - | - | - | - |
| G8_ha3-a | + | - | - | - | - | - | - |
| G8_ha3-b | - | + | - | - | - | - | - |
| G8_ha3-c | - | - | + | - | - | - | - |
| G8_ha3-d | - | - | - | + | - | - | - |
| G8_ha3-e | - | - | - | - | + | - | - |
| G8_ha3-f | - | - | - | - | - | + | - |
| G8_ha3-g | - | - | - | - | - | - | + |

Pour chaque variation de G₈_ha3-0, un angle dièdre Φ passe de +58 à -58 rad.

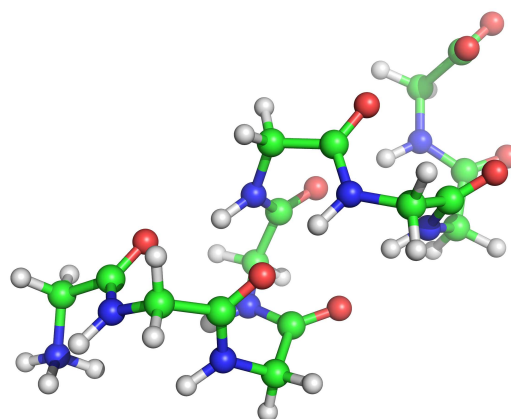


FIG. 4.1: Géométrie de G₈_ha3-0, structure hélicoïdale de la forme zwitterionique de G₈.

| structure | G8_ha3-0 | G8_ha3-a | G8_ha3-b | G8_ha3-c | G8_ha3-d | G8_ha3-e | G8_ha3-f | G8_ha3-g |
|-----------|----------|----------|----------|----------|----------|----------|----------|----------|
| G8_ha3-0 | 0,00 | 0,97 | 2,03 | 2,45 | 2,66 | 2,51 | 1,98 | 0,76 |
| G8_ha3-a | 0,97 | 0,00 | 1,81 | 2,69 | 2,89 | 2,59 | 2,16 | 1,26 |
| G8_ha3-b | 2,03 | 1,81 | 0,00 | 2,51 | 3,69 | 3,09 | 2,45 | 2,22 |
| G8_ha3-c | 2,45 | 2,69 | 2,51 | 0,00 | 2,41 | 3,56 | 2,97 | 2,46 |
| G8_ha3-d | 2,66 | 2,89 | 3,69 | 2,41 | 0,00 | 2,45 | 3,74 | 2,73 |
| G8_ha3-e | 2,51 | 2,59 | 3,09 | 3,56 | 2,45 | 0,00 | 2,63 | 2,76 |
| G8_ha3-f | 1,98 | 2,16 | 2,45 | 2,97 | 3,74 | 2,63 | 0,00 | 2,04 |
| G8_ha3-g | 0,76 | 1,26 | 2,22 | 2,46 | 2,73 | 2,76 | 2,04 | 0,00 |

TAB. 4.1: Influence de variations de dièdres sur le RMSD entre des structures de G₈

d'un seul angle dièdre peut avoir des conséquences très différentes sur l'indice de dissimilarité mesuré.

Comme la modification d'un dièdre peut être responsable d'un changement important en terme de structure et énergie, cet exemple montre qu'une comparaison des dièdres ne peut pas être directement corrélée à une comparaison des distances atomiques.

Traitement de l'équivalence par symétrie locale

La plupart des programmes de comparaison de structures par RMSD nécessite que les atomes soient écrits dans le même ordre dans leur fichier respectif. Les positions atomiques sont comparées « ligne par ligne ». Cela a pour conséquence une possible surévaluation de la différence entre les structures en cas de rotation conservant la symétrie. Par exemple, une rotation de NH₃ ou de CH₃ de $\frac{2\pi}{3}$ rad ne modifie pas la géométrie d'une structure mais

| Rotation | NH ₃ ⁺ | COO ⁻ | COO ⁻ & NH ₃ ⁺ |
|---------------------------------------|------------------------------|------------------|---|
| RMSD tous atomes | 0,38 | 0,37 | 0,53 |
| RMSD pondéré par la masse tous atomes | 0,13 | 0,54 | 0,55 |
| RMSD sans H | 0,00 | 0,51 | 0,51 |
| RMSD pondéré par la masse sans H | 0,00 | 0,55 | 0,55 |

TAB. 4.2: Influence d'une rotation conservant la symétrie sur le RMSD en Å de la structure hélicoïdale de G₈ dans sa forme zwitterionique

fausse la valeur du RMSD de cette structure vis-à-vis d'une autre, si le calcul n'en tient pas compte. Cette considération ne s'applique pas qu'aux atomes d'hydrogène. Il en va de même d'autres groupements chimiques : COO⁻, C₆H₅, CH(CH₃)₂, C(NH₂)₂... avec des angles de π rad dans les cas cités. Dans le cas de la structure hélicoïdale du système G₈ dans sa forme zwitterionique (Tab. 4.2), la rotation d'un groupement NH₃ ne modifie évidemment la valeur du RMSD que s'il tient compte des hydrogène. Dans ce cas, la valeur peut atteindre 0,38 Å pour une simple rotation conservant la symétrie. En pondérant le RMSD par la masse des atomes, cette valeur diminue mais demeure supérieure à 0,1 Å. Les rotations de COO⁻ conservant la symétrie provoquent des surévaluations du RMSD de l'ordre de 0,5 Å si on ne considère pas les hydrogène ou si on pondère le RMSD par la masse des atomes. Si on considère les atomes d'hydrogène dans le calcul du RMSD unitaire, c'est-à-dire sans pondération massique, la valeur du RMSD est ramenée à 0,37 Å par l'augmentation de la proportion d'atomes inchangés dont le nombre passe de 31 à 57.

Pour ces raisons, nous avons décidé de prendre en compte toutes les rotations possibles conservant la symétrie locale. Un RMSD est alors calculé pour chacune de ces rotations. Le RMSD réel est la valeur du RMSD la plus faible obtenue *via* les rotations.

Le réordonnement des atomes est une tâche longue. On modifie l'ordre des atomes mais aussi leur connectivité, qui doit être connue pour identifier les rotations conservant la symétrie.

Une solution serait de tester tous les ordres atomiques possibles mais ce serait trop lourd. Un test sur le système G₈-Na⁺ a montré que l'essai de $19,6 \times 10^6$ ordres occupe 73 Go de disque dur et environ 3 heures CPU alors qu'il existe $60! \approx 8,3 \times 10^{81}$ ordres possibles pour un système de 60 atomes !

Pour que la recherche automatique de l'ordre se fasse sans intervention humaine, il faudrait un script plus élaboré qui tienne compte des noms des atomes (ou des éléments) et de

leurs connectivités. Certains algorithmes permettent la reconnaissance des symétries dans les molécules¹⁴ et les complexes¹⁵.

Dans notre procédure, nous avons choisi de laisser à l'utilisateur le choix des permutations à considérer dans le calcul du RMSD.

Traitement d'une symétrie non-locale : le cas des polyglycines sodées

La glycine a pour particularité d'avoir un hydrogène pour chaîne latérale. Ainsi, son C_α n'est pas asymétrique contrairement à celui des autres acides aminés. En conséquence, la glycine est achirale et il en va de même en ce qui concerne les polyglycines. Si deux structures de polyglycine sodée sont l'image l'une de l'autre dans un miroir (Fig. 4.2), elles sont pratiquement toujours perçues comme étant différentes alors qu'elles présentent exactement les mêmes propriétés géométriques et *a fortiori* les mêmes interactions. Il n'y a en effet, *a priori*, ni de raison qu'une structure de peptide possède un plan de symétrie, ni qu'elle le conserve ou en acquiert un en la présence d'un cation, cas d'intérêt dans ce travail. Si elles sont exactement l'image l'une de l'autre dans un miroir alors elles sont exactement dégénérées et sont donc considérées comme redondantes par l'étape de discernement sur la base des énergies. Si par contre, les deux structures ne sont pas rigoureusement l'image exacte l'une de l'autre, il n'y a pas de raison particulière pour qu'elles soient dégénérées à la précision du critère fixé. Dans ce cas, les deux structures sont considérées comme étant différentes bien qu'elles puissent être quasiment superposables à l'image l'une de l'autre. Comme on suppose que deux structures images l'une de l'autre dans un miroir ne seront pas responsables de spectres IR différents, nous considérerons ces structures comme redondantes. Nous appliquons alors les mêmes fonctions de dissimilarité avec les mêmes critères que dans le cas général. La comparaison entre une structure *a* et une structure *b* est effectuée de manière directe entre *a* et *b*, puis entre la structure *b* et l'image dans un miroir de *a*. Après avoir testé les deux possibilités, la grandeur de comparaison à considérer est la plus proche de zéro.

La démarche appliquée pour filtrer les structures considérées redondantes est présentée dans la figure 4.3. Pour chaque structure, toutes les rotations conservant la symétrie sont prises en compte pour engendrer des structures équivalentes. Dans les cas plus fréquents

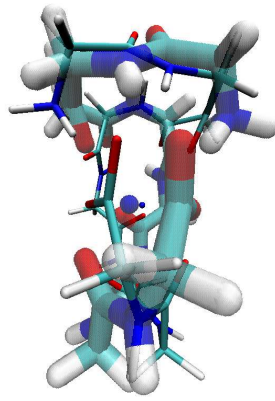


FIG. 4.2: Polyglycines, images dans un miroir

ne présentant pas d'asymétrie à l'instar des polyglycines, sodées ou non, les structures équivalentes par réflexion dans un miroir sont engendrées pour être comparées aux autres. Les structures sont testées une à une dans l'ordre énergétique croissant (de la plus stable à la moins stable) et à chaque fois, un RMSD temporaire est calculé en considérant les structures équivalentes (dont l'ordre des atomes varie mais qui sont superposables). Le RMSD authentique est le RMSD temporaire le plus petit puisque les autres sont artificiellement surévalués. Si le RMSD entre deux structures comparées est suffisamment faible, en dessous d'un seuil arbitraire, nous considérons la seconde structure (plus haute en énergie) comme étant redondante avec la structure de référence à laquelle elle est comparée. Dans ce cas, cette structure est sortie de la liste et n'est plus comparée aux autres. Idéalement, deux structures vraiment identiques ont un RMSD de 0 Å. Nous définissons alors cette valeur comme valeur cible, la valeur à atteindre, ainsi qu'une valeur seuil jusqu'à laquelle nous considérons comme identiques les deux structures. Le choix de la valeur seuil dépend du système ; nous avons souvent utilisé 0,1 Å.

Nous obtenons ainsi une liste des structures conservées selon les critères de similarité que nous imposons sur le RMSD ainsi que l'information énergétique qui leur est associée.

Influence de la taille du système sur le RMSD

L'indice de dissimilarité donné par le RMSD pour une modification structurale locale dépend non seulement de l'importance de celle-ci mais aussi de la taille du système. Le

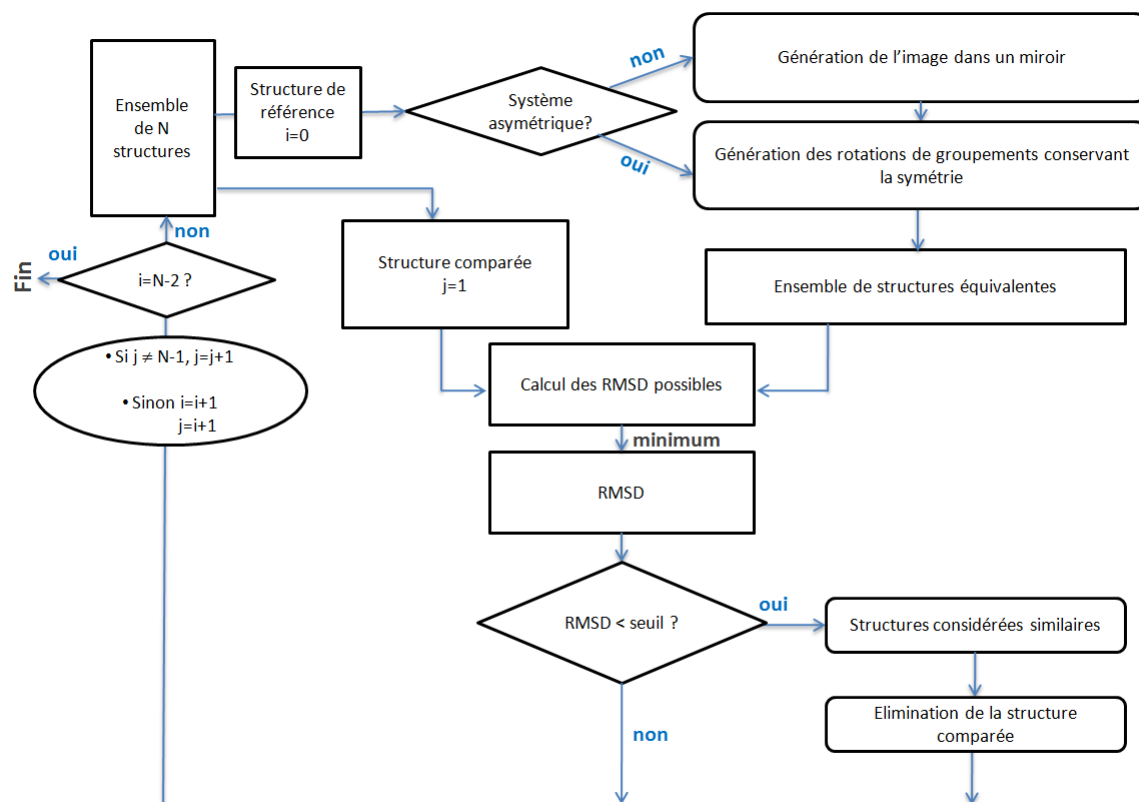


FIG. 4.3: Filtrage des structures redondantes

tableau 4.3 illustre cette dépendance en considérant des modifications structurales localisées aux terminaisons NH_3^+ et COO^- de polyglycines composées de 1 à 8 résidus. À titre de comparaison avec les exemples précédents, cette série reprend des rotations conservant la symétrie : $\frac{2\pi}{3}$ rad pour NH_3^+ et π rad pour COO^- . Pour qu'elles soient non-nulles, les valeurs de RMSD présentées sont calculées directement, sans prendre en compte la symétrie locale, mais ces conclusions sont valables aussi pour le RMSD tenant compte de la symétrie locale, appliqué à des rotations ne conservant pas la symétrie. Les RMSD présentés ne sont pas pondérés par les masses atomiques et considèrent tous les atomes. Le poids statistique d'une modification locale décroît quand le nombre d'atomes du système augmente. Cela s'explique par le facteur $\frac{1}{\sqrt{N}}$ de la définition du RMSD. Ce terme est justifié et la dépendance du RMSD vis-à-vis de la taille du système doit être considérée lors du choix de la valeur seuil pour le critère d'identité.

| système | nombre d'atomes | NH ₃ ⁺ | COO ⁻ | COO ⁻ & NH ₃ ⁺ |
|----------------|-----------------|------------------------------|------------------|---|
| G ₁ | 10 | 0,55 | 0,86 | 0,94 |
| G ₂ | 17 | 0,46 | 0,68 | 0,82 |
| G ₃ | 24 | 0,41 | 0,58 | 0,70 |
| G ₄ | 31 | 0,36 | 0,51 | 0,62 |
| G ₅ | 38 | 0,33 | 0,46 | 0,56 |
| G ₆ | 45 | 0,30 | 0,42 | 0,52 |
| G ₇ | 52 | 0,28 | 0,39 | 0,48 |
| G ₈ | 59 | 0,26 | 0,37 | 0,45 |

TAB. 4.3: RMSD en Å sur tous les atomes pour une rotation locale en fonction de la taille de la polyglycine G_n

4.2 Tri et groupement de structures

Après discrimination des structures considérées comme étant redondantes, le nombre de structures présentes demeure souvent conséquent. Pour faciliter l'interprétation chimique, nous opérons un groupement des structures en familles. Tout comme le discernement de structure, leur rassemblement en familles est opéré sur des critères structuraux. Ces familles peuvent être définies par des indices globaux de dissimilarité (basés sur RMSD, UME... des distances atomiques par exemple) ou par la proximité de certaines propriétés géométriques définies comme des distances caractéristiques ou des dièdres. Dans le premier cas, l'identité et la représentation d'une famille sont plus abstraites. Dans le second, la représentation d'une famille est clairement définie mais limitée par le nombre de contraintes géométriques choisies.

Pour ces deux possibilités, les limites des familles sont fixées par des seuils comme pour l'étape de discernement. S'il s'agit de RMSD ou d'UME, la valeur cible est 0 Å. Le seuil familial est alors plus élevé que le seuil de dissimilarité utilisé lors de l'étape de discernement ; la reconnaissance d'un individu est plus restrictive que l'appartenance à une famille. Nous avons toujours choisi pour seuil de RMSD délimitant l'appartenance à une famille une valeur supérieure à 0,1 Å mais ce choix est fortement dépendant du système. Si par contre, l'appartenance à une famille est basée sur des données structurales locales, la définition des familles est alors limitée par les bornes d'un intervalle cible pour chaque caractéristique structurale considérée.

Les résultats du tri et la répartition en familles des structures issues de la deuxième génération

| total | uniques | [4,9;5,4] | [5,4;5,9] | [5,9;6,4] | [6,4;6,9] |
|-------|---------|-----------|-----------|-----------|-----------|
| 2 322 | 1 603 | 337 | 878 | 707 | 98 |

TAB. 4.4: Nombre de structures issues du REMD de A₁₂-Na⁺ après tri et exemple de regroupement en familles selon le rayon de giration en Å

du REMD de A₁₂-Na⁺ sont illustrés dans le tableau 4.4. 12 000 structures ont été sauvegardées durant la simulation REMD. Après un premier filtrage sur l'énergie à 10⁻⁴ kcal.mol⁻¹ près, il en reste 2 322. Sur ces 2 322 structures, 1 603 sont considérées comme étant uniques, car leur RMSD avec les autres dépasse 0,1 Å. Le tableau présente la répartition de ces 1 603 en quatre familles définies par le rayon de giration* des structures. La structure la plus stable, en hélice, est caractérisée par un rayon de giration de 6,3 Å. La famille dont le rayon de giration est supérieur à 6,4 Å regroupe des structures partiellement hélicoïdales mais avec des extrémités dépliées. Les deux autres familles regroupent des structures ayant un caractère plus globulaire, avec un sodium ayant un degré de coordination au peptide plus élevé.

Le regroupement en familles de structures ou *clustering* peut être effectué par différents algorithmes disponibles dans des suites logicielles comme CHARMM¹⁶, AMBER¹⁷, XCluster¹⁸, MaxCluster¹⁹, ou DALI¹².

Nous pouvons regrouper des structures en familles sur des critères de ressemblance globale mais si nous nous intéressons à des systèmes dont la structure expérimentale n'est pas connue (par RMN, diffraction des rayons X ou des neutrons...), ce regroupement n'est pas suffisant. Pour assigner une structure à une observation expérimentale, il convient de filtrer les structures sur des caractéristiques plus précises. Nous préférons utiliser nos scripts adaptés au format de structures de Tinker pour regrouper les structures sur la base de propriétés physiques (rayon de giration, moment dipolaire) ou géométriques comme des angles dièdres ou des distances d'interaction interatomiques (liaisons hydrogène, coordination au cation). Lorsque cela s'avère nécessaire, nous testons les permutations d'atomes ou réflexion des structures dans un miroir comme cela a été fait lors de l'étape de discernement. Par la suite, en fonction des systèmes et des informations que nous voudrions mettre en évidence, nous préciserons nos choix concernant le tri et le groupement des structures en familles.

*distance moyenne des N atomes de coordonnées i au centroïde d'un système,

$$R_g = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - r_{\text{centroïde}})^2}$$

Bibliographie

- [1] M. E. KARPEN, D. J. TOBIAS et C. L. BROOKS : Statistical clustering techniques for the analysis of long molecular dynamics trajectories : analysis of 2.2-ns trajectories of YPGDV. *Biochemistry*, 32:412–420, 1993.
- [2] D. MAO, E. WACHTER et B. A. WALLACE : Folding of the mitochondrial proton adenosine-triphosphatase proteolipid channel in phospholipid vesicles. *Biochemistry*, 21:4960–4968, 1982.
- [3] M. R. HANLON, R. R. BEGUM, R. J. NEWBOLD, D. WHITFORD et B. A. WALLACE : In vitro membrane-inserted conformation of the cytochrome b(5) tail. *Biochem. J.*, 352:117–124, 2000.
- [4] M. R. BETANCOURT et J. SKOLNICK : Universal similarity measure for comparing protein structures. *Biopolymers*, 59:305–309, 2001.
- [5] K. L. DAMM et H. A. CARLSON : Gaussian-weighted RMSD superposition of proteins : a structural comparison for flexible proteins and predicted protein structures. *Biophys. J.*, 90:4558–4573, 2006.
- [6] D. R. FERRO et J. HERMANS : A different best rigid-body molecular fit routine. *Acta Crystallogr., Sect. A : Found. Crystallogr.*, 33:345–347, 1977.
- [7] W. KABSCH : A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. A : Found. Crystallogr.*, 32:922–923, 1976.
- [8] W. KABSCH : A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. A : Found. Crystallogr.*, 34:827–828, 1978.
- [9] W. HUMPHREY, A. DALKE et K. SCHULTEN : VMD – Visual Molecular Dynamics. *J. Mol. Graphics*, 14:33–38, 1996.

- [10] R. DIAMOND : A note on the rotational superposition problem. *Acta Crystallogr., Sect. A : Found. Crystallogr.*, 44:211–216, 1988.
- [11] S. K. KEARSLEY : An algorithm for the simultaneous superposition of a structural series. *J. Comput. Chem.*, 11:1187–1192, 1990.
- [12] L. HOLM : Protein Structure Comparison by Alignment of Distance Matrices. *J. Mol. Biol.*, 233:123–138, 1993.
- [13] D. BARTHEL, J. HIRST, J. BLAZEWICZ, E. BURKE et N. KRASNOGOR : ProCKSI : a decision support system for Protein (Structure) Comparison, Knowledge, Similarity and Information. *BMC Bioinformatics*, 8(1):416, 2007.
- [14] J. M. IVANOV et G. SCHÜÜRMAN : Simple Algorithms for Determining the Molecular Symmetry. *J. Chem. Inf. and Comput. Sci.*, 39:728–737, 1999.
- [15] W. CHEN, J. HUANG et M. K. GILSON : Identification of Symmetries in Molecules and Complexes. *J. Chem. Inf. Comput. Sci.*, 44:1301–1313, 2004.
- [16] B. R. BROOKS, III BROOKS, C. L., A. D. MACKERELL, Jr., L. NILSSON, R. J. PETRELLA, B. ROUX, Y. WON, G. ARCHONTIS, C. BARTELS, S. BORESCH, A. CAFLISCH, L. CAVES, Q. CUI, A. R. DINNER, M. FEIG, S. FISCHER, J. GAO, M. HODOSCEK, W. IM, K. KUCZERA, T. LAZARIDIS, J. MA, V. OVCHINNIKOV, E. PACI, R. W. PASTOR, C. B. POST, J. Z. PU, M. SCHAEFER, B. TIDOR, R. M. VENABLE, H. L. WOODCOCK, X. WU, W. YANG, D. M. YORK et M. KARPLUS : CHARMM : The Biomolecular Simulation Program. *J. Comput. Chem.*, 30:1545–1614, 2009.
- [17] D. A. CASE, T. A. DARDEN, T. E. CHEATHAM, C. L. SIMMERLING, J. WANG, R. E. DUKE, R. LUO, M. CROWLEY, R. C. WALKER, W. ZHANG, Jr MERZ, K. M., B. WANG, S. HAYIK, A. ROITBERG, G. SEABRA, I. KOLOSSVARY, K. F. WONG, F. PAESANI, J. VANICEK, X. WU, S. R. BROZELL, T. STEINBRECHER, H. GOHLKE, L. YANG, C. TAN, J. MONGAN, V. HORNAK, G. CUI, D. H. MATHEWS, M. G. SEETIN, C. SAGUI, V. BABIN et P. A. KOLLMAN : *AMBER 10*. University of California, San Francisco, 2008.
- [18] P. S. SHENKIN et Q. D. MCDONALD : Cluster analysis of molecular conformations. *J. Comput. Chem.*, 15:899–916, 1994.
- [19] A. HERBERT : MaxCluster. <http://www.sbg.bio.ic.ac.uk/~maxcluster>, visité le 19 avril 2010. Structural Bioinformatics Group, Imperial College, London.

Chapitre 5

Simulation de spectre infrarouge

Un oscillateur quantique est caractérisé par un spectre d'énergies et de fonctions d'onde. Dans l'approximation harmonique, la constante de force k est proportionnelle à la dérivée seconde de l'énergie E par rapport à la coordonnée d'oscillation x :

$$E = \frac{1}{2}kx^2 \text{ avec } k = \frac{\partial^2 E}{\partial x^2} \quad (5.1)$$

Pour une molécule de N atomes, comprenant $3N - 6$ oscillateurs, la matrice des dérivées secondes de l'énergie, appelée Hessien, peut être diagonalisée pour obtenir les constantes de force de $3N - 6$ modes « normaux ». La fréquence ν_i en est déduite facilement pour tout mode normal i en fonction de la constante de force k_i et de la masse réduite m_i correspondantes, selon l'expression suivante :

$$\nu_i = \frac{1}{2\pi} \sqrt{\frac{k_i}{m_i}} \quad (5.2)$$

La plupart des spectres calculés, décrits dans la littérature, sont obtenus dans cette approximation « statique » harmonique. Il est possible d'ajouter des termes correctifs décrivant l'anharmonicité et les couplages des modes normaux de vibration (p. 94). Lorsque des méthodes quantiques fiables sont utilisées pour le calcul de l'énergie et de ses dérivées, ces corrections sont difficiles à mettre en œuvre et limitées jusqu'ici à des molécules d'une dizaine d'atomes au maximum. L'utilisation d'un champ de force classique peut permettre de contourner cette limite. La facilité du calcul de l'énergie et du moment dipolaire per-

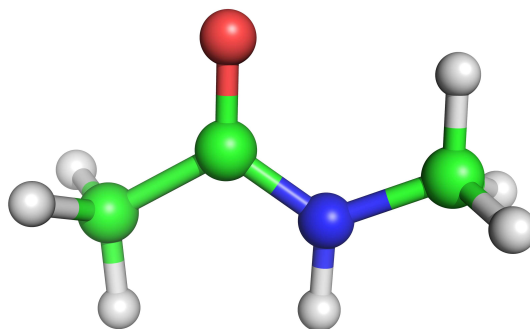


FIG. 5.1: N-méthylacétamide (NMA)

met d'abandonner l'approximation harmonique, que ce soit pour le calcul des fréquences ou pour celui des intensités. Cette approche est l'objet principal de ce chapitre. Nous utiliserons le N-méthylacétamide (NMA, Fig 5.1) comme modèle de la liaison peptidique. Sa liaison amide est commencée et terminée par un groupement méthyle en lieu et place des carbone α des peptides. Ce modèle de liaison peptidique permet d'entrevoir nos capacités à reproduire les bandes d'absorption correspondant aux modes « amide ». Pour le spectre d'absorption du NMA, nous nous intéressons aux modes amides :

- A qui correspond à l'élongation d'une liaison N-H ;
- I qui est majoritairement constitué de l'élongation d'une ou de plusieurs liaison C=O ;
- II, composée principalement de la déformation angulaire de la liaison N-H dans le plan du squelette peptidique (N-C-C α) et des élongations C α -N et dans une moindre mesure N-C.

5.1 Approximation harmonique : exemple du N-méthylacétamide

De nombreux travaux sur le N-méthylacétamide (NMA), expérimentaux et théoriques, sont décrits dans la littérature. Il existe en particulier une étude récente¹ en jets moléculaires fournissant des éléments de spectre IR de la molécule isolée. Dans la logique AMOEBA, nous utiliserons également des calculs de spectres statiques MP2 réalisés pour l'occasion, à comparer aux valeurs expérimentales ou pour servir de point de comparaison lorsque les valeurs expérimentales ne sont pas disponibles.

| mode | <i>exp</i> [1] | MP2/aug-cc-pVTZ | MP2/SVP | AMOEBA0 | AMOEBA mod. |
|------|----------------|-----------------|---------|---------|-------------|
| A | 3 510 | 3 535 | 3 554 | 3 488 | 3 533 |
| I | 1 722 | 1 723 | 1 740 | 1 703 | 1 719 |

TAB. 5.1: Fréquences infrarouge statiques et expérimentales du N-méthylacétamide [1] jet moléculaire réalisé par Albrecht¹

Pour le calcul de spectres statiques dans le formalisme d'AMOEBA, nous avons noté une réponse inadéquate des fréquences de vibration à l'environnement (liaison hydrogène...). Nous avons alors modifié les paramètres électrostatiques du champ de force en utilisant les multipôles obtenus par DMA (p. 39) à partir d'une molécule isolée de NMA. La comparaison avec des spectres expérimentaux et *ab initio* harmoniques (Tab. 5.1) montre une nécessité de modification de certains paramètres. Ainsi, une modification de la constante de force de la liaison N-H de 487 à 500 kcal.mol⁻¹.Å⁻² permet de ramener la fréquence associée au mode d'élongation N-H de 3 488 cm⁻¹ à une valeur beaucoup plus proche des références *ab initio* : 3 533 cm⁻¹ en AMOEBA avec le nouveau jeu de paramètres contre 3 535 cm⁻¹ en MP2/aug-cc-pVTZ avec un facteur d'échelle de 0,945 et 3 554 cm⁻¹ en MP2/SVP avec un facteur d'échelle de 0,952. Hobza a montré que dans certains cas, augmenter la taille de la base utilisée pour un calcul MP2 n'apporte pas d'amélioration voire détériore les caractéristiques énergétiques et structurelles obtenues avec une base plus petite² car l'augmentation de la base peut aller contre une compensation d'erreur. MP2 avec une base moins étendue, SVP conduit à des fréquences plus cohérentes que MP2/aug-cc-pVTZ avec les spectres expérimentaux du dimère de NMA¹ pour les modes de vibration amide A, I et II dont nous discuterons dans la section 5.4 (p. 108). Notons qu'il est courant de trouver dans la littérature l'utilisation d'un facteur d'échelle pour chaque bande pour plus de précision sur les fréquences. Le seul moyen de tester la fiabilité des méthodes est donc de disposer de données expérimentales telles que les spectres IR en phase gazeuse. Lorsque les données expérimentales ne sont pas disponibles, il faudrait pouvoir avoir recours aux méthodes CC2 ou mieux, CCSD(T) ce qui est faisable pour l'énergétique mais très limitant pour les optimisations de géométrie et les calculs de spectre. Lorsque les valeurs expérimentales ne sont pas connues, les fréquences MP2 seront utilisées à titre de comparaison des tendances mais ne constituent pas des données de référence de grande précision en comparaison à ces spectres expérimentaux qui sont bien résolus par rapport à l'IRMPD pour des peptides tels que G₈.

Comme la masse de l'hydrogène est très faible, on attend un effet important de l'anharmonicité sur la fréquence de l'élongation N-H³. La surestimation par rapport à la valeur expérimentale (3510 cm⁻¹) de 23 cm⁻¹ avec AMOEBA mod., et plus importante dans le cas des valeurs MP2 est donc au moins partiellement due à l'absence de traitement de l'anharmonicité dans les calculs de spectres statiques effectués à 0 K. En effet, la validité de ces spectres obtenus directement par diagonalisation du Hessien suppose que la température du système étudié est de 0 K afin que l'approximation consistant à ne considérer que la structure d'énergie minimale et que l'approximation harmonique soient adéquates. Les effets d'anharmonicité deviennent importants au fur-et-à mesure que la température du système augmente ; la structure s'éloigne de plus en plus de son équilibre. La température vibrationnelle du NMA dans le travail expérimental cité¹ est estimée varier au moins entre 50 et 300 K.

5.2 Traitement statique de l'anharmonicité

Plusieurs effets tels que l'anharmonicité, les bandes de combinaison ou l'interaction avec l'environnement chimique par exemple, sont à prendre en compte lors de l'interprétation d'un spectre expérimental, c'est-à-dire l'attribution des bandes à des modes de vibration. Les calculs quantiques de fréquences harmoniques sont d'une grande aide dans cette interprétation et la différence entre les fréquences calculées et expérimentales est en général réduite par l'utilisation d'un facteur d'échelle approprié⁴.

Par ailleurs, plusieurs modèles ont été développés pour prendre en compte explicitement l'anharmonicité dans le calcul des fréquences. Les plus connus sont basés sur la résolution de l'équation de Schrödinger vibrationnelle par convergence d'un champ auto-cohérent vibrationnel (VSCF) en développant l'énergie potentielle en série de Taylor en coordonnées normales autour de la géométrie d'équilibre du système⁵⁻⁹. Barone propose une correction de l'anharmonicité en employant des calculs de perturbation du second ordre¹⁰ nommés VPT2 par similarité avec la méthode MP2. L'interaction de configuration vibrationnelle, ou VCI¹¹⁻¹³ constitue une méthode variationnelle pouvant être étendue jusqu'à devenir exacte mais est bien plus lourde, le temps de calcul augmentant drastiquement avec le nombre des atomes.

Le couplage de telles méthodes avec des calculs fiables de structure électronique permet aujourd'hui de calculer des fréquences de vibration en très bon accord avec les données expérimentales, de l'ordre de 5 à 15 cm^{-1} . Cependant, ces méthodes sont généralement applicables pour des systèmes chimiques simples contenant moins d'une dizaine d'atomes. À titre d'exemple, une étude récente vient d'être proposée par Pouchan et coll. sur la glycine radicalaire, système flexible d'intérêt biologique comprenant 10 atomes.³

5.3 Calcul de spectre par transformée de Fourier de la fonction d'autocorrélation du dipôle

Les spectres calculés à partir du Hessien dans une approche « statique » ne permettent pas la prise en compte directe des effets de température sur la position et l'élargissement des bandes ou encore la combinaison ou les harmoniques des modes normaux. La dynamique moléculaire permet de dépasser ces limites. Nous décrivons dans les paragraphes suivants les bases théoriques sur lesquelles s'appuient de telles méthodes, l'implémentation que nous avons choisie et sa première application au calcul du spectre infrarouge du N-méthylacétamide.

5.3.1 Bases théoriques

Théorème de Wiener-Khinchin

Le théorème de Wiener-Khinchin¹⁴ 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 5.3.

$$R_f(t) = \frac{\langle f(0)f(t) \rangle}{\langle f(0)f(0) \rangle} \quad (5.3)$$

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 τ de la manière suivante :

$$R_f(\tau) = \int_{-\infty}^{\infty} f(t)f(t+\tau)dt \quad (5.4)$$

Pour un signal de durée finie, celle-ci est divisée en portions de longueur T . T 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 :

$$R_f(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(t)f(t+\tau)dt \quad (5.5)$$

Dans la mesure où le processus étudié n'est pas strictement stationnaire, l'égalité 5.5 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

$$S_f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_f(\tau)e^{-j2\pi\omega\tau}d\tau \quad (5.6)$$

comme la transformée de Fourier de $R_f(\tau)$, la fonction d'autocorrélation du signal aléatoire $f(t)$ (équ. 5.6).

Applicabilité à la spectroscopie infrarouge

La spectroscopie infrarouge est basée sur l'interaction entre la matière et un rayonnement électromagnétique allant d'environ 10 cm^{-1} pour le lointain infrarouge à environ $13\,000 \text{ cm}^{-1}$ pour le proche infrarouge. Le champ électrique du rayonnement interagit avec le dipôle de la molécule ou, de manière plus générale, du système considéré. Si la fréquence d'oscillation du champ électrique porté par le rayonnement correspond à une fréquence vibrationnelle (ou rotationnelle dans le cas du lointain infrarouge), le système absorbe l'énergie du rayonnement, augmentant ainsi l'amplitude des variations du moment dipolaire et des vibrations de la molécule.

Le moment dipolaire, tout comme le jeu de vitesses atomiques instantanées, est un proces-

sus dont le caractère stochastique est dû au mouvement thermique du système. Si la durée d'une simulation de dynamique moléculaire est suffisante, *i.e.* si l'équilibre thermodynamique est atteint, ce processus stochastique est stationnaire au sens large car sa moyenne sur une durée fixée est indépendante du point de départ.

Par application du théorème de Wiener-Khinchin, on peut calculer le spectre d'absorption infrarouge par transformée de Fourier de la fonction d'autocorrélation (ACF) :

- des vitesses atomiques (VACF) ;
- du moment dipolaire (DACF).

En 1977, Noid et coll. ont utilisé le DACF pour calculer un spectre à partir d'une trajectoire de dynamique moléculaire¹⁵. La dynamique moléculaire classique a été utilisée pour la simulation de spectres rotationnels purs ou rovibrationnels de CO¹⁶ en phase gazeuse. Les avantages du DACF en terme d'anharmonicité et de largeur de bande sur l'analyse des coordonnées normales illustrés par divers travaux^{3,17} expliquent l'augmentation de son utilisation ces dernières années.

La fonction d'autocorrélation des vitesses (VACF) permet la simulation de spectre infrarouge à partir de données facilement extractibles d'une trajectoire de dynamique moléculaire propagée avec les logiciels les plus répandus. De plus, les vitesses sont déjà calculées à chaque pas de la trajectoire quand l'intégrateur de Beeman est utilisé. Pour ces raisons, la méthode VACF est relativement facile à mettre en œuvre. Elle ne nécessite que le stockage de données déjà calculées au cours d'une trajectoire et d'effectuer la transformée de Fourier de leur fonction d'autocorrélation. Les spectres simulés par VACF incluent les effets thermiques d'élargissement de bandes et d'anharmonicité¹⁸ et l'éventuel besoin de traiter les transitions structurelles rapides par rapport à la longueur des trajectoires desquelles les spectres sont extraits. Cependant, les fluctuations des vitesses atomiques ne représentent pas à elles seules la totalité de l'évolution du moment dipolaire. Elles n'ont pas toutes la même influence sur ce dernier et c'est pour quoi les intensités relatives des différents pics d'absorption obtenus en VACF n'ont pas de signification physique directe.

Si au contraire, 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.

Le moment dipolaire $\vec{\mu}$ est défini classiquement pour un système à N particules comme la somme des produits de la charge q_i et du vecteur position \vec{r}_i de chaque particule i .

$$\vec{\mu} = \sum_{i=1}^N q_i \vec{r}_i \quad (5.7)$$

En mécanique quantique, pour un système neutre, le moment dipolaire $\vec{\mu}$ est exprimé par la somme de termes dus aux noyaux et de termes dus aux électrons :

$$\vec{\mu} = \sum_{i=1}^N Z_i \vec{R}_i - \sum_{j=1}^{N_e} \langle \Psi | \vec{r}_j | \Psi \rangle \quad (5.8)$$

Pour un système de charge totale Q non nulle, la définition quantique dépend du choix des coordonnées de son origine. Si l'origine est déplacée d'un vecteur \vec{r}' , on a :

$$\vec{\mu} = \sum_{i=1}^N Z_i \vec{R}_i - \sum_{j=1}^{N_e} \langle \Psi | \vec{r}_j | \Psi \rangle - Q \vec{r}' \quad (5.9)$$

Les trajectoires de dynamique moléculaire sont discrétisées en temps. Par conséquent il faut aussi discrétiser l'équation de la fonction d'autocorrélation (équ. 5.5) appliquée au moment dipolaire $\vec{\mu}$ pour un intervalle de temps m exprimé en nombre de pas de temps devient $C(m)$ tel que :

$$C(m) = \frac{1}{n} \sum_{i=1}^n \vec{\mu}(m+i) \times \vec{\mu}(i) \quad \text{avec} \quad n \leq m \quad (5.10)$$

avec n la longueur (en nombre de pas) des portions utilisées et $\vec{\mu}(i)$, le vecteur dipôle au pas de temps i . La densité spectrale est obtenue par transformée de Fourier discrète de cette fonction d'autocorrélation discrétisée $C(m)$.

5.3.2 Mise en œuvre et exploitation

Les spectres infrarouge peuvent être, nous l'avons vu, simulés à partir des vitesses ou des moments dipolaires instantanés calculés au cours d'une trajectoire de dynamique moléculaire.

Tinker permet de stocker les vitesses instantanées d'une trajectoire. Le post-traitement de ces données et l'application du théorème de Wiener-Khinchin permettent la simulation de

spectres infrarouge par transformée de Fourier de la fonction d'autocorrélation (ACF).

Tinker a été modifié afin de calculer le moment dipolaire instantané du système et de le stocker dans une archive spécifique à la trajectoire propagée. L'augmentation de temps CPU due au calcul du moment dipolaire et à son écriture sur le disque dépend de la taille du système étudié. Pour des simulations de 100 000 pas, l'activation de l'option de sauvegarde du dipôle à chaque pas correspond à un accroissement du temps d'exécution (*user time* + *system time*) de l'ordre de :

- 21% dans le cas de H₂O ;
- 35% pour le N-méthylacétamide (12 atomes) ;
- 59% pour la dodéca-alanine sodée, A₁₂⁺ composée de 124 atomes.

On constate que l'augmentation du temps CPU dû à l'extraction du dipôle s'accroît avec le nombre d'atomes mais aussi que cet accroissement ralentit avec celui-ci. Les temps de simulation demeurent relativement courts : 1 h39 min pour une trajectoire de A₁₂-Na⁺ de 100 000 pas effectuée par un bi-processeur AMD Opteron™ Dual Core 2220 (mais exécution sur un seul cœur) à 2,8 GHz. La taille de l'archive de moments dipolaires créée est indépendante du nombre d'atomes du système et représente moins de 4 Mo pour 100 000 pas. Cette archive est utilisée en entrée d'une procédure de calcul de spectre par DACF.

Le programme Fourier de A. Kohlmeyer¹⁹ 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.

La limite classique du mouvement des noyaux introduit une approximation. La formule²⁰ pour le coefficient d'absorption infrarouge par unité de longueur dans l'approximation du dipôle (E1)* est donnée en fonction de la densité spectrale I par :

$$\alpha(\omega) = \frac{4\pi^2\omega}{3V\hbar cn(\omega)}(1 - e^{-\beta\hbar\omega})I(\omega) \quad (5.11)$$

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

*approximation selon laquelle on néglige la réaction de multipôles magnétiques et on ne considère pas la réaction des multipôles d'ordre supérieur à celui du dipôle électrique (1).

Selon une approximation quasi-classique, on peut introduire des facteurs de correction quantique²¹. Le coefficient d'absorption corrigé, $\alpha_{QC}(\omega)$ est alors exprimable en fonction du facteur de correction quantique, $Q_{QC}(\omega)$ et de la densité spectrale dans l'approximation classique, $I_{CL}(\omega)$.

$$\alpha_{QC}(\omega) = \frac{4\pi^2\omega}{3V\hbar cn(\omega)}(1 - e^{-\beta\hbar\omega})Q_{QC}(\omega)I_{CL}(\omega) \quad (5.12)$$

Ces corrections quantiques permettent de satisfaire au principe d'« équilibre détaillé » (équ. 5.13)

$$I(\omega) = e^{\beta\hbar\omega}I(-\omega) \quad (5.13)$$

Dans la limite classique, la condition d'équilibre détaillé se ramène à : $\lim_{T \rightarrow \infty} e^{-\beta\hbar\omega} = 1$ et ainsi à la parité de la fonction I .

Fourier laisse en outre à l'utilisateur le choix du préfacteur de correction des intensités. À chaque fréquence du fichier de sortie sont associées 5 valeurs d'intensités.

1. La première correspond directement à la densité spectrale, I_{CL} dans l'approximation classique.
2. La deuxième donne le coefficient d'absorption dans la limite classique, dite de Gordon, c.-à-d. sans correction quantique.
3. Le coefficient d'absorption avec correction de Kubo ou approximation standard, α_{SA} , peut s'exprimer comme :

$$\alpha_{SA} = \omega \tanh(\beta\hbar\omega/2)I_{CL} \quad (5.14)$$

4. L'utilisation du facteur de correction quantique harmonique $Q_{HA} = \frac{\beta\hbar\omega}{1 - e^{-\beta\hbar\omega}}$, conduit à un coefficient d'absorption α_{HA} tel que :

$$\alpha_{HA} = \omega\beta\hbar\omega I_{CL} \quad (5.15)$$

5. Le coefficient d'absorption après correction par le facteur de Schofield²², Q_{SC} :

$$Q_{SC} = e^{\frac{\beta\hbar\omega}{2}} \quad (5.16)$$

D'autres facteurs de correction quantique existent et plus de détails sont disponibles dans le livre *Ab initio molecular dynamics : basic theory and advanced methods* de Dominik Marx et Jürg Hutte²³. Cet ouvrage et d'autres études amènent à la conclusion qu'il est préférable d'utiliser la correction harmonique pour une simulation de spectroscopie infrarouge^{21,24-27}. Après comparaison, c'est cette correction que nous avons toujours utilisée.

À une fréquence donnée, l'intensité du signal dépend de la longueur de la trajectoire de dynamique moléculaire. Koszykowski, Noyd et Marcus^{15,28} ont montré que la longueur des trajectoires simulées influence la largeur à la base du signal mais sans modifier les rapports de hauteur des pics. Le programme Fourier, en normalisant les spectres simulés, fournit des valeurs d'intensités des *maxima* des bandes théoriquement indépendantes du temps de simulation.

La procédure DACF basée sur une dynamique moléculaire effectuée par Tinker avec le champ de force AMOEBA est testée sur le N-méthylacétamide (NMA). Un des avantages de ce modèle de liaison peptidique est qu'il ne possède qu'un mode d'élongation dans les hautes fréquences (au dessus de $3\,400\text{ cm}^{-1}$). De même, il ne présente qu'une bande amide I et une amide II. Pour la simulation des spectres par DACF, nous utilisons AMOEBA avec des paramètres modifiés (p. 93) pour approcher les fréquences expérimentales et obtenues par calcul statique MP2/SVP du spectre infrarouge du NMA. La simplicité du système nous permet de tracer directement son spectre infrarouge obtenu *via* DACF, sans le lisser (p. 105). Le mode d'élongation N-H de ce modèle servira à illustrer l'influence de divers paramètres des trajectoires de dynamique moléculaire simulées dans le but de produire un spectre infrarouge.

Pour propager des trajectoires à température constante, trois thermostats sont disponibles dans Tinker. Chacun introduit un biais dans la trajectoire. Cependant, l'effet de ce biais sur les spectres issus des trajectoires est très dépendant du thermostat choisi. Nous avons testé ces trois thermostats (Fig. 5.2). Le thermostat de Berendsen²⁹ introduit un dédoublement bien résolu du pic d'élongation N-H sur une amplitude de moins de 10 cm^{-1} . Ce dédoublement est aussi observé pour les pics d'élongation symétrique et asymétrique du groupement NH_2 de l'acétamide, *i.e.* en l'absence du méthyle porté par l'azote. Par contre, le N-méthylformamide, obtenu en remplaçant le groupement méthyle du côté du carbonyle par un hydrogène, ne semble pas présenter de dédoublement du pic d'élongation N-H. À 50 K, le méthyle lié à

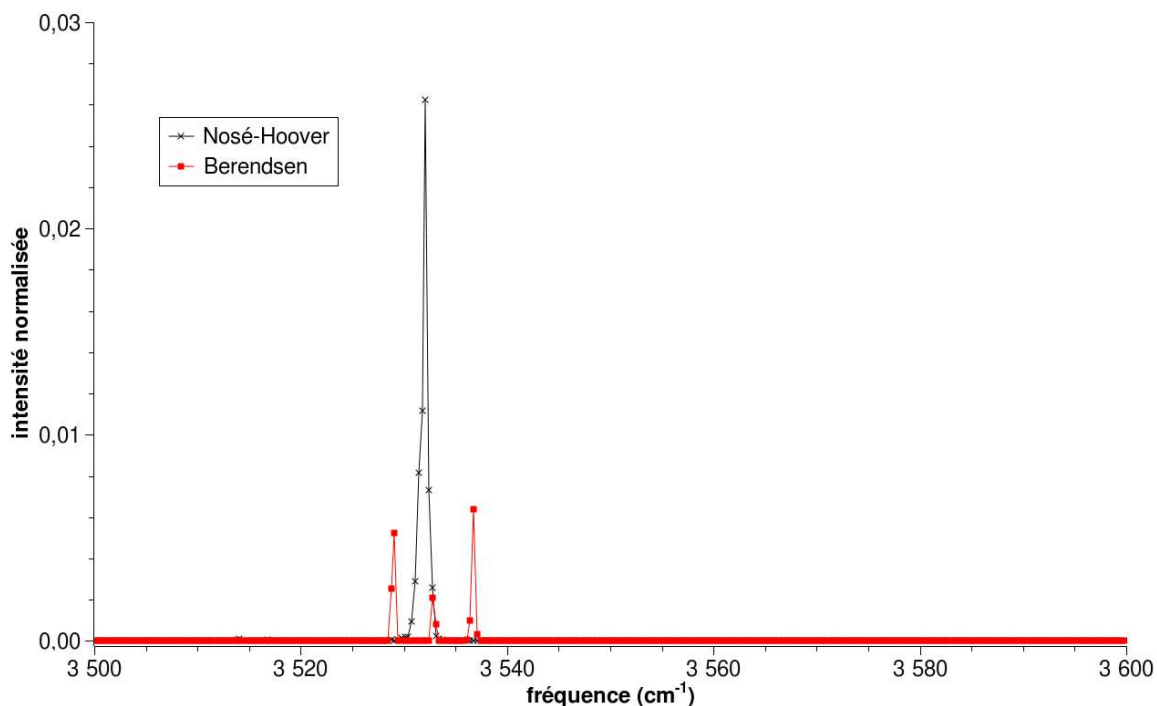


FIG. 5.2: Effets du thermostat utilisé sur le spectre simulé de l'élongation N-H du N-méthylacétamide

l'azote est bloqué en position éclipsée avec le carbonyle, alors que le méthyle du carbonyle est libre et évolue continuellement avec seulement une légère préférence pour la position éclipsée avec l'oxygène.

Avec l'utilisation du thermostat de Nosé-Hoover^{30,31}, un seul pic est observé. Cela tend à montrer que le thermostat de Berendsen avec les options par défaut de Tinker (temps de couplage de 0,1 ps...) est responsable d'un *flying ice cube effect*³² : le rééchantillonnage régulier et brutal des vitesses pour maintenir la température induit un transfert d'énergie vers les modes de faible fréquence, notamment les rotations globales du système, mais pas uniquement. Cet effet aurait lieu en dépit de la procédure d'arrêt de translation et rotation du système régulièrement effectué par Tinker. Les bandes multiples observées dans le spectre infrarouge seraient donc des bandes de combinaison de l'élongation N-H et de rotation de groupement méthyle, qui n'est pas une rotation d'ensemble et n'est donc pas affecté par les mesures de Tinker pour lutter contre le *flying ice cube effect*.

En comparaison avec ces deux thermostats, celui d'Andersen³³ a aussi été testé. Les spectres issus de trajectoires dans lesquelles ce thermostat a été utilisé sont si bruités que l'on ne distingue pas de pic. Les options par défaut de Tinker pour le thermostat d'Andersen peuvent

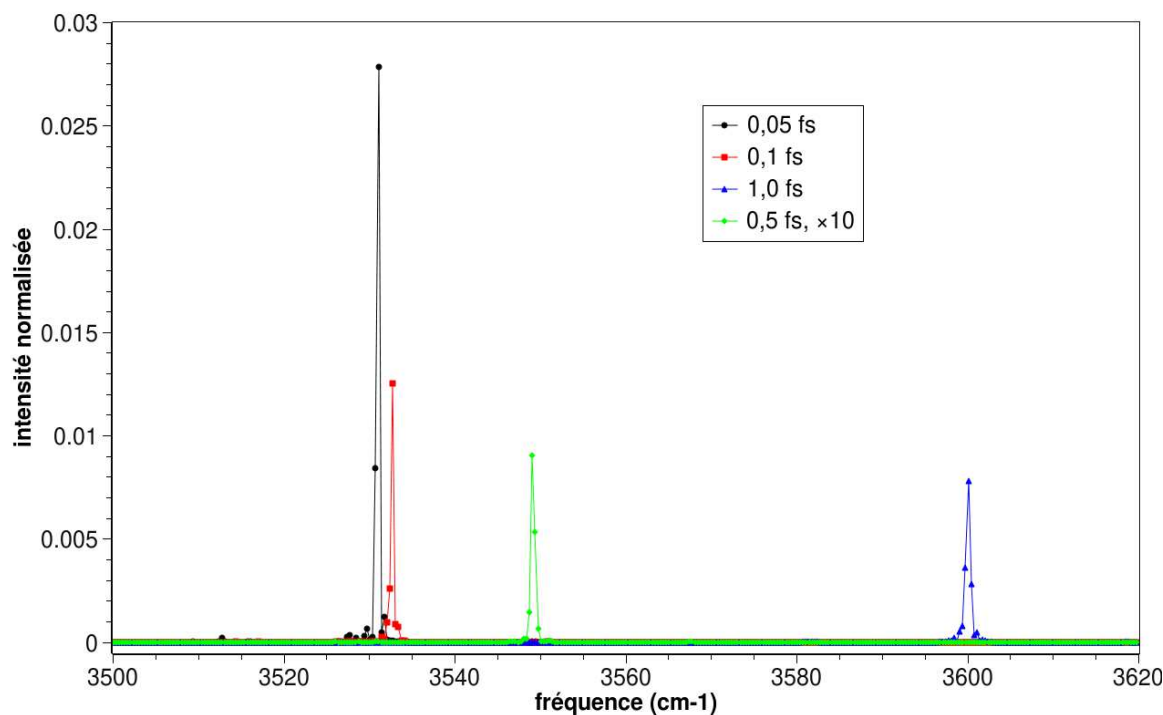


FIG. 5.3: Effets du pas de temps de la trajectoire sur le spectre simulé de l'élongation N-H du N-méthylacétamide à 20 K

avoir une part de responsabilité dans ce constat mais nous considérons que le thermostat de Nosé-Hoover est le meilleur (parmi ceux disponibles dans Tinker) pour la simulation des spectres *via* DACF. C'est donc le thermostat de Nosé-Hoover que nous utilisons pour calculer des spectres dynamiques.

La figure 5.3 superpose des spectres obtenus à partir des 100 dernières ps de trajectoires de 200 ps propagées avec 4 pas de temps allant de 0,05 à 1 fs. La différence entre les pics correspondant aux pas de temps de 0,05 et 0,1 fs est inférieure à 2 cm^{-1} . Par contre, avec des pas de temps plus élevés, le décalage des pics vers les hautes fréquences augmente à environ 18 cm^{-1} pour un pas de 0,5 fs et va jusque près de 70 cm^{-1} pour le pas de temps de 1 fs. Dans le dernier cas, le pas de temps utilisé pour le calcul de la trajectoire est clairement trop élevé pour la prise en compte des vibrations à haute fréquence comme l'élongation N-H dans le cas présent. En revanche, le spectre simulé avec un pas de 0,1 fs est satisfaisant et il s'agira du pas de temps utilisé par la suite pour le calcul des trajectoires dans une optique de calcul de spectres. On peut penser qu'un pas plus grand resterait acceptable pour d'autres régions spectrales comme celle des « empreintes digitales » ($1\ 000\text{--}2\ 000 \text{ cm}^{-1}$).

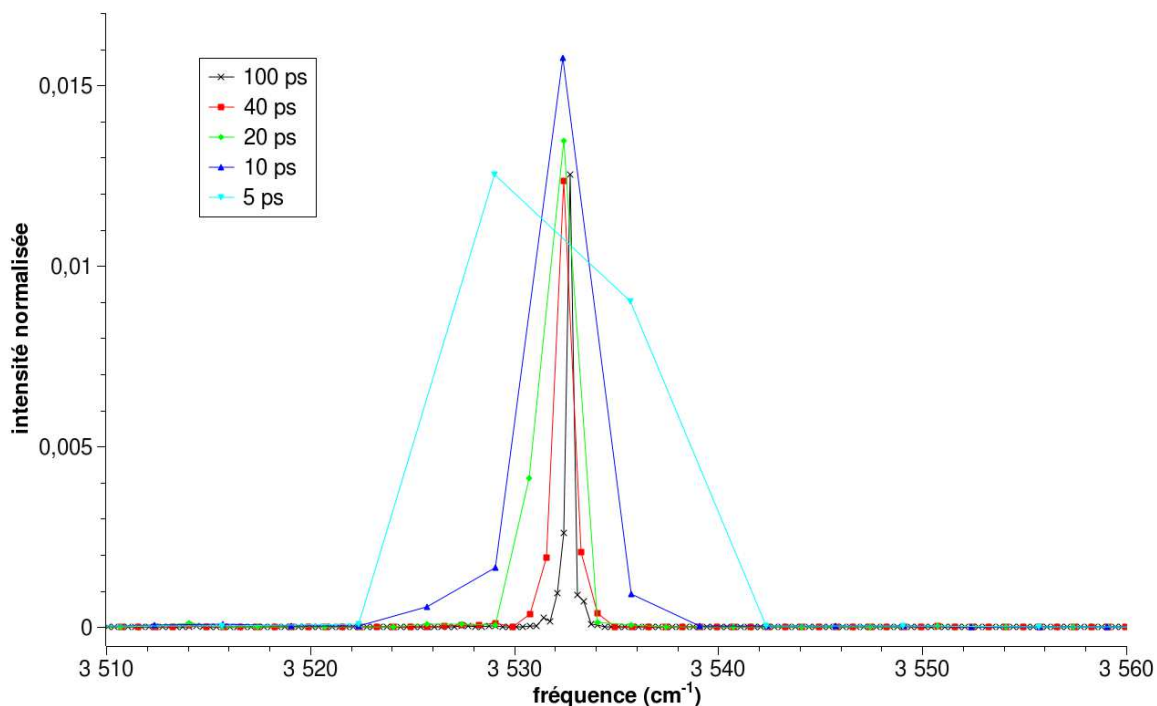


FIG. 5.4: Effets de la durée utile de trajectoire sur le spectre simulé, exemple de l'élongation N-H du N-méthylacétamide à 20 K

Exploitation des spectres obtenus par DACF

Plus le nombre de points utilisés pour l'autocorrélation est élevé, plus sa transformée de Fourier comporte également de points. À pas de temps d'intégration de la dynamique constant, cela signifie qu'utiliser une portion de trajectoire plus longue permet l'obtention d'un spectre plus fin. En étudiant les vibrations du NMA au cours d'une trajectoire de 200 ps effectuée à 20 K, nous avons comparé les spectres obtenus par DACF (Fig. 5.4) en utilisant les dernières 100, 40, 20 et 10 et 5 ps. Les spectres correspondants aux 40 dernières ps et aux 100 dernières ps sont proches l'un de l'autre. Le pic obtenu sur les 20 dernières ps reste très proche du pic effectué sur 100 ps ! La différence en intensité normalisée s'accroît de plus en plus lorsque diminue la durée de la portion de trajectoire utilisée. Ainsi l'élargissement dû à l'exploitation uniquement des 10 dernières ps excède celui observé pour une durée double alors qu'avec 5 ps de trajectoire utilisée, on observe une bande s'étalant d'environ 3 522 à 3 542 cm^{-1} .

Kleinhesselink et coll. ont montré l'importance de la longueur de la trajectoire sur la convergence de la fonction d'autocorrélation³⁴. Ils ont, de plus, discuté de l'efficacité d'effectuer la transformée de Fourier sur la fonction d'autocorrélation tronquée. Dans le but

de s'affranchir des problèmes liés aux « queues » des fonctions d'autocorrélation, une alternative est de ne pas effectuer le calcul de l'autocorrélation dans l'espace réel mais dans l'espace de Fourier. C'est ce qui nous permet, en utilisant le programme Fourier, d'obtenir un spectre de puissance plus stable quant au bruit statistique et dont les intensités sont toutes positives (ou nulles).

Contrairement à ceux dits statiques, les spectres infrarouge obtenus par DACF ne sont pas constitués de pics mais d'un ensemble de points répartis de façon homogène sur la plage de fréquence étudiée. Plus longue est la trajectoire à partir de laquelle le spectre est calculé, et plus celui-ci comporte de points. Comme les spectres expérimentaux, les spectres simulés sont parasités par un bruit. Il existe différents algorithmes permettant de limiter le bruit et de lisser des spectres. Pour le lissage des spectres simulés, nous utilisons l'algorithme de Savitzky-Golay^{35,36}. Cet algorithme est utilisé en spectroscopie infrarouge pour lisser les spectres expérimentaux comme théoriques³⁷⁻³⁹. Chaque algorithme de lissage introduit une distorsion du signal dans le spectre lissé. L'algorithme de Savitzky-Golay a l'avantage notable de conserver dans une certaine mesure la hauteur des pics.

Les effets thermiques sur les spectres obtenus par DACF sont visibles dans la figure 5.5. Le spectre du NMA est simulé avec un thermostat de Nosé-Hoover fixé à 20, 50, 100, 150 et 200 K à partir de trajectoires propagées pendant 200 ps avec un pas de 0,1 fs par l'intégrateur de Beeman. Les intensités normalisées aux températures considérées sont, dans cette gamme de fréquence très différentes entre elles. Afin de les visualiser sur un même graphique, les intensités rencontrées aux températures 100, 150 et 200 K ont été multipliées respectivement par 10, 4 et 8. Les différences d'intensité entre ces températures sont reproduites sur plusieurs trajectoires, ce qui n'aide pas à la compréhension du phénomène. On peut cependant constater un élargissement de la bande avec l'élévation de la température. De plus, un autre effet de l'anharmonicité est visible : à l'exception du spectre à 100 K, la fréquence du maximum d'absorption se déplace vers le rouge.

Les spectres obtenus par DACF ne permettent pas d'établir une correspondance directe entre une bande calculée et un mode normal. L'attribution des bandes d'absorption par décomposition des vibrations en coordonnées internes ou cartésiennes⁴⁰⁻⁴² constitue une perspective pour mieux exploiter les spectres obtenus par DACF.

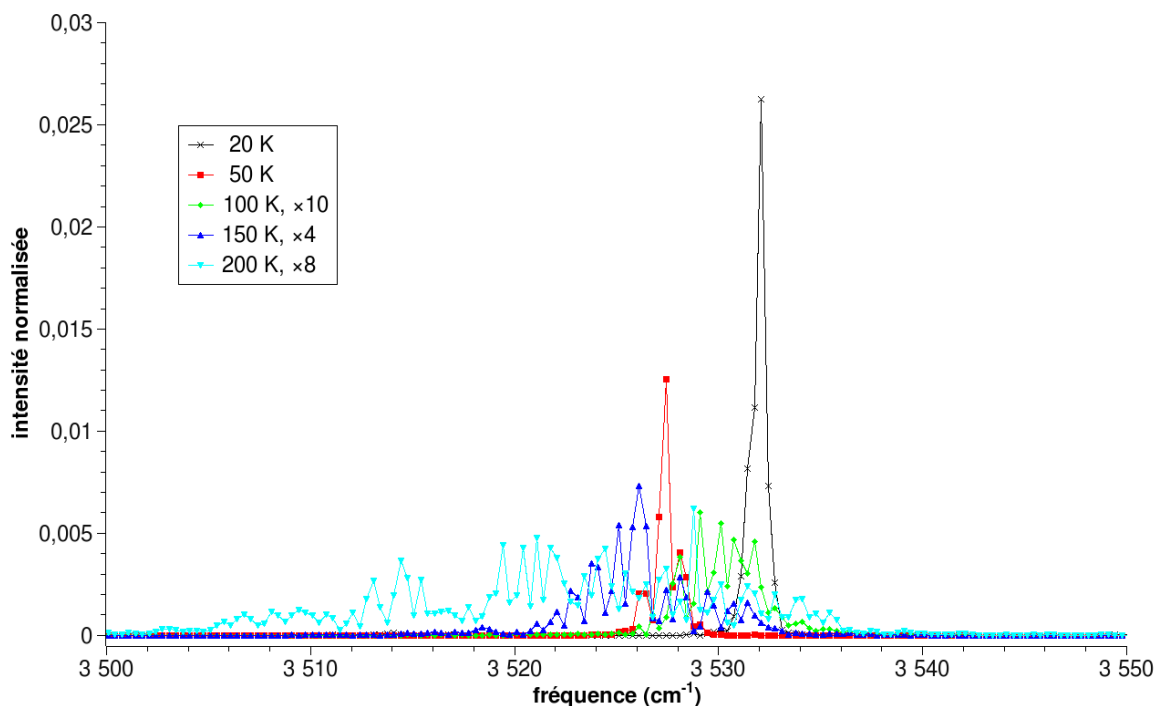


FIG. 5.5: Effets de températures sur l'élongation N-H du N-méthylacétamide

5.3.3 Spectres DACF et transitions structurales

En plus de l'approximation double-harmonique, un spectre statique ne caractérise qu'une structure au fond d'un puits d'énergie, soit correspond à une température de 0 K. Les spectres dynamiques obtenus par DACF permettent de s'affranchir de ces limites.

Si au cours d'une trajectoire, une transition structurale est observée, il faut décider si le DACF doit être effectué sur la totalité ou sur une partie seulement de la trajectoire. L'avantage multiconformationnel du DACF n'est réel que si la trajectoire est suffisamment longue pour échantillonner de manière statistiquement satisfaisante les conformations d'intérêt. Si le système subit des transitions alternant des structures avec des poids statistiques corrects, le DACF sur l'ensemble de la trajectoire fournit un spectre multiconformationnel.

Si une seule transition est observée, ou de manière plus générale, si les structures rencontrées n'ont pas un poids statistique correct car la durée de simulation est trop courte, ou du fait des limitations de la méthode utilisée pour le calcul des forces, alors un calcul de spectre par DACF sur l'ensemble de la trajectoire serait faussé par le défaut de la statistique. Dans un tel cas, il peut être plus intéressant de n'exploiter la trajectoire que par portions durant lesquelles aucune transition n'est observée, si la longueur de chaque portion est suf-

fisante pour échantillonner le bassin auquel elle correspond. Les spectres obtenus par DACF sur ces portions ne sont pas multiconformationnels mais chacun d'eux est approprié en tant que spectre associé à une conformation donnée.

Si plusieurs structures sont pertinentes et ont chacune une stabilité cinétique importante, le spectre expérimental devrait correspondre à une combinaison des spectres monoconformationnels associés à chacune de ces structures puisque la durée de la transition structurale est alors négligeable devant la durée de vie des structures de basse énergie. Bien que les stabilités des différentes structures soient trop grandes pour être interconverties avec une pondération statistique respectée durant une trajectoire de dynamique moléculaire, la propagation de trajectoires pour toutes les structures d'intérêt est une solution pour la simulation de spectres comparables à l'expérience.

En revanche, si l'état de transition permettant l'interconversion entre les structures de plus basse énergie est suffisamment accessible, le temps passé par le système entre les *minima* peut devenir assez important pour qu'une approche dynamique de la simulation de spectres infrarouge soit nécessaire. Gaigeot et coll.⁴³ ont montré que la dialanine protonée en phase gazeuse subit une isomérisation de manière continue et répétée au cours de leur simulation de dynamique moléculaire Car-Parinello. Le spectre obtenu ne peut alors pas être représenté par une somme pondérée des spectres des *minima* explorés.

La durée d'une trajectoire doit être augmentée en fonction de la lenteur ou de la rareté des transitions structurales à prendre en compte. Les durées des trajectoires propagées sont toutefois limitées et cette restriction est d'autant plus importante que le pas de temps d'écriture des données est choisi à 0,1 fs. En conséquence, une simulation de 200 ps est constituée de 2×10^6 points, alors que pour des systèmes ayant un nombre élevé de degrés de liberté comme les peptides, les temps des mouvements de repliement mis en jeu peuvent largement dépasser les centaines de ps. L'utilisation d'un champ de force permet tout de même d'engendrer des trajectoires de plusieurs ns mais cela n'est pas toujours suffisant selon la complexité du système considéré.

Une autre limite dans notre utilisation du calcul de spectre par DACF provient de l'utilisation de la mécanique moléculaire. Aucun des champs de force disponibles pour les molécules d'intérêt biologique ne permet, à l'heure actuelle, d'effectuer des modifications

(formation ou rupture) de liaisons covalentes. AMOEBA ne permet donc pas de décrire de telles transitions structurales. Pour de tels systèmes, le DACF nécessiterait l'utilisation de chimie quantique⁴³ ou de champs de forces réactifs. Le développement de tels champs de forces est poursuivi depuis les travaux de Stillinger et David sur H₂O en 1978⁴⁴ mais concerne principalement les hydrocarbures. Les principaux champs de forces réactifs, AI-REBO^{45,46} et ReaxFF⁴⁷ ne possèdent pas à ce jour de paramètres adaptés aux peptides. Des efforts ont toutefois été faits pour traiter des transferts de proton dans des peptides au moyen d'un champ de force⁴⁸ basé sur Amber96⁴⁹ qui ne traite pas explicitement la polarisation. L'utilisation de la mécanique moléculaire *via* un champ de force polarisable, pour la simulation de spectres *via* DACF pour des systèmes peptidiques ne peut, à l'heure actuelle pas traiter ce genre de problème.

En dehors de ces limites, intrinsèques ou non, la dynamique moléculaire classique ou quantique offre, grâce au DACF, un outil puissant pour la simulation de spectre infrarouge aux températures expérimentales.

5.4 Application du DACF

Après avoir mis au point une méthode de calcul de spectre d'absorption IR du NMA isolé par DACF, nous avons suivi le déplacement de ses bandes d'absorption lorsque le NMA est en interaction avec d'autres molécules pour étudier les capacités du DACF à représenter ces déplacements en fonction de l'environnement. Le but est d'évaluer les capacités de la méthodologie DACF en combinaison avec le champ de force AMOEBA, que nous nommerons « DACF AMOEBA ». Afin de limiter les comparaisons aux modes de vibrations les plus intéressants dans un squelette peptidique, seuls les modes amide A, I et II sont discutés. Quelques paramètres ont été modifiés par rapport aux paramètres originaux d'AMOEBA. Les multipôles proviennent d'une analyse de distribution multipolaire basée sur un calcul MP2/aug-cc-pVTZ d'une molécule de NMA isolée (cf p. 39). La modification des constantes de force d'élongation N-H et de déformation angulaire H-N-CO permet rapidement et simplement de rapprocher les fréquences amide A, I et II calculées *via* le Hessien AMOEBA de celles calculées avec le Hessien MP2 (Tab. 5.2). La méthode *AMOEBA mod.* résulte de l'utilisation du champ de force AMOEBA après ces modifica-

| mode | exp1 | exp2 | exp3 | MP2/aug-cc-pVTZ | MP2/SVP | AMOEBAA0 | AMOEBAA mod. |
|------|-------|------------|------|-----------------|---------|----------|--------------|
| A | 3 510 | - | - | 3535 | 3554 | 3488 | 3533 |
| I | 1 722 | 1731, 1714 | 1707 | 1723 | 1740 | 1703 | 1719 |
| II | - | 1499 | 1511 | 1541 | 1508 | 1541 | 1499 |

exp1¹ jet-FTIR estimant une température rotationnelle inférieure à 30 K et des températures vibrationnelles comprises entre 50 et 300 K

exp2⁵¹ en phase gazeuse à 100 K

exp3⁵² dans une matrice d'argon à 20 K

TAB. 5.2: Fréquences infrarouge statiques et expérimentales du N-méthylacétamide

tions. L'utilisation des paramètres originaux d'AMOEBAA constitue la méthode AMOEBAA0. Par simplicité, « AMOEBAA » désignera le champ de force AMOEBAA avec paramètres modifiés. Les valeurs MP2/aug-cc-pVTZ sont échelonnées à :

- 0,9557 pour le mode amide A, *i.e.* l'élongation N-H comme conseillé par Wilson et coll.⁵⁰ pour les hautes fréquences ;
- 0,983 pour les modes amide I et II, ce qui rapproche ces valeurs et les expérimentales.

Les fréquences MP2/SVP sont données après application d'un facteur d'échelle de 0,952. Les fréquences MP2 et les facteurs d'échelle associés sont discutés un peu plus loin. En AMOEBAA0, le *stretch* N-H du NMA isolé a une fréquence sous-estimée respectivement de 22 et 66 cm⁻¹ par rapport aux valeurs expérimentale et MP2/SVP. La fréquence amide I est trop basse d'une vingtaine de cm⁻¹, en fonction des estimations. Par contre, la fréquence de la bande amide II est surestimée. Les modifications de paramètres ont permis de passer d'une surestimation en AMOEBAA0 par rapport aux valeurs expérimentales de plus de 40 cm⁻¹ (33 cm⁻¹ par rapport au MP2/SVP) à l'égalité avec la fréquence expérimentale de Kubelka⁵¹. En parallèle, la bande amide I est déplacée de 16 cm⁻¹ vers le bleu, la rendant ainsi plus proche de la fréquence MP2 et de la plupart des valeurs expérimentales.

Pour représenter des environnements variés, quatre systèmes sont considérés en utilisant le NMA comme point de comparaison pour les déplacements spectraux dus aux interactions non covalentes. Le NMA en interaction avec Na⁺ permet l'étude des déplacements des bandes amide A et amide I et II quand le carbonyle est attaché à ce cation, cas d'intérêt biologique et pertinent pour des études expérimentales par IRMPD de peptides attachés à un cation sodium, décrites dans d'autres chapitres. Le système NMA-H₂O permet de discuter de l'effet d'une liaison hydrogène de micro-solvatation sur l'élongation N-H. Enfin,

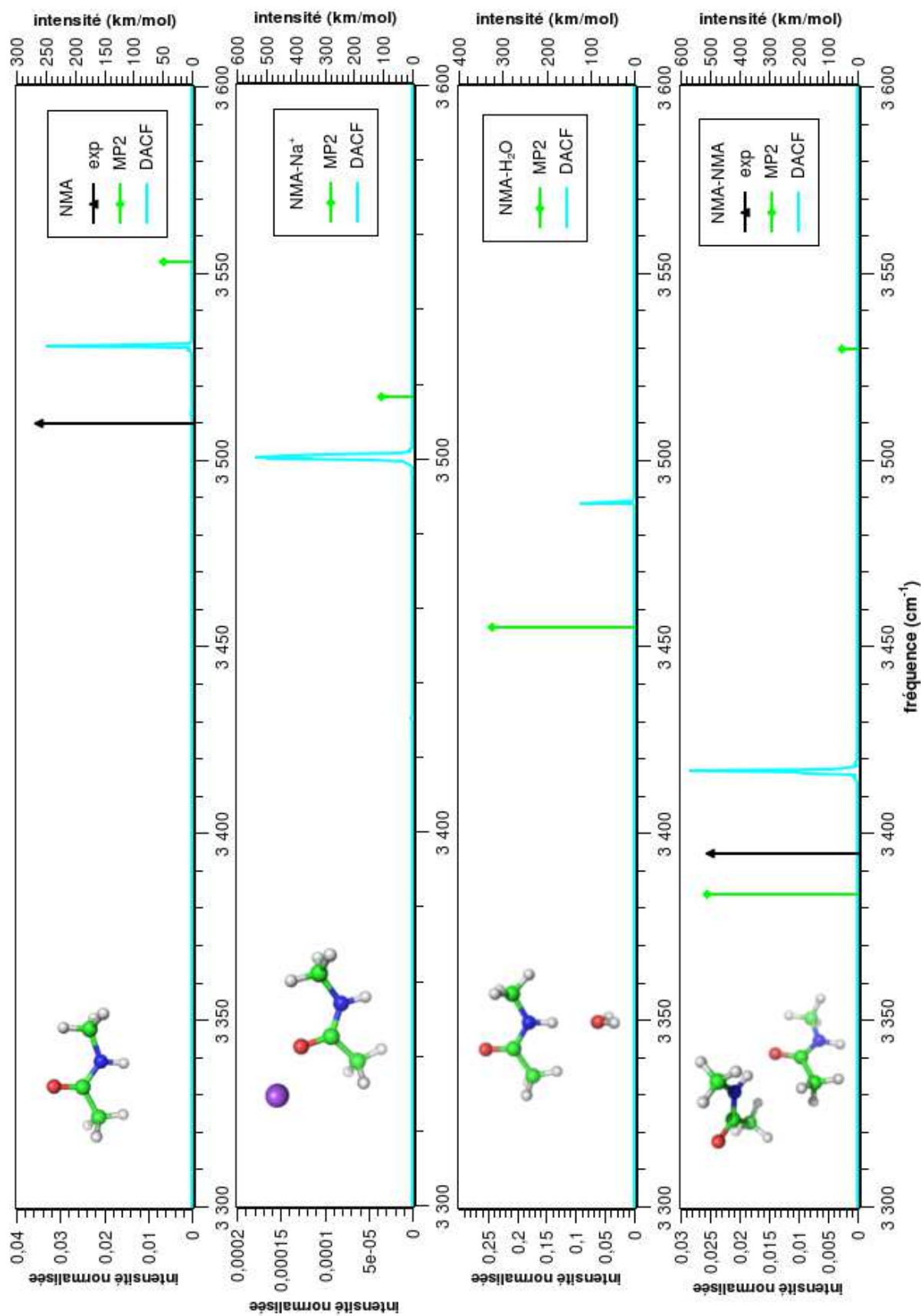
le dimère, *i.e.* le NMA en interaction avec une autre molécule de NMA est un système présentant à la fois un CO et un NH libres, un CO et un NH liés et est particulièrement pertinent pour évaluer les paramètres d'AMOEBa que nous avons modifiés pour le terme électrostatique, notamment. Les spectres simulés par DACF AMOEBa ici sont tous issus de trajectoires de dynamique moléculaire utilisant les paramètres modifiés du champ de force AMOEBa. Elles sont propagées pendant 200 ps par un algorithme de Beeman, avec un pas de 0,1 fs dans l'ensemble thermodynamique NPT, avec le thermostat de Nosé-Hoover. Pour le calcul des spectres, uniquement les 100 dernières ps sont considérées, bien après que l'on puisse considérer le système comme thermalisé.

Le stretch N-H est un mode qui, nous l'avons vu, est isolé dans sa gamme de fréquence. Mises à part les élongations O-H en cas d'interaction avec H₂O, il ne peut y avoir aucune ambiguïté quant à l'interprétation des bandes expérimentales ou théoriques. La figure 5.6 montre les bandes d'absorption N-H en fonction de leur situation. Les données expérimentales étant limitées, des spectres calculés au niveau MP2/SVP servent de point de comparaison, bien qu'ils ne prennent pas en compte les effets dus à la température, présents même à basse température et que la justesse des fréquences obtenues soit sujette à discussion (utilisation d'un seul facteur pour toutes les bandes...). Notons aussi que la température de simulation de 20 K est *a priori* inférieure aux quelques températures des expériences citées, qui ne sont précisément connues. Le facteur d'échelle de 0,952 est trop grand pour permettre une superposition avec la fréquence expérimentale de 3510 cm⁻¹ dans le cas d'une molécule isolée de NMA. La valeur MP2/SVP est alors de 3554 cm⁻¹ mais la fréquence du N-H lié dans le cas du dimère de NMA, expérimentalement à 3395 cm⁻¹ est elle, sous-estimée de 12 cm⁻¹ en MP2/SVP. L'exemple du dimère montre la difficulté de modéliser les *stretch* N-H avec un seul facteur d'échelle appliqué à un calcul MP2. À 20 K, la fréquence du mode amide A calculée par DACF en utilisant AMOEBa se positionne 20 cm⁻¹ au dessus de la fréquence expérimentale et 21 cm⁻¹ en dessous de la valeur MP2. Dans le cas du dimère, la fréquence expérimentale d'élongation du N-H lié est déplacée vers le rouge de 115 cm⁻¹ par rapport au N-H libre. Ce déplacement est d'environ 110 cm⁻¹ en calcul statique AMOEBa et 113 en DACF/AMOEBa contre respectivement 166 et 147 cm⁻¹ en MP2 avec les bases aug-cc-pVTZ et SVP. C'est la principale raison pour laquelle une base plus petite que aug-cc-pVTZ a été choisie pour effectuer des comparaisons. Des compensations d'erreur amènent ici à de meilleures fréquences.

Le déplacement des bandes obtenu par DACF avec AMOEBA à 20 K est en bon accord avec les données expérimentales du monomère et du dimère de NMA. En AMOEBA comme en MP2, le pic correspondant à l'élongation du N-H libre est peu déplacée par rapport au monomère : moins de 10 cm^{-1} en AMOEBA et moins de 25 cm^{-1} en MP2. Le spectre expérimental ne permet pas de confirmer la présence d'une absorption sensible par ce mode de vibration. L'intensité de ce mode en DACF est très faible et à peine visible en comparaison avec la bande du N-H lié. Ceci n'est pas seulement en accord avec le spectre expérimental mais aussi avec les intensités MP2 qui présentent un rapport d'environ 1/10.

Pour les systèmes de NMA en interaction avec un cation sodium et NMA avec une molécule d'eau, ne disposant pas de données expérimentales, nous ne pouvons que comparer les valeurs aux fréquences MP2. Les déplacements du mode d'élongation N-H observés en MP2/SVP, respectivement de 37 et 98 cm^{-1} sont suivis par des déplacements en DACF de 30 et 42 cm^{-1} . Les résultats présentés sur la bande amide A sont donc très satisfaisants par rapport aux données expérimentales et sont en accord avec les tendances des fréquences MP2 dans le cas des systèmes pour lesquelles aucune donnée expérimentale n'est disponible.

Les trajectoires de dynamique moléculaire utilisées pour simuler les spectres DACF du *stretch* N-H permettent aussi de tracer les spectres dans la gamme des bandes amide I et II (Fig. 5.7). L'attribution des bandes est effectuée par comparaison au spectre statique AMOEBA, en supposant qu'à 20 K, les bandes ne seront que très peu déplacées par la température. Dans le cas du monomère, la bande amide I DACF de $1\,719\text{ cm}^{-1}$ est moins de 5 cm^{-1} au dessous de la bande expérimentale et une vingtaine de cm^{-1} en dessous du pic MP2 (Tab. 5.2). L'analyse vibrationnelle variationnelle utilisée par Kaledin et Bowman prédit une fréquence de $1\,727\text{ cm}^{-1}$. La valeur obtenue par DACF AMOEBA se démarque des $1\,609\text{ cm}^{-1}$ obtenus avant par Gageot et coll. à partir d'un DACF effectué sur une trajectoire CPMD⁴¹ et qui est ramenée à $1\,712\text{ cm}^{-1}$ après échelonnage par un facteur de $1,064$ ⁵³ en phase gazeuse. D'après les spectres expérimentaux d'Albrecht et coll., la bande amide I du CO lié dans le dimère est déplacée de $1\,722$ (dans le monomère) à $1\,690\text{ cm}^{-1}$. Ce déplacement de 32 cm^{-1} est reproduit à 3 cm^{-1} près en MP2/SVP. Le déplacement vers le rouge obtenu en DACF AMOEBA est d'une vingtaine de cm^{-1} et est par conséquent moins important de 12 cm^{-1} qu'expérimentalement. Dans le spectre expérimental, on ne distingue



Les positions des *maxima* des bandes expérimentales pour NMA et NMA-NMA sont tirées des travaux d'Albrecht et coll.¹

FIG. 5.6: Déplacement de la bande de l'élongation N-H du N-méthylacétamide à 20 K

pas la fréquence du mode amide I du CO libre qui est probablement de relativement faible intensité et recouvert par la bande amide I du monomère, qui est plus peuplée. Un faible déplacement de la bande amide I libre entre le dimère et le monomère de moins de 10 cm^{-1} en DACF/AMOEBE et inférieur à 2 cm^{-1} en MP2/SVP de la bande amide I libre entre le dimère et le monomère est en accord avec l'observation expérimentale. En MP2, la bande amide I est déplacée vers le rouge de 66 cm^{-1} pour NMA- Na^+ et n'est pas déplacée de plus d'un cm^{-1} pour NMA- H_2O . Le comportement de la bande amide I en DACF/AMOEBE suit de près celui observé en MP2/SVP. Il est 7 cm^{-1} inférieur à la valeur *ab initio* dans le cas du NMA en interaction avec le sodium et prévoit un déplacement d'environ 6 cm^{-1} pour le NMA avec une molécule d'eau quand MP2 estime un déplacement inférieur à l'unité.

Parmi les systèmes étudiés, le monomère est le seul à disposer de fréquences expérimentales connues pour le mode amide II. Expérimentalement, cette bande est située à $1\,499\text{ cm}^{-1}$. La bande obtenue en DACF/AMOEBE recouvre cette valeur et son maximum n'en diffère pas de plus 2 ou 3 cm^{-1} . La valeur MP2/SVP surestime de 10 cm^{-1} la fréquence de ce mode.

Les attributions de Paul Tavan et coll.⁵⁴ dont les simulations situent le mode amide II à $1\,508\text{ cm}^{-1}$ en BP86/TZVP et $1\,498\text{ cm}^{-1}$ en dynamique moléculaire Car-Parrinello avec la même fonctionnelle, sont en accord avec l'attribution expérimentale $1\,499\text{ cm}^{-1}$ et le résultat obtenu en VCI MP2/aug-cc-pVTZ par Kaledin et Bowman de $1\,516$ ou $1\,519\text{ cm}^{-1}$ selon le niveau de couplage de modes qu'ils ont utilisé. Tout comme le spectre que nous obtenons par DACF/AMOEBE, le spectre obtenu par Tavan et coll. à partir d'une trajectoire de dynamique moléculaire utilisant un champ de force polarisable paramétré sur des données issues de calculs DFT BP86/TZVP est proche du spectre expérimental. Un spectre calculé *via* DACF/CPMD par Gageot et coll.⁴¹, sous-estime la fréquence de la bande amide II de manière importante et la place à $1\,458\text{ cm}^{-1}$ (ou $1\,551\text{ cm}^{-1}$ après échelonnage).

Lorsque le NMA interagit avec un autre corps, la différence entre les fréquences MP2 et DACF/AMOEBE s'accroît considérablement. En MP2, ces bandes sont systématiquement déplacées vers le bleu, respectivement de 55, 35 et 30 cm^{-1} pour NMA- Na^+ , NMA- H_2O et le dimère, NMA-NMA. On peut noter que le mode amide II correspondant au N-H lié, dans le cas du dimère, est déplacé vers le bleu d'encre 15 cm^{-1} par rapport au mode amide II du N-H libre. En AMOEBE/DACF, la bande du amide II du N-H libre n'est déplacée que de 6 cm^{-1} vers le bleu par rapport au monomère et celle du N-H lié est déplacée de 5 cm^{-1}

vers le rouge. Les fréquences MP2 et DACF/AMOEBAs sont très différentes, mais l'écart de 11 cm^{-1} entre les deux bandes amides en DACF est proche des 15 cm^{-1} obtenus en MP2. En présence du sodium, le déplacement en DACF/AMOEBAs de la bande amide II du NMA est de 20 cm^{-1} , quand MP2 l'évalue à 55 cm^{-1} . En interaction avec une molécule, cette bande est déplacée de 3 cm^{-1} vers le rouge, alors que le déplacement calculé en MP2/SVP de 35 cm^{-1} vers les hautes fréquences. Pour l'ensemble des systèmes, la bande amide II est globalement dans la bonne zone de fréquence.

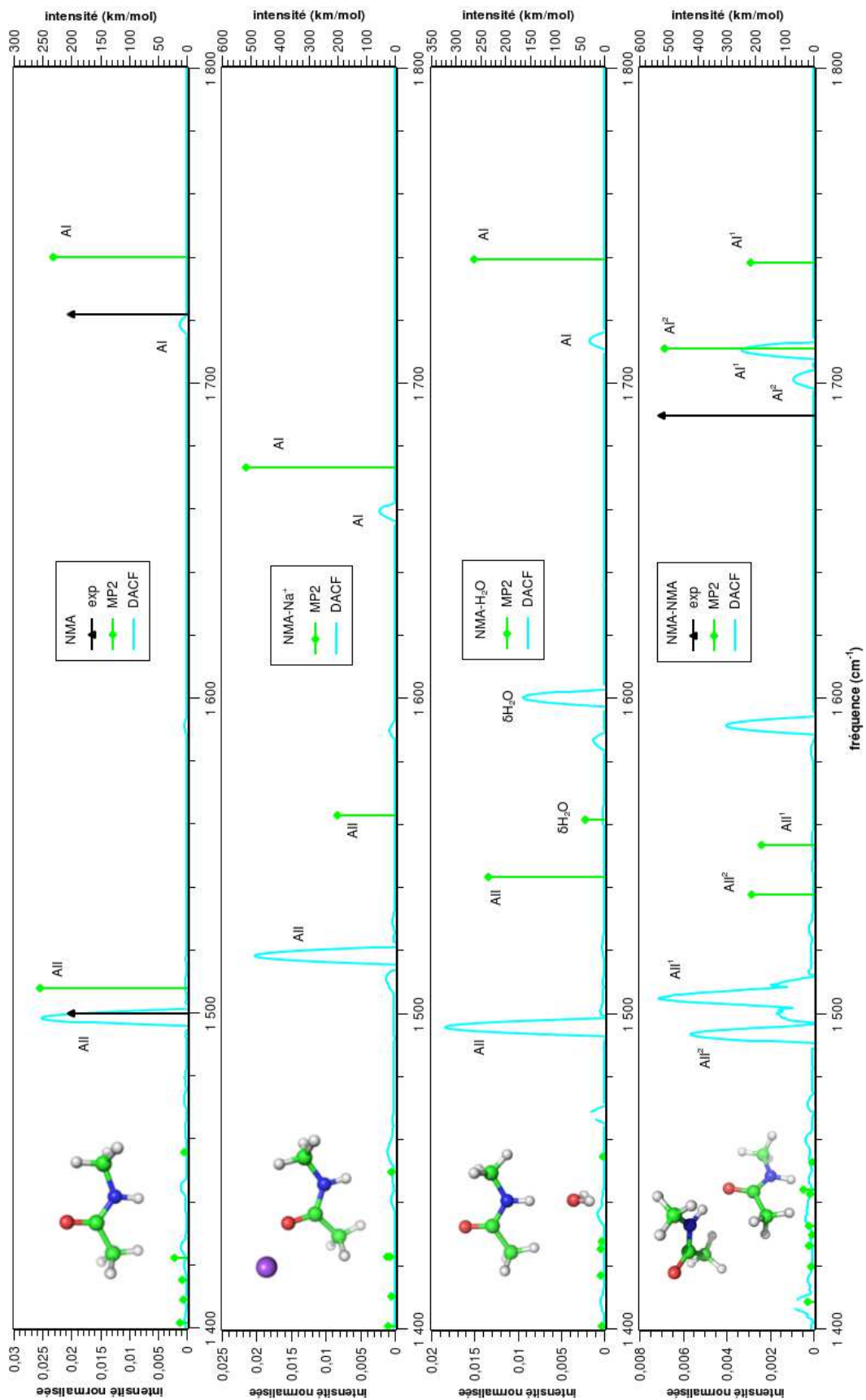
Enfin, les pics observés autour de 1590 cm^{-1} pour les 4 systèmes mais qui sont intenses pour le dimère, sont attribués à des mouvements de parapluie des groupements méthyle, principalement celui lié à l'azote. Le couplage de ces mouvements avec une déformation angulaire du N-H dans le plan pourrait expliquer l'intensité observée. La fréquence élevée de ces modes en AMOEBAs pourrait provenir d'une surestimation de la constante de force, auquel cas une modification de celle-ci la déplacerait.

Si les facteurs d'échelle du MP2 et les paramètres d'AMOEBAs utilisés sont discutables, les modes amide II simulés par les deux méthodes se situent à plus de 40 cm^{-1} l'un de l'autre, à l'exception du monomère pour lequel l'écart entre les deux méthodes est limité à 3 cm^{-1} .

En résumé, la démarche DACF AMOEBAs se montre capable de modéliser correctement à la fois la position des bandes amide A, I et II. Elle permet en outre d'en évaluer les déplacements dus aux différentes interactions auxquelles le NMA peut être exposé. Ces déplacements sont qualitativement et quantitativement satisfaisants. Ils reproduisent exceptionnellement bien l'expérimental pour un champ de force. Une étude plus approfondie pourrait peut-être révéler un problème provenant des paramètres des méthyles ou la nécessité d'affiner les paramètres modifiés. On peut éventuellement envisager la modification de quelque paramètre supplémentaire.

Bien que l'on manque de données expérimentales, la méthodologie DACF associée au formalisme d'AMOEBAs est prometteuse pour la simulation de spectres infrarouge des peptides. Le traitement de l'interaction avec le cation sodium notamment, nécessite une comparaison expérimentale pour être pleinement validée.

Il serait intéressant de calculer des spectres *via* DACF/AMOEBAs avec paramètres modifiés pour des cas dans lesquels les déplacements vers le rouge sont plus importants (plusieurs



Les positions des *maxima* des bandes expérimentales pour NMA et NMA-NMA sont tirées des travaux de Kubelka et coll.⁵¹

FIG. 5.7: Déplacement des bandes amide I et II du N-méthylacétamide (notées respectivement AI et AII)

centaines de cm^{-1}) afin d'évaluer les limitations d'AMOEBa en terme d'influence de l'environnement sur des modes de vibration. Ces limites pourraient provenir du formalisme d'AMOEBa qui ne décrit pas explicitement une des composantes de la liaison hydrogène, le transfert de charge. Un modèle de charges atomiques fluctuantes⁵⁵ pourrait encore augmenter les possibilités de réponse des modes de vibration à l'environnement chimique dans le cas d'un champ de force classique.

Bibliographie

- [1] M. ALBRECHT, C. A. RICE et M. A. SUHM : Elementary Peptide Motifs in the Gas Phase : FTIR Aggregation Study of Formamide, Acetamide, N-Methylformamide, and N-Methylacetamide. *J. Phys. Chem. A*, 112:7530–7542, 2008.
- [2] M. KOLÁŘ et P. HOBZA : Accurate Theoretical Determination of the Structure of Aromatic Complexes Is Complicated : The Phenol Dimer and Phenol ··· Methanol Cases. *J. Phys. Chem. A*, 111:5851–5854, 2007.
- [3] P. CARBONNIÉRE, A. DARGELOS, I. CIOFINI, C. ADAMO et C. POUCHAN : Vibrational analysis of glycine radical : a comparative ab initio static and dynamic study. *Phys. Chem. Chem. Phys.*, 11:4375–4384, 2009.
- [4] P. PULAY : *Modern Electronic Structure Theory, Part II*, page 1191. World Scientific, Singapore, 1995.
- [5] S. CARTER, J. M. BOWMAN et N. C. HANDY : Extensions and tests of ”multimode” : a code to obtain accurate vibration/rotation energies of many-mode molecules. *Theor. Chem. Acc.*, 100:191–198, 1998.
- [6] J. M. BOWMAN : Self-consistent field energies and wavefunctions for coupled oscillators. *J. Chem. Phys.*, 68:608–611, 1978.
- [7] K. M. CHRISTOFFEL et J. M. BOWMAN : Investigations of self-consistent field, SCF CI and virtual state configuration interaction vibrational energies for a model three-mode system. *Chem. Phys. Lett.*, 85:220–224, 1982.
- [8] G. D. CARNEY, L. L. SPRANDEL et C. W. KERN : Variational Approaches to Vibration-Rotation Spectroscopy for Polyatomic Molecules. *Adv. Chem. Phys.*, 37:305–379, 2007.

- [9] L. S. NORRIS, M. A. RATNER, A. E. ROITBERG et R. B. GERBER : Møller-Plesset perturbation theory applied to vibrational problems. *J. Chem. Phys.*, 105:11261–11267, 1996.
- [10] V. BARONE : Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.*, 122:014108, 2005.
- [11] R. B. GERBER et M. A. RATNER : A semiclassical self-consistent field (SC SCF) approximation for eigenvalues of coupled-vibration systems. *Chem. Phys. Lett.*, 68: 195–198, 1979.
- [12] O. CHRISTIANSEN, J. KONGSTED, M. J. PATERSON et J. M. LUIS : Linear response functions for a vibrational configuration interaction state. *J. Chem. Phys.*, 125:214309, 2006.
- [13] P. CASSAM-CHENAÏ et J. LIÉVIN : The VMFCI method : A flexible tool for solving the molecular vibration problem. *J. Comput. Chem.*, 27:627–640, 2006.
- [14] Donald Allan MCQUARRIE : *Statistical Mechanics*, pages 553–561. HarperCollinsPublishers, 1976.
- [15] D. W. NOID, M. L. KOSZYKOWSKI et R. A. MARCUS : A spectral analysis method of obtaining molecular spectra from classical trajectories. *J. Chem. Phys.*, 67:404–408, 1977.
- [16] P. H. BERENS et K. R. WILSON : Molecular dynamics and spectra. I. Diatomic rotation and vibration. *J. Chem. Phys.*, 74:4872–4882, 1981.
- [17] D. A. TILBROOK, B. J. HOWLIN et I. HAMERTON : Assessment of two methods for application in the prediction of the infrared spectra of polymers. *Comput. Theor. Polym. Sci.*, 11:287–302, 2001.
- [18] M. R. JOHNSON et H. P. TROMMSDORFF : Vibrational spectra of crystalline formic and acetic acid isotopologues by inelastic neutron scattering and numerical simulations. *Chem. Phys.*, 355:118–122, 2009.
- [19] A. KOHLMAYER et H. FORBERT : Fourier-Version 21.02.2006.
- [20] D. A. MCQUARRIE : *Statistical Mechanics*, pages 471–376. HarperCollinsPublishers, 1976.

- [21] R. RAMÍREZ, T. LÓPEZ-CIUDAD, P. K. PADMANABHAN et D. MARX : Quantum corrections to classical time-correlation functions : Hydrogen bonding and anharmonic floppy modes. *J. Chem. Phys.*, 121:3973–3983, 2004.
- [22] P. SCHOFIELD : Space-Time Correlation Function Formalism for Slow Neutron Scattering. *Phys. Rev. Lett.*, 4:239–240, 1960.
- [23] D. MARX et Hutte J. : *Ab initio molecular dynamics : basic theory and advanced methods*, pages 260–266. Cambridge University Press, 2009.
- [24] R. IFTIMIE et M. E. TUCKERMAN : Decomposing total IR spectra of aqueous systems into solute and solvent contributions : A computational approach using maximally localized Wannier orbitals. *J. Chem. Phys.*, 122:214508, 2005.
- [25] C. P. LAWRENCE, A. NAKAYAMA, N. MAKRI et J. L. SKINNER : Quantum dynamics in simple fluids. *J. Chem. Phys.*, 120:6621–6624, 2004.
- [26] M. SCHMITZ et P. TAVAN : Vibrational spectra from atomic fluctuations in dynamics simulations. II. Solvent-induced frequency fluctuations at femtosecond time resolution. *J. Chem. Phys.*, 121:12247–12258, 2004.
- [27] B. JOALLAND, M. RAPACIOLI, A. SIMON, C. JOBLIN, C. J. MARSDEN et F. SPIEGELMAN : Molecular Dynamics Simulations of Anharmonic Infrared Spectra of [SiPAH]⁺ π -Complexes. *J. Phys. Chem. A*, 2010. à paraître.
- [28] D. W. KOSZYKOWSKI, M. L. Noid et R. A. MARCUS : Semiclassical theory of intensities of vibrational fundamentals, overtones, and combination bands. *J. Phys. Chem.*, 86:2113–2117, 1982.
- [29] H. J. C. BERENDSEN, J. P. M. POSTMA, W. F. van GUNSTEREN, A. DI NOLA et J. R. HAAK : Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
- [30] S. NOSÉ : A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511–519, 1984.
- [31] W. G. HOOVER : Canonical dynamics : Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.
- [32] S. C. HARVEY, R. K.-Z. TAN et T. E. III CHEATHAM : The flying ice cube : Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.*, 19:726–740, 1998.

- [33] H. C. ANDERSEN : Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 1980.
- [34] D. KLEINHESSELINK et M. WOLFSBERG : The evaluation of power spectra in molecular dynamics simulations of anharmonic solids and surfaces. *Surf. Sci.*, 262:189–207, 1992.
- [35] A. SAVITZKY et M. J. E. GOLAY : Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.*, 36:1627–1639, 1964.
- [36] J. STEINIER, Y. TERMONIA et J. DELTOUR : Comments on smoothing and differentiation of data by simplified least square procedure. *Anal. Chem.*, 44:1906–1909, 1972.
- [37] S. MORITA, H. SHINZAWA, R. TSENKOVA, I. NODA et Y. OZAKI : Computational simulations and a practical application of moving-window two-dimensional correlation spectroscopy. *J. Mol. Struct.*, 799:111–120, 2006. Applications of Two-Dimensional correlation spectroscopy, A Collection of Papers Presented at the Third International Symposium on Two-Dimensional Correlation Spectroscopy (2DCOS-3) Delavan, Wisconsin, USA, August 12-14, 2005.
- [38] J. P. LARENTZOS, J. A. GREATHOUSE et R. T. CYGAN : An ab Initio and Classical Molecular Dynamics Investigation of the Structural and Vibrational Properties of Talc and Pyrophyllite. *J. Phys. Chem. C*, 111:12752–12759, 2007.
- [39] P. K. MANKOO et T. KEYES : Induction model for molecular electrostatics : Application to the infrared spectroscopy of CO liquid. *J. Chem. Phys.*, 124:204503, 2006.
- [40] E. HENSSGE, D. DUMONT, D. FISCHER et D. BOUGEARD : Analysis of infrared and Raman spectra calculated by molecular dynamics. *J. Mol. Struct.*, 482-483:491–496, 1999.
- [41] M.-P. GAIGEOT, M. MARTINEZ et R. VUILLEUMIER : Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations : application to small peptides. *Mol. Phys.*, 105:2857–2878, 2007.
- [42] M.-P. GAIGEOT : Unravelling the Conformational Dynamics of the Aqueous Alanine Dipeptide with First-Principle Molecular Dynamics. *J. Phys. Chem. B*, 113:10059–10062, 2009.

- [43] D. C. MARINICA, G. GRÉGOIRE, C. DESFRANÇOIS, J. P. SCHERMANN, D. BORGIS et M. P. GAIGEOT : Ab Initio Molecular Dynamics of Protonated Dialanine and Comparison to Infrared Multiphoton Dissociation Experiments. *J. Phys. Chem. A*, 110:8802–8810, 2006.
- [44] F. H. STILLINGER et C. W. DAVID : Polarization model for water and its ionic dissociation products. *J. Chem. Phys.*, 69:1473–1484, 1978.
- [45] D. W. BRENNER : The Art and Science of an Analytic Potential. *Phys. Status Solidi B*, 217:23–40, 2000.
- [46] D. W. BRENNER, O. A. SHENDEROVA, J. A. HARRISON, S. J. STUART, B. NI et S. B. SINNOTT : A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. *J. Phys. Condens. Matter*, 14:783, 2002.
- [47] A. C. T. van DUIN, S. DASGUPTA, F. LORANT et III GODDARD, W. A. : ReaxFF : A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A*, 105:9396–9409, 2001.
- [48] F. CALVO et Ph. DUGOURD : Theoretical Evidence for Temperature-induced Proton Mobility in Isolated Lysine-rich Polyalanines. *J. Phys. Chem. A*, 112:4679–4687, 2008.
- [49] P. KOLLMAN, R. DIXON, W. CORNELL, T. FOX, C. CHIPOT et A. POHORILLE : *The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of ab Initio calculations and experimental data*, volume 3, pages 83–96. Springer, 1997.
- [50] P. SINHA, S. E. BOESCH, C. GU, R. A. WHEELER et A. K. WILSON : Harmonic Vibrational Frequencies : Scaling Factors for HF, B3LYP, and MP2 Methods in Combination with Correlation Consistent Basis Sets. *J. Phys. Chem. A*, 108:9213–9217, 2004.
- [51] J. KUBELKA et T. A. KEIDERLING : Ab Initio Calculation of Amide Carbonyl Stretch Vibrational Frequencies in Solution with Modified Basis Sets. 1. N-Methyl Acetamide. *J. Phys. Chem. A*, 105:10922–10928, 2001.
- [52] S. ATAKA, H. TAKEUCHI et M. TASUMI : Infrared studies of the less stable cis form of N-methylformamide and N-methylacetamide in low-temperature nitrogen matrices and vibrational analyses of the trans and cis forms of these molecules. *J. Mol. Struct.*, 113:147–160, 1984.

- [53] M.-P. GAIGEOT, R. VUILLEUMIER, M. SPRIK et D. BORGIS : Infrared Spectroscopy of N-Methylacetamide Revisited by ab Initio Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, 1:772–789, 2005.
- [54] V. SCHULTHEIS, R. REICHHOLD, B. SCHROPP et P. TAVAN : A Polarizable Force Field for Computing the Infrared Spectra of the Polypeptide Backbone. *J. Phys. Chem. B*, pages 12217–12230, 2008.
- [55] C. CLAVAGUÉRA, F. CALVO et J.-P. DOGNON : Theoretical study of the hydrated Gd^{3+} ion : Structure, dynamics, and charge transfer. *J. Chem. Phys.*, 124:074505, 2006.

Deuxième partie

Applications

Les méthodes décrites dans les chapitres précédents ont été appliquées à la simulation de spectres infrarouge de peptides attachés à un cation sodium en phase gazeuse. Il fallait d'abord nous doter des moyens de modéliser correctement leur structure et leur énergétique. La recherche des structures de basse énergie est un processus long et il convient d'utiliser des méthodes exécutables dans des temps raisonnables. Nous avons opté pour l'utilisation du champ de force polarisable AMOEBA, incluant des termes d'anharmonicité, un fin traitement de l'électrostatique et les effets à n corps qui sont importants pour les peptides que nous étudions. AMOEBA ne disposait pas de paramètre pour les terminaisons neutres des peptides. Nous avons dû ajouter et modifier les paramètres des terminaisons et de certaines torsions du squelette peptidique. L'absence de chaîne latérale permet d'effectuer, sur les polyglycines des calculs quantiques de niveau élevé nécessaires pour servir de référence au champ de force. L'article p. 127 présente les originalités structurales des structures de polyglycines sodées en phase gazeuse. Les résultats obtenus avec AMOEBA sont comparés à ceux obtenus avec la fonctionnelle de la densité B3LYP et les méthodes post-Hartree-Fock MP2 et CC2. Des trajectoires de dynamique moléculaire ont conduit à des structures plus stables que les structures obtenues auparavant.

Des désaccords persistants entre les spectres infrarouge calculés et certains spectres IRMPD expérimentaux indiquaient que ces résultats de dynamique moléculaire ne suffisaient pas pour obtenir les structures les plus stables. Nous avons utilisé AMOEBA en couplage avec une méthode efficace d'exploration de l'espace conformationnel : la dynamique par échange de réplique ou REMD (cf. p. 59). L'exploration a été réalisée sur la surface de potentiel du cas le plus complexe *a priori*, $G_8\text{-Na}^+$ (p. 141). Elle a permis de localiser de nouveaux *minima* très stables, en particulier deux structures plus basses en énergie que les autres. L'exploration a été réalisée sur la surface de potentiel. La structuration autour du sodium est telle que six des huit groupements carbonyle sont attachés au sodium. Ces deux structures d'octaglycine sodée sont caractérisées par une liaison hydrogène atypiquement forte. L'état de transition correspondant à cette interconversion est lui-même bas en énergie, suffisamment pour que la transition structurale, dans un sens comme dans l'autre, soit *a priori* rapide à température ambiante. Le spectre de l'isomère à solvation de charge, calculé au niveau MP2/SVP, est en bon accord avec le spectre IRMPD, alors qu'aucun des spectres B3LYP-D ne concorde avec l'expérience.

Pour des systèmes de taille plus importante, il convient de pouvoir utiliser une méthode moins lourde que MP2 et en laquelle on puisse avoir confiance, même dans des cas atypiques comme celui de $G_8\text{-Na}^+$. Dans l'article p. 153, nous comparons les résultats obtenus avec différentes fonctionnelles publiées ces quelques dernières années, avec le spectre expérimental et les résultats obtenus au niveau MP2/SVP. Notre but était d'établir quelle(s) fonctionnelle(s) sont capables de fournir de bonnes fréquences au moins pour les bandes amide A, I et II, que les fonction amides soient libres ou engagées dans des liaisons hydrogène. Nous cherchions aussi à savoir comment se comportent les différentes fonctionnelles vis-à-vis des deux structures présentées. La plupart des fonctionnelles testées se sont avérées incapables de reproduire les deux *minima* voisins. Seules B3LYP-D et la série des M06 (à l'exception de M06-2X) s'en sont avérées capables. Finalement, après MP2, c'est M06-L qui permet l'obtention du spectre le plus proche de l'expérimental.

Forts du bon accord entre calculs MP2 et expériences obtenu sur $G_8\text{-Na}^+$, nous avons étudié la série complète des $G_n\text{-Na}^+$ (avec n de 2 à 8). Des spectres IRMPD avaient été obtenus au centre CLIO à Orsay dans la gamme spectrale 1000-2000 cm^{-1} pour l'ensemble de ces complexes. Des spectres, parfois incomplets, ont ensuite été obtenus dans la gamme 3000-3700 cm^{-1} pour $n = 2,3,6-8$. Pour les modélisations, la méthodologie REMD/AMOEBa mise en œuvre pour $G_8\text{-Na}^+$ a été étendue à $G_6\text{-Na}^+$ et $G_7\text{-Na}^+$, les peptides plus petits ayant été explorés par des travaux antérieurs. Les résultats indiquent que jusqu'à 6 résidus, les peptides présentent tous leurs oxygène d'amide vers le cation pour maximiser l'énergie d'interaction cation-peptide. $G_6\text{-Na}^+$ a une taille suffisante pour que cet enroulement permette une liaison hydrogène directe entre ses extrémités. Des transitions spectrales par rapport aux peptides plus petits sont visibles dans les deux gammes utilisées, correspondant à cette interaction. À partir de $G_7\text{-Na}^+$, une structuration interne du peptide commence à s'établir et la structure devient plus nettement tridimensionnelle. Des coudes γ et β permettent de combiner liaisons hydrogène et interactions peptide-sodium.

Chapitre 6

**Structure et énergétique de peptides
attachés à un cation métallique :
pertinence d'un champ de force
polarisable**

Structural, energetic and dynamical properties of sodiated oligo-glycines: relevance of a polarizable force field

David Semrouni, Gilles Ohanessian and Carine Clavaguéra*

Received 18th November 2009, Accepted 10th February 2010

First published as an Advance Article on the web 26th February 2010

DOI: 10.1039/b924317h

Oligo-glycine peptides (from two to ten residues) complexed to the sodium ion were studied by quantum chemical and molecular mechanics calculations to understand their structural and energetic properties. Modeling such systems required the use of a polarizable force field and AMOEBA, as developed by Ren and Ponder [*J. Comput. Chem.*, 2002, **23**, 1497], was chosen. Some electrostatic and torsional parameters were re-optimized using a rigorous procedure and validated against both geometric and energetic *ab initio* data in the gas phase. Molecular dynamics simulations were performed on seven sodiated octa-glycine (G_8) structures. Structural transitions were generally observed (with the notable exception of the α -helix), leading to new structures that were further proved by *ab initio* calculations to be of low energies. The main result is that for $G_8\text{-Na}^+$, there is a compromise between sodium peptide interactions and multiple hydrogen bonding. The accuracy achieved with AMOEBA demonstrates the potential of this force field for the realistic modeling of gaseous peptides.

1. Introduction

Sodium is one of the most important metal ions in biological processes. It is well-known for its role in ionic channels to regulate energy metabolism through transmembrane concentration gradients. Na^+ also plays a structural role by stabilizing proteins by electrostatic effects such as charge–charge, charge–dipole and charge– π interactions.¹ Long-range effects are observed, for instance when the metal charge induces the stabilization of a helical motif by interaction with its macro-dipole. Although these interactions are affected by the solvent in biological conditions, it is important to obtain a detailed picture of the intrinsic properties of Na^+ in its local environment. Gas phase experiments have been instrumental in establishing the energetics of sodium interactions with organic molecules and amino acids.^{2–8} More recently, sodium–peptide interactions in the gas phase have started to be unraveled as well, using a variety of mass spectrometric methods.^{9–18} An interesting feature of such studies is that observable experimental data such as binding energies or infrared vibration frequencies can be directly compared to computed results, most often obtained at the post-Hartree–Fock or DFT level. This combination of mass spectrometry and quantum chemistry has been most successful for small species; however, it cannot tackle biomolecular properties in general. In order to obtain reliable theoretical results on peptides with more than 5–10 residues, other approaches must be used. Indeed, quantum calculations are limited by the size of the molecules, the complexity of the potential energy surfaces and the harmonic approximation in the computation of vibrational frequencies. While molecular mechanics has been developed to deal with

larger cases, usual biological force fields have been proven to be inadequate to obtain an accurate description of peptides interacting with a metal cation.^{19–23} A force field including electrostatic energy based on multipole moments and many-body effects is required to reproduce both the structural and the energetic data of such systems.^{22,24–27} Several force fields were tested for reproducing amino-acid conformational energies by Jensen *et al.*²⁸ The force fields including an electrostatic energy based on partial charges only are able to reproduce about half of the conformations identified using accurate quantum mechanical methods, while the AMOEBA polarizable force field^{24,29} can reach 80%. It is the purpose of this paper to describe our recent progress along such lines.

The present study deals with oligo-glycine peptides attached to a sodium ion. The number of residues ranges from two to ten and the systems are termed $G_n\text{-Na}^+$ ($n = 2, 3, 4, 8$ and 10) herein. Peptides with up to ten residues provide a compromise, maintaining both realistic structural complexity and the ability to perform high level quantum chemistry calculations as references. Single-point coupled-cluster calculations were carried out at the CC2 level for all peptide sizes including $G_{10}\text{-Na}^+$. This allows us to establish the reliability of the MP2 level for future investigations on still larger cases. The polarizable AMOEBA force field^{22,24,29} was adapted to study sodiated peptides in the gas phase and extraction of new torsion parameters for the backbone is described. The AMOEBA force field was shown³⁰ to reproduce accurately bulk water properties and hydration of K^+ , Na^+ , Cl^+ and their salts. It is expected to be well suited for describing hydrated biomolecules as well. Herein we seek to extend its range of application to gaseous charged peptides in order to interpret mass spectrometry results. An extensive validation of these parameters is reported on sodium–peptide interaction energies for two, three and four residues. Geometries and relative energies of several conformers are also considered in detail

Laboratoire des Mécanismes Réactionnels, Département de Chimie, Ecole Polytechnique, CNRS, 91128 Palaiseau Cedex, France.
E-mail: carine.clavaguera@dmr.polytechnique.fr

for all peptide sizes. In this manner the performances of AMOEBA can be evaluated on structures which have been established to be of low energies by other methods. We wanted to set out the challenge of exploring potential energy surfaces extensively, rather than limiting the evaluation of AMOEBA performances to a limited set of pre-defined structures. To this aim, long molecular dynamics simulations using AMOEBA were carried out to follow the time evolution of seven structures of $G_8\text{-Na}^+$ at room temperature. This leads to new stable structures being predicted, and in turn *ab initio* calculations are carried out to check the accuracy of the AMOEBA relative energies.

2. Methods

2.1 Quantum chemistry calculations

2.1.1 Structures and energetics. For all but the smaller cases, determination of the low energy structures must rely on a systematic procedure. We have used Monte Carlo sampling and the Amber force field as described elsewhere.^{13,14} Because the Amber energetics is unreliable for such systems, quantum chemical calculations had to be run for numerous structures. DFT B3LYP/6-31G(d) geometry optimizations were performed and for the several most stable conformers, vibrational frequencies were calculated at the same level to confirm that the structures are true minima on the potential energy surfaces. Single point energies were obtained at the B3LYP/6-311+G(2d,2p) level. These calculations were performed using the Gaussian03 package.³¹ For each $G_n\text{-Na}^+$, the low energy subset of these structures was selected for post-Hartree-Fock calculations and AMOEBA study.

Following the logic of the AMOEBA force field and its parameter set determination, MP2 energies were used as references. To this end, geometry re-optimizations were performed at the RI-MP2/SVP level and single point energy calculations were obtained both at RI-MP2/TZVPP and RI-CC2/TZVPP levels. These calculations were done for $G_2\text{-Na}^+$, $G_3\text{-Na}^+$ and $G_4\text{-Na}^+$. For $G_8\text{-Na}^+$ and $G_{10}\text{-Na}^+$, single point RI-MP2/TZVPP and RI-CC2/TZVPP energy calculations were performed at the B3LYP geometries. The post-Hartree-Fock calculations were performed with the Turbomole 5.10 package^{32,33} to take advantage of the resolution-of-the-identity (RI) approximation implemented. The coupled-cluster CC2 level is expected to provide an accurate reference against which all other methods can be compared.³⁴⁻³⁶

All these quantum chemistry calculations provide a database of geometries, sodium-peptide interaction energies and relative energies, with the aim to validate the parameters of the AMOEBA force field. RI-MP2 and RI-CC2 give most of the time comparable results on energies while DFT/B3LYP results can be sometimes far from the reference RI-CC2 ones.

2.1.2 DFT partitioning of the binding energy. The Morokuma-type decomposition of the binding energy into steric and orbital interaction terms³⁷ was used to obtain a detailed picture of the sodium-peptide interactions in the $G_2\text{-Na}^+$ and $G_8\text{-Na}^+$ cases. For the sake of efficiency, the calculations were done using the TZ2P Slater basis sets and the BLYP functional in the ADF 2007.01 package.³⁸ For the $G_2\text{-Na}^+$ structures, the BLYP

binding energies with respect to the $(G_2 + \text{Na}^+)$ ionic dissociation limit are given in Table 1. They are close to the B3LYP, RI-MP2 and RI-CC2 interaction energies given in Table 2. The steric contribution corresponds to the sum of electrostatic and Pauli repulsion contributions. The part of the orbital interaction energy is far from being negligible, about 35% of the total binding energy conferring a non-simple electrostatic character to the interaction between the peptide and the sodium ion. The Voronoi deformation density (VDD) charges were calculated in order to evaluate the flow of electron density from the peptide fragment to the ion due to the interaction.³⁹ For the sodium ion, the VDD is about $-0.07 e$, *i.e.* there is only a small charge transfer to the peptide. Accordingly, for such non-covalent interactions, the main orbital DFT contribution is likely to arise from polarization effects. This result gives confidence in the use of the AMOEBA polarizable force field without an explicit charge transfer term for the study of peptides complexed to the sodium ion. This many-body effect is implicitly taken into account by the other components of the binding energy.

2.2 The AMOEBA force field

2.2.1 Choice of the polarizable force field. The AMOEBA polarizable force field was chosen for its good performances in reproducing cluster, liquid and solid bulk properties, and because its numerical cost remains relatively modest. It has previously been used successfully to determine solvation free energies for K^+ , Na^+ , Cl^- , Ca^{2+} and Mg^{2+} ions in liquid water and formamide,^{30,40} to calculate trypsin-ligand binding free energies from explicit and implicit solvent simulations,⁴¹ to investigate the role of polarization effects in the alanine dipeptide,²⁹ and to establish the role of polarization effects on energies in oligoacene crystals.⁴² In the context of the present study in the gas phase, AMOEBA also has the appealing feature that it is entirely based on *ab initio* data, avoiding the need for experimental information that is not available.

The analytical form of the force field is based on second-order exchange-perturbation theory of intermolecular forces. The electrostatic component incorporates not only a fixed partial charge, but also a dipole and a quadrupole on each atom as derived from quantum mechanical calculations.²² Many-body polarization effects are explicitly treated using a self-consistent dipole polarization procedure.²⁹ Repulsion-dispersion interactions between pairs of nonbonded atoms are represented by a buffered 14-7 potential. The intramolecular

Table 1 Morokuma decomposition of the sodium-peptide interaction energies (kcal mol^{-1}) computed at the BLYP/TZVPP level. BE stands for G_2/Na^+ and G_8/Na^+ binding energies computed without relaxation of the G_2 and G_8 structures at dissociation^a

| | BE | Steric | Orbital Interaction | %Orb |
|----------------------|--------|--------|---------------------|------|
| $G_2\text{-tri}$ | -63.2 | -40.7 | -22.5 | 36 |
| $G_2\text{-II}$ | -59.2 | -40.7 | -18.5 | 31 |
| $G_2\text{-IV}$ | -54.7 | -37.6 | -17.2 | 31 |
| $G_8\text{-2}$ | -99.9 | -61.6 | -38.3 | 38 |
| $G_8\text{-compact}$ | -116.9 | -74.6 | -42.2 | 36 |
| $G_8\text{-helix}$ | -98.6 | -64.6 | -34.1 | 35 |

^a See Fig. 4 and 7.

Table 2 Interaction energies in kcal mol⁻¹ between the sodium ion and the peptide for G₂-Na⁺, G₃-Na⁺ and G₄-Na⁺ (kcal mol⁻¹).^a The peptide structure is the same as that in the sodium complex at equilibrium

| E_{int} | B3LYP ^b | MP2 ^c | CC2 ^c | AMOEBA ^d |
|-------------------------|--------------------|------------------|------------------|---------------------|
| G ₂ _tri | -64.2 | -63.1 | -62.8 | -62.4 |
| G ₂ _II | -59.9 | -57.4 | -57.0 | -56.5 |
| G ₂ _IV | -56.2 | -55.3 | -55.1 | -53.9 |
| G ₃ _3s | | -73.8 | -73.1 | -76.7 |
| G ₃ _3sb | | -74.3 | -73.7 | -76.0 |
| G ₃ _3s_ring | | -70.1 | -69.8 | -72.7 |
| G ₃ _4s | | -82.0 | -81.5 | -82.5 |
| G ₄ _4s | | -91.0 | -90.3 | -94.9 |
| G ₄ _5s | | -97.5 | -96.9 | -99.5 |
| G ₄ _5sb | | -97.5 | -95.8 | -97.2 |
| G ₄ _3s_ring | | -78.9 | -78.4 | -73.0 |

^a See Fig. 4–6. ^b At B3LYP geometries. ^c At MP2 geometries. ^d At AMOEBA geometries.

valence terms consist in bond stretchings, angle bendings and torsions with the MM3 force field energy form, including anharmonicity effects with higher-order deviations from ideal bond lengths and angles.⁴³ Additional valence terms are used to model the coupling between stretching and bending, with a Urey–Bradley functional form.

Several studies have pointed out that quantum effects are important for strong hydrogen bonds, in liquid water for example.^{44,45} Such effects may be implicitly included in empirical force field simulations. The role of quantum effects was highlighted by path integral molecular dynamics simulations associated to a polarizable force field by Paesani and Voth⁴⁵ on macroscopic properties such as density or enthalpy of vaporization.

2.2.2 Parameter adjustment. We used most of the AMOEBA parameters for the glycine residue and the sodium ion. However, some adjustments were necessary to study non-zwitterionic species in the gas phase^{9–17} with neutral COOH and NH₂ terminations. For the COOH group, we chose the parameters of the carboxylic group of the aspartic acid already available in the AMOEBA parameter set. For the NH₂ terminal group, we have used the multipole and van der Waals parameters of the N-terminal ammonium group. Quantum chemistry calculations at MP2 level with various basis sets including triple- ζ basis sets were performed on the ethylamine model. Based on Mulliken charges, the charges were set to -0.75 for the nitrogen atom and 0.35 for each hydrogen as the best compromise for a good accuracy on interaction energies (see Table 2). At this stage, several problems remained to reproduce *ab initio* geometries and energies of G₂-Na⁺. A detailed analysis revealed that these problems arose from torsions.

The backbone geometry of the peptide chain is a function of three important dihedral angles:

ψ : N-C _{α} -C-N to control the distance between the nitrogen atoms of two adjacent residues;

ϕ : C-N-C _{α} -C to control the distance between adjacent carbonyl groups, which is a more constrained torsion;

ω : C _{α} -C-N-C _{α} , the dihedral angle around the peptide bond.

The N_{ter}-C _{α} -C-N and C-N-C _{α} -C_{ter} dihedral angles must be distinguished from the general dihedrals ψ and ϕ because they involve different atom types in the peptide terminations, while ω is always the same. We performed an analytic optimization of the parameters of two specific torsions: the N-C _{α} -C-N (ψ backbone) and the N-C _{α} -C_{ter}-OH of the C-terminal residue torsions. Note that the ψ torsional parameters were set to zero in the original force field.

The adjustment of each of the ψ and N-C _{α} -C_{ter}-OH torsion parameters was done in the following procedure by first building reference potential energy curves at RI-MP2/TZVPP level in the case of the G₃ peptide model. To evaluate the role of diffuse functions, a number of energy calculations were also performed at the RI-MP2/aug-cc-pVTZ level. As displayed later in Fig. 2 and 3, the energetic profiles are similar at both levels of calculations. Each potential energy curve was built for the same G₃_3s conformer of G₃-Na⁺ without the sodium ion (see Fig. 5a) by varying the corresponding dihedral angle from -175 to 180°: nine regularly spaced points were calculated around each of the -120, -60, 0, 60, 120 and 180° points of the potential. This is necessary to have a good representation of the energetic profile not only for the G₃_3s conformer but also for various conformers at different peptide sizes.

Fig. 1 illustrates the repartition of the N-C _{α} -C-N torsion values for the G _{n} -Na⁺ conformers presented in this work. It is clear that there are clustering regions around *e.g.* -60°, however the majority of the points lie outside these regions. Due to this broad distribution, from -180 to 180°, a fit on the energetic profile is a better way than a fit on energy minima to ensure the transferability of the parameters. Then, an analytical parameter optimization was independently performed for the N-C _{α} -C-N and N-C _{α} -C_{ter}-OH torsions. Consequently, the peptide geometry was not relaxed in order to avoid AMOEBA vs. MP2 structural differences. As these potential energy curves are not directly usable in an analytic optimization process, each reference curve has been written so that it only represents the corresponding energetic torsion component. This way the analytic expression of the torsion energetic component will be used straightforwardly to reproduce the *ab initio* reference potential. We call $E_{\text{component}}$ the difference between RI-MP2/TZVPP and AMOEBA_torsionless relative energies (eqn (1)):

$$E_{\text{component}} = E_{\text{MP2}} - E_{\text{torsionless}} \quad (1)$$

Assuming that MP2 and AMOEBA are in good agreement for all terms besides this torsion, the difference essentially reduces to this torsion's contribution. In the AMOEBA formalism, as in the MM3 formalism, the dihedral potential energy function is implemented as a truncated Fourier series (eqn (2)):

$$E_{\text{torsion}} = \frac{1}{2} \sum_{n=1}^3 V_n [1 + \cos(n\alpha_i - \Phi_i)] \quad (2)$$

At this point, V_1 , V_2 and V_3 were optimized for E_{torsion} to fit $E_{\text{component}}$. This way, the AMOEBA potential energy should get closer to the RI-MP2/TZVPP one. A nonlinear least squares regression has been performed in order to optimize these parameters. A Levenberg–Marquardt algorithm^{46,47} was used as it is fast and efficient in curve fitting.

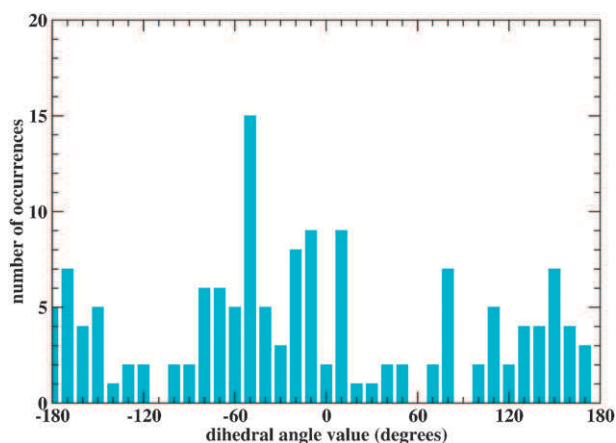


Fig. 1 N-C $_{\alpha}$ -C-N dihedral repartition on the most stable conformers of G $_n$ -Na $^{+}$.

The first optimized torsion was the N-C $_{\alpha}$ -C-N one as defined above. The RI-MP2 energetic profile (Fig. 2) presents a high energetic barrier corresponding to a seven-membered C $_7$ motif with a very short contact (1.5 Å) because the rest of the geometry is not relaxed. We anticipate it very hard to reproduce such an interaction and a perfect agreement in this area is not a necessity for the fit. The V_2 parameter was fixed to zero in the optimization procedure to avoid favoring this problematic area. This constraint is adequate since there is no bond conjugation along the N-C $_{\alpha}$ -C-N dihedral. It is important to mention that the rest of the energy profile corresponds to the most recurrent cases in the peptide structures. V_1 and V_3 parameters were optimized within the [-5; +5] kcal mol $^{-1}$ interval to avoid unphysical solutions. The final values are -4.838 and 0.685 kcal mol $^{-1}$ for V_1 and V_3 , respectively. The resulting energetic profile exhibits a clear improvement (see Fig. 2). The same procedure was applied to the fit of the N-C $_{\alpha}$ -C $_{\text{ter}}$ -OH torsional parameters. Fig. 3 shows a smaller magnitude of energy variation than for the N-C $_{\alpha}$ -C-N one. The three parameters were limited to the [-5; +5] kcal mol $^{-1}$ interval during the procedure and the optimized values are -1.276, 4.256 and -0.637 kcal mol $^{-1}$, respectively. By homology with the N-C $_{\alpha}$ -C-N N $^{-1}$ -C $_{\alpha}$ -C-O torsion pair, the parameters of the N-C $_{\alpha}$ -C $_{\text{ter}}$ -O $_{\text{ter}}$ torsion, associated to the N-C $_{\alpha}$ -C $_{\text{ter}}$ -O torsion, were set to zero.

3. Results

In the previous section, the parameters were validated by comparison to *ab initio* results on the sample of conformations composing the reference MP2 potential energy surface. Now it is crucial to validate the parameters on new configurations on both energetic and geometric criteria. In this work, the sodium-biomolecule interactions are characterized in the case of sodiated-oligoglycines G $_n$ -Na $^{+}$, $n = 2, 3, 4, 8$ and 10 . For $n = 2-4$, the low energy structures are mostly determined by peptide chelation by Na $^{+}$. These relatively simple cases are used for detailed calibration of the force field. For $n = 8$ and 10 , the potential energy surfaces are highly complex and even the global characteristics of the lowest energy structures can no longer be guessed qualitatively. From the perspective of

demonstrating the force field performance, this complexity may be exploited in a two-pronged strategy: first, the force field and *ab initio* results are compared on a set of structures whose stability has been inferred independently of the force field. Second, the force field is used for an extensive, unbiased exploration of the potential energy surface, and the validity of its predictions of other stable structures is directly assessed by subjecting them to *ab initio* calculations.

Due to the strong attachment of the sodium ion to the peptide, distances of the chelating atoms to Na are good structural indices and will be used below together with hydrogen bonds. Residues are numbered from the N to the C-terminus and the atom number corresponds to the number of the related residue. Hydrogen atoms are labelled from the atom to which they are covalently bound. OH and HO correspond to O and H atoms in the C-terminal hydroxyl group, respectively. HN1a and HN1b are the two hydrogens of the N-terminus amine.

3.1 Structures and energies

Total interaction energies are reported in Table 2 to quantify the G $_2$, G $_3$ and G $_4$ peptide interaction with the Na $^{+}$ ion. AMOEBA interaction energies are close to MP2 and CC2 ones thanks to the combination of a high level description of the intramolecular interactions in the force field, a multipolar distribution for the electrostatic component and an explicit self-consistent polarization term. The global error on the interaction energy by comparison with CC2 is smaller than 5% except in the G $_4$ -3s $_r$ ring system for which the error is about 7% due to large changes in the geometry as discussed below. Beyond this essential validation, the force field is also required to reproduce relative energies and geometries for all peptide sizes. This is discussed below for each peptide size. Relative energies of G $_n$ -Na $^{+}$ structures are given in Table 3. The AMOEBA energies correspond to optimized geometries with the AMOEBA force field from a MP2 structure for the G $_2$ -Na $^{+}$, G $_3$ -Na $^{+}$ and G $_4$ -Na $^{+}$ systems and a B3LYP structure for the G $_8$ -Na $^{+}$ and G $_{10}$ -Na $^{+}$ systems.

G $_2$ -Na $^{+}$. The sodium ion favors electrostatic interactions by binding to the polar carbonyl groups and the terminal amine to a lesser extent. Three conformers of low energy were identified in ref. 12 and 48 and are represented in Fig. 4. The G $_2$ -tri conformer (Fig. 4a) has a coordination of sodium with the two oxygens and the terminal amine (O1, O2 and N1). In structure G $_2$ -II (Fig. 4b), the sodium ion is coordinated only to the two oxygens (O1 and O2), while the N atom of the terminal amine interacts with the hydrogen of the peptide amide forming a five-membered C $_5$ motif. Finally, in conformer G $_2$ -IV (Fig. 4c), the sodium is coordinated by the terminal amine and the peptidic carbonyl (N1 and O1). The second residue exhibits a hydrogen bond between the amide hydrogen and the carbonyl oxygen of the acid group forming a C $_5$ cycle. The three geometries are well reproduced with the AMOEBA force field, especially for the distances between Na and the coordinated oxygens (see Fig. 4). The G $_2$ -tri conformer is the most stable at the CC2 level (0.7 and 3.5 kcal mol $^{-1}$ more stable than G $_2$ -II and G $_2$ -IV, respectively). The AMOEBA force field overstabilizes the G $_2$ -II conformer by

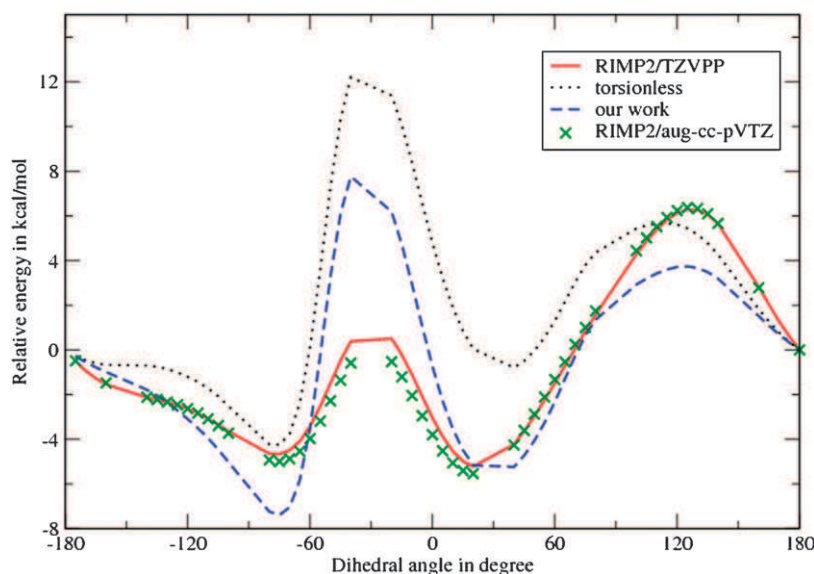


Fig. 2 Relative energetic profiles for G_3_3s along the $N-C_\alpha-C-N$ dihedral angle without relaxation (RI-MP2 in solid lines, torsionless in dotted lines and present work in dashed lines).

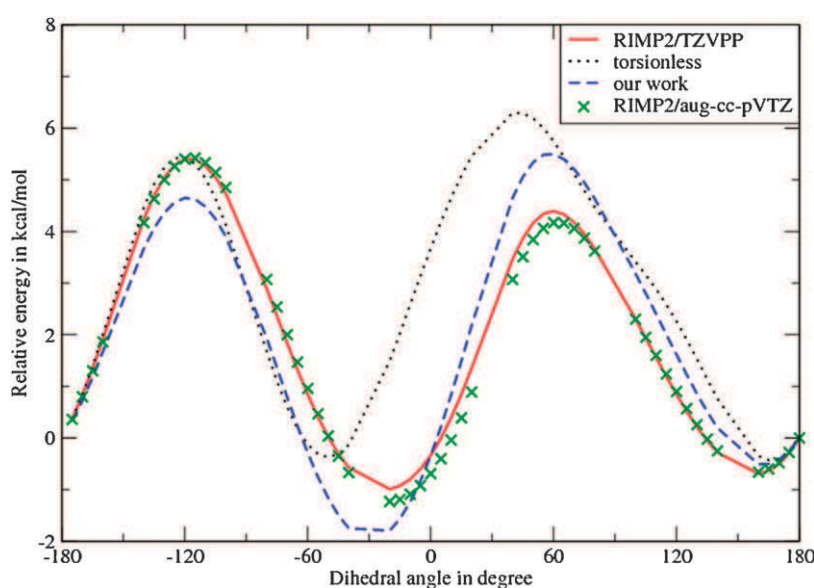


Fig. 3 Relative energetic profiles for G_3_3s along the $N-C_\alpha-C_{ter}-OH$ dihedral angle without relaxation (RI-MP2 in solid lines, torsionless in dotted lines and present work in dashed lines).

$0.9 \text{ kcal mol}^{-1}$. This behavior is similar at the DFT/B3LYP level. Furthermore, the G_2_{IV} structure is overstabilized by $6.5 \text{ kcal mol}^{-1}$ in AMOEBA. This is partially due to an uncommon energetic contribution of the $C=O$ bond-stretching term in the carboxyl group. In the MP2 structure, this $C=O$ bond distance is shorter by 0.038 \AA than its AMOEBA ideal value. After the AMOEBA geometry optimization, the difference is reduced to 0.001 \AA , allowing for a stabilization of $1.1 \text{ kcal mol}^{-1}$ for this single bond-stretching term. The energy decreases during the AMOEBA optimizations (starting from the MP2 structures) are 3.5 , 0.9 and $1.2 \text{ kcal mol}^{-1}$ for G_2_{IV} , G_2_{II} and G_2_{tri} , respectively. This energetic phenomenon can be considered as an exception in our set of peptides.

G_3-Na^+ . Low energy structures were taken from ref. 14. As for the G_2-Na^+ conformers, the G_3-Na^+ structures maximize the peptide-ion electrostatic and polarization interactions with all three carbonyls coordinated to sodium. In addition, the terminal amine is bound to Na^+ in the G_3_4s conformer (Fig. 5c). For G_3_3s and G_3_3sb structures (Fig. 5a and b), the difference lies in the orientation of the central $C=O$ bond interacting with the cation, relative to the other two. In both cases the carbonyls are oriented towards Na^+ , but the difference of conformation leads to a more compact structure for G_3_3sb . Conformer $G_3_3s_{ring}$ also bears three carbonyl-sodium interactions (Fig. 5d), yet it has a more compact structure since the C-terminal acid is a hydrogen bond donor towards the N-terminal amine. The three carbonyls are

Table 3 Relative energies in kcal mol⁻¹ of G_n-Na⁺ structures, with root mean square deviation in kcal mol⁻¹ (RMSD) for each size of peptide

| Method | RI-CC2 ^a | RI-MP2 ^a | B3LYP ^b | AMOEBA ^c |
|-------------------------|---------------------|---------------------|--------------------|---------------------|
| G ₂ _tri | 0.0 | 0.0 | 0.0 | 0.0 |
| G ₂ _II | 0.7 | 0.5 | -1.3 | -0.2 |
| G ₂ _IV | 3.5 | 3.3 | 1.6 | -3.0 |
| RMSD | 0.0 | 0.1 | 1.4 | 3.3 |
| | | | | |
| G ₃ _3s | 0.0 | 0.0 | 0.0 | 0.0 |
| G ₃ _3sb | -0.8 | -0.7 | 0.4 | -3.0 |
| G ₃ _3s_ring | 0.8 | 2.0 | 7.5 | -2.5 |
| G ₃ _4s | -1.3 | -0.8 | 1.6 | 1.1 |
| RMSD | 0.0 | 0.4 | 2.5 | 0.6 |
| | | | | |
| G ₄ _4s | 0.0 | 0.0 | 0.0 | 0.0 |
| G ₄ _5s | 0.2 | 0.6 | 3.3 | 1.6 |
| G ₄ _5sb | 0.2 | 0.5 | 3.1 | 0.5 |
| G ₄ _3s_ring | 0.0 | 1.4 | 5.5 | 1.8 |
| RMSD | 0.0 | 0.5 | 2.1 | 0.6 |
| | | | | |
| G ₈ _1 | 0.0 | 0.0 | 0.0 | 0.0 |
| G ₈ _2 | -5.6 | -5.1 | -1.7 | 1.2 |
| G ₈ _3 | 4.1 | 3.4 | -0.5 | -11.2 |
| G ₈ _4 | -0.4 | 0.0 | 2.8 | 1.0 |
| G ₈ _5 | -0.9 | -1.6 | -0.6 | -6.9 |
| G ₈ _compact | 8.2 | 6.8 | 4.1 | 5.6 |
| G ₈ _helix | -2.5 | -2.7 | 3.2 | -13.7 |
| RMSD | 0.0 | 0.3 | 1.6 | 3.5 |
| | | | | |
| G ₁₀ _g1 | -0.5 | -1.1 | -3.0 | -0.6 |
| G ₁₀ _g2 | 0.0 | 0.0 | 0.0 | 0.0 |
| G ₁₀ _g3 | 6.0 | 5.1 | 1.5 | 4.5 |
| G ₁₀ _h1 | 1.5 | 0.3 | 1.4 | -6.3 |
| G ₁₀ _h2 | 4.0 | 4.1 | 1.5 | -8.5 |
| G ₁₀ _h3 | 7.1 | 6.6 | 1.0 | 4.0 |
| G ₁₀ _h4 | 3.6 | 2.3 | 1.6 | -2.2 |
| RMSD | 0.0 | 0.4 | 1.4 | 2.7 |

^a At MP2 geometries for G₂-Na⁺, G₃-Na⁺ and G₄-Na⁺, and at B3LYP geometries for G₈-Na⁺ and G₁₀-Na⁺. ^b At B3LYP geometries.

^c At AMOEBA geometries.

roughly parallel and oriented towards the metal. The G₃_3s is the reference one from which the energetic profiles were calculated to extract the parameters of the N-C_α-C-N and N-C_α-C_{ter}-OH torsions. The G₃_3s_ring AMOEBA structure re-organized to a structure close to G₃_3sb, differing only by a rotation of the N-C-C_α-N dihedral angle, giving a similar relative energy for both conformers (Fig. 5e). The G₃_3sb and G₃_4s relative energies deviate by about 2 kcal mol⁻¹ from CC2 references. The major consequence is that an energy splitting between the three more stable structures is observed in AMOEBA. However, for this peptide size, both geometries

and relative energies can be considered to be accurate enough to validate the new parameters.

G₄-Na⁺. Low energy structures were taken from ref. 14. The G₄-Na⁺ structures maximize the peptide-ion electrostatic and polarization interactions with up to all four carbonyls coordinated to sodium. This is the case for the G₄_4s conformer (Fig. 6a). In addition, the terminal amine is bound to Na⁺ in the G₄_5s and G₄_5sb conformers (Fig. 6b and c) that have similar geometries (close to being mirror images one of one another, with only a small shift for the N-terminal group). Conformer G₄_3s_ring bears only three carbonyl-sodium interactions and it has a more compact structure with a hydrogen bond between the acid and amine terminations (Fig. 6d). The three carbonyls are roughly parallel and oriented towards the metal. The four reference structures are energetically quasi-degenerate in MP2 and CC2 calculations and the most stable is G₄_4s. The global energetic and geometric behaviors are preserved with AMOEBA; the average error on relative energies is only 0.6 kcal mol⁻¹. After the AMOEBA geometry optimization step, the G₄_3s_ring keeps its characteristic conformation (Fig. 6e). However, the distances between O3, O4 and the sodium ion are increased due to the formation of a seven-membered C₇ motif between O2 and HN4. The formation of a O4-HN1 hydrogen bond is also observed. Overall, the relative energy of this structure is close to that of the reference *ab initio* structure due to energetic compensations. G₄-Na⁺ is the smallest peptide outside of the training set for new parameters. The quality of the matching between AMOEBA and CC2 results is strongly encouraging.

G₈-Na⁺. Low energy structures were obtained by an extensive conformational search as described in ref. 13 and 49. With eight glycine residues, the sodium ion can no longer accommodate binding to all carbonyl oxygens simultaneously. There is then a competition between sodium-peptide interactions and hydrogen bonding within the peptide and two general shapes can be investigated leading to globular and helix structures. One compact structure with six oxygen atoms connected to Na, five globular structures and one α -helix motif were included in the set (Fig. 7). As detailed in Table 3, most of the relative energies and geometries are well reproduced with the AMOEBA force field. The global deviation in energy is of 3.5 kcal mol⁻¹ and the largest energy errors are for the G₈_3 and G₈_helix structures. The most important differences between B3LYP and AMOEBA geometries are given below. In the G₈_1 conformer (Fig. 7a and b), dihedral variations

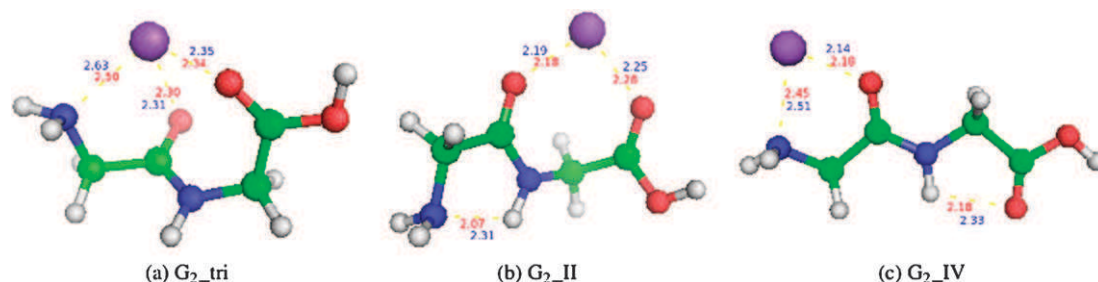


Fig. 4 G₂-Na⁺ structures (RI-MP2 distances in red, AMOEBA distances in blue).

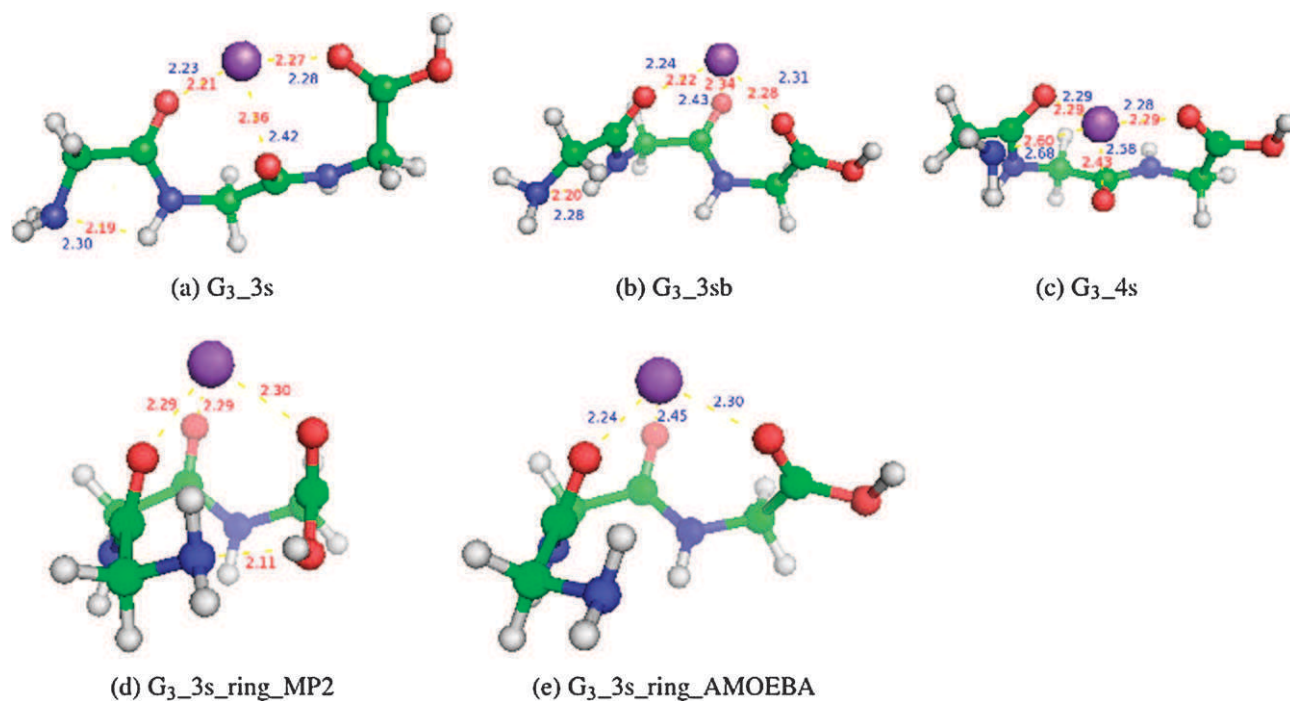


Fig. 5 $G_3\text{-Na}^+$ structures (RI-MP2 distances in red, AMOEB: distances in blue).

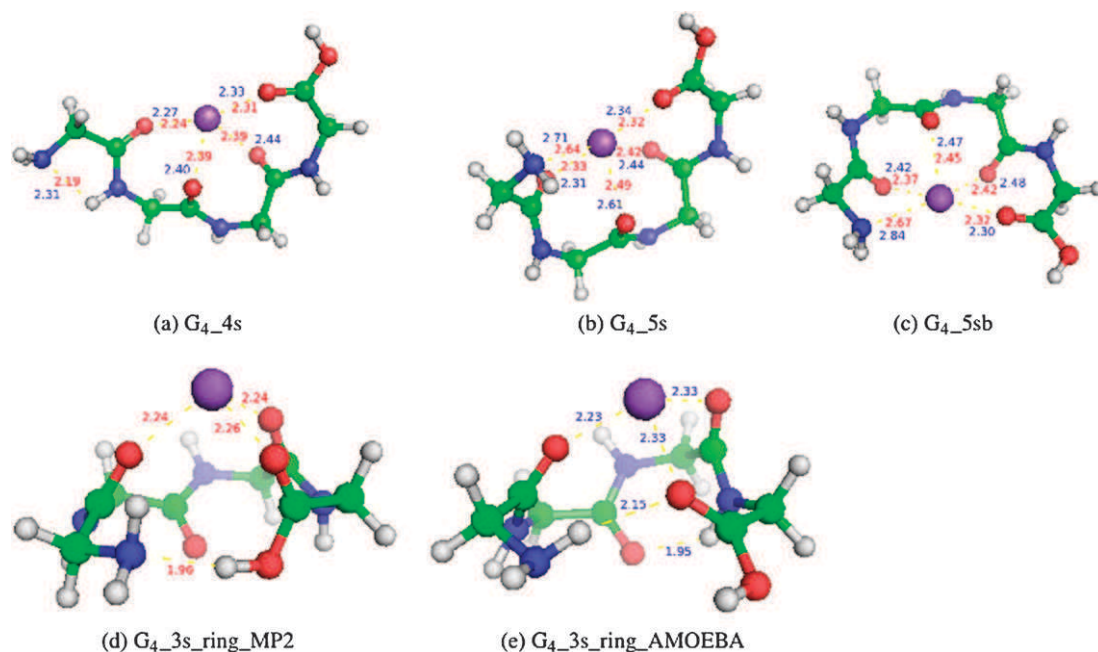


Fig. 6 $G_4\text{-Na}^+$ structures (RI-MP2 distances in red, AMOEB: distances in blue).

induce small changes in the attachment distances to the sodium ion and the formation of a hydrogen bond between HN1 and OH. The G_8_2 conformer geometry is well reproduced (Fig. 7c). The main changes hold in a shortened hydrogen bond: O2–HN4 by 0.17 Å and two elongated ones: O2–HO and HN2–N1 by 0.45 and 0.22 Å. The major change in the G_8_3 conformer (Fig. 7d and e) is the O1–OH hydrogen bond break to create a seven-membered C_7 motif between O1 and HN3. Another C_7 motif is created between O2 and HN4. The

four Na–O distances for the carbonyl groups connected to Na (O3, O6, O7 and O8 atoms) are shortened by 0.11 Å in average in the AMOEB: vs. B3LYP structures. Furthermore, O4 becomes linked to Na with a distance of 2.45 Å. For the G_8_4 (Fig. 7f) conformer, all the Na^+ coordination sites and all hydrogen bonds are maintained. These characteristic distances are globally shorter with AMOEB: except for the N1–Na and O1–HO ones. The geometry of the G_8_5 conformer (Fig. 7g) is preserved in spite of a coordination

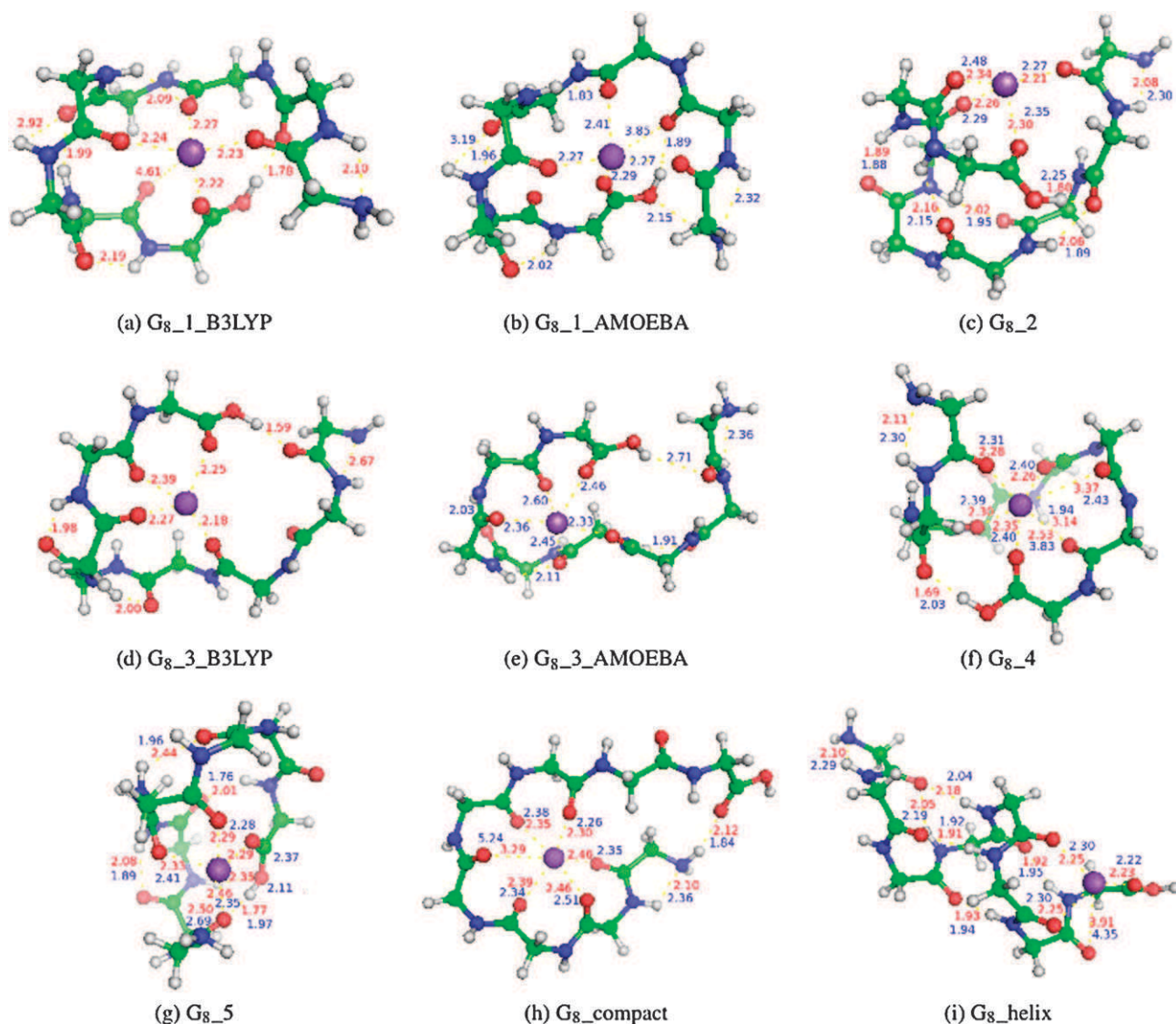


Fig. 7 $G_8\text{-Na}^+$ structures (B3LYP distances in red, AMOEBA distances in blue).

exchange to Na from O7 to O6. The O7 atom is then involved in a hydrogen bond with HN5. The geometry of the $G_8\text{-compact}$ structure (Fig. 7h) is well reproduced with five oxygen atoms coordinated to Na. In the $G_8\text{-helix}$ conformer (Fig. 7i), we see a rotation in the C9_N10_C10_C10 dihedral from -79 to -121 degrees, but this has no influence on the global α -helix structure.

Most structures are satisfactorily reproduced at the AMOEBA level. The multiple local interactions which may be built to stabilize a structure are sometimes slightly different, with *e.g.* a hydrogen bond replacing a sodium attachment site. Overall, the agreement between AMOEBA and CC2 or MP2 is rather good, within a few kcal mol^{-1} , except for two cases with errors in the $10\text{--}15 \text{ kcal mol}^{-1}$ range. The behavior of each structure will be also explored in molecular dynamics in detail in the last section.

$G_{10}\text{-Na}^+$. Low energy structures were obtained by an extensive conformational search as described in ref. 13 and 49. RI-MP2 and RI-CC2 energy calculations were performed

on B3LYP optimized geometries for eight structures including helix-type as well as globular shapes. $G_{10}\text{-g1}$, $G_{10}\text{-g2}$, $G_{10}\text{-g3}$ have globular shapes, $G_{10}\text{-h1}$, $G_{10}\text{-h2}$ and $G_{10}\text{-h3}$ are helix-motifs, $G_{10}\text{-h4}$ is a helix with a globular head. Most geometries are well reproduced with AMOEBA (Fig. 8). In the relative energies, global behavior is preserved with a RMSD of $2.7 \text{ kcal mol}^{-1}$ but one case, the $G_{10}\text{-h2}$ helix structure provides a large error of $12.5 \text{ kcal mol}^{-1}$. It can be noticed that some large errors are also provided by B3LYP calculations. The main differences between AMOEBA and B3LYP optimized structures are described below. The structure used as a reference in energy is $G_{10}\text{-g2}$ (Fig. 8b). In this conformer, the O1, O2, O8, O9 and O10 atoms are connected to Na in the B3LYP geometry but the O1 and O9 atoms are disconnected in AMOEBA (with Na–O distances of 2.57 vs. 3.58 \AA and 2.43 vs. 3.31 \AA , respectively). The N1–HN2, O1–HN4, O3–HN5, O4–HN6, O5–HN8, O6–HN10 and O7–HN9 hydrogen bonds are present in both B3LYP and AMOEBA structures and in addition, the O9–HN1 one is created in AMOEBA (Fig. 8c). As already seen for $G_8\text{-Na}^+$ structures,

there is a slightly different trade-off between sodium attachment sites and hydrogen bonding in the B3LYP and AMOEBA structures. In the G_{10_g1} conformer (Fig. 8a), all coordination sites to the sodium are preserved with AMOEBA: O3, O8, O9 and O10 with just a small increase in the distances (0.07 Å in average). Eight hydrogen bonds exist in the B3LYP structure: N1–HN2, O1–HN4, O2–HN5, O4–HN7, O4–HO, O5–HN8, O6–HN10 and O7–HN9, the O5–HN8 one is created in addition in AMOEBA. No global change is observed for the G_{10_g3} structure (Fig. 8d) with the O1, O3, O7 and O9 atoms connected to Na and the N1–HN2, O4–HN6, O5–HN7, O5–HN8, O7–HN9 and O8–HN5 hydrogen bond network. The G_{10_h1} has a structure well organized in an α -helix shape which is fully preserved in AMOEBA (Fig. 8e). The global geometry corresponds to the O7, O8 and O10 atoms in the sodium coordination sphere and the N1–HN2, O1–HN4, O1–HN5, O2–HN6, O3–HN7, O4–HN7, O5–HN9 and O6–HN10 hydrogen bond network. In the G_{10_h2} structure (Fig. 8f), the sodium ion is tetra-coordinated by N1, O3, O4 and O5 atoms with the O1–HN3, O2–HN8, O6–HN2, O7–HN9, O8–HN4, O8–HN5, O9–HN6 and O10–HN7 hydrogen bonds. Only small variations in these characteristic distances are observed after the AMOEBA geometry optimization step. In the G_{10_h3} conformer (Fig. 8g), O1, O3, O5 and O6 are connected to Na in B3LYP and AMOEBA structures, the major difference being an increase of the O3–Na distance in AMOEBA from 2.28 to 2.53 Å. The hydrogen bond network is preserved with the N1–HN2, O4–HN9, O7–HN3, O8–HN10, O9–HO, O10–HN7 and O10–HN6 bonds. Moreover, the O7–HN4 bond appears in AMOEBA. The O9–HO distance involved in a C_7 motif is very short in B3LYP (1.65 Å) and is not well reproduced with the force field (2.03 Å). The helix conformation of the G_{10_h4} structure is built on the N1–HN2, O1–HN4, O2–HN5, O2–HN6, O3–HN7, O6–HN9, O8–HN10 hydrogen bond network (Fig. 8h). In addition, the O4, O5, O7, O9 and O10 atoms are connected to Na in the B3LYP structure. The main difference in the AMOEBA geometry is the decoordination of O9.

Associated with these minor structural changes, AMOEBA energetics are generally in good agreement with reference coupled-cluster values, with only one large error of 12.5 kcal mol⁻¹.

3.2 Molecular dynamics simulations on G_8 -Na⁺

Molecular dynamics at finite temperature is an established way to explore potential energy surfaces. It was used in combination with the AMOEBA force field to extend the conformational search on the G_8 -Na⁺ peptide. Simulations were carried out with the TINKER software package⁵⁰ from all seven AMOEBA optimized structures previously presented. The Beeman algorithm was used for the propagation of dynamical trajectories. The simulations were performed at constant temperature (300 K) with a Berendsen thermostat and a 1 fs time step, for a total simulation time of 20 ns per trajectory. Configurations were recorded every 10 fs. For each G_8 -Na⁺ conformer, the thermalization period is about 10 ps. Fig. 9 and 10 display the variations of potential energy and of the Na–X (X = O or N) distances as a function of time.

The G_8 -helix structure in α -helix conformation remains stable along the entire simulation with three oxygen atoms (O5, O6 and O8) strongly connected to the sodium ion (the average distances are 2.37 Å, 2.33 Å and 2.27 Å, respectively) and a strong hydrogen bond network. Fig. 9 presents a flat potential energy behaviour which is preserved even if the simulation is extended to 40 ns. An electrostatic interaction is sometimes observed between the N-terminal amine group and HN4, but most of the time the N-terminus has a large mobility. The structure with this N1–HN4 interaction is metastable. It does not collapse during an AMOEBA optimization and exists for a few hundred of ps during the trajectory.

For the G_{8_1} conformer, the energy variations in Fig. 10 are associated with structural transitions which can be followed by considering the variations of Na–X distances shown in the upper part of the Figure. After the thermalization period, O1, O2, O3 and O8 are connected to the sodium ion with the HN1–O6, HN3–O4, HN7–O3, HN7–O4, HN8–O6 hydrogen bond network. HN3–O5 is gained after the breaking of the C_7 -motif previously binding HN5 to O3 and participates to the observed molecule “squeezing”. Three potential energy plateaus are observed along the simulation (see Fig. 10). The first transition is observed at 2 ns decreasing the potential energy by 11 kcal mol⁻¹ from the thermalized one. At this stage, O1, O2, O5 and O8 are connected to Na with the HN1a–O3, HN5–O1, HN5–O2, HN7–O4, HN6–O4, HN3–O7 and HN8–O6 hydrogen bonds. When O2 begins to be connected to Na, O3 is decoordinated. Large fluctuations are observed for the O6 and O7 atoms, and it can be considered that these two atoms are not in the coordination sphere of the sodium ion. The second structural transition is at 3.2 ns with a potential energy stabilization of 9 kcal mol⁻¹. At this stage, N1 is definitively linked to the Na atom and a concomitant stronger Na–O1 interaction is established. At the end of the simulation, there are interchanges between two almost identical structures with four or five atoms coordinated to Na⁺. The main difference between those two structures is the coordination of O5 to the sodium ion in only one of them. In both cases, N1, O1, O6 and O8 atoms are in the first coordination sphere of the cation with the HN1b–O5, N5–O2, HN7–O3, HN8–O6, HN3–O7 hydrogen bond network. The final structure looks more compact because of the coordination of N1 to sodium and the exchange of a hydrogen bond (HN1b–O7 is replaced by HN1b–O5). Similar major transitions were observed for the G_{8_2} , G_{8_3} , G_{8_4} , G_{8_5} and compact conformers to reach an energetically stable structure that is then kept along the rest of the trajectory. These five new conformations exhibit a more “compact” structure than their initial geometries, combining a globular shape with a hydrogen bond network in the N-terminus side.

The AMOEBA molecular dynamics simulations pointed out a new series of conformers for G_8 -Na⁺ and at the end of the trajectories, each reoptimized final conformation is lower in energy than the initial one, except for the helix that keeps its initial conformation. The energetic order was also changed: the new G_{8_1} conformation is now the lowest in energy, followed by the new G_{8_2} structure at 2.4 kcal mol⁻¹. The final validation step consisted in geometry reoptimizations at the RI-MP2 level to ensure that these structures are low in

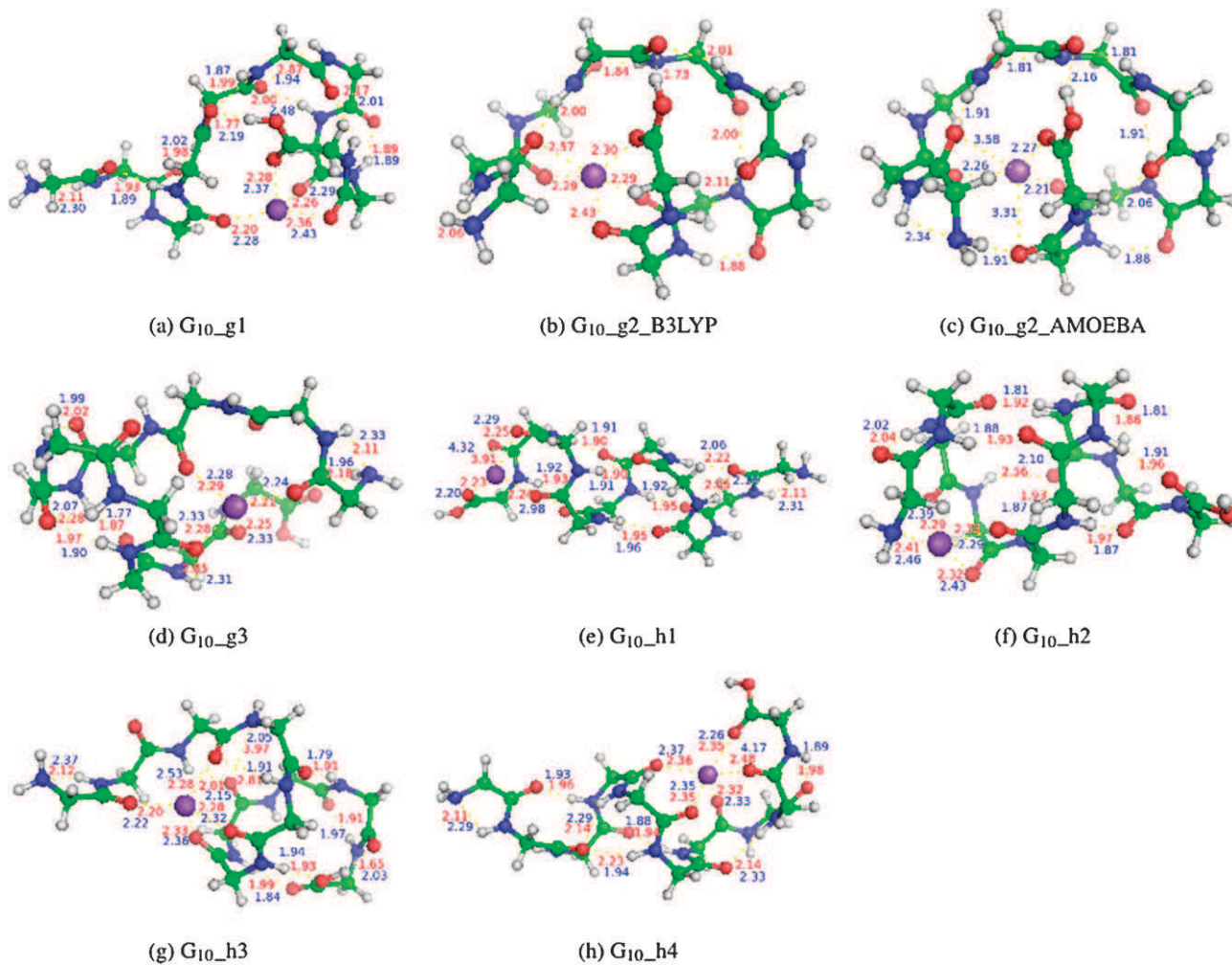


Fig. 8 G₁₀-Na⁺ structures (B3LYP distances in red, AMOEBA distances in blue).

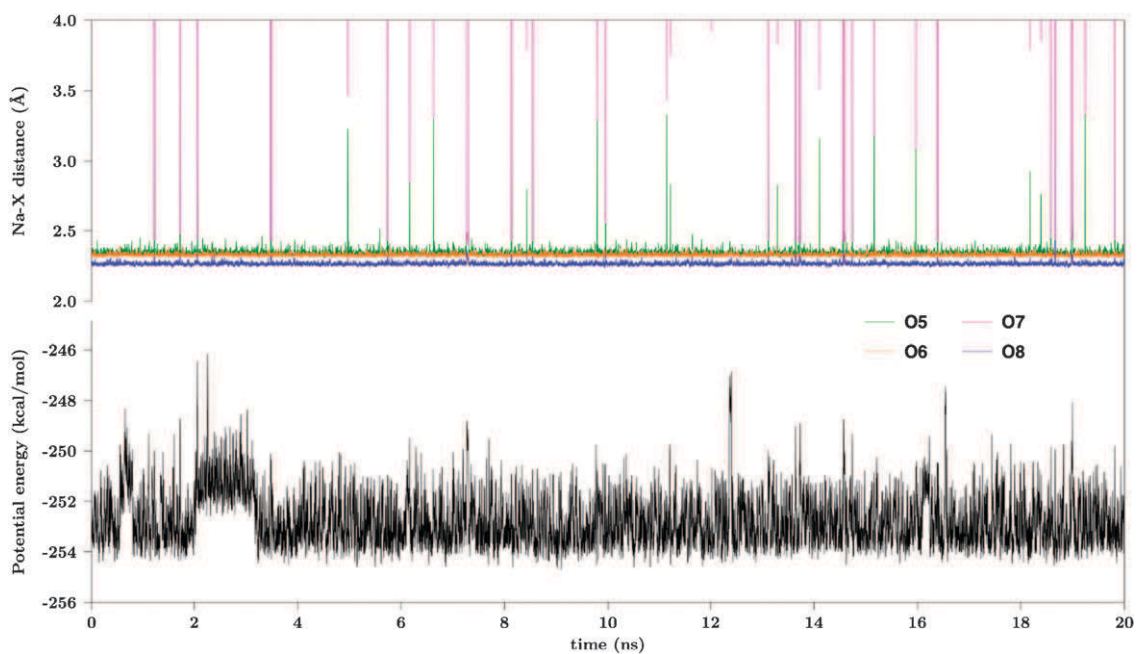


Fig. 9 Time evolution of the short-time averaged Na-X distances and of the potential energy for the G₈-helix conformer.

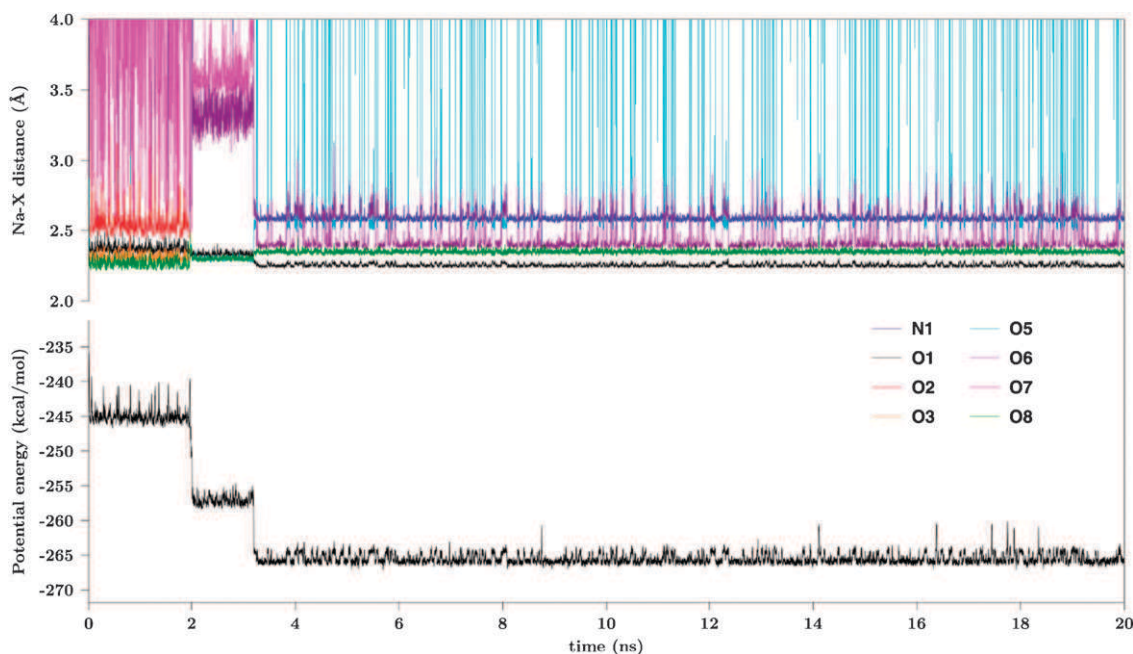


Fig. 10 Time evolution of the short-time averaged Na–X distances and of the potential energy for the G_8_1 conformer.

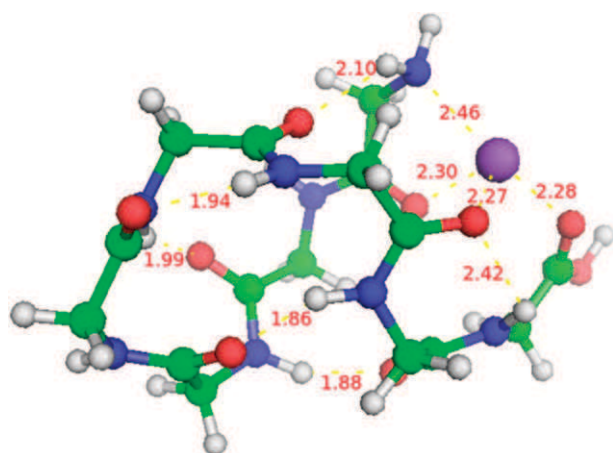


Fig. 11 G_8_1 _new conformation of the G_8 - Na^+ peptide (RI-MP2 geometry).

energy. The geometry reoptimization of the seven initial B3LYP structures was also necessary for direct comparison.

The *ab initio* results are presented in Table 4 giving the relative RI-MP2/TZVPP energies of all 13 conformers for which the geometry was optimized at the RI-MP2/SVP level, and the number and name of atoms coordinated to the sodium ion. The relative energies of the new structures confirm that they are low in energy on the potential energy surface of G_8 - Na^+ . Interestingly, the preferred coordination number is 4, *i.e.* smaller than in a number of initial structures, with 5 to 7 hydrogen bonds. The N-terminus amine is coordinated for the most stable structures in the new series and the C-terminus carbonyl is also coordinated for most of the cases. Due to the fluxionality of the peptide, the sodium ion does not have a preferred localization in the most stable structures as it is either in an internal or external position. The G_8_1 _new conformer is the lowest over the entire series, with a globular shape around the Na^+ ion with four coordinated atoms, and a strong hydrogen bond network on the N-terminus side (see Fig. 11). These results indicate that the competition between complexation to Na^+ and internal structuration *via* multiple hydrogen bonding results in a subtle energy balance.

Table 4 RI-MP2 relative energies in kcal mol^{-1} of the initial and final conformers of G_8 - Na^+ with the coordination number and the atoms coordinated to the sodium ion. All at RI-MP2 geometries

| Conformer | RI-MP2/TZVPP | Coordination | Coord. sites |
|--------------------|--------------|--------------|------------------------|
| G_8_1 | 0.0 | 4 | O1, O3, O5, O8 |
| G_8_2 | -5.8 | 4 | O1, O6, O7, O8 |
| G_8_3 | 2.7 | 4 | O3, O6, O7, O8 |
| G_8_4 | -1.7 | 5 | N1, O1, O4, O5, O8 |
| G_8_5 | -4.9 | 6 | O1, O3, O5, O6, O7, O8 |
| G_8 _compact | 4.7 | 5 | O1, O2, O3, O5, O6 |
| G_8 _helix | -3.5 | 3 | O5, O6, O8 |
| G_8_1 _new | -7.6 | 4 | N1, O1, O6, O8 |
| G_8_2 _new | -4.3 | 4 | N1, O1, O4, O8 |
| G_8_3 _new | 8.6 | 3 | O2, O3, O7 |
| G_8_4 _new | -4.8 | 5 | N1, O1, O4, O5, O8 |
| G_8_5 _new | -3.4 | 3 | O3, O4, O7 |
| G_8 _compact_new | -3.2 | 5 | N1, O1, O3, O5, O6 |

4. Conclusion

We have used the AMOEBA force field with some parameter extensions to study the potential energy surface of $G_n\text{-Na}^+$ with up to ten residues. Accurate quantum mechanical levels were used as reference. While the coupled-cluster CC2 level was used herein, the MP2 level has been found to be reliable enough to be used as reference in future studies. The AMOEBA force field provides energetics that are generally reliable, with errors on relative energies of a couple of kcal mol^{-1} , except in a few problematic cases. In order to go beyond the mere calibration of the force field on pre-determined structures, several molecular dynamics trajectories were generated for $G_8\text{-Na}^+$, yielding new low-energy structures. MP2 calculations confirmed that these structures are highly stable, with a new most stable structure identified. These molecular dynamics studies on $G_8\text{-Na}^+$ complex displayed significant fluxionality, since several structural transitions were observed at room temperature. Based on the backbone torsional energy profiles shown in Fig. 2 and 3, and the binding energies computed at the AMOEBA and *ab initio* levels, we believe that the energy barriers for structural transitions are reasonably well described by the modified AMOEBA force field. Our results indicate that trajectories of tens of nanoseconds (or more) are necessary for a meaningful exploration of the potential energy surface. The fact that several trajectories starting from very different structures lead to globular structures with both multiple sodium-peptide strong interactions and an extended network of hydrogen bonds lends support to the conclusion that such structures are very stable for $G_8\text{-Na}^+$. Given this successful identification of new low energy structures *via* MD simulations using AMOEBA, more refined procedures such as replica-exchange MD may now be carried out in order to determine the absolute minimum of such complex potential energy surfaces. Further studies are also in progress to investigate the role of the environment by including explicit solvent. The stepwise hydration thermodynamics of sodiated peptides will be used as a structure-revealing property by combining mass spectrometry experiments and molecular modeling.

Acknowledgements

Financial support by the European Commission EPITOPES project (Electron Plus Infrared TO Probe and Elucidate Structures, EC project 15367) funded through the NEST (New and Emerging Science and Technology) program is gratefully acknowledged. This work was supported by a grant of computer time at the Institut du Développement et des Ressources en Informatique Scientifique (IDRIS, project 0543) and at the Centre Informatique National de l'Enseignement Supérieur (CINES, project dcm2335).

References

- 1 J. C. Ma and D. A. Dougherty, *Chem. Rev.*, 1997, **97**, 1303–1324.
- 2 S. Hoyau, K. Norrman, T. B. McMahon and G. Ohanessian, *J. Am. Chem. Soc.*, 1999, **121**, 8864–8875.
- 3 P. B. Armentrout and M. T. Rodgers, *J. Phys. Chem. A*, 2000, **104**, 2238–2247.
- 4 T. B. McMahon and G. Ohanessian, *Chem.–Eur. J.*, 2000, **6**, 2931–2941.
- 5 M. M. Kish, G. Ohanessian and C. Wesdemiotis, *Int. J. Mass Spectrom.*, 2003, **227**, 509–524.
- 6 P. Wang, G. Ohanessian and C. Wesdemiotis, *Int. J. Mass Spectrom.*, 2008, **269**, 34–45.
- 7 P. Armentrout, A. Gabriel and R. Moision, *Int. J. Mass Spectrom.*, 2009, **283**, 56–68.
- 8 A. S. Lemoff, M. F. Bush and E. R. Williams, *J. Am. Chem. Soc.*, 2003, **125**, 13576–13584.
- 9 T. Wyttenbach, J. E. Bushnell and M. T. Bowers, *J. Am. Chem. Soc.*, 1998, **120**, 5098–5103.
- 10 M. Kohtani, B. S. Kinnear and M. F. Jarrold, *J. Am. Chem. Soc.*, 2000, **122**, 12377–12378.
- 11 H. A. Cox, R. R. Julian, S.-W. Lee and J. L. Beauchamp, *J. Am. Chem. Soc.*, 2004, **126**, 6485–6490.
- 12 M. M. Kish, C. Wesdemiotis and G. Ohanessian, *J. Phys. Chem. B*, 2004, **108**, 3086–3091.
- 13 C. Kapota and G. Ohanessian, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3744–3755.
- 14 P. Wang, C. Wesdemiotis, C. Kapota and G. Ohanessian, *J. Am. Soc. Mass Spectrom.*, 2007, **18**, 541–552.
- 15 O. P. Balaj, C. Kapota, J. Lemaire and G. Ohanessian, *Int. J. Mass Spectrom.*, 2008, **269**, 196–209.
- 16 N. C. Polfer, J. Oomens and R. C. Dunbar, *ChemPhysChem*, 2008, **9**, 579–589.
- 17 S. J. Ye and P. B. Armentrout, *J. Phys. Chem. A*, 2008, **112**, 3587–3596.
- 18 J. S. Prell, M. Demireva, J. Oomens and E. R. Williams, *J. Am. Chem. Soc.*, 2009, **131**, 1232–1242.
- 19 M. P. Gaigeot, R. Vuilleumier, M. Sprik and D. Borgis, *J. Chem. Theory Comput.*, 2005, **1**, 772–789.
- 20 H. Valdes, K. Pluhackova, M. Pitonak, J. Rezac and P. Hobza, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2747–2757.
- 21 M. P. Gaigeot, *J. Phys. Chem. B*, 2009, **113**, 10059–10062.
- 22 T. D. Rasmussen, P. Ren, J. W. Ponder and F. Jensen, *Int. J. Quantum Chem.*, 2007, **107**, 1390–1395.
- 23 P. Cieplak, F.-Y. Dupradeau, Y. Duan and J. Wang, *J. Phys.: Condens. Matter*, 2009, **21**, 333102 (21pp).
- 24 J. W. Ponder and D. A. Case, *Advances in Protein Chemistry*, Academic Press, London, UK, 2003, vol. 66, pp. 27–85.
- 25 N. Gresh, G. A. Cisneros, T. A. Darden and J.-P. Piquemal, *J. Chem. Theory Comput.*, 2007, **3**, 1960–1986.
- 26 G. A. Cisneros, S. N.-I. Tholander, O. Parisel, T. A. Darden, D. Elking, L. Perera and J. P. Piquemal, *Int. J. Quantum Chem.*, 2008, **108**, 1905–1912.
- 27 G. A. Cisneros, T. A. Darden, N. Gresh, P. Reinhardt, O. Parisel, J. Pilmé and J. P. Piquemal, in *Design of next generation polarizable force fields from ab initio computations: beyond point charges*, ed. D. M. York and T.-S. Lee, Springer Verlag, Dordrecht, the Netherlands, 2009.
- 28 J. Kaminsky and F. Jensen, *J. Chem. Theory Comput.*, 2007, **3**, 1774–1788.
- 29 P. Y. Ren and J. W. Ponder, *J. Comput. Chem.*, 2002, **23**, 1497–1506.
- 30 A. Grossfield, P. Y. Ren and J. W. Ponder, *J. Am. Chem. Soc.*, 2003, **125**, 15671–15682.
- 31 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03, Revision C.02*, Gaussian Inc., Wallingford, CT, 2004.

-
- 32 R. Ahlrichs, M. Bar, M. Haser, H. Horn and C. Kolmel, *Chem. Phys. Lett.*, 1989, **162**, 165–169.
- 33 C. Hattig and F. Weigend, *J. Chem. Phys.*, 2000, **113**, 5154–5161.
- 34 O. Christiansen, H. Koch and P. Jorgensen, *Chem. Phys. Lett.*, 1995, **243**, 409–418.
- 35 C. Hattig, A. Hellweg and A. Kohn, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1159–1169.
- 36 T. Fleig, S. Knecht and C. Haettig, *J. Phys. Chem. A*, 2007, **111**, 5482–5491.
- 37 F. M. Bickelhaupt and E. J. Baerends, in *Rev. Comput. Chem.*, ed. K. B. Lipkowitz and D. B. Boyd, Wiley-VCH, New York, 2000, vol. 15, pp. 1–86; F. M. Bickelhaupt, *Rev. Comput. Chem.*, 2000, **15**, 1.
- 38 *ADF⁸: Amsterdam Density Functional software, version 2007.01*, SCM, Vrije Universiteit, Amsterdam, The Netherlands, <http://www.scm.com>.
- 39 C. F. Guerra, J. W. Handgraaf, E. J. Baerends and F. M. Bickelhaupt, *J. Comput. Chem.*, 2004, **25**, 189–210.
- 40 J.-P. Piquemal, L. Perera, G. A. Cisneros, P. Ren, L. G. Pedersen and T. A. Darden, *J. Chem. Phys.*, 2006, **125**, 054511.
- 41 D. Jiao, J. Zhang, R. E. Duke, G. Li, M. J. Schnieders and P. Ren, *J. Comput. Chem.*, 2009, **30**, 1701–1711.
- 42 J. E. Norton and J.-L. Brédas, *J. Am. Chem. Soc.*, 2008, **130**, 12377–12384.
- 43 N. L. Allinger, Y. H. Yuh and J. H. Liu, *J. Am. Chem. Soc.*, 1989, **111**, 8551–8566.
- 44 J. A. Morrone and R. Car, *Phys. Rev. Lett.*, 2008, **101**, 017801.
- 45 F. Paesani and G. A. Voth, *J. Phys. Chem. B*, 2009, **113**, 5702–5719.
- 46 K. E. Garbow, B. S. Hillstrome and J. J. More, *Documentation for Minpack Subroutine LMDIF Double Precision Version.*, Argonne National Laboratory, Argonne, IL, 1980.
- 47 D. W. Marquardt, *SIAM J. Appl. Math.*, 1963, **11**, 431–441.
- 48 B. A. Cerda, S. Hoyau, G. Ohanessian and C. Wesdemiotis, *J. Am. Chem. Soc.*, 1998, **120**, 2437–2448.
- 49 C. Kapota, PhD Thesis, Ecole Polytechnique, Palaiseau, France, 2005.
- 50 J. W. Ponder, *TINKER, Software Tools for Molecular Design, Version 4.2*, Washington University School of Medicine, Saint Louis, MO, 2004, <http://dasher.wustl.edu/tinker>.

Chapitre 7

Structure de l'octa-glycine sodée

Structure of Sodiated Octa-Glycine: IRMPD Spectroscopy and Molecular Modeling

David Semrouni,^a O. Petru Balaj,^{a,*} Florent Calvo,^b Catarina F. Correia,^{a,†} Carine Clavaguéra,^a and Gilles Ohanessian^a

^a Laboratoire des Mécanismes Réactionnels, Ecole Polytechnique and CNRS, Palaiseau, France

^b Laboratoire de Spectrométrie Ionique et Moléculaire, Université de Lyon I and CNRS, Villeurbanne, France

The structure of the sodiated peptide GGGGGGGG-Na⁺ or G₈-Na⁺ was investigated by infrared multiple photon dissociation (IRMPD) spectroscopy and a combination of theoretical methods. IRMPD was carried out in both the fingerprint and N-H/O-H stretching regions. Modeling used the polarizable force field AMOEBA in conjunction with the replica-exchange molecular dynamics (REMD) method, allowing an efficient exploration of the potential energy surface. Geometries and energetics were further refined at B3LYP-D and MP2 quantum chemical levels. The IRMPD spectra indicate that there is no free C-terminus OH and that several N-Hs are free of hydrogen bonding, while several others are bound, however not very strongly. The structure must then be either of the charge solvation (CS) type with a hydrogen-bound acidic OH, or a salt bridge (SB). Extensive REMD searches generated several low-energy structures of both types. The most stable structures of each type are computed to be very close in energy. The computed energy barrier separating these structures is small enough that G₈-Na⁺ is likely fluxional with easy proton transfer between the two peptide termini. There is, however, good agreement between experiment and computations in the entire spectral range for the CS isomer only, which thus appears to be the most likely structure of G₈-Na⁺ at room temperature. (J Am Soc Mass Spectrom 2010, 21, 728–738) © 2010 American Society for Mass Spectrometry

The biological importance of sodium in performing or facilitating essential biological processes, such as neurotransmission, osmotic balance, and cellular metabolism is well documented [1–3]. Mass spectrometric methods have been used extensively to provide insight into peptide sequences [4, 5] starting from sodium-cationized species, however with considerable debate as to the structure of the parent species and the fragmentation mechanisms [6–8]. In this context, sodiated oligoglycines have been used in the last decade as a valuable testing ground for new experimental developments designed to obtain refined energetic and/or structural data. These include ion mobility measurements for global shape information [9, 10], H/D exchange extent and kinetics for isomeric/conformational content [11], the kinetic [12, 13] and the threshold collision induced decomposition [14] methods for thermochemical measurements, as well as infrared multiple photon dissociation (IRMPD) spectroscopy [15] for identification of functional groups and their

interactions. All these studies have been complemented by extensive molecular modeling as required for translating experimental data into properties of specific molecular structures.

Oligoglycines owe their value as model peptides to their relative simplicity. While the number of residues is an obvious source of conformational complexity, the absence of side chains limits the number of factors shaping their structures and energies. On the one hand, the main components of sodium-molecule interactions are electrostatic and polarization, favoring metal binding to the carbonyl oxygens. On the other hand, internal structuration of the peptides is dominated by the construction of a network of hydrogen bonds. The smaller peptides with 2–5 residues tend to wrap around the metal ion to maximize the binding interactions [13]. In larger peptides, a competition takes place between metal-peptide binding and internal stabilization of the peptide through multiple hydrogen bonds [16]. As a result, the most stable structures are much more difficult to predict.

It is the purpose of this paper to report our investigations on one such case, sodium-octaglycine or G₈-Na⁺. We used IRMPD spectroscopy to obtain fingerprints of the functional groups present in G₈-Na⁺, and of their environments. This is carried out in two energy ranges, 1000–1900 and 3000–3700 cm⁻¹. Interpretation of the results required extensive molecular modeling

Address reprint requests to Dr. G. Ohanessian or Dr. C. Clavaguéra, DCMR, Ecole Polytechnique-CNRS, 91128 Palaiseau Cedex, France. E-mail: gilles.ohanessian@polytechnique.fr; carine.clavaguera@dcmr.polytechnique.fr

* Current address: Institut für Physikalische Chemie der Christian-Albrechts-Universität zu Kiel, Olshausenstrasse 40, D-24098 Kiel, Germany/

† Current address: Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, Université d'Evry Val d'Essonne et CNRS, boulevard F. Mitterrand, 91025 Evry Cedex, France.

using a variety of methods, ranging from extensive sampling with a polarizable force field to quantum chemical calculations using density functional theory and the *ab initio* MP2 method. We first describe the computational and experimental methods used, then we describe the IRMPD spectra and computed IR absorption spectra, and combine them to ascribe a structure to gaseous $G_8\text{-Na}^+$ at 298 K.

Computational Methods

Replica-Exchange Molecular Dynamics with the AMOEBA Force Field

The salt-bridge and charge-solvated isomers of $G_8\text{-Na}^+$ have been modeled using the polarizable AMOEBA force field [17, 18], which was shown to provide reliable structural and energetic properties for sodiated oligoglycine peptides [16]. In this force field, the electrostatic component incorporates a fixed partial charge, a dipole, and a quadrupole on each atom as derived from quantum mechanical calculations. Many-body polarization effects are explicitly treated using a self-consistent induced dipole polarization procedure. The conformational landscapes were explored using molecular dynamics (MD) simulations, improved with the replica-exchange (RE) strategy. The REMD approach is now widely used as a powerful tool for sampling the potential energy surface of biomolecular systems [19–21]. Briefly, the idea is to perform simultaneously several MD trajectories at various temperatures, occasionally attempting an exchange between two configurations of neighboring trajectories, and accepting this exchange with a Metropolis probability [22]. The success of the REMD method comes from the very broad sampling at high temperatures, where barrier crossing is signifi-

cantly enhanced, combined with an efficient annealing to low-energy structures by successive exchanges with lower temperature replicas. Its popularity is due to the use of conventional simulations in canonical ensembles, rather than nonphysical biases, as well as straightforward parallelization.

The present REMD simulations have been performed with eight replicas allocated according to a geometric progression in the range 200–600 K, that is 200, 234, 274, 320, 375, 438, 513, and 600 K. This progression maximizes the successive overlap between energy distributions (at least for harmonic systems), which is a necessary condition for successfully exchanging configurations. Each molecular dynamics trajectory was propagated with a time step of 1 fs, and the temperature was controlled with a Berendsen thermostat with a coupling time of 0.1 ps. Exchanges of a random pair of adjacent configurations were attempted every 10 ps, and configurations were saved every 4 ps for further local optimization with AMOEBA. The total simulation time for each replica was taken as 4 ns, and the REMD simulations were repeated twice, with all replicas starting each time from the most stable structure found so far. The present simulations with the AMOEBA force field were performed with the Tinker software package [23], recently interfaced for replica-exchange MD simulations by Penev and coworkers [24].

Figure 1 shows the distributions of potential energies obtained for the SB complex during the third REMD iteration. These histograms show a broad overlap for the four highest temperatures, but the four replicas at lowest temperatures seem to overlap poorly. However, looking at the instantaneous index of the replica initially set to the lowest temperature replica (200 K), we find a very satisfactory number of successful exchanges

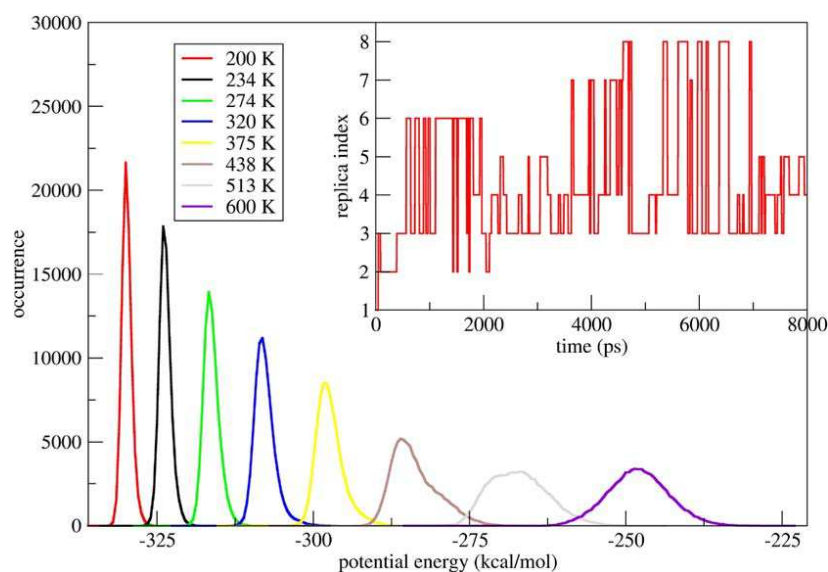


Figure 1. Histograms of potential energy visited during the REMD simulations of the SB system. The inset shows the replica index corresponding to 200 K, indicating that all replicas proceed several times at the lowest temperature within 8 ns.

with other replicas, up to the highest temperature of 600 K. This ensures that our temperature schedule will allow a correct communication between replicas, hence a much improved sampling with respect to conventional MD.

Three iterations of REMD simulations appeared necessary for the SB system, as new minima lower in energy than those previously obtained were found after the first run. The first iteration started from a structure constructed from a previously obtained CS isomer (G_8_1 , see Table 1) by transferring the proton from the C- to the N-terminus. The second iteration was initiated with the lowest energy minimum obtained during the first, while the third iteration was simply a restart from the final configurations of the second iteration. The same computational effort was devoted to the CS system, however, no conformer lower in energy than minima obtained previously by conventional MD simulations [16] could be located.

The most stable configurations collected during the REMD explorations were further subjected to electronic structure calculations to determine the best candidates for comparison to experiment.

Quantum Chemical Calculations

For the conformations selected from REMD simulations, geometry optimizations were performed at DFT-D level with the B3LYP functional and a SVP basis set. The B3LYP functional has been shown to be unre-

liable for relative energies of $G_8\text{-Na}^+$ conformers [16] and $A_n\text{-H}^+$ conformers ($n = 3\text{-}5,7$) [35]. Thus, we used the general empirical dispersion correction proposed by Grimme for density functional calculations [25, 26]. The resolution-of-the-identity (RI) approximation was employed for molecular orbital two-particle integrals. The errors made within this approximation, with optimized auxiliary basis sets, are generally negligible compared with errors due to the one-electron basis set incompleteness. RI-B3LYP-D and RI-MP2 energy calculations on DFT-D optimized geometries were performed with a triple- ζ TZVPP basis set. Vibrational frequencies were calculated at the RI-B3LYP-D/SVP level. The calculated band intensities were convoluted assuming a Lorentzian profile with a 15 cm^{-1} width. The vibrational frequencies were scaled by a factor 0.96 in the entire spectral range. While it is often considered that different scaling factors should be used below 2000 and above 3000 cm^{-1} , reasonable matching was found to occur with a single scaling factor in the present case. The value of 0.96 is a good compromise for B3LYP-D/SVP frequencies in the amide A region as shown by Bouteiller et al. [27]. Finally, the relative Gibbs free energies at 298 K were determined at the B3LYP-D/TZVPP//B3LYP-D/SVP and MP2/TZVPP//B3LYP-D/SVP levels of theory. In the cases of the lowest salt bridge and charge solvation structures, RI-MP2/SVP geometry optimizations and vibrational frequency calculations were additionally carried out. A scaling factor of 0.943 was applied to MP2 frequencies. All quantum chemical calculations were performed with the TURBOMOLE 5.10 package [28].

Experimental

The ions were generated by electrospray ionization and analyzed using two experimental setups, to record IRMPD spectra in two different photon energy ranges. The $1000\text{--}1800\text{ cm}^{-1}$ fingerprint region generally provides information on the identity of the functional groups in the molecule and to some extent on their environment, while the $2900\text{--}3600\text{ cm}^{-1}$ range is where the O–H and N–H stretching motions are expected to be IR-active with frequencies that are highly sensitive to the existence and strength of hydrogen bonds. In the first case, we used a quadrupole ion trap mass spectrometer (modified Bruker Esquire 3000+) [29, 30] coupled to the Free Electron Laser (FEL) at the Centre Laser Infrarouge d'Orsay (CLIO). In the second case, we used an FT-ICR mass spectrometer (Bruker APEX IV) equipped with a 7 T superconducting magnet, coupled to an OPO/OPA laser also at CLIO [31, 32].

Operations of the FEL and OPO/OPA lasers, coupled to either of the above mentioned mass spectrometers, have been detailed elsewhere [29–32]. We provide below the specific parameters used to obtain the results described herein.

Table 1. Relative energies (kcal/mol) of the lower-energy $G_8\text{-Na}^+$ structures at B3LYP-D and MP2 levels. All results are at B3LYP-D optimized geometries except for CS_1. ΔE_0 stands for relative electronic energies and ΔG for relative Gibbs free energies

| Name | ΔE_0 | | ΔG | |
|----------------|--------------|--------------------------|------------|--------------------------|
| | B3LYP-D | MP2 | B3LYP-D | MP2 |
| $G_8_compact$ | 15.80 | 16.14 | 12.10 | 12.44 |
| G_8_1 | 12.76 | 11.96 | 12.06 | 11.25 |
| G_8_helix | 11.35 | 8.33 | 10.64 | 7.61 |
| G_8_1N | 3.77 | 4.22 | 6.85 | 7.30 |
| G_8_5 | 8.38 | 7.12 | 5.88 | 4.63 |
| CS_10 | 6.56 | 9.73 | 10.12 | 13.29 |
| CS_9 | 12.95 | 14.00 | 12.16 | 13.21 |
| CS_8 | 13.46 | 14.85 | 10.41 | 11.79 |
| CS_7 | 1.17 | 3.98 | 3.93 | 6.74 |
| CS_6 | 5.03 | 6.02 | 5.47 | 6.46 |
| CS_5 | 1.68 | 2.26 | 4.65 | 5.23 |
| CS_4 | −0.50 | 1.75 | 2.89 | 5.14 |
| CS_3 | 3.46 | 4.19 | 4.14 | 4.86 |
| CS_2 | 3.40 | 3.29 | 4.24 | 4.13 |
| CS_1 | 0.82 | 1.32 (1.40) ^a | 1.46 | 1.96 (0.49) ^a |
| SB_5 | 12.89 | 12.46 | 14.85 | 14.42 |
| SB_4 | 7.55 | 8.46 | 8.34 | 9.26 |
| SB_3 | 6.11 | 6.30 | 8.53 | 8.73 |
| SB_2 | 5.09 | 4.48 | 7.39 | 6.78 |
| SB_1 | 0.00 | 0.00 | 0.00 | 0.00 |

^aUsing MP2 geometries and vibrational frequencies.

IRMPD Experiments in the Fingerprint Region

The experimental set-up is based on a modified Bruker Esquire 3000+ ion trap mass spectrometer. To allow the irradiation of the ions with the focused CLIO beam, a 0.7 mm hole was drilled in the ring electrode. The ions generated by an electrospray ion source are transferred to the trap by two octopoles. There, the ions are stabilized by collisions with the He bath gas at a pressure of 10^{-3} mbar. After a rapid thermalization, the ions are confined in a very small volume in the center of the trap. The laser beam was focused in the center of the trap using a 500 mm focal length ZnSe focal lens. The laser beam position was optimized to maximize the fragmentation and allowed for a greater overlap with the ion cloud, resulting in increased fragmentation efficiency than in the ICR cell.

The IR FEL CLIO light is delivered in 8 μs long macropulses fired at a repetition rate of 25 Hz. Each macropulse contains 500 micropulses, each a few picoseconds long. The mean IR power was about 500 mW corresponding to micropulse and macropulse energies of 40 μJ and 20 mJ, respectively. For the experiments described herein, we have used an electron energy of 45 MeV, which allows access to wavelengths in the 1000–1900 cm^{-1} range. In a specific experiment, the 1900–2400 cm^{-1} was searched with a laser power of ca. 600 mW. No absorption band could be detected.

Following a 0.5 s relaxation time, the ions are mass selected and fragmented by macropulses of IR light. Ion fragmentation is observable with only 3 macropulses, but in the experiments presented herein the spectra have been recorded with 10 macropulses. The conditions were tuned for maximum fragmentation efficiency while avoiding saturation.

IRMPD Experiments in the 3000–3600 cm^{-1} Region

The ions are irradiated with the beam generated by a OPO/OPA laser (LaserVision, Bellevue, WA, USA) pumped by a 25 Hz Nd:Yag Splitlight 600 (Innolas Laser GmbH, Krailling, Germany). The laser beam enters the FT-ICR cell along the magnet axis and interacts with the ions trapped into 10^{-10} mbar vacuum. A spherical mirror with 2 m focal distance, placed between the laser and the cell, generates an almost constant diameter of the laser beam in the cell region. Once trapped into the ICR cell, the ions were irradiated for 20 s before signal detection. Each point in the spectra was measured only once, without any signal averaging. For maximum fragmentation efficiency, the IRMPD spectra were split in sections no larger than 300–400 cm^{-1} , and the laser power was optimized before each measurement for the range we intended to scan.

IRMPD fragmentation spectra are plotted using the fragmentation yield R defined as: $R = -\ln[\sum I_{\text{Fragment}} / (I_{\text{Parent}} + \sum I_{\text{Fragment}})]$. The FEL fingerprint spectrum was obtained by including the signal intensities for 24

fragment ions, while the OPO/OPA spectrum was based on six fragment ion intensities.

The spectra in both photon energy ranges were smoothed using a weighted polynomial regression smoothing algorithm.

Sample Preparation

Gaseous ions were obtained by electrospray of diluted sample solutions. First, 15 mL peptide stock solutions were prepared using 6 mL 10% formic acid and 9 mL $\text{H}_2\text{O}/\text{MeOH}$ 50:50 to dissolve the 10 mg of solid peptide; 10–50 μL of stock solution (10^{-3} M) were mixed 1:1 with NaCl solution (10^{-3} M) and diluted to 1.0 mL with $\text{H}_2\text{O}/\text{MeOH}$ 50:50 with 1% formic acid. This protocol produced samples with concentrations of peptide in the μM range, leading to the formation of the sodiated peptide ion as the dominant species in the gas-phase.

Results and Discussion

The IRMPD spectra of $G_8\text{-Na}^+$ in the two spectral ranges investigated are shown in Figure 2 and Figure 4, together with those computed for some of the lowest energy structures shown in Figure 3. Band assignment can be made straightforwardly in some cases, for instance the amide I and amide II massifs in the 1650–1750 and 1480–560 cm^{-1} ranges, respectively. The absence of a free carboxylic acid O–H stretching band near 3550 cm^{-1} [33, 34] is equally noteworthy. Yet other band assignments and the relation of these experimental patterns to one or several molecular structures require modeling results.

Previous molecular dynamics searches [16] were carried out only from a CS structure, which was intentionally chosen not to be among the most stable ($G_8\text{-compact}$, see Table 1). While the trajectories obtained at different temperatures showed the potential for sampling other areas of the potential energy surface and did locate new low-energy structures, the AMOEBA force field cannot describe the breaking or formation of covalent bonds and, therefore, SB structures could not be sampled. The REMD simulations carried out in the present work were started from both CS and SB structures, using appropriate parameters for the different functional groups. The efficiency of the REMD computational strategy is illustrated in Figure 5, where we show the energy of the replica at 274 K in the first REMD iteration of SB structures, after local minimization with the AMOEBA force field as a function of simulation time. For comparison, the minimized energy obtained from a regular MD trajectory at the same temperature and initialized from the same configuration is also displayed. Clearly, the set of isomers sampled by regular molecular dynamics (without exchanges) is rather limited, indicating that this MD trajectory was stuck in metastable regions of the landscape.

The REMD sampling stage provided large databases of candidate structures, from which about 40, all lying in a 5 kcal/mol window, were chosen for further

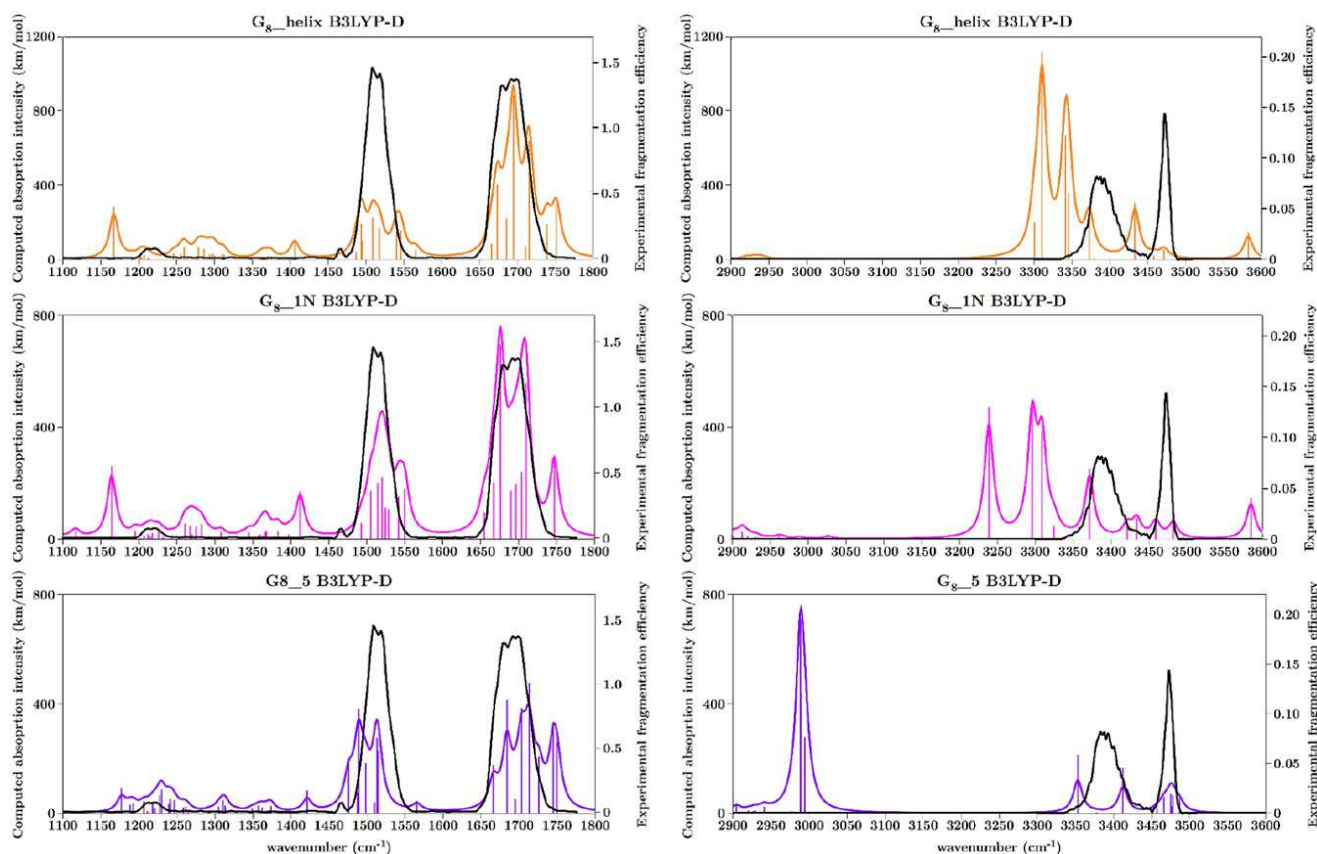


Figure 2. IRMPD spectra of $G_8\text{-Na}^+$ and computed IR absorption spectra for the $G_8\text{-helix}$, $G_8\text{-1N}$, and $G_8\text{-5}$ structures at the B3LYP-D/SVP level using a scaling factor of 0.96.

re-optimization by quantum chemistry methods. Structural refinement generated a significant number of duplicates, including mirror image pairs, which were eliminated. Structures were considered to be duplicates when the root mean square difference of their cartesian structures was smaller than 0.1 Å. In the course of DFT-D re-optimization, some of the SB structures generated with AMOEBA collapsed to CS geometries. Since such rearrangements involve proton transfer, the resulting structures all have the C-terminus O–H in strong hydrogen bonding interaction. The final energetics confirmed the adequacy of the AMOEBA force field to describe $G_8\text{-Na}^+$ as the energy orders were nearly identical. It is noteworthy that most of these low-energy structures escaped initial searches based on chemical intuition.

The description below uses the sequence numbering from the N- to the C-terminus with superscripts, i.e., NH^2 and CO^2 stand for the peptidic N–H and C=O bonds of the second residue. Special cases are for residue 1 where the N–H bonds are in a NH_2 rather than an amide group and for residue 8 for which CO^8 is an acidic rather than an amide C=O. Residue 8 also bears the single O–H bond of $G_8\text{-Na}^+$.

The final energetic results are given in Table 1 at the B3LYP-D//B3LYP-D and MP2//B3LYP-D levels, except for SB_1 and CS_1 for which MP2//MP2 results

are also available. The first five structures in Table 1 have been described previously [16]. Their names have been kept the same for clarity. Other structures (CS_1-CS_10 and SB_1-SB_5) are new and are denoted by their CS or SB type plus their number in MP2//B3LYP-D free-energy order within the CS and SB families. Inspection of Table 1 indicates that B3LYP-D//B3LYP-D and MP2//B3LYP-D relative energies are in very good agreement. Most values are within 1 kcal/mol of one another and the largest discrepancies are ca. 3 kcal/mol. The same holds true for free energies since thermal and entropy contributions are based on B3LYP-D geometries and frequencies in both cases. The stability order is only slightly different when comparing B3LYP-D//B3LYP-D and MP2//B3LYP-D results, as well as when comparing relative energies and free energies at a given computational level, so that differences will not be commented further. The discussion below will be based on MP2//B3LYP-D free energies unless otherwise indicated.

Elimination of Structural Families on the Basis of Their Computed Spectra

The structures that were subjected to full quantum chemical study are widely different according to the

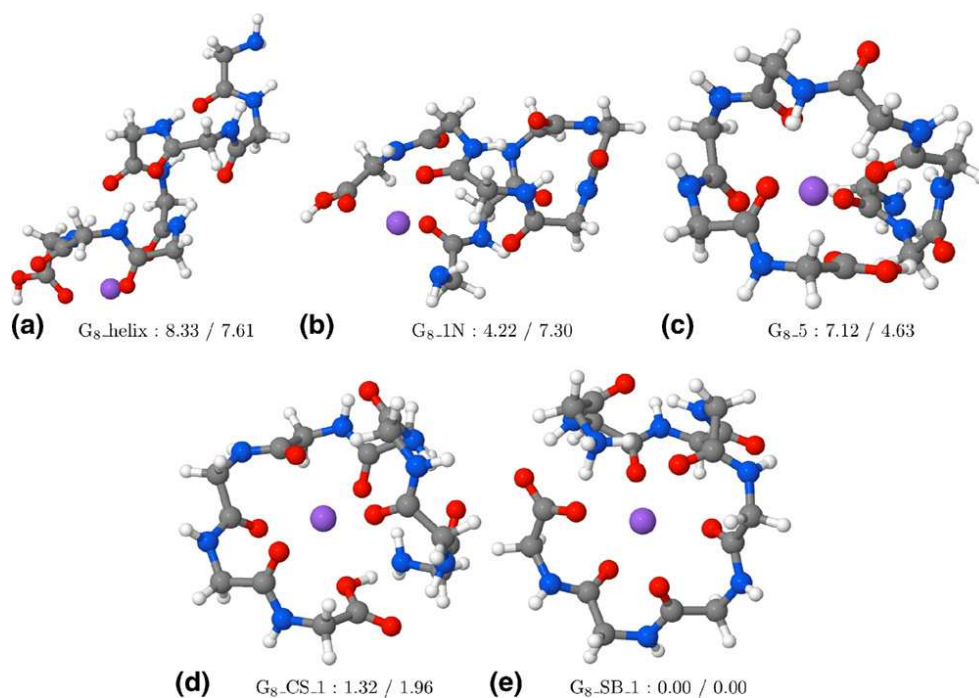


Figure 3. $G_8\text{-Na}^+$ optimized structures (see text for labeling). Relative MP2//B3LYP energies in kcal/mol are reported as electronic energies/Gibbs free energies.

number of carbonyl oxygens attached to sodium (with Na–O distances of 2.7 Å or less), which varies from 3 to 6, and to the number and strength of hydrogen bonds. Some of them may be easily shown not to be good candidates for the experimentally observed $G_8\text{-Na}^+$ ions, based on the comparison of the computed and experimental spectra. In this section we describe some of these cases, illustrating how the final set of structures was selected.

The most stable α -helical structure G_8_helix is shown in Figure 3. The sodium is bound at the C-terminus and interacts with CO⁵, CO⁶, and CO⁸. It has the C-terminus OH free of any interaction, leading to O–H stretch and COH bend peaks at 3583 and 1167 cm^{-1} , respectively, which are both absent from the IRMPD spectrum (see Figure 2). The free O–H stretch has been previously identified for smaller protonated or cationized amino acids and peptides [33–36]. The COH bend has been used as a diagnostic tool as well [15, 37–42]. Among peptidic N–Hs, NH⁵ interacts weakly with CO¹ with an amide A (N–H stretching) frequency at 3433 cm^{-1} , lying just in between the experimental bands. Furthermore, the N–H–O=C hydrogen bonding pattern typical of a helix leads to a group of amide A frequencies that are stronger than that of NH⁵, in agreement with those of the helical peptides $\text{Ac-FA}_n\text{-K-H}^+$ ($n = 5, 10$) in the 3300–3500 cm^{-1} region [36], and thus are red-shifted in the present case compared with the IRMPD feature at 3340–3440 cm^{-1} by ca. 70 cm^{-1} . Clearly, the α -helical structure may be excluded for $G_8\text{-Na}^+$. This is in agreement with its computed free-energy of 7.6 kcal/mol relative to that of the most stable structure (see Table 1).

Structure G_8_1N is another case that can be easily eliminated. Although very different from the α -helix, it also has a free OH, which is incompatible with the IRMPD spectrum. Some of the amide A bands are even more red-shifted than for the α -helix, leading to more inconsistency with IRMPD. Its relative free-energy of G_8_1N of 7.3 kcal/mol is also unfavorable, leading us to discard it.

A third example is structure G_8_5 , in which there are five peptidic N–Hs with no hydrogen bonds, with amide A frequencies between 3466 and 3487 cm^{-1} , in good agreement with the IRMPD band at 3460–3490 cm^{-1} . There are also two bands, the N–H⁵ stretch (involved in a C₁₁ motif with CO⁷) and the symmetric NH₂ stretch (involved in a C₅ motif with NH²) at 3412 and 3378 cm^{-1} , respectively, which are in good agreement with the IRMPD massif between 3350 and 3440 cm^{-1} . This would lead to the narrow experimental feature being the convolution of five bands, while the broader feature would encompass only two, which may not be realistic. The carboxylic O–H is hydrogen-bonded to the CO², accounting for the lack of IRMPD bands near 1160 and 3550 cm^{-1} discussed above. However, this hydrogen bond is strong enough that the O–H stretch is computed to be red-shifted to 2989 cm^{-1} , a band that is clearly absent from the IRMPD spectrum. In addition, the amide I massif is computed to be too wide with two bands near 1750 cm^{-1} , which are clearly at odds with the experimental result.

The previous discussion shows that although the resolution of the IRMPD spectrum recorded at room temperature is low, band positions carry enough information to make a severe selection among low-energy

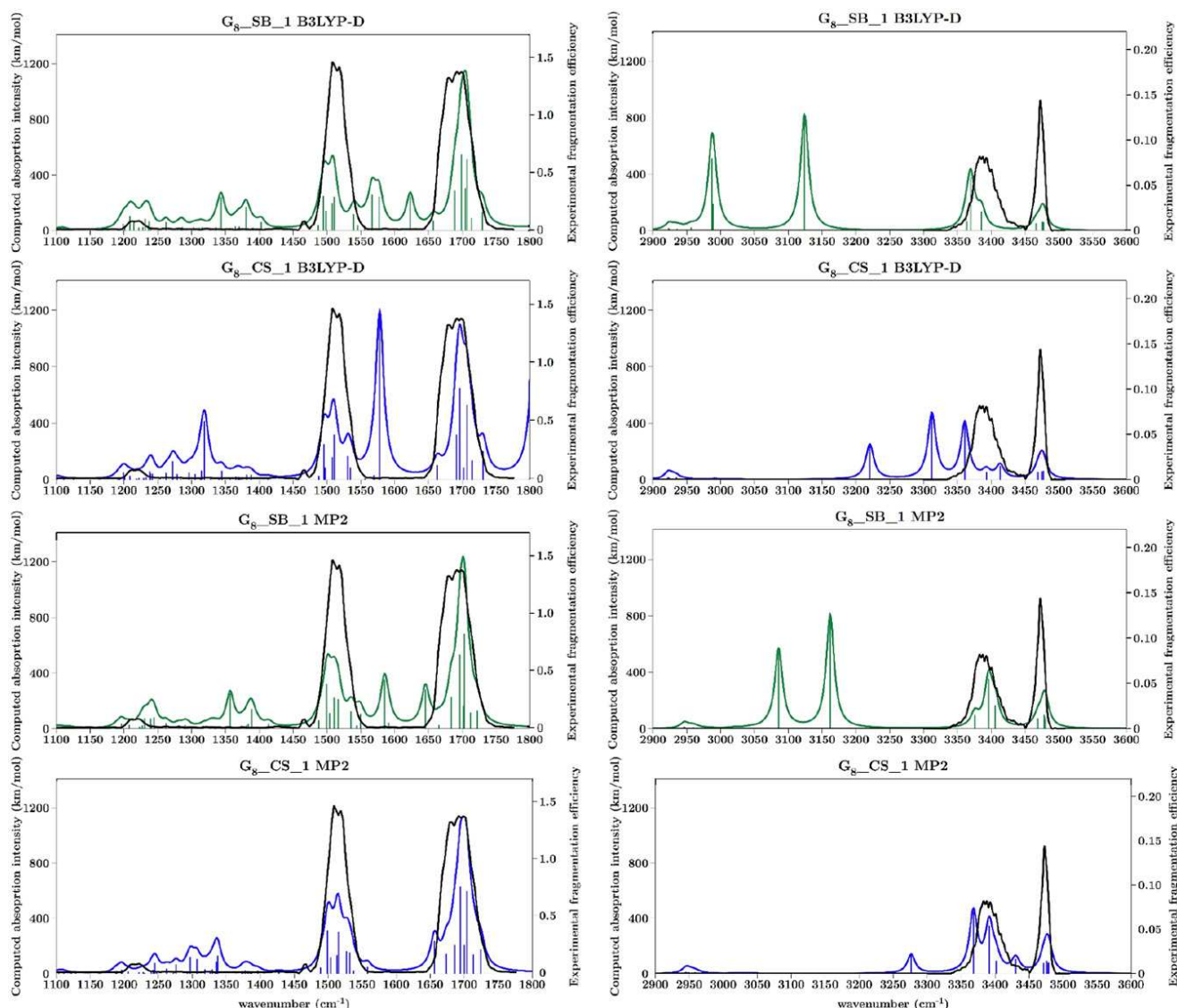


Figure 4. IRMPD spectra of $G_8\text{-Na}^+$ and computed IR absorption spectra for the CS_1 and SB_1 low-energy structures at the B3LYP-D/SVP level using a scaling factor of 0.96, and at MP2/SVP level using a scaling factor of 0.943.

structures. In the following section, we compare the experimental IRMPD spectrum with spectra computed for the two most stable structures. Both satisfy the criteria established above: lack of free O–H and adequate distribution of hydrogen bonds for peptidic N–Hs. The IR spectra of some of the other structures computed at the B3LYP-D level are fairly similar. Owing to the rather high computational level of the present investigation, the lowest-energy structures obtained in this work should be the most abundantly populated, if not exclusively formed in the experiment, although experimental trapping of gaseous ions into local minima cannot be ruled out completely. The two most stable structures happen to be one CS and one SB isomer, which are separated from the others by a significant energy gap (especially for the SB series, see Table 1).

Comparison of IRMPD and DFT Computed Spectra for the Two Lowest Energy Structures

The most stable SB and CS structures are shown in Figure 3, and their computed spectra are shown in Figure 4. We focus on B3LYP-D spectra in this section. The two structures are very similar to each other, with the same set of carbonyl oxygens bound to sodium and the same network of hydrogen bonds except, of course, for the two termini. The peptidic bonds NH^3 , NH^6 , NH^7 , and NH^8 are free, leading in both cases to four bands in the $3468\text{--}3478\text{ cm}^{-1}$ range, in good agreement with the IRMPD feature at $3460\text{--}3490\text{ cm}^{-1}$. The other bands in this region correspond to H-bound peptidic N–Hs and to amine and ammonium bands in CS_1 and SB_1, respectively. Some of them fit reasonably well with the IRMPD broad band at $3340\text{--}3440$

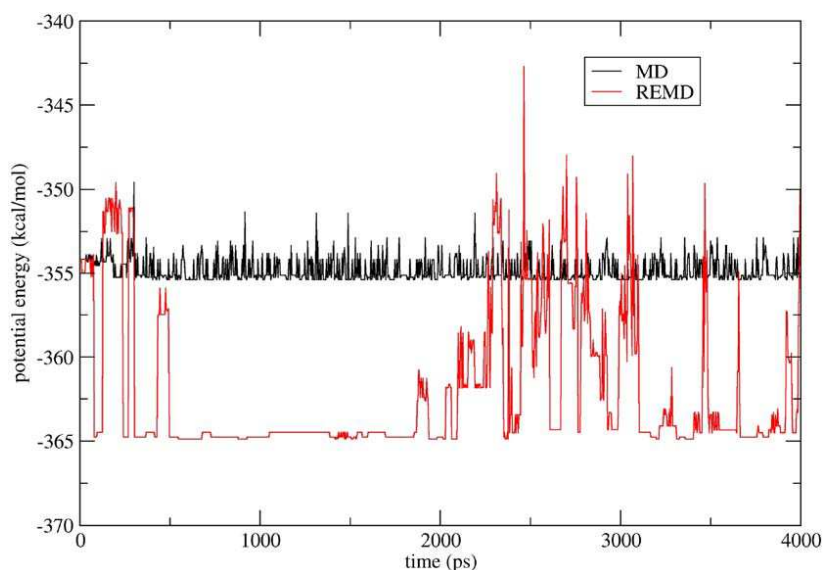


Figure 5. Potential energy obtained after local re-optimization during the first iteration of REMD (in red) and regular MD (in black) simulations of the SB isomer at 274 K, as a function of time. The two simulations were initialized from the same configuration.

cm^{-1} (three in each case), while others appear at significantly lower frequencies. In the SB_1 isomer, the ammonium group has one free N–H bond (with a computed frequency of 3364 cm^{-1}), one bound to CO^4 (N–H—O=C of 1.79 \AA , N–H stretching frequency of 2980 cm^{-1}), and the third bound to an oxygen of the carboxylate at the C terminus with a very small N–H—O=C distance of 1.44 \AA . The latter N–H bond is significantly elongated at 1.13 \AA . This unusually strong hydrogen bond, together with the electrostatic attraction between the ammonium and the carboxylate, leads to a massive red-shift of the N–H stretch frequency to 1890 cm^{-1} . In the CS_1 isomer, one N–H of the N-terminus NH_2 is free while the other is bound to CO^4 although with a distance of 2.0 \AA that is larger than in SB_1 as expected for a non-ionic hydrogen bond. The asymmetric and symmetric NH_2 stretch frequencies are computed at 3393 and 3221 cm^{-1} , respectively. The C-terminus O–H bond interacts strongly with the amine nitrogen (O–H of 1.09 \AA , O–H–N hydrogen bond of 1.47 \AA), leading here again to a massive red-shift of the O–H stretch frequency to 1809 cm^{-1} . The proximity of this frequency with that of the carboxylic C=O and their close interaction lead to a strong out-of-phase mixing in the 1809 cm^{-1} band, and to the appearance of a band corresponding to their in-phase mixing at 1578 cm^{-1} . This band emerges between the amide I and II massifs, a feature that is absent from the IRMPD spectrum. In the same region, the SB_1 isomer displays three HNH bending modes of the ammonium group at 1569 , 1577 , and 1624 cm^{-1} . Clearly, both cases have many features in good agreement with the IRMPD spectrum, yet significant discrepancies between theory and experiment also exist in both spectral ranges, such that attribution of the spectrum to either structure is not possible.

Comparison of IRMPD and MP2 Computed Spectra for the Two Lowest Energy Structures

The previous discrepancies may also occur because the nonbonded interactions, which are present in the folded structures described above, are not well described by the B3LYP functional complemented by the empirical correction used because the ionic hydrogen bonds are particularly strong in the present case. To improve the computational level, we repeated the geometry optimizations and frequency calculations of SB_1 and CS_1 at the RI-MP2/SVP level. Both minima could be located at this level as well, with some structural differences. In SB_1, the MP2 N–H/NH–O distances are $1.12/1.45 \text{ \AA}$ versus $1.13/1.44 \text{ \AA}$ with B3LYP-D. In CS_1, the MP2 O–H/OH–N distances are $1.07/1.51 \text{ \AA}$, while they are $1.09/1.47 \text{ \AA}$ with B3LYP-D. Although small, these differences are not negligible. In all cases, the MP2 covalent bond lengths are less elongated, and the hydrogen bonding distances are larger than at the B3LYP-D level. Since N–H and O–H stretching frequencies are strongly sensitive to hydrogen bonding, it is not surprising that such changes lead to important differences in the computed spectra, inducing a series of blue shifts in band positions for the corresponding stretching modes. The MP2//MP2 free-energy difference between the two structures is negligible at 0.49 kcal/mol . The computed IR spectra of CS_1 and SB_1 at the B3LYP-D and MP2 levels are shown in Figure 4. The IR spectrum of SB_1 in the fingerprint region is not much different at the MP2 and B3LYP-D levels, except that the ammonium N–H band is shifted from 1890 to 1947 cm^{-1} . Because of limitations on the electron energy range, IRMPD experiments are not usually carried out in the high-energy end of the fingerprint region. We did scan the $1900\text{--}2400 \text{ cm}^{-1}$ range in a pass where the laser power was

nearly constant at 0.6 W. No band was detected in this region. However, such extremely strong hydrogen bonding is expected to be associated to large anharmonicity such that harmonic computations are not expected to be particularly reliable, as discussed recently for deprotonated glutamic and aspartic acids [37].

While the amide I and amide II broad bands are well positioned in both cases, blue shifts of the HNH bending modes of the ammonium at 1585, 1591, and 1646 cm^{-1} still lead to bands in between the amide bands that are not observed experimentally. In the high-frequency region, the three amide A frequencies of H-bound N–Hs are slightly blue-shifted such that they agree very well with the IRMPD band at 3340–3440 cm^{-1} . However the ammonium N–H stretches, although significantly blue-shifted, still show exceedingly low frequencies with respect to experiment. Analogously large red shifts have been previously described for N–H stretches of ammonium groups that are hydrogen-bonded to carbonyl oxygens, with bands in the 3000–3200 cm^{-1} area [33, 35]. Careful re-scanning after laser power optimization in this region confirmed the absence of any experimental band.

Blue shifts are also observed in the MP2 versus the B3LYP-D spectrum of CS_1. The O–H stretch frequency is displaced from 1809 to 1950 cm^{-1} , without significant coupling with the carboxylic C=O stretch and disappearance of the band at 1578 cm^{-1} . Again, there is no equivalent band in the experimental spectrum. The carboxylic C=O stretch at 1656 cm^{-1} is within the amide I massif, while the NH_2 bend at 1558 cm^{-1} is within the amide II. Thus, experiments and computations match very well in the fingerprint region. Amide A bands of the two strongly H-bound N–Hs are blue-shifted by ca. 50 cm^{-1} , such that they fit well with the experimental band at 3340–3440 cm^{-1} . The symmetric NH_2 stretch, although shifted by 50 cm^{-1} as well, is the only band not seen experimentally. Its weak intensity is probably responsible for this lack of observation. Overall, the MP2 spectrum of CS_1 is in very good agreement with the IRMPD spectrum. One may then conclude that this structure is the most likely candidate for the room-temperature structure of $\text{G}_8\text{-Na}^+$.

Discussion

To characterize these very strong interactions in more detail, the transition-state for interconversion of CS_1 and SB_1, i.e., proton transfer between the N- and C-termini was determined at the B3LYP-D/SVP level. It was found to be very similar to the SB_1 and CS_1 minima it connects, with N–H and O–H distances of 1.310 and 1.201 Å, respectively. The activation barrier is computed to be 0.3 and -0.06 kcal/mol at the B3LYP-D//B3LYP-D and MP2//B3LYP-D levels, respectively, relative to the CS_1 minimum. Thus, one would expect that easy proton transfer takes place and that the overall structure may not be described properly as either SB_1 or CS_1. Our results indicate, however, that the com-

puted spectrum of CS_1 fits well with the experimental spectrum.

The salt bridge versus charge solvation structures of cationized amino acids have been a matter of intense research using IRMPD spectroscopy in the last few years [38]. This has been extended to cationized peptides recently [38, 39, 41–43]. Most cases of cationized peptides for which a SB structure was established contain arginine, a highly basic residue that favors proton transfer [38, 39, 41]. Only when a divalent ion was attached to the AA dipeptide was a SB structure found to be favored, while the alkali cationized di- and tripeptides favored a CS isomer [15, 42, 43]. Apparently, the present results are the first to point to a high stability for a salt bridge in a cationized peptide without any basic residue and limited electrostatic perturbation induced by an alkali cation.

Our results indicate that there may be no general way to distinguish between CS and SB isomers in peptides with more than a few residues on the basis of their IRMPD spectra, especially in the fingerprint region. While carboxylic O–H bonds that are free of hydrogen bonding are simple to detect (O–H stretching and COH bending frequencies near 3550 and 1160 cm^{-1} , respectively, plus carboxylic C=O stretching frequency near or above 1750 cm^{-1}), it is unlikely that a carboxylic acid remains free of interaction in peptides of growing size. Any carboxylic acid involved in strong hydrogen bonding with, e.g., an amine, may be rather difficult to distinguish from its isomeric carboxylate in strong hydrogen bonding with, e.g., an ammonium, especially when the growing size of the peptide makes the amide I and II massifs grow broader. This is illustrated by the very similar amide I and II massifs computed for the CS_1 and SB_1 isomers of $\text{G}_8\text{-Na}^+$ (see Figure 4). Some of the ammonium bending bands appear near 1400 cm^{-1} when free but are blue-shifted by hydrogen bonding [42] and may then be hidden, at least partially, by the amide II massif. Our results point to the possibility that ammonium N–H stretches appear below ca. 3300 cm^{-1} as a result of strong, ionic hydrogen bonding. Such bands would then be typical of salt bridges, assuming that CS isomers cannot give rise to analogous shifts. Firm conclusions along these lines await more comprehensive experimental results on ionic peptides. In any event, the present results indicate that although $\text{G}_8\text{-Na}^+$ appears to have a CS structure, salt bridges may be highly stable even for peptides of growing size in which the metal charge can be efficiently solvated by carbonyl oxygens, even without basic residues. Dynamical studies would be particularly welcome on such systems with a shared proton [44], however, current limitations of density functionals (see below) make such calculations of limited use at least in the present case.

Although B3LYP-D provides energetics in very good agreement with MP2, it appears to yield structures that are slightly too compact, generating significant errors on IR band positions whenever there are strong hydro-

gen bonds. In such cases, red shifts of ca. 50 cm⁻¹ with respect to MP2 values have been obtained for G₈-Na⁺. This trend indicates that some of the results reported may not be free of similar errors. More work is needed to establish a reliable quantum chemical level of calculation that is able to yield reliable IR spectra for still larger biomolecules. Since vibrational frequency calculations cannot be carried out at the MP2 level for peptides much larger than G₈, the most promising track remains DFT now. However, the adequate density functional for treating charged molecules with very strong hydrogen bonds remains to be identified. Work is also underway on oligoglycines of different sizes to establish if the structures probed experimentally are of the charge solvation or salt bridge type and to understand their dynamical behavior in more detail.

Conclusions

IRMPD spectra have been obtained for G₈-Na⁺ at room temperature in the 1000–1900 and 3000–3700 cm⁻¹ regions. The results indicate that the C-terminus OH group is either hydrogen bound or absent. The 3000–3700 cm⁻¹ spectrum is indicative of a network of N–H bonds, most of which are either free or not very strongly hydrogen-bound. To gain more structural insight, extensive molecular modeling was carried out. A powerful approach coupling the polarizable force field AMOEBA and replica exchange molecular dynamics was used to explore the conformation landscape of G₈-Na⁺ efficiently. Both charge solvation and salt bridge structures were identified and refined at the B3LYP-D and MP2 quantum chemical levels. It turned out that the two structures with the most favorable free energies at 298 K are a CS/SB pair related by a simple proton transfer between the N- and C-termini with a high degree of peptidic C=O coordination to the sodium. The activation barrier for interconversion of these structures was computed to be quite small. It was only at the MP2 level that good agreement between experiments and computations could be obtained, yielding a charge solvation structure as the most likely candidate for the gas-phase structure of G₈-Na⁺ at 298 K.

Acknowledgments

O.P.B. thanks Ecole Polytechnique for a 1 year post-doctoral fellowship and the European Union (EU) commission for a 1 year post-doctoral support through the STREP 15,637 “EPITOPES” funded by the New and Emerging Science and Technology program of the 6th pluriannual framework of EU. C.F.C. thanks the Fundação para a Ciência e a Tecnologia for a 2-year post-doctoral fellowship. Implementation of the REMD method was supported by a travel grant to D.S. and C.C. from GDR 2758 funded by CNRS. This work was granted access to the HPC resources of [CCRT/CINES/IDRIS] under the allocation x2009085107 made by GENCI (Grand Equipement National de Calcul Intensif).

References

1. Michell A. R., Ed. *The Clinical Biology of Sodium: The Physiology and Pathophysiology of Sodium in Mammals*; Pergamon: Oxford, UK, 1995; p. 16.
2. Skou, J. C.; Norby, J. G.; Maunsbach, A. B.; Esmann, M. In *Progress in Clinical and Biological Research*; Vol. CCLXVIII; Alan R. Liss, Inc.: New York, 1987, pp 401–632.
3. Aidley, D. J.; Stanfield, P. R. *Ion Channels: Molecules in Action*; Cambridge University Press: Cambridge, 1994, pp 1–58.
4. Teesch, L. M.; Orlando, R. C.; Adams, J. Location of the Alkali Metal Ion in Gas-Phase Peptide Complexes. *J. Am. Chem. Soc.* **1991**, *113*, 3668–2675.
5. Lin, T.; Payne, A. H.; Glish, G. L. Dissociation Pathways of Alkali-Cationized Peptides: Opportunities for C-Terminal Peptide Sequencing. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 497–504.
6. Feng, W. Y.; Gronert, S.; Fletcher, K. A.; Warren, A.; Lebrilla, C. B. The Mechanism of C-Terminal Fragmentations in Alkali Metal Ion Complexes of Peptides. *Int. J. Mass Spectrom.* **2003**, *222*, 117–134.
7. Bulleigh, K.; Howard, A.; Do, T.; Wu, Q.; Anbalagan, V.; van Stipdonk, M. Investigation of Intramolecular Proton Migration in a Series of Model, Metal-Cationized Tripeptides Using In Situ Generation of an Isotope Label. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 227–232.
8. Lee, S. W.; Kim, H. S.; Beauchamp, J. L. Salt Bridge Chemistry Applied to Gas-Phase Peptide Sequencing: Selective Fragmentation of Sodiated Peptide Ions Adjacent to Aspartic Acid Residues. *J. Am. Chem. Soc.* **1998**, *120*, 3188–3195.
9. Wyttenbach, T.; Bushnell, J. E.; Bowers, M. T. Salt Bridge Structures in the Absence of Solvent? The Case for the Oligoglycines. *J. Am. Chem. Soc.* **1998**, *120*, 5098–5103.
10. Kohtani, M.; Jarrold, M. F.; Wee, S.; O’Hair, R. A. J. Metal-Ion Interactions with Polyalanine Peptides. *J. Phys. Chem. B* **2004**, *108*, 6093–6097.
11. Cox, H. A.; Julian, R. R.; Lee, S.-W.; Beauchamp, J. L. Gas-Phase H/D Exchange of Sodiated Glycine Oligomers with ND₃: Exchange Kinetics Do Not Reflect Parent Ion Structures. *J. Am. Chem. Soc.* **2004**, *126*, 6485–6490.
12. Kish, M. M.; Wesdemiotis, C.; Ohanessian, G. The Sodium Ion Affinity of Glycylglycine. *J. Phys. Chem. B* **2004**, *108*, 3086–3091.
13. Wang, P.; Kapota, C.; Ohanessian, G.; Wesdemiotis, C. The Sodium Ion Affinities of Simple Di-, Tri-, and Tetrapeptides. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 541–552.
14. Ye, S. J.; Armentrout, P. B. Absolute Thermodynamic Measurements of Alkali Metal Cation Interactions with a Simple Dipeptide and Tripeptide. *J. Phys. Chem. A* **2008**, *112*, 3587–3596.
15. Balaj, O. P.; Kapota, C.; Lemaire, J.; Ohanessian, G. Vibrational Signatures of Sodiated Oligopeptides (GG-Na⁺, AA-Na⁺, GGG-Na⁺, and AAA-Na⁺) in the Gas Phase. *Int. J. Mass Spectrom.* **2008**, *269*, 196–209.
16. Semrouni, D.; Ohanessian, G.; Clavaguera, C. Structural, Energetic, and Dynamical Properties of Sodiated Oligoglycines: Relevance of a Polarizable Force Field. *Phys. Chem. Chem. Phys.* **2010**, DOI: 10.1039/b924317h.
17. Ren, P. Y.; Ponder, J. W. Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanic Calculations. *J. Comput. Chem.* **2002**, *23*, 1497–1506.
18. Grossfield, A.; Ren, P. Y.; Ponder, J. W. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *J. Am. Chem. Soc.* **2003**, *125*, 15671–15682.
19. Voelz, V. A.; Shell, M. S.; Dill, K. A. Predicting Peptide Structures in Native Proteins from Physical Simulations of Fragments. *PLoS Comp. Biol.* **2009**, *5*(2), e1000281.
20. Buchete, N. V.; Hummer, G. Protein Folding Kinetics from Replica Exchange Molecular Dynamics. *Phys. Rev. E* **2008**, *77*(3), 030902.
21. Weinstock, D. S.; Narayanan, C.; Felts, A. K.; Andreic, M.; Levy, R. M.; Wu, K. P.; Baum, J. Distinguishing Between Structural Ensembles of the GB1 Peptides: REMD Simulations and NMR Experiments. *J. Am. Chem. Soc.* **2007**, *129*, 4858–4859.
22. Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
23. Ponder, J. W. TINKER, Software Tools for Molecular Design, Version 4.2, **2004**.
24. Penev, E. S.; Lampoudi, S.; Shea, J.-E. TiReX: Replica-Exchange Molecular Dynamics Using TINKER. *Comp. Phys. Commun.* **2009**, *180*, 2013–2019.
25. Grimme, S. Accurate Description of van der Waals Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.
26. Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
27. Bouteiller, Y.; Pouilly, J.-C.; Desfrancois, C.; Grégoire, G. Evaluation of MP2, DFT, and DFT-D Methods for the Prediction of Infrared Spectra of Peptides. *J. Phys. Chem. A* **2009**, *113*, 6301–6307.
28. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169. For the current version, see <http://www.turbomole.com>.
29. MacAleese, L.; Simon, A.; McMahon, T. B.; Ortega, J. M.; Scuderi, D.; Lemaire, J.; Maitre, P. Mid-IR Spectroscopy of Protonated Leucine

- Methyl Ester Performed with an FTICR or a Paul Type Ion-Trap. *Int. J. Mass Spectrom.* **2006**, *249*, 14–20.
30. MacAleese, L.; Maitre, P. Infrared Spectroscopy of Organometallic Ions in the Gas Phase: From Model to Real World Complexes. *Mass Spectrom. Rev.* **2007**, *26*, 583–605.
 31. Bakker, J. M.; Besson, T.; Lemaire, J.; Scuderi, D.; Maitre, P. Gas-Phase Structure of a π -Allyl-Palladium Complex: Efficient Infrared Spectroscopy in a 7 T Fourier Transform Mass Spectrometer. *J. Phys. Chem. A* **2007**, *111*, 13415–13424.
 32. Bakker, J. M.; Sinha, R. K.; Besson, T.; Brugnara, M.; Tosi, P.; Salpin, J. Y.; Maitre, P. Tautomerism of Uracil Probed Via Infrared Spectroscopy of Singly Hydrated Protonated Uracil. *J. Phys. Chem. A* **2008**, *112*, 12393–12400.
 33. Stearns, J. A.; Mercier, S.; Seaiby, C.; Guidi, M.; Boyarkin, O.; Rizzo, T. R. Conformation-Specific Spectroscopy and Photodissociation of Cold, Protonated Tyrosine and Phenylalanine. *J. Am. Chem. Soc.* **2007**, *129*, 11814–11820.
 34. Bush, M. F.; O'Brien, J. T.; Prell, J. S.; Saykally, R. J.; Williams, E. R. Infrared Spectroscopy of Cationized Arginine in the Gas Phase: Direct Evidence for the Transition from Nonzwitterionic to Zwitterionic Structure. *J. Am. Chem. Soc.* **2007**, *129*, 1612–1622.
 35. Vaden, T. D.; de Boer, T. S. J. A.; Simons, J. P.; Snoek, L. C.; Suhai, S.; Paizs, B. Vibrational Spectroscopy and Conformational Structure of Protonated Polyalanine Peptides Isolated in the Gas Phase. *J. Phys. Chem. A* **2008**, *112*, 4608–4616.
 36. Stearns, J. A.; Seaiby, C.; Boyarkin, O.; Rizzo, T. R. Spectroscopy and Conformational Preferences of Gas-Phase Helices. *Phys. Chem., Chem. Phys.* **2009**, *11*, 125–132.
 37. Oomens, J.; Steill, J. D.; Redlich, B. Gas-Phase IR Spectroscopy of Deprotonated Amino Acids. *J. Am. Chem. Soc.* **2009**, *131*, 4310–4319.
 38. Prell, J. S.; O'Brien, J. T.; Steill, J. D.; Oomens, J.; Williams, E. R. Structures of Protonated Dipeptides: The Role of Arginine in Stabilizing Salt Bridges. *J. Am. Chem. Soc.* **2009**, *131*, 11442–11449.
 39. Prell, J. S.; Demireva, M.; Oomens, J.; Williams, E. R. Role of Sequence in Salt-Bridge Formation for Alkali Metal Cationized GlyArg and ArgGly Investigated with IRMPD Spectroscopy and Theory. *J. Am. Chem. Soc.* **2009**, *131*, 1232–1242.
 40. Polfer, N. C.; Oomens, J. Vibrational Spectroscopy of Bare and Solvated Ionic Complexes of Biological Relevance. *Mass Spectrom. Rev.* **2009**, *28*, 468–494.
 41. Polfer, N. C.; Paizs, B.; Snoek, L. C.; Compagnon, I.; Suhai, S.; Meijer, G.; von Helden, G.; Oomens, J. Infrared Fingerprint Spectroscopy and Theoretical Studies of Potassium Ion Tagged Amino Acids and Peptides in the Gas Phase. *J. Am. Chem. Soc.* **2005**, *127*, 8571–8579.
 42. Polfer, N. C.; Oomens, J.; Dunbar, R. C. Alkali Metal Complexes of the Dipeptides PheAla and AlaPhe: IRMPD Spectroscopy. *Chem. Phys. Chem.* **2008**, *9*, 579–589.
 43. Dunbar, R. C.; Steill, J. D.; Polfer, N. C.; Oomens, J. Peptide Length, Steric Effects, and Ion Solvation Govern Zwitterion Stabilization in Barium-Chelated Di- and Tripeptides Dunbar. *J. Phys. Chem. B* **2009**, *113*, 10552–10554.
 44. Cimas, A.; Vaden, T. D.; de Boer, T. S. J. A.; Snoek, L. C.; Gageot, M. P. Vibrational Spectra of Small Protonated Peptides from Finite Temperature MD Simulations and IRMPD Spectroscopy. *J. Chem. Theor. Comput.* **2009**, *5*, 1068–1078.

Chapitre 8

Évaluation de fonctionnelles de la densité pour la prédiction du spectre IR de G₈-Na⁺

Assessment of density functionals for predicting the infrared spectrum of sodiated octa-glycine

David Semrouni^a, Carine Clavaguéra^{a,*}, Jean-Pierre Dognon^b, Gilles Ohanessian^{a,*}

^a*Laboratoire des Mécanismes Réactionnels, Département de Chimie, Ecole Polytechnique, CNRS, 91128 Palaiseau Cedex, France*

^b*CEA/SACLAY, UMR 3299 CEA/CNRS SIS2M, Laboratoire de chimie de coordination des éléments f, F-91191 Gif-sur-Yvette, France*

Abstract

The sodiated peptide GGGGGGGG-Na⁺ or G₈-Na⁺ has a remarkable structure with a highly coordinated sodium ion and an acidic OH that is strongly hydrogen-bound with the N-terminus. The presence of the sodium ion makes this hydrogen bond unusually strong and makes proton transfer easy, leading to an equally stable, salt bridge isomer. The performances of a variety of density functionals in describing the geometries, energetics and infrared spectra of these two isomers were investigated. Usual density functionals were tested and moreover, more recent functionals such as dispersion-corrected ones and Truhlar's M06 series were also considered. The computed infrared spectra are compared with ab initio results and InfraRed Multiple Photon Dissociation (IRMPD) experiments. Two functionals in the M06 series have been proved to be quite efficient. A large number of functionals seems to be inadequate to compute infrared spectra for peptide in the amide N-H stretching region. In addition, a detailed analysis of the sodium-peptide interaction and of the hydrogen bond between the two peptide terminations points out a distinct electronic structure for the two isomers.

Keywords: density functional theory, infrared spectra, peptides

*Corresponding authors

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

1. Introduction

In the last years, InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy has emerged as an efficient way to obtain infrared signatures of gaseous molecular and versatile ions. It has been applied successfully to organic, inorganic and biological species, enabling the distinction between isomers and, in some instances between conformers [1, 2, 3, 4]. Since the species under scrutiny are gaseous ions rather than neutral molecules in the condensed phase, the large databases of infrared frequencies that have been gathered over the years, are useful indicators but can not be used as they stand. Assigning IRMPD bands to specific vibrational modes therefore relies heavily on molecular modeling. In most cases, this is done using quantum chemical methods based on the double harmonic approximation (mechanics and electrical). The purpose of the work described below is to provide a panorama of such methods for one specific test cases, sodiated octa-glycine $G_8\text{-Na}^+$, that has been characterized in detail recently [5].

Oligo-glycines G_n owe their value as model peptides to their relative simplicity due to the absence of side chains. Since Na^+ makes essentially electrostatic and polarization interactions with little or no charge transfer, small $G_n\text{-Na}^+$ complex involve peptide folding around the sodium ion to maximize interactions of amide oxygens with Na^+ [6, 7, 8]. In medium-size peptides, a delicate balance occurs between metal-peptide binding and internal stabilization of the peptide through multiple hydrogen bonds. As a result, the most stable structures are much more difficult to predict when the number of residues increases. In a previous study, we investigated the structure of the sodiated octa-glycine $G_8\text{-Na}^+$ [5]. IRMPD spectra have been obtained at room temperature in the $1000\text{-}1900\text{ cm}^{-1}$ and $2900\text{-}3700\text{ cm}^{-1}$ regions. The coupling of the polarizable force field AMOEBA and the replica-exchange molecular dynamics (REMD) method allowed for an efficient exploration of the potential energy surface. Geometries, energetics and IR spectra were obtained at quantum chemical levels including DFT with dispersion effects and MP2. $G_8\text{-Na}^+$ is a medium-size peptide and has the advantage of combining various non-covalent interactions such as a very strong hydrogen bond between the N- and the C-termini, seven-membered C_7 and ten-membered

C₁₀ motifs, and several sodium-amide interactions. Based on our previous results showing that only MP2/SVP calculations can match the IRMPD spectra, the main idea behind this work is to evaluate the ability of recent DFT methods to reproduce the experimental spectra.

The description of long-range interactions in term of geometry, energy and infrared spectrum remains a challenge of theoretical chemistry. On the one hand, the standard DFT functionals fails to describe dispersion interactions [9]. On the other hand, the MP2 theory provides an approximate treatment of dispersion, and remains tractable for systems involving up to c.a. 100 atoms, on the condition that the resolution of identity (RI) approximation is used. Several studies have been carried out to evaluate the ability of MP2 to describe energetic and structure but only few of them deal with vibrational frequency estimates [10, 11]. It has been shown that in some cases, MP2 overestimates the dispersion effects with extended basis sets as indicated by overly compact structures [10].

In the last years, the DFT-D approach has been developed by adding an empirical dispersion term for which the parameters are specific to each functional. It has been proved to be an efficient way to describe dispersion effects [12]. Recently, Truhlar’s group have proposed new functionals, called the M06 series, that have been parameterized to take into account various physical effects such as non-covalent interactions. They have been shown to provide a real improvement on the description of chemical properties in DFT [13]. M06-L is the meta-GGA local functional in the series, derived from the M05 one and including a self-interaction correction [14]. M06 is a meta hybrid functional with 27% of Hartree-Fock exchange for general purpose [15]. M06-2X and M06-HF are specific meta hybrid functionals with 54% and 100% of Hartree-Fock exchange, respectively. M06-2X has improved performance for thermochemistry [16] and M06-HF is recommended to provide the correct asymptotic behavior of the exchange-correlation potential for charge-transfer problems, for example [17]. The B2PLYP double-hybrid density functional has been also presented in which the semi-local GGA correlation is replaced by a standard MP2-type perturbation theory correction using Kohn-Sham orbitals and eigenvalues [18]. Moreover, long-range

dispersion effects are added by a R^{-6} classical potential in the B2PLYP-D functional [19]. However, these functionals are CPU time consuming and most of time limited to single-point energy calculations. Korth and Grimme have built "artificial" and randomly generated molecules containing 8 atoms to test a variety of density functionals [20]. The CCSD(T) / complete basis set level was taken as reference for energetics. They showed that the double-hybrid functionals such as B2PLYP-D perform best, followed by the M06-2X meta-hybrid functional, to predict relative energies for thermochemical benchmarking purpose. The assignment of infrared spectra of peptides in the gas phase was investigated by MP2 theory and DFT with and without dispersion corrections by Bouteiller et al. [21]. They pointed out that MP2 (with SVP basis set) outperforms a set of the density functionals with an average error of 5 cm^{-1} . B3LYP-D and B97-D functionals were also recommended for DFT vibrational frequency computations and specific scaling factors were proposed for each method and each modes (NH_2 , NH and CO stretches, NH_2 and NH in-plane bends). The theoretical treatment of dispersive forces was investigated by evaluating geometries, energies, and frequencies in the case of aromatic-aromatic interactions in isolated capped dipeptides by Gloaguen et al. [22]. The authors recommended the use of the B97-D functional to obtain reliable infrared spectra.

In the present work, we report our investigations in the assessment of a variety of density functionals for predicting the structure and the infrared spectrum of sodiated octa-glycine. Due to their close relative energies, both charge solvation (CS) with a hydrogen-bound acidic OH and salt bridge (SB) isomers are considered for relative energetics. Understanding of their electronic structure is obtained thanks to energy decomposition analysis, Hirshfeld and Voronoi Deformation Density (VDD) charges and electron localization function (ELF) analysis to determine explicitly the nature of the hydrogen bond.

2. Theoretical calculations

RI-MP2 calculations have been previously reported on this system[5]. Density functional theory (DFT) calculations were performed using various functionals with the TURBOMOLE program package[23, 24], except for the functionals of the M06 series. Geometry optimizations were performed at DFT levels using the resolution-of-the-identity[25, 26] (RI) approximation with the split valence plus polarization double- ζ (SVP) basis set. One GGA functional, i.e. PBE[27], one meta-GGA functional, i.e. TPSS[28] and three hybrid functionals, i.e. TPSSH[29], PBE0[30] and B3LYP[31] were used. Furthermore, B3LYP has been shown to give inaccurate energetics for G_8-Na^+ [32] and A_n-H^+ conformers[33]. Schwabe and Grimme have also shown the role of the lack of dispersion effects by observing an increase of the errors while the system sizes get larger[19]. Accordingly, four density functionals corrected by the empiric dispersion term proposed by Grimme[34] were used: B3LYP-D, PBE-D, TPSS-D and B97-D[35, 36], see (equations (1),(2)).

$$E_{DFT-D} = E_{DFT} + E_{disp} \quad (1)$$

$$E_{disp} = -s_6 \sum_j \sum_{i \neq j} f_{damp}(R_{ij}) \frac{C_{6(ij)}}{R_{ij}^6} \quad (2)$$

$$\text{with } f_{damp}(R_{ij}) = \frac{1}{1+e^{-d(R_{ij}/R_r-1)}} \text{ and } C_{6(ij)} = \sqrt{C_{6(i)}C_{6(j)}}$$

with $C_{6(ij)}$, the dispersion coefficient for atom pair ij and the damping function depending on the damping factor d , the interatomic distance R_{ij} and R_r , the sum of these atoms vdW radii.

RI-DFT and RI-MP2 energy calculations on the optimized geometries, and RI-CC2 energy calculations on the RI-MP2 geometries were performed with the valence plus polarization triple- ζ (TZVPP) basis set. Vibrational frequencies were calculated at the same level as the geometry optimization without the RI approximation. The calculated band intensities were convoluted assuming a Lorentzian profile with a full width at half maximum of 15 cm^{-1} . The vibrational frequencies were scaled by a factor depending on the method. The scaling factors are provided in Table 1. For MP2

and some functionals, one scaling factor can be applied in the entire spectral range and has been determined on the adjustment of the free stretch N-H band. For the other functionals, different scaling factors have to be used below 2000 cm^{-1} (named the fingerprint region) and above 3000 cm^{-1} . In these cases, the scaling factor used in the fingerprint region was determined such as to optimize the amide I - amide II frequency difference. Finally, the relative enthalpies and Gibbs free energies at 298 K were determined at the DFT/TZVPP//DFT/SVP, MP2/TZVPP//MP2/SVP and CC2/TZVPP//MP2/SVP levels of theory.

Geometry optimizations were also performed on both isomers using the M06-series functionals [14, 15, 16, 17], i.e. M06, M06-L, M06-2X and M06-HF, associated to the cc-pVDZ basis set, with the GAMESS program package[37, 38]. Single-point energy calculations were performed at the optimized geometries with the cc-pVTZ basis set.

Finally, geometries and vibrational frequencies were computed using the SSB-D functional with the DZP basis set in the ADF 2009.01 program package [39, 40]. The SSB-D functional is based on a small correction of the exchange part of the PBE functional combined with the gradient expansion exchange term from Keal and Tozer and Grimme’s dispersion correction. The SSB-D functional was designed for biological applications in order to improve the representation of hydrogen bonding, $\pi - \pi$ stacking, spin-state splitting, accuracy of geometries, reaction barriers.[41]

Energy decomposition analysis (EDA) based on the works by Morokuma [42, 43], and Ziegler and Rauk [44] was carried out using the ADF 2009.01 program package [39, 40] at the TURBOMOLE geometries with the M06-L functional. The interaction energy ΔE_{int} between the peptide and the sodium ion is decomposed into electrostatic, Pauli repulsion and orbital mixing components as in equation (3):

$$\Delta E_{int} = \Delta E_{elstat} + \Delta E_{Pauli} + \Delta E_{orb} \quad (3)$$

Slater-type orbitals (STOs) were employed as basis functions in SCF calculations. The TZ2P basis sets used have triple- ζ quality augmented by two sets of polarization functions. Detailed information on chemical bonding was provided with an electron localization function (ELF) analysis of the ADF results using the DGrid 4.5 program package [45].

3. Results and discussion

3.1. Structures and energetics of $G_8\text{-Na}^+$

The lowest energy CS and SB isomers were found to be the two most stable structures for the $G_8\text{-Na}^+$ complex. They are stabilized by a hydrogen bond between the N-terminus and the C-terminus[5]. The two structures obtained at the MP2/SVP level are shown in Figure 1. They are very similar to each other and the major difference is located on the nature of the terminations, neutral or charged. At the MP2 level of theory, we were able to locate the two isomers as minima on the potential energy surface. However, in most of the DFT calculations, it was impossible to locate one of the isomers. Indeed, only B3LYP-D, M06, M06-L and M06-HF functionals supplied both isomers. All the other functionals (i.e. PBE, TPSS, TPSSH, PBE0, B3LYP, PBE-D, TPSS-D, SSB-D and B97-D) provided only the SB structure with a systematic proton transfer occurring during the geometry optimization of the CS one. The M06-2X results are noteworthy with a proton transfer from SB to CS, locating only the CS isomer. To exclude a possible basis set effect due to the lack of diffuse functions, the aug-cc-pVDZ basis set was used in cases of B97-D and PBE-D functionals, however this not prevent the proton transfer from CS to SB. It is not possible to exclude strictly the existence of a second isomer in any of these cases, however our results indicate that it is at best a very shallow minimum. The fact that the most reliable density functionals agree with MP2 in predicting the existence of two minima lends confidence into this results, and point to deficiencies in other density functionals.

The description of the structures is given using the sequence numbering from the N- to the C-terminus with superscripts, i.e., NH^2 and CO^2 stand for

the peptidic N-H and C=O bonds of the second residue. For residue 1, the N-H bonds are in the NH₂ group rather than an amide group and for residue 8, CO⁸ is an acidic rather than an amide C=O. The structure of G₈-Na⁺ is characterized by a high coordination for the sodium ion by five peptidic carbonyl groups and the C-OH or C-O⁻ at the C-terminus. The acidic C=O is hydrogen-bound to the peptidic NH². A strong direct interaction exists between the N- and C-termini, either ⁺N-H...O⁻ (SB isomer) or N...H-O (CS isomer). In both isomers, the second N-H at the N-terminus is bound to CO⁴ and in the SB case, the third one is free. NH⁴ interacts through a hydrogen bond in a ten-membered C₁₀ motif with CO¹ and NH⁵ is involved in a hydrogen bond in a seven-membered C₇ motif with CO³. The other peptidic N-H (i.e. NH³, NH⁶, NH⁷ and NH⁸) are free. Relative energies, enthalpies and free energies at 298 K of the CS and SB isomers are given in Table 2 taking the CS isomer as the reference. In most cases, the SB isomer has the lowest energy. It is the case at the CC2/TZVPP level with a ΔG of -1.3 kcal/mol which we consider as the reference. MP2 values differ from the CC2 ones by less than 1 kcal/mol. The B3LYP-D and M06 series energies are consistent with the post-Hartree-Fock ones, giving confidence in these DFT calculations. B3LYP-D and M06-L energies are very close to the CC2 ones, while M06 gives CS as more stable than SB. All free energy differences are in a ± 2 kcal/mol window with respect to the CC2 values. We consider that both systems are about equally stable in the limit of the calculation accuracy.

3.2. Analysis of the electronic structure of the CS and SB isomers

Geometric, energetic and population analysis parameters are provided in Table 3 at the M06-L level. The total bonding energy is given with respect to the (G₈ + Na⁺) ionic dissociation limit (the peptide structure being the same as that in the sodium complex at equilibrium) with an absolute value larger for SB than for CS. Electrostatic interactions are slightly stronger in SB due to the presence of the charges on terminations corresponding to 77% of the total bonding energy in comparison with 74% for CS. Orbital interactions are of the same magnitude in both systems with 37% and 36% of

the total bonding energy for CS and SB respectively. No strong difference can be pointed out to distinguish the isomers because the interaction energy is a global property while their differences are highly localized: the chemical nature of the terminations does not have a strong influence on the total sodium-peptide binding energy. Other theoretical frameworks must be used to focus on the terminations such as population analyses. Hirshfeld and Voronoi Deformation Density (VDD) atomic charges are both based on the molecular electron density. They were proved to be less dependent upon basis sets than Mulliken charges and often give similar results[46]. The VDD calculates the flow of electron density between fragments due to bond formation. The population analyses on the N-terminus and C-terminus groups reflect their chemical differences. The global charges correspond to the sum of the atomic charges at the termination groups defined as $\text{CO}_2\text{H}/\text{NH}_2$ in the CS isomer and $\text{CO}_2^-/\text{NH}_3^+$ in the SB one. For SB, a clear charge separation is found with a positive charge for the N-terminus (+0.32 or +0.38 electron) and a negative charge for C-terminus (-0.47 or -0.37 electron). The terminations are close to neutral in CS with charges around +0.1 and -0.1. These DFT population frameworks are able to discriminate between the two isomers in spite of the very similar geometries and interactions with the sodium ion.

The electron localization function (ELF) provides a "picture" of the electronic structure showing the regions of the molecular space where the electron pairs localize. It is possible to split the molecular space into basins corresponding to the core electrons and to the shared and unshared (lone pair) electron density of the valence shell. The core basins are centered on the nuclei and the valence basins are in the remaining space. The notation typically used is $C(A)$ for the core basin of atom A, $V(A)$ for a lone pair of atom A, $V(A, B)$ for a two center bond between A and B [47]. The ELF function ranges from 0 (same spin pairs) to 1 (opposite spin pairs) and the local maxima of ELF define localization domains. In previous studies, a topological description of the hydrogen bond was proposed confirming that it is possible to evaluate the strength of the hydrogen bonds on the basis of the ELF criteria [48, 49]. The formation of the hydrogen bond changes

the properties of the ELF basins. It enables to discuss the bonding in both the CS and SB isomers. The CS isomer is characterized by the formation of an OH \cdots N bond between the hydrogen atom and the nitrogen lone pair. The hydrogen atom is bound to O and located in the V(O,H) basin which has a boundary with the nitrogen lone pair basin V(N). As a consequence, the formation of a V(N,H) basin is observed (top of Figure 2). For the SB isomer where a proton transfer occurs, the V(O,H) basin turns into a lone pair basin V(O) and the hydrogen atom is relocated in a V(N,H) basin with a boundary with V(O). A V(O,H) basin is created as an indication of the presence of a hydrogen bond with the oxygen atom (bottom of Figure 2).

The smallest value of the ELF maximum in SB (0.90) with respect to the CS one (0.95) corresponds to the delocalization of the spin-paired electron density. The ELF distribution along the O \cdots H \cdots N bond path shows lower values for the SB isomer (minima: 0.32) than for the CS isomer (minima: 0.46). This can be observed in the 2D picture of the cut plane through the O \cdots H \cdots N atoms of interest for both isomers in Figure 2. Although relatively small, these differences remain significant as already discussed in the literature (see ref. [49] for example). So, we can expect a weaker hydrogen bond in the SB isomer. These results indicate that the two isomers have clearly distinct electronic properties and are not just two numerical minima on a very flat portion of the potential energy surface.

3.3. Spectrum analysis

Comparison of computed spectra for the SB

The SB isomer is available for almost all the functionals and the resulting spectra for this structure are provided in Figures 3, 4 and 5. It was already demonstrated that the structure of G₈-Na⁺ cannot be a salt bridge on the basis of the IRMPD spectrum[5] and MP2 results. Several significant bands were used to discriminate between the two isomers: the presence of intense bands due to HNH bending modes of the ammonium in between the experimental amide I and II bands and two intense bands for N-H stretches of NH₃⁺ that are hydrogen-bonded to carbonyl oxygens in the 3000-3200 cm⁻¹ area. Although this structure is not consistent with IRMPD results, it is the one

which provides the largest data for the assessment of density functionals to reproduce infrared spectra. The MP2 spectrum has been taken as reference to establish the performance of a large diversity of functionals; the IRMPD bands are also shown in order to make visual comparison easier. Because of the breadth of the amide I and II bands for an octa-peptide, most functionals yield spectra that are in relatively good agreement with the MP2 reference in the fingerprint region. The higher sensitivity of frequencies to hydrogen bonding in the N-H stretch region makes this region more revealing.

The B3LYP-D and MP2 spectra are very similar (Figure 3), the major difference exists in the N-H stretch range with a red-shift of the intense NH_3^+ stretch bands due to stronger hydrogen bonds with carbonyl groups. This shift involves an artificial coupling of one band with C-H stretch modes, resulting in an intense doublet just below 3000 cm^{-1} . The B3LYP spectrum is comparable to the B3LYP-D one with a larger red-shift of the NH_3^+ bands giving a coupling of both bands with C-H stretch modes (Figure 4) and a small red-shift of the H-bound peptidic N-H bands. These results point out the role of dispersion effects, treated here by including an empirical correction in the popular B3LYP functional. The B97-D spectrum (Figure 4) matches the B3LYP one with a red-shift of the H-bound N-H stretch modes, particularly for one band situated outside of the IRMPD band. This result emphasizes some limitations of this dispersion-corrected GGA functional possibly due to the absence of Hartree-Fock exchange.

PBE, PBE-D and PBE0 spectra (Figure 4) are very similar to each other, especially for the amide I and amide II range with a small red-shift of the amide II band. For the three functionals, the N-H stretch modes of the ammonium are red-shifted between 2900 and 3050 cm^{-1} with a minor improvement in PBE-D. Moreover, the H-bound N-H band is not in agreement with the MP2 one, even if the position of the band is slightly better in PBE0 thanks to the presence of Hartree-Fock exchange. We can conclude that this series of functionals is ineffective to reproduce the vibrational frequencies in the N-H stretch region with a unique scaling factor.

TPSS, TPSS-D, TPSSH and SSB-D spectra (Figure 5) are very similar to each other in the fingerprint region and in agreement with the MP2 spec-

trum. In the N-H/O-H stretch range, the TPSS spectrum presents the same features than the PBE one. The TPSS-D and TPSSH spectra gives a slight better agreement thanks to the inclusion of dispersion effects or Hartree-Fock exchange. As the SSB-D functional was adjusted on hydrogen-bound systems, the spectrum presents an improvement in the reproduction of the ammonium N-H stretch modes in comparison with MP2 results. For all these functionals, as for the PBE series, the position of the free N-H stretch band is well reproduced but not for the H-bound N-H modes. They cannot be recommended for frequency calculations in this range.

In the M06 series, the M06 spectrum shows a red-shift of the amide II band and different shifts of the N-H stretch bands of NH_3^+ . The M06-L spectrum is very similar to the MP2 one in all the spectral range. The M06-HF functional provides a good agreement in the fingerprint region but a red-shift of the H-bound N-H stretch modes. In this series, the M06-L gives the best agreement moderated by the spread of the free N-H band with a red-shift about 60 cm^{-1} for the weakly-bound band.

In conclusion, these results highlight the sensitivity of DFT vibrational frequency calculations to the inclusion of Hartree-Fock exchange or dispersion effects. Some functionals built to study hydrogen-bond complexes are not able to reproduce the MP2 data even for some of the weakly bound N-H stretches for such an uncommon system. The recent meta-GGA M06-L functional provides encouraging results in comparison with the MP2 ones.

In addition to the spectra shown in Figures 3-5, the very strong hydrogen bond between one of the N-H bonds at the ammonium terminus and one of the oxygens at the carboxylate terminus leads to a spectacular red-shift of more than 1000 cm^{-1} at some levels. Figure 6 presents the computed stretch frequency for this ammonium N-H as a function of the $\text{NH}\cdots\text{O}$ hydrogen bond distance. It is clear that there is a strong correlation between these two quantities. DFT-D and MP2 results provide short distances with a large red-shift for the N-H stretch band. B3LYP and other standard functionals are intermediate. On the contrary, M06 and M06-L give a longer hydrogen bond, and as a consequence, a weaker interaction with a N-H mode between 2250 and 2350 cm^{-1} . This behavior points out the difficulty to evaluate the

position of this red-shifted band by standard harmonic calculations. Furthermore, adding the dispersion term involves a smaller blue-shift of the other two $^+\text{N-H}$ bands, due to a weaker interaction of the N-terminus with CO^4 . This can be seen by comparing the spectra of TPSS and TPSS-D, B3LYP and B3LYP-D, and with a lesser effect, PBE and PBE-D. Other correlations were attempted between the N-H bond distance and the respective stretch mode frequency for free peptidic NH groups, and also between the $\text{Na}\cdots\text{O}$ distance and the concerned C=O or N-H stretch frequencies. However, the variations in these distances are too small to be significant and no clear new correlation was identified.

Comparison of computed spectra for the CS

As described in section 3.1, the CS isomer could be found with only five functionals. The resulting IR spectra are given in Figure 7. The MP2 spectrum of CS was proved to be in very good agreement with the IRMPD spectrum[5]. The amide I and amide II massifs are clearly identifiable in the 1650-1750 and 1480-1560 cm^{-1} ranges, respectively. The IRMPD band in the 3460-3490 cm^{-1} range is characteristic of several free N-H modes. The experimental broad band at 3340-3440 cm^{-1} corresponds to weakly H-bound N-H or NH_2 bands. The B3LYP-D spectrum presents an intense band in between the amide I and II bands corresponding to a strong mixing of the O-H stretch and the carboxylic C=O stretch modes. Moreover, a red-shift of the H-bound N-H stretch modes is observed in worse agreement with the IRMPD spectrum than the MP2 one. In the M06 series, the free N-H bands are spread out but mainly around the experimental position. The M06 spectrum displays the same problem as for SB with a red-shift of the amide II band. The rest of the spectrum is well reproduced. The M06-2X functional, which was unable to locate the SB isomer, is in good agreement with the IRMPD spectrum except for one C=O stretch band outside the amide I band. The M06-HF spectrum is also in rather good agreement and the unique failure is the red-shift of the H-bound N-H stretch band by 50 cm^{-1} as already seen for the SB spectrum. As for the SB case, the M06-L spectrum is in good agreement with the IRMPD spectrum, nearly as good

as the MP2 one except for a spread of the free N-H bands. This functional gives also the best agreement with the MP2 data in the SB case.

4. Conclusions

Some successes and failures of DFT methods to study the structure, energetics and spectroscopy of peptides were presented in the case of the sodiated octa-glycine peptide. A large number of functionals are not able to locate both the charge solvated and the salt bridge isomers of lowest energy. Only B3LYP-D and the M06 series (except for M06-2X) of functionals succeed due to a combination of Hartree-Fock exchange and dispersion effects in the first case, and a diverse set of data in the parameterization for the new meta-GGA derived functionals. DFT energy decomposition analysis at the M06-L level attests that the major part of the bonding energy between the peptide and the sodium ion comes from electrostatic interaction. Population analyses based on the electron density allow to differentiate SB from CS based on realistic values of the termination charges. The ELF analysis of M06-L densities results shows a large stability due to the strong hydrogen bond in the CS structure. The M06 series performs best among density functionals to reproduce the spectra in comparison with MP2 and IRMPD spectra. The best results are obtained with the M06-L GGA functional giving the CS and SB structures close in energy and a good agreement with the experimental spectrum. The PBE and TPSS series are not able to give the good positions of the weakly hydrogen-bound N-H stretch bands and cannot be recommended to compute IR spectra for peptides in this region.

Acknowledgments

This work was granted access to the HPC resources of [CCRT/CINES/IDRIS] under the allocation x2009085107 made by GENCI (Grand Equipement National de Calcul Intensif).

References

- [1] L. MacAleese, P. Maître, Infrared spectroscopy of organometallic ions in the gas phase: from model to real world complexes, *Mass Spectrom. Rev.* 26 (4) (2007) 583–605.
- [2] K. R. Asmis, J. Sauer, Mass-selective vibrational spectroscopy of vanadium oxide cluster ions, *Mass Spectrom. Rev.* 26 (4) (2007) 542–562.
- [3] N. C. Polfer, J. Oomens, Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance, *Mass Spectrom. Rev.* 28 (3) (2009) 468–494.
- [4] T. D. Fridgen, Infrared consequence spectroscopy of gaseous protonated and metal ion cationized complexes, *Mass Spectrom. Rev.* 28 (4) (2009) 586–607.
- [5] D. Semrouni, O. P. Balaj, F. Calvo, C. F. Correia, C. Clavaguéra, G. Ohanessian, Structure of Sodiated Octa-Glycine: IRMPD Spectroscopy and Molecular Modeling, *J. Am. Soc. Mass Spectrom.* 21 (2010) 728–738.
- [6] M. M. Kish, C. Wesdemiotis, G. Ohanessian, The Sodium Ion Affinity of Glycylglycine, *J. Phys. Chem. B* 108 (9) (2004) 3086–3091.
- [7] P. Wang, C. Wesdemiotis, C. Kapota, G. Ohanessian, The Sodium Ion Affinities of Simple Di-, Tri-, and Tetrapeptides, *J. Am. Soc. Mass Spectrom.* 18 (3) (2007) 541–552.
- [8] O. P. Balaj, C. Kapota, J. Lemaire, G. Ohanessian, Vibrational signatures of sodiated oligopeptides (GG-Na^+ , GGG-Na^+ , AA-Na^+ and AAA-Na^+) in the gas phase, *Int. J. Mass Spectrom.* 269 (3) (2008) 196–209.
- [9] K. E. Riley, M. Pitoňák, P. Jurečka, P. Hobza, Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories, *Chem. Rev. ASAP* (2010) doi: 10.1021/cr1000173.

- [10] J. Cerny, P. Hobza, Non-covalent interactions in biomacromolecules, *Phys. Chem. Chem. Phys.* 9 (2007) 5291–5303.
- [11] K. E. Riley, P. Hobza, Assessment of the MP2 Method, along with Several Basis Sets, for the Computation of Interaction Energies of Biologically Relevant Hydrogen Bonded and Dispersion Bound Complexes, *J. Phys. Chem. A* 111 (33) (2007) 8257–8263.
- [12] T. Schwabe, S. Grimme, Theoretical Thermodynamics for Large Molecules: Walking the Thin Line between Accuracy and Computational Cost, *Acc. Chem. Res.* 41 (4) (2008) 569–579.
- [13] Y. Zhao, D. G. Truhlar, Density Functionals with Broad Applicability in Chemistry, *Acc. Chem. Res.* 41 (2) (2008) 157–167.
- [14] Y. Zhao, D. G. Truhlar, A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions, *J. Chem. Phys.* 125 (19) (2006) 194101–18.
- [15] Y. Zhao, D. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals, *Theo. Chem. Acc.* 120 (1) (2008) 215–241.
- [16] Y. Zhao, D. G. Truhlar, Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions, *J. Chem. Theory Comput.* 4 (11) (2008) 1849–1868.
- [17] Y. Zhao, D. G. Truhlar, Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States, *J. Phys. Chem. A* 110 (49) (2006) 13126–13130.

- [18] S. Grimme, Semiempirical hybrid density functional with perturbative second-order correlation, *J. Chem. Phys.* 124 (3) 034108.
- [19] T. Schwabe, S. Grimme, Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability, *Phys. Chem. Chem. Phys.* 9 (2007) 3397–3406.
- [20] M. Korth, S. Grimme, “Mindless” DFT Benchmarking, *J. Chem. Theory Comput.* 5 (4) (2009) 993–1003.
- [21] Y. Bouteiller, J. C. Pouilly, C. Desfrancois, G. Grégoire, Evaluation of MP2, DFT, and DFT-D Methods for the Prediction of Infrared Spectra of Peptides, *J. Phys. Chem. A* 113 (22) (2009) 6301–6307.
- [22] E. Gloaguen, H. Valdes, F. Pagliarulo, R. Pollet, B. Tardivel, P. Hobza, F. Piuzzi, M. Mons, Experimental and Theoretical Investigation of the Aromatic-Aromatic Interaction in Isolated Capped Dipeptides, *J. Phys. Chem. A* 114 (9) (2009) 2973–2982.
- [23] R. Ahlrichs, M. Bar, M. Haser, H. Horn, C. Kolmel, Electronic-structure calculations on workstation computers - The program system TURBO-MOLE, *Chem. Phys. Lett.* 162 (3) (1989) 165–169.
- [24] C. Hattig, F. Weigend, CC2 excitation energy calculations on large molecules using the resolution of the identity approximation, *J. Chem. Phys.* 113 (13) (2000) 5154–5161.
- [25] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, Auxiliary basis sets to approximate Coulomb potentials, *Chem. Phys. Lett.* 242 (6) (1995) 652 – 660.
- [26] K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs, Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials, *Theo. Chem. Acc.* 97 (1997) 119–124.
- [27] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple, *Phys. Rev. Lett.* 77 (1996) 3865–3868.

- [28] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids, *Phys. Rev. Lett.* 91 (14) (2003) 146401.
- [29] V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes, *J. Chem. Phys.* 119 (23) (2003) 12129–12137.
- [30] J. P. Perdew, M. Ernzerhof, K. Burke, Rationale for mixing exact exchange with density functional approximations, *J. Chem. Phys.* 105 (1996) 9982–9985.
- [31] A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.* 98 (7) (1993) 5648–5652.
- [32] D. Semrouni, G. Ohanessian, C. Clavaguéra, Structural, energetic and dynamical properties of sodiated oligoglycines: relevance of a polarizable force field, *Phys. Chem. Chem. Phys.* 12 (2010) 3450–3462.
- [33] T. D. Vaden, T. S. J. A. de Boer, J. P. Simons, L. C. Snoek, S. Suhai, B. Paizs, Vibrational Spectroscopy and Conformational Structure of Protonated Polyalanine Peptides Isolated in the Gas Phase, *J. Phys. Chem. A* 112 (20) (2008) 4608–4616.
- [34] S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion contribution, *J. Comput. Chem.* 27 (15) (2006) 1787–1799.
- [35] A. D. Becke, Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals, *J. Chem. Phys.* 107 (20) (1997) 8554–8560.
- [36] S. Grimme, Accurate description of van der waals complexes by density functional theory including empirical corrections, *J. Comput. Chem.* 25 (12) (2004) 1463–1473.

- [37] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, General Atomic and Molecular Electronic Structure System, *J. Comput. Chem.* 14 (1993) 1347–1363.
- [38] M. S. Gordon, M. W. Schmidt, *Advances in electronic structure theory: GAMESS a decade later*, Elsevier, Amsterdam, 1167–1189, 2005.
- [39] ADF2009.01, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, <http://www.scm.com>, 2009.
- [40] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. F. Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler, *Chemistry with ADF*, *J. Comput. Chem.* 22 (2001) 931–967.
- [41] M. Swart, M. Sola, F. M. Bickelhaupt, A new all-round density functional based on spin states and S_N2 barriers, *J. Chem. Phys.* 131 (9) (2009) 094103.
- [42] K. Morokuma, *Molecular Orbital Studies of Hydrogen Bonds. III*, *J. Chem. Phys.* 55 (3) (1971) 1236–1244.
- [43] K. Kitaura, K. Morokuma, A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation, *Int. J. Quantum Chem.* 10 (2) (1976) 325–340.
- [44] T. Ziegler, A. Rauk, On the calculation of bonding energies by the Hartree Fock Slater method, *Theor. Chem. Acta* 46 (1977) 1–10.
- [45] M. Kohout, Program DGrid, version 4.5, Max-Planck Institut für Chemische Physik fester Stoffe, Dresden, 2009.
- [46] C. F. Guerra, J. W. Handgraaf, E. J. Baerends, F. M. Bickelhaupt, Voronoi deformation density (VDD) charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD methods for charge analysis, *J. Comput. Chem.* 25 (2) (2004) 189–210.

- [47] S. Noury, B. Silvi, R. J. Gillespie, Chemical Bonding in Hypervalent Molecules: Is the Octet Rule Relevant?, *Inorg. Chem.* 41 (8) (2002) 2164–2172.
- [48] M. E. Alikhani, F. Fuster, B. Silvi, What can tell the topological analysis of ELF on hydrogen bonding?, *Structural Chemistry* 16 (3) (2005) 203–210.
- [49] I. Drebuschak, S. Kozlova, Hydrogen bond in the $\text{H}_3\text{O}(\text{Ph}_3\text{PO})^{3+}$ complex. Features of the electron density distribution, *J. Struct. Chem.* 51 (1) (2010) 166–169.

Table 1: Scaling factors used for the two spectral ranges (i.e 1100-1800 cm^{-1} and 2900-3600 cm^{-1}) at MP2 and various DFT levels

| Method | 1100-1800 cm^{-1} | 2900-3600 cm^{-1} |
|---------------|----------------------------|----------------------------|
| MP2 | | 0.943 |
| B3LYP/B3LYP-D | | 0.960 |
| B97-D | 0.985 | 0.975 |
| PBE/PBE-D | 0.985 | 0.980 |
| PBE0 | 0.940 | 0.945 |
| TPSS/TPSS-D | 0.985 | 0.980 |
| TPSSH | 0.960 | 0.965 |
| SSB-D | 0.975 | 0.968 |
| M06-L | | 0.958 |
| M06 | 0.955 | 0.960 |
| M06-2X | | 0.950 |
| M06-HF | 0.945 | 0.950 |

Table 2: Energies, enthalpies and free energies at 298 K (in kcal/mol) of the SB relative to the CS isomer at various quantum chemistry levels.

| Method | ΔE | ΔH | ΔG |
|------------------|------------|------------|------------|
| CC2 ^a | -2.2 | -1.9 | -1.3 |
| MP2 | -1.4 | -1.1 | -0.5 |
| B3LYP-D | -0.8 | -0.5 | -1.5 |
| M06 | +0.6 | +1.6 | +0.5 |
| M06-L | -0.8 | -0.5 | -0.6 |
| M06-HF | -2.1 | -2.1 | -2.7 |

^a at MP2 geometries

ΔG calculated using MP2 frequencies

Table 3: Characteristic distances for the N-terminus/C-terminus interaction in Å, bonding Energy (BE) analysis without relaxation of the G₈ structure at dissociation (All values in kcal/mol). Pauli + Electrostatic + Orbital = BE, and population analysis.

| | CS M06-L | SB M06-L |
|-----------------|----------|----------|
| d(O-H) | 1.056 | 1.551 |
| d(N-H) | 1.584 | 1.094 |
| BE | -106.5 | -110.1 |
| Pauli repulsion | 11.0 | 14.4 |
| Electrostatic | -78.6 | -85.1 |
| Orbital | -38.9 | -39.4 |
| VDD Nter | 0.13 | 0.32 |
| VDD Cter | -0.06 | -0.47 |
| Hirshfeld Nter | 0.12 | 0.38 |
| Hirshfeld Cter | -0.07 | -0.37 |
| ELF attractor | 0.95 | 0.90 |

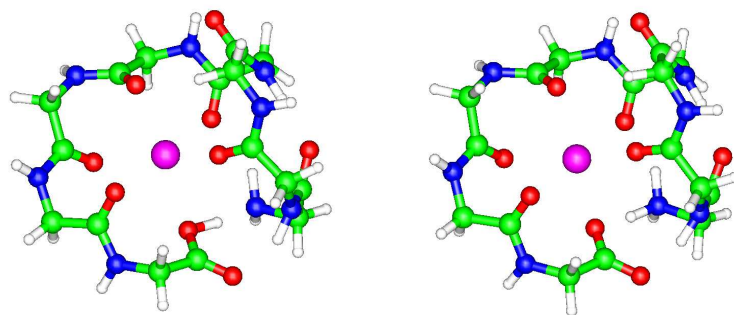


Figure 1: CS (left side) and SB (right side) MP2 geometries

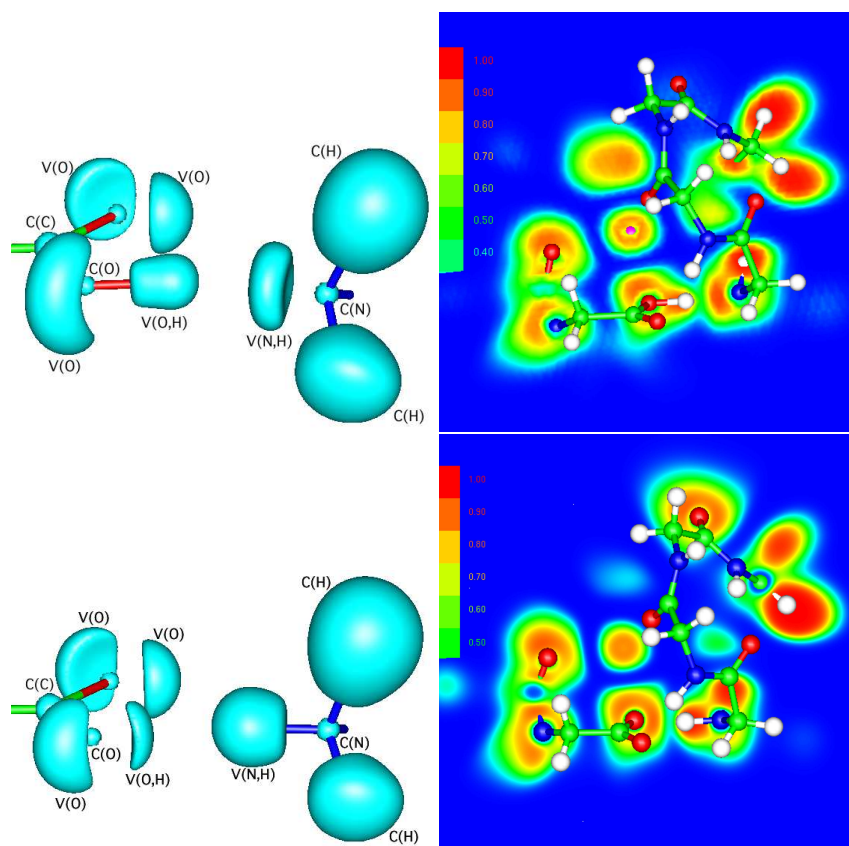


Figure 2: ELF localization domains and cut-plane ELF representations for CS (top) and SB (bottom) isomers

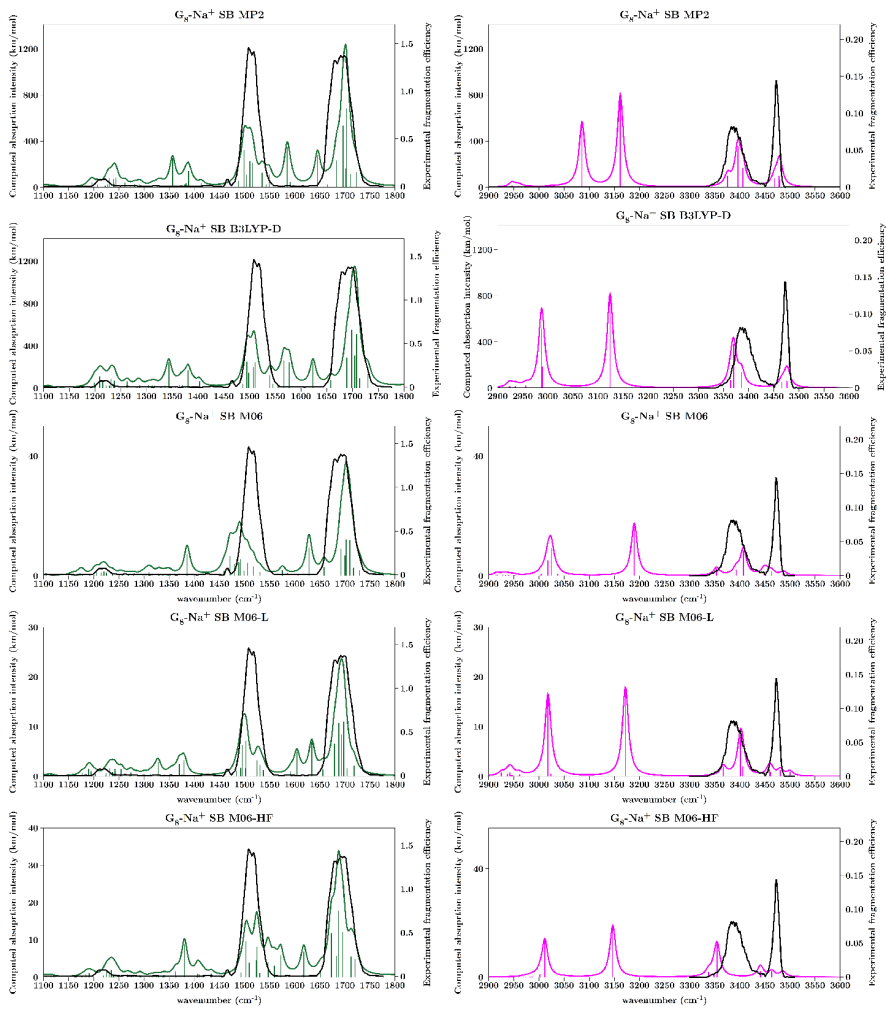


Figure 3: $G_8\text{-Na}^+$ SB spectra at MP2, B3LYP-D, M06, M06-L and M06-HF levels.

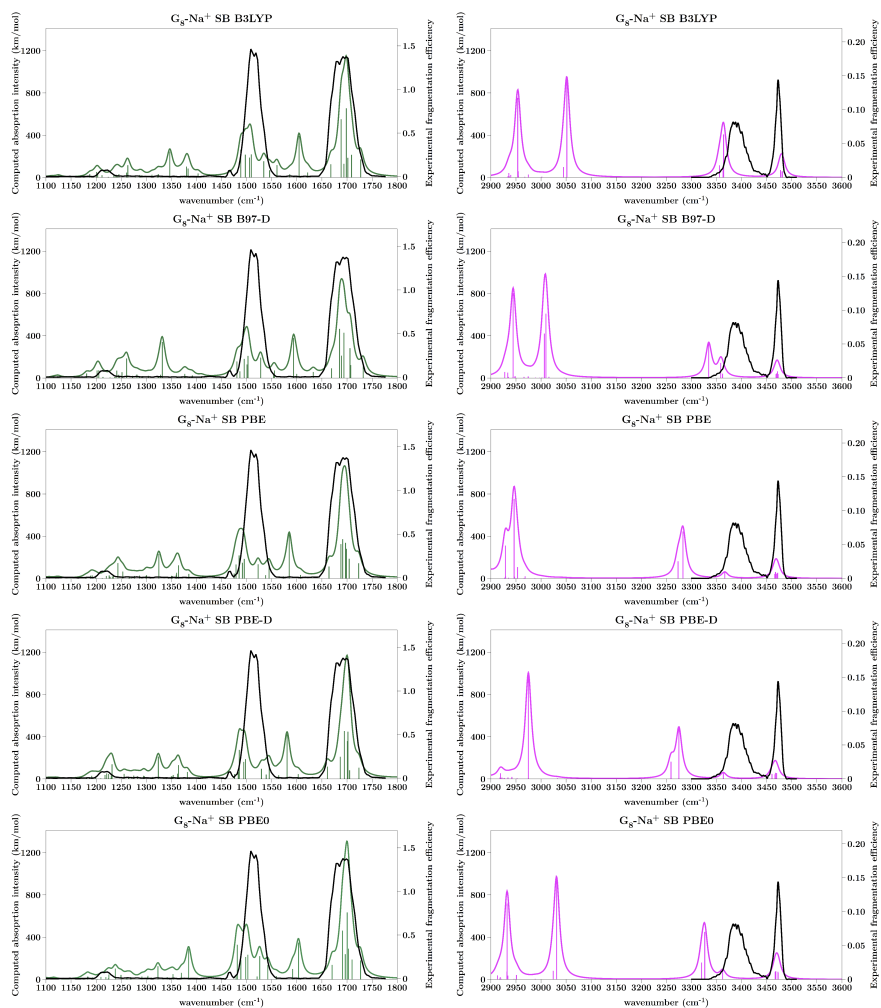


Figure 4: $G_8\text{-Na}^+$ SB spectra at B3LYP, B97-D, PBE, PBE-D and PBE0 levels.

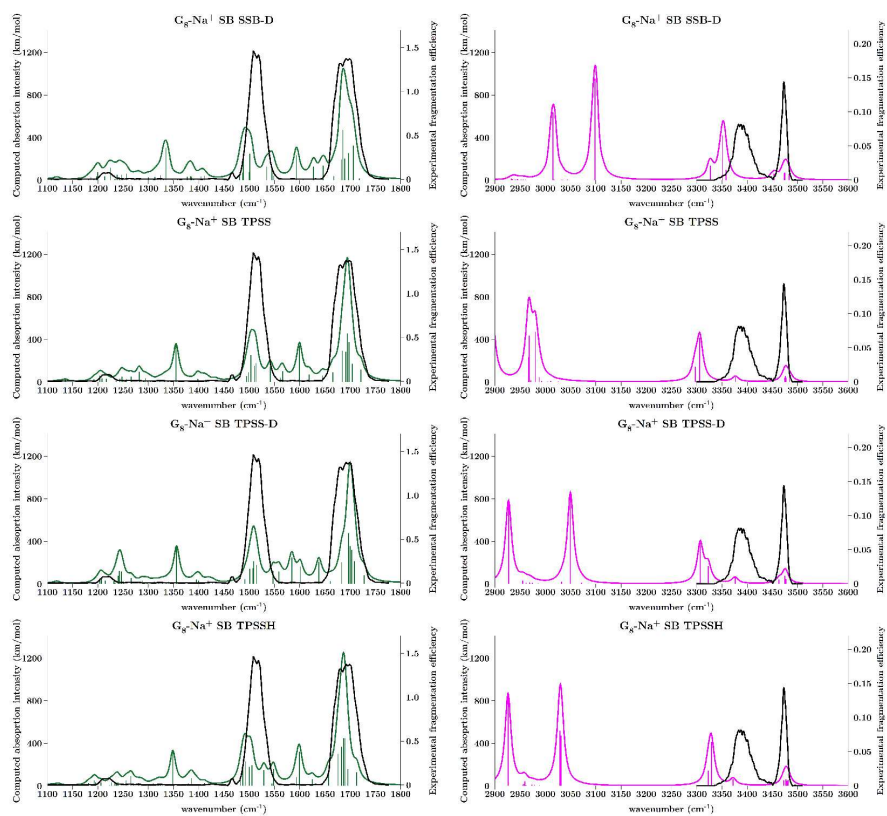


Figure 5: G_8-Na^+ SB spectra at SSB-D, TPSS, TPSS-D and TPSSH levels.

Chapitre 9

Spectroscopie IRMPD et modélisation de polyglycines sodées

STRUCTURE OF SODIATED POLYGLYCINES

O. Petru Balaj,^{a,†} David Semrouni,^a Joël Lemaire,^b
Edith Nicol,^a Carine Clavaguéra^{*a}, Gilles Ohanessian^{*a}

^a : Laboratoire des Mécanismes Réactionnels, Ecole Polytechnique and CNRS, 91128 Palaiseau Cedex, France

^b : Laboratoire de Chimie Physique, Université de Paris-Sud 11 and CNRS, 91405 Orsay Cedex, France

Running title : IRMPD spectroscopy and modeling of sodiated polyglycines

[†] Present address : Institut für Physikalische Chemie der Christian-Albrechts-Universität zu Kiel, Olshausenstrasse 40, D-24098 Kiel, Germany

Abstract

The folding of peptides about a sodium ion in the absence of solvent has been investigated in detail using InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy and a combination of theoretical methods. IRMPD was carried out on the sodiated polyglycines $G_n\text{-Na}^+$ ($n=2-8$), in both the fingerprint and N-H/O-H stretching regions. Potential energy surfaces were explored using the Replica-Exchange Molecular Dynamics (REMD) method and the polarizable force field AMOEBA. Geometries, energetics and computed IR spectra were further refined at the MP2 quantum chemical level with coupled cluster calculations being used to generate reference data for energies. It is this combination of experimental and computational approaches that enables deciphering of most structural details. The most stable structures of the small peptides up to $G_6\text{-Na}^+$ maximize the metal-peptide interaction with all peptidic C=Os bound to sodium, implying strain that is maximum for $G_6\text{-Na}^+$. The increased flexibility of larger peptides leads to more complex folding involving smaller strain with six Na^+ binding sites and internal peptide structuration via γ or β turns. This structural transition observed between $G_6\text{-Na}^+$ and $G_7\text{-Na}^+$ leads to structures with a sodium coordination that becomes tri-dimensional and thus closer to that of sodium sites in proteins. The most favourable hydrogen bonding pattern is found to involve direct interaction between the peptide termini, opening the way to salt bridge formation. While accurate quantum chemical calculations predict the salt bridge and charge solvation structures to become essentially degenerate for $G_8\text{-Na}^+$, there is good agreement between experiment and computations in the entire spectral range for the charge solvation isomer only.

Introduction

InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy has recently become an invaluable tool for obtaining insight into the structures of gaseous ions, be them organic, organometallic, inorganic or biological [1-7]. Mass spectrometric methods provide very high sensitivity for gaseous ion detection and tandem mass spectrometry [8] often provides significant structural evidence for small molecules and have been successfully applied to peptide sequencing [9]. While distinction between isomers is often made possible, the 3D structures can hardly be probed by fragmentation patterns so that more direct and general access to structural information was desirable. The development of IRMPD spectroscopy within a mass spectrometer has turned out to be just such a tool, yielding vibrational signatures that are similar to those obtained by more traditional IR spectroscopy. Gaseous ions are formed and mass-selected in order to record a spectrum for each species of interest. Since the number density of ions in the high vacuum of a mass spectrometer is much too small to record absorption spectra, action spectroscopy is used instead. Single or multiple photon absorption provides enough energy to trigger ion fragmentation when the photon energy is in resonance with a vibrational band of the molecular ion. Recording the fragment ion signal(s) as a function of photon energy, taking advantage of the high sensitivity of ion detection, thus provides access to the vibrational spectrum. The structures of a number of molecular ions have been established in this manner. IRMPD has also been used successfully to obtain structural information on gaseous peptides of growing size [10-13] and on proteins [14-15], however full structural characterization has remained difficult.

In the work described herein, we have used IRMPD spectroscopy to obtain fingerprints of the functional groups present in sodiated polyglycines $G_n\text{-Na}^+$ bearing two to eight residues ($n=2-8$). Because of the biological importance of the sodium ion and also because of their value as model peptides, sodiated oligo-glycines have been used in the last decade as a valuable testing ground for new experimental developments designed to obtain refined energetic and/or structural data for peptides in the gas phase. These include ion mobility measurements for global shape information [17,18], H/D exchange extent and kinetics for isomeric/conformational content [19], the kinetic method [20,21]

and the threshold collision induced decomposition [22] methods for thermochemical measurements, as well as IRMPD spectroscopy for identification of the functional groups as perturbed by specific environments [23, 24]. Most of these studies (except for ion mobility) have been confined to the smaller members of the series, a limitation which we overcome here.

We consider the series of $G_n\text{-Na}^+$ ($n=2-8$) as a tool to improve the connexion between IR spectra and structures by probing the gradual changes of the vibrational signatures as the peptide size grows larger. This approach has been followed recently for $G_n\text{-H}^+$ ($n=1-5$) [13] in the 1000-2000 cm^{-1} region and $A_n\text{-H}^+$ ($n=3-5,7$) [16] in the 2600-3700 cm^{-1} region. In the present work, we aim at making this a more powerful approach by combining the two spectral ranges, 1000-1900 and 3000-3700 cm^{-1} . This allows for probing fingerprint as well as N-H/O-H stretch responses. Interpretation of the results required extensive molecular modeling using a variety of methods, ranging from extensive structure sampling with a polarizable force field to quantum chemical calculations using density functional theory and the ab initio MP2 and CC2 methods.

We first describe the computational and experimental methods used. Then we describe the IRMPD spectra and computed IR absorption spectra and combine them to ascribe structures to gaseous $G_n\text{-Na}^+$ at 298 K and discuss the trends in structures from two to eight residues.

Computational Methods

Replica-Exchange Molecular Dynamics with the AMOEBA force field

The methods used have been described in details recently for the case of $G_8\text{-Na}^+$ [24] and are only briefly recalled below. For $G_4\text{-Na}^+$ and $G_5\text{-Na}^+$, previous conformational searches were used to select low-energy structures [21, 25]. For $G_6\text{-Na}^+$ and $G_7\text{-Na}^+$, both the salt-bridge and charge-solvated isomers have been modeled using the polarizable AMOEBA force field [26], which was shown to provide reliable structural and energetic properties for sodiated oligoglycine peptides [27]. The conformational landscapes were explored using molecular dynamics (MD) simulations, improved with the replica-

exchange (RE) strategy. The REMD approach is now widely used as a powerful tool for sampling the potential energy surface of biomolecular systems [12, 24, 28-31]. Briefly, the idea is to perform simultaneously several MD trajectories at various temperatures called replicas, occasionally attempting an exchange between two configurations of neighboring temperatures, and accepting this exchange with a Metropolis probability. REMD simulations for G₆-Na⁺ and G₇-Na⁺ have been performed with 8 replicas allocated according to a geometric progression in the range 200—600 K. This allows combining high temperatures to overcome energy barriers and low temperatures to explore the low energy regions in detail. This procedure was started from both charge solvation (CS) and salt bridge (SB) structures since different force field parameters must be used for the different functional groups: NH₂/CO₂H for CS vs. NH₃⁺/CO₂⁻ for SB [27]. More details on the REMD simulations are provided as supporting information.

The present simulations with the AMOEBA force field were performed with the Tinker software package [32] interfaced with TiReX [33].

Quantum chemical calculations

The most stable configurations collected during the REMD explorations were further subjected to electronic structure calculations to determine the best candidates for comparison to experiment. Geometry optimizations were performed at the DFT-D level with the B3LYP functional and the SVP basis set. The B3LYP functional has been shown to be unreliable for relative energies of G₈-Na⁺ conformers [34]. Thus we used the empirical correction for dispersion interactions proposed by Grimme for density functional calculations [35,36]. The resolution-of-the-identity (RI) approximation was employed for molecular orbital two-particle integrals. On the basis of relative energies computed at the RI-B3LYP-D/SVP level, infrared spectra were calculated at the RI-MP2/SVP level after geometry re-optimization. The MP2 vibrational frequencies were scaled by a factor of 0.943 in the entire spectral range [24]. While it is often considered that different scaling factors should be used below 2000 and above 3000 cm⁻¹ [37], reasonable matching was found to occur for all peptides with the single scaling factor previously optimized for G₈-Na⁺ [24]. In order to further check this result, additional calculations of vibrational spectra were carried out at the RI-MP2/SVP level for G₂-Na⁺ and the scaling factor was

determined so as to obtain good agreement with the experimental band at 3557 cm^{-1} , since this band is narrow and can only correspond to the O-H stretch vibration. The good agreement between scaling factors determined by two independent procedures on two different peptides reinforces our conclusion that a single scaling factor is sufficient at the MP2/SVP level. The calculated band intensities were convoluted assuming a Lorentzian profile with a full width at half maximum of 15 cm^{-1} . Both individual and convoluted peaks are shown in the figures below. Refined relative energies and Gibbs free energies at 298 K were obtained at the MP2/TZVPP//MP2/SVP and CC2/TZVPP//MP2/SVP levels. All quantum chemical calculations were performed with the TURBOMOLE 6.1 package [38].

Experimental methods

The ions have been generated by electrospray ionization and analyzed using two experimental setups, to record IRMPD spectra in two different photon energy ranges. The $1000\text{-}1800\text{ cm}^{-1}$ fingerprint region generally provides information on the identity of the functional groups in the molecule and to some extent on their environment, while the $2900\text{-}3600\text{ cm}^{-1}$ range is where the O-H and N-H stretching motions are expected to be IR-active with frequencies that are highly sensitive to the existence and strength of hydrogen bonds. In the first case, we used a quadrupole ion trap mass spectrometer (modified Bruker Esquire 3000+) [39, 40] coupled to the Free Electron Laser (FEL) at the Centre Laser Infrarouge d'Orsay (CLIO). In the second case, we used an FT-ICR mass spectrometer (Bruker APEX IV) equipped with a 7 T superconducting magnet, coupled to an OPO/OPA laser also at CLIO [41, 42].

Operations of the FEL and OPO/OPA lasers, coupled to either of the above mentioned mass spectrometers, have been detailed elsewhere [39-42]. We provide below the specific parameters used to obtain the results described herein.

IRMPD experiments in the fingerprint region

The experimental set-up is based on a modified Bruker Esquire 3000+ ion trap mass spectrometer. To allow the irradiation of the ions with the focused CLIO beam, a 0.7 mm

hole was drilled in the ring electrode. The ions generated by an electrospray ion source, are transferred to the trap by two octopoles. There, the ions are stabilized by collisions with the He bath gas at a pressure of 10^{-3} mbar. After rapid thermalization the ions are confined in a very small volume in the center of the trap. The laser beam was focused in the center of the trap using a 500 mm ZnSe focal lens. The laser beam position was optimized in order to maximize the fragmentation and allows for a greater overlap with the ion cloud, resulting in increased fragmentation efficiency than in the ICR cell.

The IR FEL CLIO light is delivered in 8 μ s long macropulses fired at a repetition rate of 25 Hz. Each macropulse contains 500 picosecond long micropulses. The mean IR power was about 500 mW corresponding to micropulse and macropulse energies of 40 μ J and 20 mJ, respectively. For the experiments described herein, we have used an electron energy of 45 MeV which allows access to wavelengths in the 1000-1900 cm^{-1} range.

Following a 0.5 s relaxation time, the ions are mass selected and fragmented by macropulses of IR light. Ion fragmentation is generally visible with only a few macropulses, however the spectra described below have been recorded with up to 10 macropulses. The conditions were tuned for maximum fragmentation efficiency while avoiding saturation.

IRMPD experiments in the 3000-3600 cm^{-1} region

The ions are irradiated with the beam generated by a OPO/OPA laser (LaserVision) pumped by a 25 Hz Nd:Yag (Innolas Splitlight 600). The laser beam enters the FT-ICR cell along the magnet axis and interacts with the ions trapped into 10^{-10} mbar vacuum. A spherical mirror with 2 m focal distance, placed between the laser and the cell, generates an almost constant diameter of the laser beam in the cell region. Once trapped into the ICR cell, the ions were irradiated for up to 20 seconds prior to signal detection. Each point in the spectra was measured only once, without any signal averaging. For maximum fragmentation efficiency, the IRMPD spectra were splitted in sections no larger than 300-400 cm^{-1} and the laser power was optimized before measurement in each section.

IRMPD fragmentation spectra are plotted using the fragmentation yield R defined as: $R = -\log[I_{\text{Parent}}/(I_{\text{Parent}}+\sum I_{\text{Fragment}})]$. The FEL fingerprint spectra were obtained by

including the signal intensities for a number of fragment ions (see below), while the OPO/OPA spectra were based on fewer fragment ion intensities, up to 6 for $G_8\text{-Na}^+$.

Sample preparation

Gaseous ions were obtained by electrospray of diluted sample solutions. First, 15 mL peptide stock solutions were prepared using 6 mL 10% formic acid and 9 mL $\text{H}_2\text{O}/\text{MeOH}$ 50:50 in order to dissolve 10 mg of solid peptide. 10-50 μL of stock solution (10^{-3} M) were mixed 1:1 with NaCl solution (10^{-3} M) and diluted to 1.0 mL with $\text{H}_2\text{O}/\text{MeOH}$ 50:50 with 1% formic acid. This protocol produced samples with concentrations of peptide in the μM range, leading to the formation of the sodiated peptide ion as the dominant species in the gas phase.

Results and discussion

Experimental results

Fragmentation of $G_n\text{-Na}^+$ ions under FEL irradiation generates several series of ions with smaller mass-to-charge ratios, with intensity patterns that are similar to those obtained through low energy collision-induced dissociation [42]. In addition to eliminations of H_2O and NH_3 , the most intense series corresponds to the successive losses of entire G residues leading to a $[\text{b}_n+\text{OH}+\text{Na}]^+$ series. Other series observed include $[\text{b}_n\text{-H}+\text{Na}]^+$, $[\text{b}_n\text{-NH}_4+\text{Na}]^+$ and $[\text{a}_n\text{-H}+\text{Na}]^+$, however some of the smaller members of each series are generally not observed. All peaks with intensities significantly above noise level were included in the calculation of the fragmentation yields. For the largest ion, $G_8\text{-Na}^+$, 24 fragment ion intensities were used. Irradiation with the OPO/OPA laser, although for much larger durations, yielded much fewer fragments, e.g. 6 fragments were used for calculating the fragmentation yields of $G_8\text{-Na}^+$.

The IRMPD spectra of $G_n\text{-Na}^+$ ($n=4-8$) in the fingerprint range are shown in Figure 1. The spectra for $n=2$ and 3 have been reported previously [23]. Spectra in the 3000-3700 cm^{-1} region are shown in Figure 2 for $n=2, 3, 6, 7$ and 8. All attempts at recording spectra in this region for $n=4$ and 5 remained unsuccessful. It should be noted that fragmentation

ratios were small in several cases, so that only the most intense bands were detected in the spectra reported for $n=2, 3$ and 6 in Figure 2. The information gleaned is nevertheless important.

Based on the literature on IRMPD spectra of ionic peptides [n], band assignment can be made straightforwardly in some cases, for instance the amide I and amide II massifs in the $1650-1780$ and $1480-1570$ cm^{-1} ranges, respectively (see Figure 1). The small band near 1170 cm^{-1} that is visible in the spectra of $G_4\text{-Na}^+$ and $G_5\text{-Na}^+$ but absent from those of $G_6\text{-Na}^+$, $G_7\text{-Na}^+$ and $G_8\text{-Na}^+$ (Figure 1) has been attributed to the carboxylic C-O-H bending motion when devoid of hydrogen bonding [15, 37-42]. This indicates that some significant changes must have occurred at the C terminus between $n=5$ and 6 . This is consistent with the band between 3550 and 3560 cm^{-1} that is visible in the spectra of $G_2\text{-Na}^+$ and $G_3\text{-Na}^+$ but absent from those of $G_6\text{-Na}^+$, $G_7\text{-Na}^+$ and $G_8\text{-Na}^+$ (Figure 2), although the lack of spectra for $G_4\text{-Na}^+$ and $G_5\text{-Na}^+$ forbids locating the structural transition in this case. This band can be attributed to the stretching of the C-terminus carboxylic O-H when not involved in a hydrogen bond [16]. Combining the observations near 1170 and 3560 cm^{-1} leads to the conclusion that the C-terminus is free for $n=2-5$ and hydrogen bound for $n\geq 6$. Contrary to the previous two, the small band near $1220-1230$ cm^{-1} is common to all spectra. It arises from the coupling of CH_2 wagging coupled to neighbouring CNH bending, a mode sometimes denoted as amide III. This band is clearly common to all peptides and must be largely insensitive to conformation. Finally the massifs observed previously at $3350-3400$ and $3460-3490$ cm^{-1} for $G_8\text{-Na}^+$ are also visible for $G_7\text{-Na}^+$ (see Figure 2), indicating that the latter involves N-H bonds that are either free of hydrogen bonding or weakly hydrogen-bound as described recently for $G_8\text{-Na}^+$. As mentioned above, the free N-H feature is the only one visible for $G_6\text{-Na}^+$, however the very low fragmentation ratio observed in this case makes structural assignment less straightforward. In such cases and in general for drawing a relation between these experimental patterns and one or several molecular structures, modeling results are required. For the series of spectra described herein, a sound analysis of band position as a function of peptide size must also be based on modeling.

Computational results: low energy structures

The description below uses the sequence numbering from the N- to the C-terminus with superscripts, i.e. NH² and CO² stand for the peptidic N-H and C=O bonds of the second residue. Special cases are for residue 1 where the N-H bonds are in a NH₂ rather than an amide group and for the residue at the C-terminus for which COⁿ is an acidic rather than an amide C=O. It also bears the single O-H bond of G_n-Na⁺. Figures 3-6 display the lowest energy structures of G_n-Na⁺ with n=4-7, together with their relative energies at the MP2/TZVPP//MP2/SVP and CC2/TZVPP//MP2/SVP levels and their MP2/SVP computed IR spectrum. It may be noted that in no case was an SB structure found to be the most stable for G_n-Na⁺ (n=4-7), while the most stable CS and SB isomers are essentially degenerate for G₈-Na⁺.

Exploration of the potential energy surfaces have been described previously for n=2, 3 and 4 [21]. The results indicate that the dominant effect is the maximization of the peptide-sodium ion interaction via coordination of amide oxygens to Na⁺. The three most stable structures of G₄-Na⁺ are shown in Figure 3. They are very similar to those reported previously [21] except that after re-optimization at the MP2/SVP level there remained three unique structures instead of four. Structures 4s and 5s have the four amide oxygens at a distance of less than 2.5 Å from Na⁺ and differ from each other only by the conformation of the N terminus. 4s is more stable than 5s by 1.0 kcal/mol at the CC2/TZVPP level, due to the hydrogen bond between the N terminus and NH². The 3s structure has the CO² free from sodium interaction and a hydrogen bond between the N and C termini. This interaction leads to a blue shift of the acidic C=O stretch and COH bend such that the two experimental bands at 1170 and 1760 cm⁻¹ no longer have a computed counterpart, making it easy to exclude the experimental formation of such a structure.

All three lowest energy structures of G₅-Na⁺ (see Figure 4) have the five amide oxygens interacting with Na⁺, however the peptide is not long enough to enable direct interaction between its termini. Thus 5s1 and 5s2 are somewhat similar and computed to be degenerate, only differing in the positions of the successive C=O bonds relative to the average molecular plane. 6s is only slightly different from 5s2 with the amine terminus folded towards Na⁺ in 6s instead of NH² in 5s2, lying 1.5 kcal/mol higher in energy at the CC2/TZVPP level. The next most stable structure is 3 kcal/mol higher in energy than 6s,

thus its formation as well as that of less stable structures is unlikely. A number of them could also be ruled out by comparison of their computed IR spectra with the experimental IRMPD spectrum, as discussed above for $G_4\text{-Na}^+$ 3s.

$G_6\text{-Na}^+$ is a special case since the peptide folded around the ion, with all six amide oxygen interaction with Na^+ , is now long enough to establish a $\text{CO-H}\cdots\text{NH}_2$ hydrogen bond between its termini. The amide C=O bonds are alternately above and below the average molecular plane. This maximization of the peptide-cation interaction leads to a single structure (see Figure 5) being much more stable than all others, except that the terminal amine may be on either side of the average molecular plane, leading to identical $\text{CO-H}\cdots\text{NH}_2$ hydrogen bond lengths and a free energy difference of only 0.6 kcal/mol. The next three structures also have six-coordination of sodium. They differ from the most stable by either non-alternating orientations of the C=O 's with respect to the average plane (+ 4.5 kcal/mol) leading to more distortion from overall planarity, hydroxyl rather than carbonyl oxygen coordination at the C-terminus (+4.6 kcal/mol) or have a reversed hydrogen bond between the termini, i.e. $\text{N-H}\cdots\text{O}^6$ together with $\text{O-H}\cdots\text{O}^1$ instead of $\text{CO-H}\cdots\text{N}^1\text{H}_2$ (+6.5 kcal/mol).

For $n=7$ and larger, the conformational space becomes much more complex since the peptide is long enough to fold around the ion and at the same time establish hydrogen bonds. The lowest energy structure is quite different from that of $G_6\text{-Na}^+$, with 5 C=O s bound to sodium and the C terminus OH bound to CO^2 . Its spectrum is in reasonable agreement with experiment, however the $\text{O-H}\cdots\text{O}=\text{C}$ interaction leads to a strong red shift of the O-H stretching mode and an intense band near 2930 cm^{-1} , a region that could not be scanned experimentally. It also has a series of N-H stretches that is widened by some weak hydrogen bonds, although the small fragmentation ratio does not provide a fully conclusive IRMPD spectrum in this area. A structure displaying a spectrum in better agreement with experiment in the N-H stretch region is that of 6s3, which involves six C=O s bound to Na^+ , a hydrogen bond between the termini and a C_7 interaction between CO^1 and NH^3 . In such a structure, the six free N-H bonds yield a narrower distribution of stretching bands, and only one H-bound N-H leads to a red-shifted absorption near 3380 cm^{-1} , in good agreement with experiment, with the two amine stretching bands in between. This structure is however only the 5th in energy, leading to an uncertainty about

the structure of $G_7\text{-Na}^+$ especially because of the poor signal to noise ratio of the IRMPD spectrum.

The structures for $n=8$ have been described in detail previously [24]. The two most stable are two isomers related to one another by a simple proton transfer between the N and C termini, i.e. CS and SB isomers. It is worth noting that the experimental spectra of $G_8\text{-Na}^+$ cannot be successfully reproduced using most current density functionals [Semrouni et al., submitted for publication]. Good agreement was however obtained at the MP2/SVP level, leading to its use in the present work. A careful analysis of the computed spectra for the CS and SB isomers in both spectral ranges leaves no doubt on the fact that the experimental spectrum corresponds to a CS structure.

Discussion

The superposition of spectra in Figure 1 is revealing of several features being common to all sizes while others are different. The discussion above has highlighted the presence of bands at 1170 and 1760 cm^{-1} for $G_n\text{-Na}^+$ ($n=4,5$) and their absence for $G_n\text{-Na}^+$ ($n=6-8$) and their usefulness in establishing that the larger peptides have a direct interaction between their termini. In those cases, the COH band appears as shoulder near 1500 cm^{-1} on the red side of the amide II band of $G_6\text{-Na}^+$ and it is buried inside the wider amide II bands of the larger peptides. The superposition of spectra makes it also clear that there is a global blue shift of the amide I and II bands of $G_6\text{-Na}^+$ relative to those of $G_4\text{-Na}^+$ and $G_5\text{-Na}^+$, while those of $G_7\text{-Na}^+$ and $G_8\text{-Na}^+$ are somewhat red-shifted back while being larger because they include a larger number of bands. This trend can be traced to the details of the peptide-sodium interactions. In the smaller peptides, there is good flexibility for the C=O bonds to point towards Na^+ in a nearly optimal way, including an alternation of orientation relative to the average molecular plane. As a result, the Na---O bond distances are all in the 2.2-2.4 Å range. There is much more strain in $G_6\text{-Na}^+$ in which peptide folding must adapt to enable a stabilizing interaction between the termini. This leads to a structure that is much closer to being planar, at the expense of larger Na---O distances, all but one at 2.45 ± 0.01 Å. In the larger peptides, the larger flexibility allows for some of the Na---O distances being shorter than 2.4 Å again, shifting the red side of the band to lower

frequencies. The amide II bands are also sensitive to the Na---O interactions, leading to a parallel trend. Thus even though IRMPD band shapes cannot be discussed in great detail in general, the comparison of a series of closely related cases enables subtle variations to be identified.

Previous work has shown that for small $G_n\text{-Na}^+$, salt bridges are systematically less stable than the best charge solvation structures [17, 19, 20-23]. The present results confirm this trend for $n=4-6$, but show that the energy difference reduces to 3 kcal/mol for $G_7\text{-Na}^+$ and to essentially zero for $G_8\text{-Na}^+$. The question remains open for larger peptides. Provided that there is no major structural change, i.e. that the peptide interacts with Na^+ via six amide oxygens and structures itself by C_7 and/or C_{10} hydrogen bonds, then the most stable structure will involve direct interaction between both termini and the salt bridge isomer may become more stable. Alternatively, a transition to a truly three-dimensional folding around sodium may lead to widely different structures with N and C termini that are remote from one another, in which case electrostatic stabilization would be reduced and the lowest CS structure is likely to be the most favorable.

Conclusions

IRMPD spectra have been obtained for $G_n\text{-Na}^+$ at room temperature in the 1000-1900 cm^{-1} region for $n=4-8$ and in the 3000-3700 cm^{-1} region for $n=2, 3$ and $6-8$. Even though the spectral resolution is low, a number of structurally relevant informations can be gleaned from these spectra. Bands in both regions indicate that the C-terminus OH group is free for $n=2-5$ but hydrogen bound or absent for $n=6-8$. The 3000-3700 cm^{-1} spectra are indicative of a network of N-H bonds most of which are either free or not very strongly hydrogen-bound for $n=7$ and 8 . Small but significant variations in the shapes of the wide experimental features have also been identified. Molecular modeling with extensive exploration of the potential energy surfaces with the polarizable force field AMOEBA and replica exchange molecular dynamics, together with accurate ab initio calculations were used to interpret these results. The peptides tend to wrap around the sodium ion in order to maximize electrostatic and polarization interactions. From $n=2$ to 6 , this amounts to folding so as to enable sodium-oxygen interactions with all peptidic

oxygens. For $n=6$, this is also compatible with a direct interaction between the peptide termini. From $n=7$ on, peptide folding maintains 6 sodium binding sites, which can be accommodated in a number of ways. Modeling results indicate that the best compromise involves hydrogen bonding between the N and C termini, opening the possibility of salt bridge formation. Both charge solvation and salt bridge structures can be identified computationally. While SBs are rather unstable for smaller oligoglycines, the increased flexibility of peptides of growing size tend to stabilize the lowest energy SBs relative to the best CSs, up to $n=8$ for which they are degenerate. Good agreement between experiments and computations can only be achieved for charge solvation structures. The present results provide high prospects for using IRMPD spectroscopy to obtain structural informations on larger peptides.

Acknowledgments

We thank the European Union (EU) for generous support provided through the STREP 15637 "EPITOPES" funded by the New and Emerging Science and Technology program of the 6th pluriannual framework of EU. OPB thanks Ecole Polytechnique for a one year post-doctoral fellowship and EU for a one year post-doctoral support through the "EPITOPES" project. This work was granted access to the HPC resources of [CCRT/CINES/IDRIS] under the allocation x2010085107 made by GENCI (Grand Equipement National de Calcul Intensif).

References

1. J. Oomens, B. G. Sartakov, G. Meijer, G. von Helden, Gas-phase infrared multiple photon dissociation spectroscopy of mass-selected molecular ions. *Int. J. Mass Spectrom.* **2006**, *254*, 1-19.
2. K. R. Asmis, J. Sauer, Mass-selective vibrational spectroscopy of vanadium oxide cluster ions. *Mass Spectrom. Rev.* **2007**, *26*, 542-562.
3. L. MacAleese, P. Maître, Infrared spectroscopy of organometallic ions in the gas phase : from model to real world complexes. *Mass Spectrom. Rev.* **2007**, *26*, 583-605.
4. T. Fridgen, Infrared consequence spectroscopy of gaseous protonated and metal ion cationized complexes. *Mass Spectrom. Rev.* **2009**, *28*, 586-607.
5. J. R. Eyler, Infrared multiple photon dissociation spectroscopy of ions in Penning traps. *Mass Spectrom. Rev.* **2009**, *28*, 448-467.
6. N. C. Polfer, J. Oomens, Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass Spectrom. Rev.* **2009**, *28*, 468-494.
7. T. Baer, R. C. Dunbar, Ion spectroscopy: where did it come from; where is it now; and where is it going? *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 681-693.
8. S. A. McLuckey, K. L. Busch, G. L. Glish, **1988**, *Mass spectrometry/mass spectrometry: techniques and applications of tandem mass spectrometry*. New York, N.Y: VCH Publishers. ISBN 0-89573-275-0
9. R. M. Caprioli, M. L. Gross, Eds., **2006**, *Encyclopedia of Mass Spectrometry, volumes 2 (biological applications part A) and 3 (biological applications part B)*. Elsevier.
10. K. Pagel, P. Kupser, F. Bierau, N. C. Polfer, J. D. Steill, J. Oomens, G. Meijer, B. Kokschi, G. Von Helden, *Int. J. Mass. Spectrom.* **2009**, *283*, 161-168.
11. P. Kupser, K. Pagel, J. Oomens, N. C. Polfer, B. Kokschi, G. Meijer, G. Von Helden, *J. Am. Chem. Soc.* **2010**, *132*, 2085-2093.

12. J. C. Pouilly, F. Lecomte, F. Nieujwaer, B. Manil, J.-P. Schermann, C. Desfrancois, G. Grégoire, R. Ballivian, F. Chirot, J. Lemoine, F. Calvo, R. Antoine, P. Dugourd, Combining ion mobility mass spectrometry and infrared multiphoton dissociation spectroscopy to probe the structure of gas-phase vancomycin-Ac₂^LK^DA^DA non-covalent complex *Int. J. Mass Spectrom.* **2010**, in press. <http://dx.doi.org/10.1016/j.ijms.2010.05.014>
13. R. Wu, T. B. McMahon, Protonation sites and conformations of peptides of glycine by IRMPD spectroscopy [Gly₁₋₅H⁺]. *J. Phys. Chem. B* **2009**, *113*, 8767-8775.
14. J. Oomens, N. C. Polfer, D. T. Moore, L. Van der Meer, A. G. Marshall, J. T. Eyler, G. Meijer, G. von Helden, *Phys. Chem. Chem. Phys.* **2005**, *7*, 1345-1348.
15. Y. M. E. Fung, T. Besson, J. Lemaire, P. Maître, R. A. Zubarev, *Angew. Chem. Int. Ed.* **2009**, *48*, 8340-8342.
16. T. D. Vaden, T. S. J. A. de Boer, J. P. Simons, L. C. Snoek, S. Suhai, B. Paizs, Vibrational spectroscopy and conformational structure of protonated polyalanine peptides isolated in the gas phase. *J. Phys. Chem. A* **2008**, *112*, 4608-4616.
17. T. Wytttenbach, J. E. Bushnell, M. T. Bowers, Salt bridge structures in the absence of solvent? The case for the oligoglycines. *J. Am. Chem. Soc.* **1998**, *120*, 5098-5103.
18. M. Kohtani, M. F. Jarrold, S. Wee, R. A. J. O'Hair, Metal ion interactions with polyalanine peptides. *J. Phys. Chem. B* **2004**, *108*, 6093-6097.
19. H. A. Cox, R. R. Julian, S.-W. Lee, J. L. Beauchamp, Gas-phase H/D exchange of sodiated glycine oligomers with ND₃: exchange kinetics do not reflect parent ion structures. *J. Am. Chem. Soc.* **2004**, *126*, 6485-6490.
20. M. M. Kish, C. Wesdemiotis, G. Ohanessian, The sodium ion affinity of glycyglycine. *J. Phys. Chem. B* **2004**, *108*, 3086-3091.
21. P. Wang, C. Kapota, G. Ohanessian, C. Wesdemiotis, The sodium ion affinities of simple di-, tri- and tetrapeptides. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 541.

22. S. J. Ye, P. B. Armentrout, Absolute thermodynamic measurements of alkali metal cation interactions with a simple dipeptide and tripeptide. *J. Phys. Chem. A* **2008**, *112*, 3587-3596.
23. O. P. Balaj, C. Kapota, J. Lemaire, G. Ohanessian, Vibrational signatures of sodiated oligopeptides (GG-Na⁺, AA-Na⁺, GGG-Na⁺ and AAA-Na⁺) in the gas phase. *Int. J. Mass Spectrom.* **2008**, *269*, 196-209.
24. D. Semrouni, O. P. Balaj, F. Calvo, C. F. Correia, C. Clavaguéra, G. Ohanessian, Structure of sodiated octa-glycine: IRMPD spectroscopy and molecular modeling. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 728-738.
25. C. Kapota, PhD thesis, Ecole Polytechnique (2005)
26. J. W. Ponder, C. Wu, P. Y. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio Jr., M. Head-Gordon, G. N. I. Clark, M. E. Johnson, T. Head-Gordon, Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *8*, 2549-2564.
27. D. Semrouni, G. Ohanessian, C. Clavaguéra, Structural, energetic and dynamical properties of sodiated oligoglycines: relevance of a polarizable force field. *Phys. Chem. Chem. Phys.* **2010**, *11*, 3450-3462.
28. V. A. Voelz, M. S. Shell, K. A. Dill, Predicting peptide structures in native proteins from physical simulations of fragments. *PLoS Computational Biology* **2009**, Vol. 5, issue 2, article e1000281.
29. N. V. Buchete, G. Hummer, Protein folding kinetics from replica exchange molecular dynamics *Phys. Rev. E* **2008**, *77*, 3, 030902.
30. D. S. Weinstock, C. Narayanan, A. K. Felts, M. Andrec, R. M. Levy, K. P. Wu, J. Baum, Distinguishing between structural ensembles of the GB1 peptides : REMD simulations and NMR experiments. *J. Am. Chem. Soc.* **2007**, *129*, 4858-4859.

31. Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141-151.
32. J. W. Ponder, TINKER, Software Tools for Molecular Design, Version 4.2, 2004.
33. E. S. Penev, S. Lampoudi, J.-E. Shea, TiReX : Replica-exchange molecular dynamics using Tinker. *Comp. Phys. Comm.* **2009**, *180*, 2013-2019.
34. K. E. Riley, M. Pitoňák, P. Jurečka, P. Hobza, Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories. *Chem. Rev.* **2010**, ASAP.
35. S. Grimme, Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463-1473.
36. S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787-1799.
37. Y. Bouteiller, J.-C. Pouilly, C. Desfrancois, G. Grégoire, Evaluation of MP2, DFT and DFT-D methods for the prediction of infrared spectra of peptides. *J. Phys. Chem. A* **2009**, *113*, 6301-6307.
38. R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, Electronic structure calculations on workstation computers : the program system Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165-169. For the current version: see <http://www.turbomole.com>.
39. L. MacAleese, A. Simon, T. B. McMahon, J.-M. Ortega, D. Scuderi, J. Lemaire, P. Maître, Mid-IR spectroscopy of protonated leucine methyl ester performed with an FTICR or a Paul type ion-trap. *Int. J. Mass Spectrom.* **2006**, *249*, 14.
40. L. MacAleese, P. Maître, Infrared spectroscopy of organometallic ions in the gas phase: from model to real world complexes. *Mass Spectrom. Rev.* **2007**, *26*, 583-605.
41. J. M. Bakker, T. Besson, J. Lemaire, D. Scuderi, P. Maître, Gas-phase structure of a π -allyl-palladium complex: efficient infrared spectroscopy in a 7 T Fourier transform mass spectrometer. *J. Phys. Chem. A* **2007**, *111*, 13415-13424.

42. J. M. Bakker, R. K. Sinha, T. Besson, M. Brugnara, P. Tosi, J.-Y. Salpin, P. Maître, Tautomerism of uracil probed via infrared spectroscopy of singly hydrated protonated uracil. *J. Phys. Chem. A* **2008**, *112*, 12393-12400.
43. T. Lin, A. H. Payne, G. L. Glish, Dissociation pathways of alkali-cationized peptides: opportunities for C-terminal peptide sequencing. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 497-504.
44. C. G. Atkins, K. Rajabi, E. A. L. Gillis, T. D. Fridgen, *J. Phys. Chem. A* **2008**, *112*, 10220-10225.
45. J. A. Stearns, S. Mercier, C. Seaiby, M. Guidi, O. Boyarkin, T. R. Rizzo, Conformation-specific spectroscopy and photodissociation of cold, protonated tyrosine and phenylalanine. *J. Am. Chem. Soc.* **2007**, *129*, 11814-11820.
46. M. F. Bush, J. T. O'Brien, J. S. Prell, R. J. Saykally, E. R. Williams, Infrared spectroscopy of cationized arginine in the gas phase: direct evidence for the transition from nonzwitterionic to zwitterionic structure. *J. Am. Chem. Soc.* **2007**, *129*, 1612-1622.
47. J. A. Stearns, C. Seaiby, O. Boyarkin, T. R. Rizzo, Spectroscopy and conformational preferences of gas-phase helices. *Phys. Chem. Chem. Phys.* **2009**, *11*, 125-132.
48. J. S. Prell, J. T. O'Brien, J. D. Steill, J. Oomens, E. R. Williams, Structures of protonated dipeptides: the role of arginine in stabilizing salt bridges. *J. Am. Chem. Soc.* **2009**, *131*, 11442-11449.
49. J. S. Prell, M. Demireva, J. Oomens, E. R. Williams, Role of sequence in salt-bridge formation for alkali metal cationized GlyArg and ArgGly investigated with IRMPD spectroscopy and theory. *J. Am. Chem. Soc.* **2009**, *131*, 1232-1242.
50. N. C. Polfer, J. Oomens, Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass Spectrom. Rev.* **2009**, *28*, 468-494.

51. N. C. Polfer, B. Paizs, L. C. Snoek, I. Compagnon, S. Suhai, G. Meijer, G. von Helden, J. Oomens, Infrared fingerprint spectroscopy and theoretical studies of potassium ion tagged amino acids and peptides in the gas phase. *J. Am. Chem. Soc.* **2005**, *127*, 8571-8579.
52. N. C. Polfer, J. Oomens, R. C. Dunbar, Alkali metal complexes of the dipeptides PheAla and AlaPhe: IRMPD spectroscopy. *ChemPhysChem.* **2008**, *9*, 579-589.
53. R. C. Dunbar, J. D. Steill, N. C. Polfer, J. Oomens, Peptide length, steric effects, and ion solvation govern zwitterion stabilization in barium-chelated di- and tripeptides. *J. Phys. Chem. B* **2009**, *113*, 10552-10554.
54. A. Cimas, T. D. Vaden, T. S. J. A. de Boer, L. C. Snoek, M. P. Gaigeot, Vibrational spectra of small protonated peptides from finite temperature MD simulations and IRMPD spectroscopy. *J. Chem. Theor. Comput.* **2009**, *5*, 1068-1078.

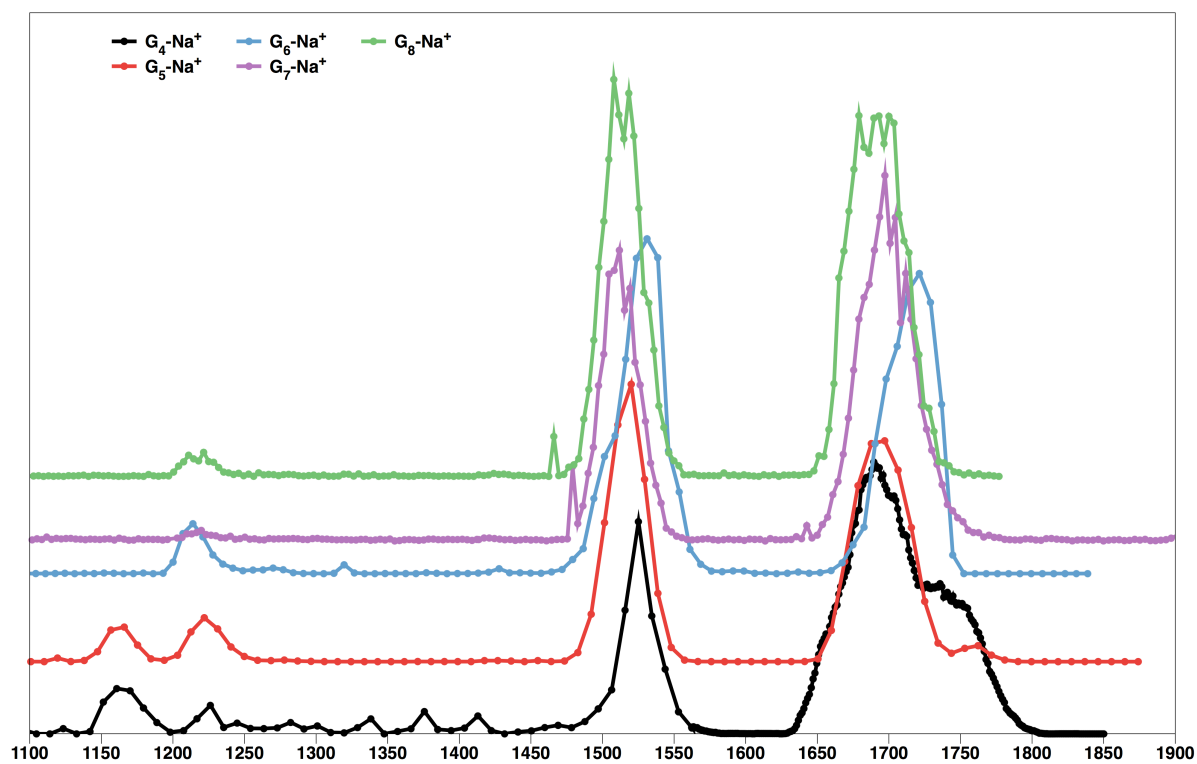


Figure 1: IRMPD spectra of $G_n\text{-Na}^+$ ($n=4-8$) in the 1100-1850 cm^{-1} range.

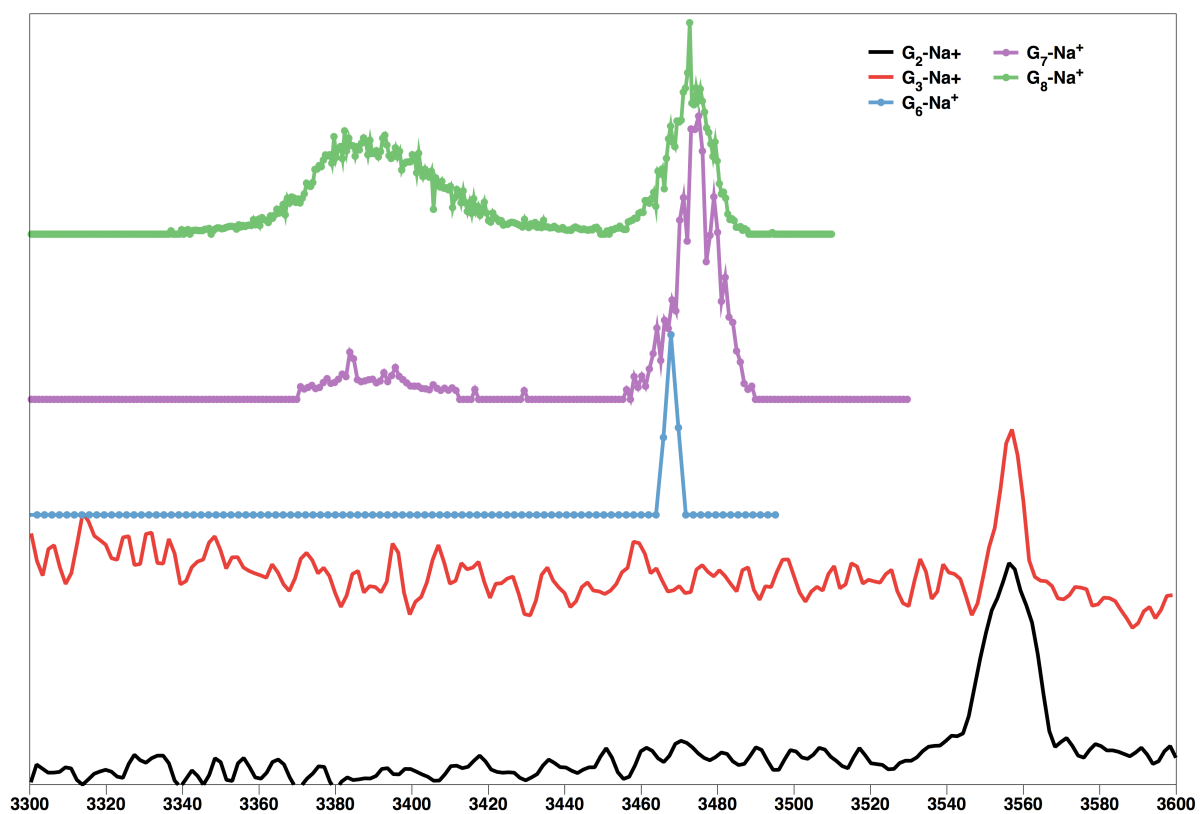


Figure 2: IRMPD spectra of $G_n\text{-Na}^+$ ($n=2, 3$ and $6-8$) in the $3300\text{-}3600\text{ cm}^{-1}$ range.

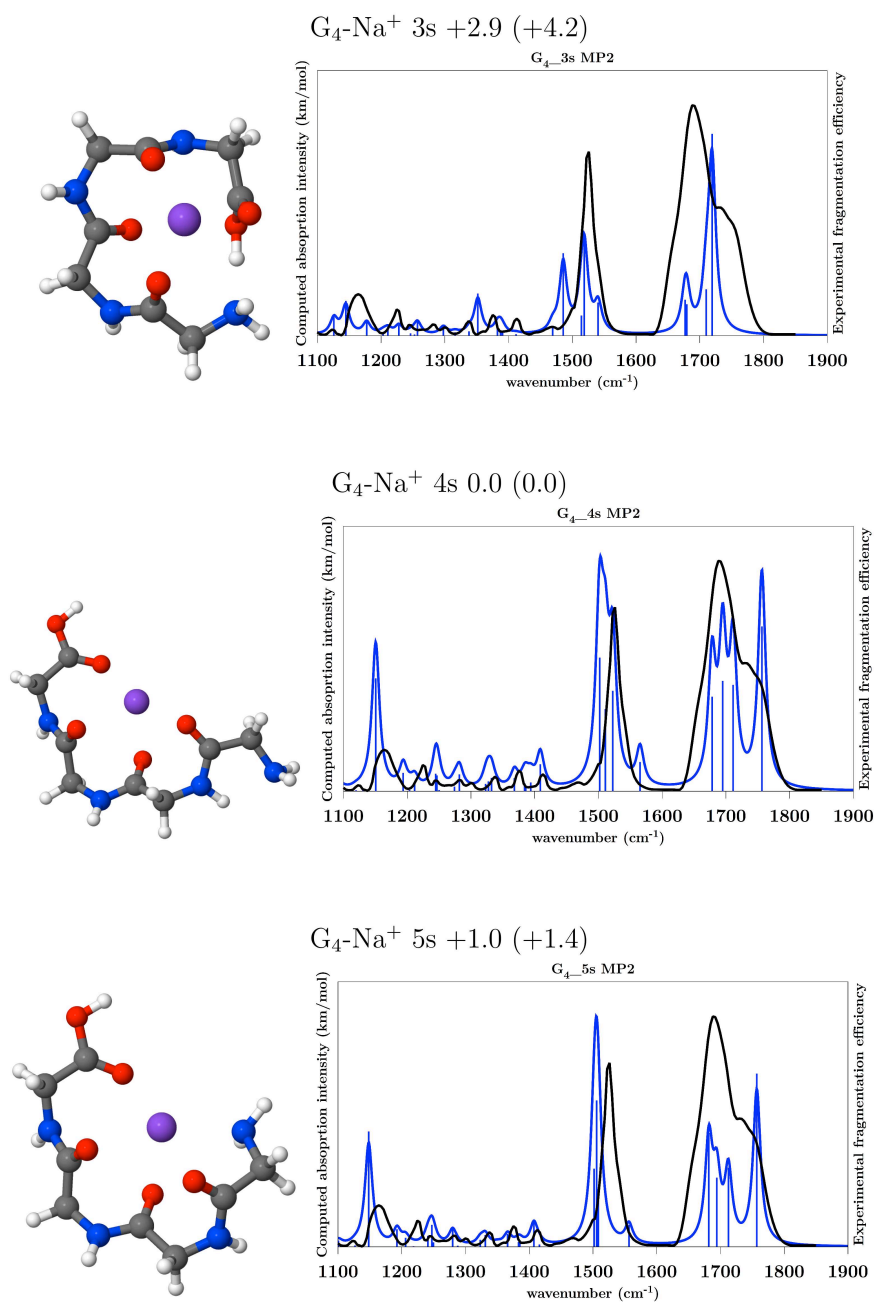


Figure 3: Most stable MP2/SVP optimized structures of $G_4\text{-Na}^+$, relative CC2//MP2 free energies in kcal/mol (with MP2//MP2 values in parentheses) and MP2/SVP computed IR spectra with IRMPD spectrum for comparison.

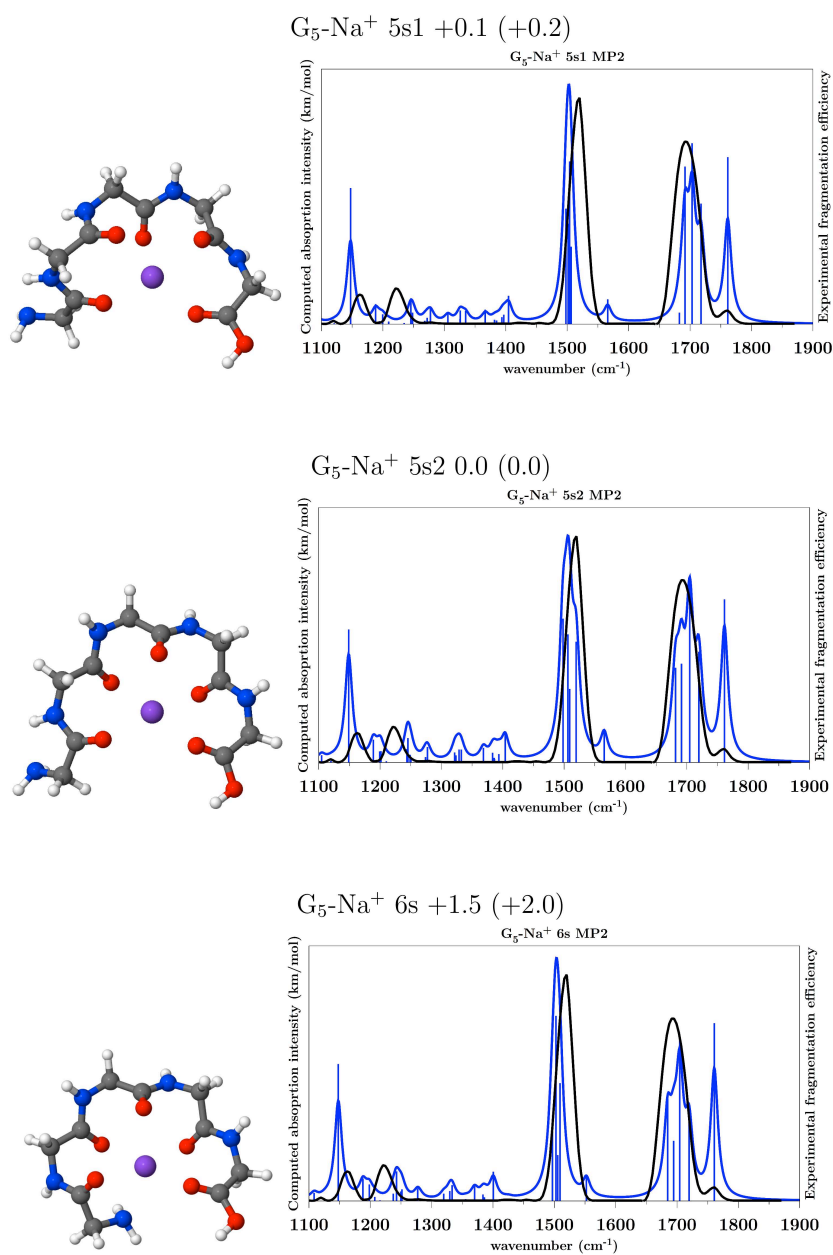


Figure 4: Most stable MP2/SVP optimized structures of $G_5\text{-Na}^+$, relative CC2//MP2 free energies in kcal/mol (with MP2//MP2 values in parentheses) and MP2/SVP computed IR spectra with IRMPD spectrum for comparison.

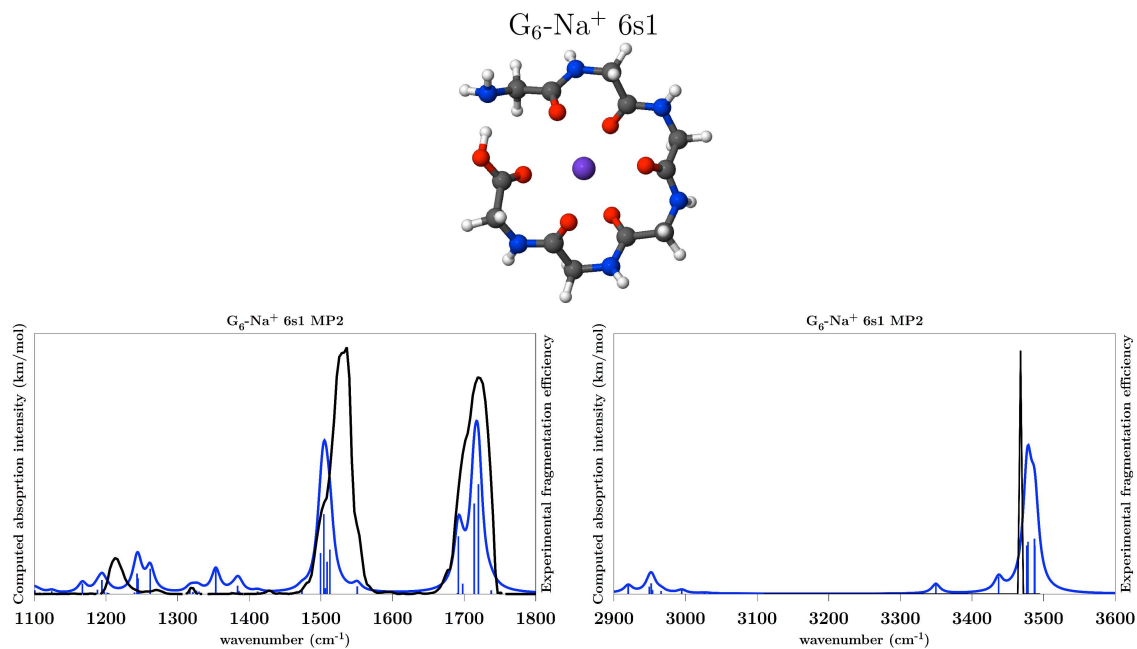


Figure 5: Most stable MP2/SVP optimized structure of $G_6\text{-Na}^+$ and MP2/SVP computed IR spectrum with IRMPD spectrum for comparison.

Figure 6: Most stable MP2/SVP optimized structures of $G_7\text{-Na}^+$, relative CC2//MP2 free energies in kcal/mol (with MP2//MP2 values in parentheses) and MP2/SVP computed IR spectra with IRMPD spectrum for comparison.

Supplementary information

REMD simulations

Each molecular dynamics trajectory in the PVT ensemble was propagated with the Beeman integrator and a time step of 1 fs, using the Berendsen algorithm with coupling times of 2.0 ps for the barostat and of 0.1 ps for the thermostat. Exchanges of random pairs of adjacent configurations were attempted every 10 ps, and all configurations were saved every 5 ps for further local optimization with AMOEBA. The total simulation time for each replica was taken as 5 ns, and the REMD simulations were repeated twice, with all replicas starting each time from the most stable structure found so far.

The REMD sampling stage provided large databases of candidate structures, from which several dozens of structures were chosen for further re-optimization by quantum chemistry methods, on the basis of their AMOEBA energy being in the lowest 5 kcal/mol window. In some instances, a proton transfer was made manually from a low energy salt bridge structure to generate a charge solvation isomer or vice versa. Structural refinement generated a significant number of duplicates, including mirror image pairs, which were eliminated. Structures were considered to be duplicates when the root mean square difference of their cartesian structures was smaller than 0.1 Å.

Conclusions et perspectives

L'objectif de notre projet était la mise en œuvre de nouvelles méthodes de modélisation pour l'identification structurale de peptides en phase gazeuse attachés à un cation métallique à partir de spectres IRMPD. Étant donné la complexité des spectres infrarouge et des espaces conformationnels considérés, l'étape cruciale a été la recherche d'une méthodologie efficace pour une bonne reproduction de leur géométrie et de leur énergétique afin d'en obtenir les structures les plus stables et, par conséquent, les plus à même d'être responsables des spectres observés expérimentalement.

Le champ de force anharmonique et polarisable AMOEBA s'est avéré pertinent pour reproduire à la fois la structure et l'énergétique de ces systèmes. Couplé à la dynamique moléculaire par échange de répliques (REMD), il nous a permis une exploration efficace de la surface d'énergie potentielle pour plusieurs systèmes de polyglycines sodées en phase gazeuse. Pour exploiter au mieux les résultats de ces recherches conformationnelles, nous avons développé et implémenté quelques outils de tri de structures.

Cette démarche, suivie par le calcul de spectres statiques, a donné des résultats comparables aux observations expérimentales et permettant l'identification structurale sur la base de spectres IRMPD pour plusieurs peptides sodés. Dans la plupart des cas, les spectres calculés avec la fonctionnelle de la densité incluant un terme de dispersion, B3LYP-D sont en accord avec les spectres expérimentaux ou aux calculs post-Hartree-Fock réalisés. Ainsi, les structures expérimentales de G_2 , G_3 , G_4 , G_5 , et G_7 , en interaction avec un cation sodium, ont pu être identifiées par comparaison avec les spectres B3LYP-D. De ces identifications, il se dégage que le peptide solvate la charge du cation avec ses carbonyles jusqu'à ce qu'il soit saturé avec 6 sites de coordination. Il s'avère aussi que ces structures sont toutes non-zwitterioniques. Après saturation de la coordination, la stabilisation interne du peptide joue un rôle plus important et on voit apparaître le début d'un réseau de liaisons hydrogène.

L'octa-glycine sodée est un exemple particulier pour lequel B3LYP-D ne permet pas le calcul d'un spectre proche de l'expérimental. Le sodium y est hexacoordiné et des liaisons hydrogène maintiennent la structure du peptide. Une partie non négligeable de sa stabilisation vient d'une interaction particulièrement forte entre les deux terminaisons du peptide. C'est pourquoi les structures de plus basse énergie proviennent de recherches conformationnelles effectuées à partir de la forme zwitterionique. Cette liaison hydrogène atypiquement courte n'est pas correctement reproduite par la plupart des fonctionnelles de la densité

testées et AMOEBA n'est pas non plus apte à la traiter. Cependant, les fréquences de vibration calculées au niveau MP2/SVP sont en très bon accord avec le spectre expérimental et confirment que la structure correspondant au spectre expérimental a bien été trouvée lors de notre recherche conformationnelle.

En parallèle à ces identifications de structures par comparaison de spectres statiques, nous avons cherché à calculer des spectres fiables par transformée de Fourier de la fonction d'Auto-Corrélation du Dipôle (DACF) afin de prendre en compte les effets de température directement dans les simulations. Le champ de force AMOEBA, du fait qu'il inclut les effets à n corps de polarisation mutuelle, permet d'obtenir des déplacements de fréquence en fonction de l'environnement local en accord avec les valeurs expérimentales disponibles. Cela a tout de même nécessité une reparamétrisation des multipôles atomiques en essayant de ne pas détériorer les excellentes énergies d'interaction intermoléculaire calculées en AMOEBA avec le jeu de paramètres original. Cette méthodologie donne de bons résultats dans la gamme 1 000–2 000 cm^{-1} ainsi que 3 000–4 000 cm^{-1} , gamme pour laquelle les spectres obtenus par DACF à partir de trajectoires CPMD peuvent poser problème. L'utilisation d'un champ de force donne plus accès à des simulations longues, permettant de coupler calcul de spectre et dynamique conformationnelle à l'échelle de quelque dizaine de ns.

L'extension de cette reparamétrisation aux systèmes peptidiques est en cours. Les polyglycines sodées sont naturellement parmi les premières applications envisagées. Il faudra pour cela réattribuer les multipôles pour les résidus glycine en position terminale ou non, dans une perspective de transférabilité et de combinaison des terminaisons chargées et neutres. . .

La description des bandes de combinaisons pour les peptides attachés à un cation est un progrès à la portée du DACF, dès que de nouveaux multipôles seront définis et validés pour les propriétés spectroscopiques. La décomposition en coordonnées internes des vibrations observées par calcul de fréquences *via* DACF serait un pas supplémentaire vers la compréhension des spectres dynamiques hors des basses températures et donc de la confrontation aux spectres expérimentaux.

Pour la recherche conformationnelle suivie de calcul de spectre, les sujets d'application ne manquent pas, qu'il s'agisse de spectres statiques ou dynamiques. Nous avons commencé l'identification structurale des polyalanines sodées en utilisant la même démarche que pour

les polyglycines en interaction avec le cation. Il s'agit dans ce cas d'identifier la taille du peptide pour laquelle se produit la transition vers une structure en hélice α . Le calcul des spectres DACF en association à un champ de force polarisable, du fait de sa rapidité et du peu de ressources informatiques demandées, peut être facilement exécuté pour un grand nombre de structures, en comparaison avec les spectres calculés avec des méthodes quantiques *ab initio* ou DFT. D'autres spectres IRMPD ou autres IR (peptides phosphorylés, Gramicidine Soviet. . .) réclament une modélisation intensive et le calcul de spectres par dynamique moléculaire en utilisant AMOEBA nous paraît pertinent. La capacité de l'approche DACF à décrire des systèmes très flexibles rend son application à des cations micro-solvatés particulièrement intéressante pour la confrontation aux spectres expérimentaux puisque les molécules de solvant peuvent y être échangées rapidement.

Table des figures

| | | |
|-----|---|----|
| 2.1 | N-méthylacétamide, repère local associé à l'atome C du groupement carbonyle | 39 |
| 2.2 | Structures du N-méthylacétamide isolé et en interaction | 44 |
| 3.1 | Simulation REMD de $A_{12}\text{-Na}^+$: structure initiale | 64 |
| 3.2 | Simulation REMD de $A_{12}\text{-na}^+$: structure la plus stable | 65 |
| 3.3 | Évolution de la température pour les trajectoires de chaque réplique de la simulation REMD, exemple de $A_{12}\text{-Na}^+$ | 66 |
| 3.4 | Recouvrement des énergies potentielles pour les différentes température du REMD, exemple de $A_{12}\text{-Na}^+$ | 66 |
| 3.5 | Évolution de l'index de réplique de chaque température, exemple de $A_{12}\text{-Na}^+$ | 67 |
| 3.6 | Évolution des distances d'attachement au sodium au cours d'une portion de la pseudo-trajectoire de plus haute température du REMD de $A_{12}\text{-Na}^+$. . . | 68 |
| 3.7 | Évolution de l'énergie potentielle des bassins : comparaison sur l'exemple de $A_{12}\text{-Na}^+$ de l'efficacité d'échantillonnage entre une pseudo-trajectoire à 298 K de REMD et une trajectoire de dynamique moléculaire conventionnelle à 300 K | 70 |
| 4.1 | Géométrie de $G_8\text{-ha3-0}$ | 82 |
| 4.2 | Polyglycines, images dans un miroir | 85 |

| | | |
|-----|---|-----|
| 4.3 | Filtrage des structures redondantes | 86 |
| 5.1 | N-méthylacétamide (NMA) | 92 |
| 5.2 | Effets du thermostat sur l'élongation N-H du N-méthylacétamide | 102 |
| 5.3 | Effets du pas de temps de la trajectoire sur le spectre simulé de l'élongation N-H du N-méthylacétamide à 20 K | 103 |
| 5.4 | Effets de la durée utile de trajectoire sur le spectre simulé, exemple de l'élongation N-H du N-méthylacétamide à 20 K | 104 |
| 5.5 | Effets de températures sur l'élongation N-H du N-méthylacétamide | 106 |
| 5.6 | Déplacement de la bande de l'élongation N-H du N-méthylacétamide à 20 K | 112 |
| 5.7 | Déplacement des bandes amide I et II du N-méthylacétamide | 115 |

Liste des tableaux

| | | |
|-----|--|-----|
| 2.1 | Influence du rayon (en Å) des atomes d'hydrogène sur les charges atomiques du N-méthyle acétamide obtenues par DMA, l'algorithme original de GDMA | 42 |
| 2.2 | Influence du rayon (en Å) des atomes d'hydrogène sur les charges atomiques du N-méthyle acétamide obtenues par l'algorithme de Lebedev | 43 |
| 2.3 | Énergies d'interaction intermoléculaire (en kcal.mol ⁻¹) de systèmes à base de N-méthylacétamide | 45 |
| 3.1 | Paramètres des simulations REMD de A ₁₂ Na ⁺ | 64 |
| 4.1 | Influence de variations de dièdres sur le RMSD entre des structures de G ₈ | 82 |
| 4.2 | Influence d'une rotation conservant la symétrie sur le RMSD en Å de la structure hélicoïdale de G ₈ dans sa forme zwitterionique | 83 |
| 4.3 | RMSD en Å sur tous les atomes pour une rotation locale en fonction de la taille de la polyglycine G _n | 87 |
| 4.4 | Nombre de structures issues du REMD de A ₁₂ -Na ⁺ après tri et exemple de regroupement en familles selon le rayon de giration en Å | 88 |
| 5.1 | Fréquences infrarouge statiques et expérimentales du N-méthylacétamide | 93 |
| 5.2 | Fréquences infrarouge statiques et expérimentales du N-méthylacétamide | 109 |

Résumé

La spectroscopie de dissociation par absorption de multiples photons infrarouge ou IRMPD permet d'obtenir les signatures vibrationnelles d'espèces chargées en phase gazeuse. L'attribution des modes de vibration pour établir une relation entre spectre et structure moléculaire n'est cependant pas triviale et nécessite le recours à la modélisation moléculaire. Ce mémoire présente une méthodologie complète de calcul de spectres vibrationnels et son application à des peptides attachés à un cation sodium. Nous commençons par l'exploration efficace de l'espace conformationnel complexe de ces systèmes, en utilisant la dynamique moléculaire par échange de répliques, pour sélectionner les structures les plus probables. Nous avons utilisé le champ de force polarisable AMOEBA pour traiter les systèmes cation-peptide, et (ré)optimisé des paramètres en fonction des problèmes abordés, notamment pour améliorer le calcul de spectres vibrationnels de composés modèles. La simulation spectrale pour les structures sélectionnées a été effectuée avec des méthodes quantiques DFT et MP2, après avoir évalué les performances d'une série de fonctionnelles, incluant ou non une correction pour les interactions de dispersion. Nous présentons finalement la transformation de Fourier de la fonction d'auto-corrélation du moment dipolaire (DACF) calculée pour une trajectoire de dynamique moléculaire classique, comme moyen de simulation de spectre infrarouge incluant les effets de température et les couplages anharmoniques nécessaires à une réelle comparaison avec l'expérience. La pertinence de la combinaison DACF/AMOEBA pour reproduire les spectres IRMPD est discutée. Cet ensemble de méthodes a été appliqué à la structure, à l'énergétique et à la spectroscopie infrarouge de poly-glycines contenant de deux à huit résidus, attachées à un cation sodium. Les spectres calculés pour les structures de plus basses énergies permettent l'attribution des bandes expérimentales, et ainsi d'établir les structures de ces complexes. L'enroulement du peptide autour de l'ion sodium est réalisé *via* les atomes d'oxygène des groupes amides. Lorsque la coordination de l'ion est saturée, la structuration du peptide est complétée par des coudes.

Mots-clés : interaction cation-peptide, champ de force polarisable, dynamique moléculaire, exploration de surface d'énergie potentielle, simulation de spectre infrarouge

Abstract

InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy allows obtaining the vibrational signatures of charged species in the gas phase. The assignment of vibrational modes to establish a relationship between a spectrum and a molecular structure is however non trivial and requires extensive molecular modeling. This thesis describes a complete methodology for the calculation of vibrational spectra and its application to peptides attached to a sodium cation. We start with the efficient exploration of the complex conformational space of these systems, using replica-exchange molecular dynamics, in order to select the most likely structures. We have used the polarizable force field AMOEBA for the cation-peptide complexes, and (re)optimized parameters whenever required, in particular for improving the calculation of vibrational spectra of model compounds. Simulation of spectra for the structures selected has been performed using DFT and MP2 quantum chemical methods. This involved evaluating the performance of several density functionals, with or without a correction for dispersion interactions. We then describe the Fourier transform of the Dipole moment Auto-Correlation Function (DACF) computed along a classical molecular dynamics trajectory as a means of simulating infrared spectra including temperature and anharmonic coupling effects that are necessary for a full comparison with experiment. We discuss the relevance of the DACF/AMOEBA combination to model IRMPD spectra. This set of methods has been applied to the structure, energetics and infrared spectroscopy of poly-glycines containing two to eight residues, attached to a sodium cation. The computed spectra for the lowest energy structures allow for the assignment of experimental bands, thus establishing the structures of these complexes. Peptide folding around the sodium ion is achieved via the amide oxygens. When the ion coordination becomes saturated, peptide structuration is completed with turns.

Keywords: cation-peptide interaction, polarizable force field, molecular dynamics, potential energy surface exploration, infrared spectrum simulation