Université de Montréal

Développement d'une fonction potentielle pour la dynamique moléculaire de biomolécules

par
Guillaume Lamoureux

Département de physique
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en physique

Août, 2004

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Développement d'une fonction potentielle
pour la dynamique moléculaire de biomolécules

présentée par :

Guillaume Lamoureux

a été évaluée par un jury composé des personnes suivantes :

Laurent J. Lewis
(président-rapporteur) _____

Benoît Roux
(directeur de recherche) _____

Alain Caillé
(codirecteur) _____

Tucker Carrington, Jr.
(membre du jury) _____

Mark E. Tuckerman
(examinateur externe) _____

Hervé Philippe
(représentant du doyen) _____

Résumé

Bien que le comportement d'un ensemble de molécules soit en principe calculable à partir de l'équation de Schrödinger, il est courant, pour simuler des biomolécules en solution et des liquides complexes, de représenter les interactions moléculaires comme une somme d'interactions atomiques dépendant uniquement des positions des atomes. Cette approximation peut être améliorée significativement en reconnaissant que les atomes se polarisent dans le champ électrique créé par leur environnement et que l'interaction électrostatique de deux atomes est affectée par la proximité d'un troisième.

Dans cette thèse, nous formulons une fonction potentielle polarisable qui décrit la réponse dipolaire des atomes à l'aide d'oscillateurs de Drude classiques. Chaque oscillateur est une particule chargée, couplée à la position d'un atome par une force de rappel harmonique, dont le déplacement sous l'effet d'un champ électrique crée un dipôle induit. L'oscillateur de Drude reproduit les principes de l'induction dipolaire lorsqu'il occupe son état de plus basse énergie de manière cohérente avec les oscillateurs voisins. Nous mettons au point un algorithme de dynamique moléculaire qui maintient les oscillateurs à leurs positions auto-cohérentes sans recourir à une approche itérative. Nous montrons que cet algorithme intègre la dynamique rapide de la réponse dipolaire presque aussi efficacement que celle, plus lente, des atomes.

Le modèle de l'oscillateur de Drude est appliqué à la molécule d'eau et calibré afin de reproduire par dynamique moléculaire quatre propriétés essentielles de l'eau liquide : l'enthalpie d'évaporation, la densité, la constante diélectrique et le coefficient d'auto-diffusion. L'étude systématique des paramètres du modèle révèle que, pour correctement décrire les propriétés diélectriques de l'eau en phase condensée, la polarisabilité moléculaire doit être renormalisée par rapport à sa valeur en phase gazeuse. Des conclusions semblables sont obtenues d'une étude de l'éthanol.

La simulation de mélanges d'éthanol et d'eau prouve que les modèles polarisables développés pour des substances pures peuvent être utilisés en combinaison.

On constate notamment que les modèles combinés reproduisent la variation irrégulière de la constante diélectrique des mélanges sur toute la plage de concentrations. Le modèle permet aussi de jeter un éclairage nouveau sur l'origine des propriétés de mélange anormales des solutions d'alcool et d'eau, et sur le phénomène de structuration des liaisons hydrogène de l'eau autour d'un groupe non-polaire.

L'étude de l'hydratation des ions alcalins et halogènes à l'aide de la fonction potentielle polarisable montre qu'il est possible de reproduire les énergies libres d'hydratation des ions alcalins sans en sacrifier les propriétés microscopiques. L'hydratation des ions halogènes montre certaines lacunes du modèle d'eau, mais fournit des pistes pour le rendre transférable à une plus grande variété de milieux. Pour les cations, le modèle polarisable reproduit correctement la dynamique d'échange des molécules d'eau entre la couche de solvatation des ions et le liquide. Une analyse de la coordinence des ions suggère les mécanismes d'échange suivants : association-dissociation $4 \rightarrow 5 \rightarrow 4$ pour le lithium et dissociation-association $6 \rightarrow 5 \rightarrow 6$ pour le sodium. Pour le potassium, le modèle polarisable prédit une coordinence fluide de 6 à 8 molécules d'eau.

MOTS CLÉS : Modélisation moléculaire, polarisabilité moléculaire, structure des liquides, hydratation hydrophobe, solvatation, eau, éthanol, lithium, sodium, potassium.

Summary

Although the behavior of a collection of molecules can in principle be computed from Schrödinger's equation, molecular interactions are often represented, for the purpose of simulating solvated biomolecules and complex liquids, as a sum of atomic interactions depending on the position of the atoms but not on their internal structure. This approximation can be significantly improved by recognizing that atoms are being polarized by the electric field generated by their environment and that the electrostatic interaction between two atoms is affected by the proximity of a third atom.

In this thesis, we formulate a polarizable potential function describing the dipolar response of atoms with the use of classical Drude oscillators. Each oscillator is a charged particle coupled to the position of an atom by a harmonic force that creates an induced dipole when shifting its position under the effect of an electric field. The Drude oscillator reproduces the principles of dipolar induction when it occupies its minimum energy state consistently with the neighboring oscillators. We design a molecular dynamics algorithm that keeps the oscillators at their auto-consistent positions without having recourse to an iterative approach. We show that this algorithm integrates the fast dynamics of the dipolar response as efficiently as the slower dynamics of the atoms.

The Drude oscillator model is applied to the water molecule and calibrated so to reproduce four salient properties of liquid water : evaporation enthalpy, density, dielectric constant, and self-diffusion constant. A systematic study of the parameters of the model reveals that, in order to correctly reproduce the dielectric properties of water in condensed phase, the molecular polarizability should be renormalized with respect to its gas-phase value. Similar conclusions are drawn from the study of ethanol.

The simulation of ethanol–water mixtures shows that polarizable models calibrated for pure compounds can be utilized in combination. For instance, it is observed that the combined water and ethanol models reproduce the specific va-

riation of the dielectric constant of mixtures for the whole range of concentrations. The model also provides new insight on the origin of the abnormal mixing properties of alcohol–water solutions, and on the structuring of hydrogen bonds of water in the vicinity of a nonpolar group.

The study of the hydration of alkali and halide ions using a polarizable potential function shows that it is possible to correctly reproduce the hydration free energies of alkali ions without sacrificing the microscopic properties. The hydration of halide ions reveals some deficiencies of the water model, but provides hints on how to improve its transferability. For cations, the polarizable model correctly reproduces the dynamics of the exchange of water molecules from the first solvation shell to the bulk liquid. A detailed analysis of the ionic coordination suggests the following water exchange mechanisms : $4 \rightarrow 5 \rightarrow 4$ association-dissociation for lithium and $6 \rightarrow 5 \rightarrow 6$ dissociation-association for sodium. For potassium, the polarizable model predicts a fluid coordination of 6 to 8 water molecules.

KEY WORDS : Molecular modeling, molecular polarizability, liquid structure, hydrophobic hydration, solvation, water, ethanol, lithium, sodium, potassium.

Table des matières

Liste des tableaux

Liste des figures

à Gaston Jeannotte, mon grand-père
qui a rêvé d'éducation

Remerciements

Je remercie Benoît Roux, qui m'a inspiré et guidé sur le parcours de ce doctorat. J'ai pu profiter de sa capacité d'aller au vif d'un sujet nouveau ou d'une question difficile, de ses vastes connaissances en biophysique, de sa profonde compréhension des rouages de la mécanique statistique et de son bon jugement vis-à-vis de leur utilisation. Il m'a généreusement introduit à son grand réseau de recherche et donné l'occasion d'y faire valoir mon travail. Je lui suis aussi reconnaissant de m'avoir accordé sa confiance et offert la liberté d'approfondir les questions périphériques qui m'ont semblé importantes. Je remercie Alexander D. MacKerell, Jr., mon collaborateur à l'University of Maryland, dont l'enthousiasme et l'intuition m'ont fourni un supplément de motivation utile dans certains passages difficiles. Je remercie aussi Alain Caillé, qui a continué de me supporter malgré son changement de carrière et m'a été de bon conseil à plusieurs reprises. C'est d'ailleurs lui qui m'a présenté à Benoît.

Je remercie mes collègues du Weill Medical College of Cornell University pour l'intérêt qu'ils ont porté à mes travaux et leur plaisir à partager leur propres recherches et réflexions. Je retiendrai mes nombreuses conversations avec Simon Bernèche, Nilesh K. Banavali, Esin Kutluay et José D. Faraldo-Goméz, ainsi qu'avec Toby W. Allen, Sergey Yu. Noskov, Wonpil Im et Deniz Sezer. À l'Université de Montréal, j'ai apprécié le contact de David Allouche, Stefan Seefeld, Pierre-Jean L'Heureux et Mark E. Casida. Je salue spécialement Simon Bernèche, Nilesh K. Banavali et Fabienne Paumet, avec qui j'ai tissé des liens d'amitié qui ont enrichi mes années à New York.

Plusieurs professeurs ont joué des rôles importants dans mon développement en tant que scientifique : Jean-René Roy à l'Université Laval, Laurent G. Caron, David Sénéchal et Claude Bourbonnais à l'Université de Sherbrooke, et John Gunn et Dennis Salahub à l'Université de Montréal. C'est avec une pointe de satisfaction que je reconnais parfois la marque de leur enseignement dans mes façons de penser et de faire.

Je remercie mes parents, Paul Lamoureux et Francine Jeannotte, pour leur affection et leur support indéfectible. Ce sont eux qui m'ont donné le goût de la science et de la poursuite de l'excellence, et alors que je prends la mesure du chemin parcouru, je leur en suis profondément reconnaissant.

Avant-propos

Les chapitres 2, 4 et 5 de cette thèse sont des articles que j'ai publiés dans des journaux avec arbitrage en tant que premier auteur. Ces articles sont reproduits tels quels, dans leur langue d'origine, et avec un minimum de changements dans leur forme. Pour cette raison, la notation mathématique n'est pas complètement uniforme. Le chapitre 6 est un manuscrit en préparation pour lequel je suis également premier auteur. Les chapitres en français, par contre, sont exclusifs à la thèse. Un article publié dont je suis deuxième auteur mais auquel j'ai fourni une contribution importante a été mis en annexe.

Malgré que cette thèse soit rattachée de près à des principes physiques et chimiques relativement fondamentaux, j'ai tenu à la présenter dans le contexte plus large de la biophysique. J'aurais pu adopter le cadre plus mesuré de la physicochimie, mais j'aurais alors occulté l'esprit de ce travail et son objectif ultime : comprendre les détails atomiques des processus biologiques.

Les atomes ne sont que l'effet de la faiblesse de notre imagination, qui aime à se reposer et à se hâter à venir à une fin dans les sous-divisions ou analyses.

Gottfried Wilhelm von Leibniz (1646-1716),
Lettre à Nicolaas Hartsœker (1710)

How small a thought it takes to fill a whole life!

Ludwig Wittgenstein (1889-1951),
Culture and Value (1980)

Chapitre 1

Introduction

1.1 Biomolécules en solution

Tout être vivant, une fois sa surface traversée, est constitué de molécules en solution. La cellule bactériale *Escherichia coli*, par exemple, est composée de 70% d'eau, 17% de protéines, 8% d'ARN, 1% d'ADN, 1% de glucides, 1% d'ions inorganiques, 1% de lipides et 1% de petites molécules organiques diverses, notamment d'acides aminés et de nucléotides [1]. La figure 1.1 en illustre la structure relativement simple : une double paroi cellulaire incrustée d'un flagelle unique et contenant le cytoplasme et le nucléoïde. Sur la figure, seuls les macromolécules et les lipides des membranes sont représentés. Ceux-ci baignent en fait dans une solution d'ions inorganiques et d'une grande variété de petites molécules organiques, invisible pour les besoins de l'illustration.

La figure 1.1—dont les proportions sont réalistes—est frappante à plusieurs points de vue. Elle présente le cytoplasme, là où se produit la majorité des réactions biochimiques, comme une « soupe » dense de protéines de tailles diverses, séparées les unes des autres par une mince couche d'eau. Le cytoplasme est en fait si dense qu'il forme presque un *gel*, au sens physique du terme, c'est-à-dire qu'il est une solution colloïdale à la limite de la solidification. Dans ce contexte, l'eau joue un rôle essentiel : parce qu'elle a une constante diélectrique élevée (79.4 fois plus grande que celle de l'air), elle atténue les interactions électrostatiques

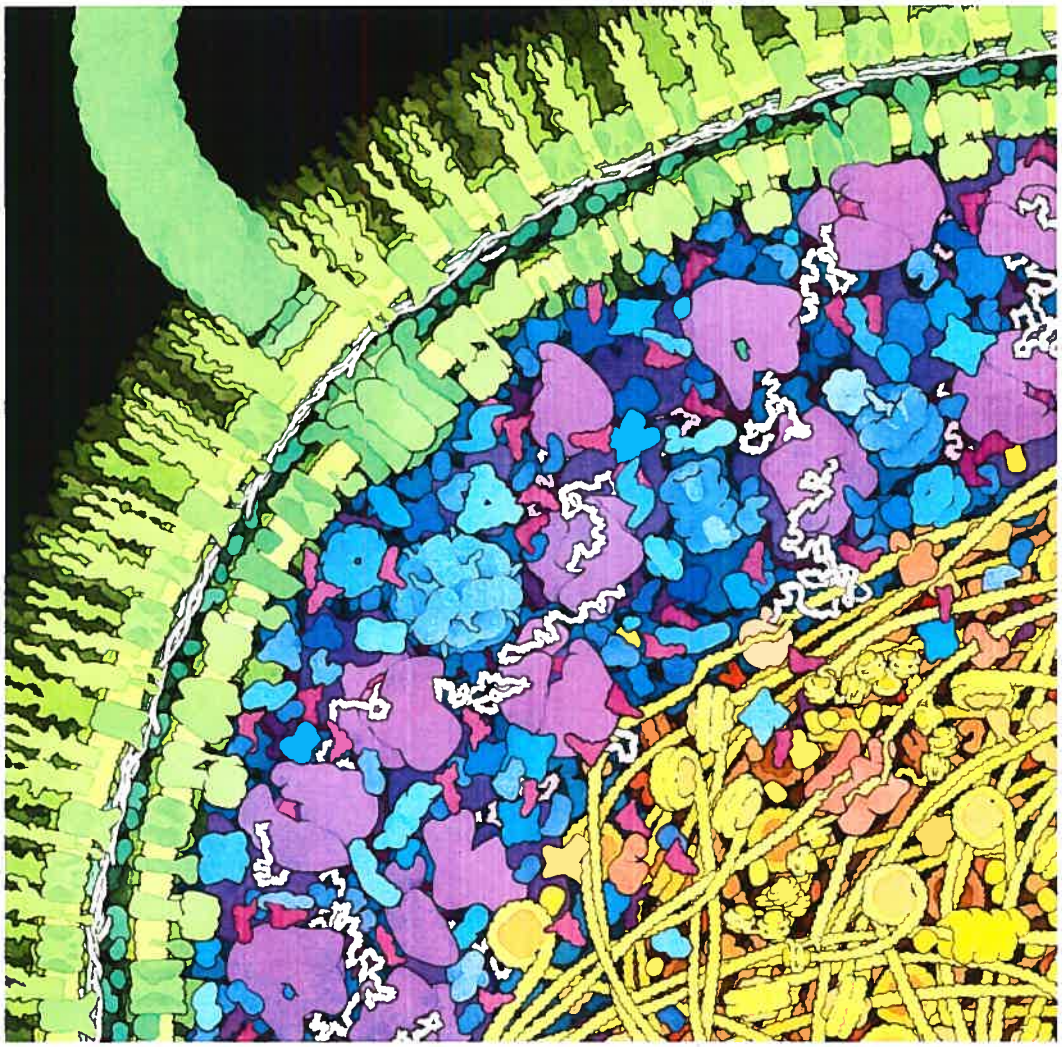

FIG. 1.1 – Illustration d'une portion de la bactérie *Escherichia coli*. Dans les tons de vert : la paroi cellulaire, formée de deux membranes dans lesquelles sont insérées des protéines transmembranaires, dont un moteur flagellaire (traversant la paroi de part en part) et le flagelle qu'il fait tourner. Dans les tons de bleu et de violet : le cytoplasme, dans lequel on voit notamment des ribosomes (les grosses molécules violettes), des ARN de transfert (les petites molécules bourgogne, en forme de L) et des ARN messagers (en blanc). Dans les tons de jaune et d'orangé : la région du nucléoïde, qui contient une seule molécule d'ADN dont certains segments sont enroulés autour de protéines HU. En rouge-orangé, on voit un complexe d'ADN-polymérase en train de répliquer un fragment de l'ADN.

non-spécifiques et permet entre les biomolécules un jeu de forces plus subtil, basé sur l'*effet hydrophobe* et sur l'association de groupes chimiques particulièrement complémentaires. Sans eau pour écranter les interactions de leurs groupes polaires ou chargés, les protéines du cytoplasme s'agglomèreraient indistinctement et perdraient leur fonction biologique.

La valeur élevée de la constante diélectrique de l'eau tient au fait que chaque molécule d'eau possède un moment dipolaire élevé qui, dans la phase liquide, s'oriente facilement dans la direction d'un champ électrique créé par un groupe chimique polaire ou chargé. Le champ électrique créé par le réarrangement des molécules neutralise presque complètement le champ original. Ainsi, l'eau en phase liquide possède deux propriétés en apparence irréconciliables : elle est fortement polaire, mais facilement réorientable. Ce paradoxe est expliqué par la façon particulière dont les molécules d'eau font et défont leurs *liaisons hydrogène*. Notons en passant que la glace (en phase 1h) possède une constante diélectrique encore plus élevée : 96.6 à 0°C. Cette valeur est due plutôt à une plus grande mobilité des protons en phase cristalline qu'en phase liquide, et à la diffusion facile de défauts orientationnels.

Une molécule d'eau est susceptible de former quatre liaisons hydrogène, en tant que donneur selon deux directions et en tant qu'accepteur selon deux autres directions. Ces liaisons forment une structure tétraédrique, illustrée à la figure 1.2. Plus généralement, une liaison hydrogène se forme entre deux atomes fortement électronégatifs, par l'intermédiaire d'un atome d'hydrogène. Le groupe qui fournit l'atome d'hydrogène s'appelle le *donneur* et le groupe qui le reçoit s'appelle l'*accepteur*. Pour être plus précis, on pourra définir la liaison hydrogène comme « une interaction directionnelle entre une liaison polaire, formée par un hydrogène et un atome donneur fortement électronégatif, et l'orbitale non-liante polaire d'un atome accepteur » [2, page 103]. Dans le contexte de la biochimie, une liaison hydrogène se forme typiquement entre l'atome d'hydrogène d'un groupe hydroxyle ($-$OH) ou amine ($\rangle$NH ou $-NH_2$) et une paire d'électrons libres d'un atome accepteur : l'oxygène d'un groupe hydroxyle ou carbonyle (O=C$\langle$), l'azote d'un groupe

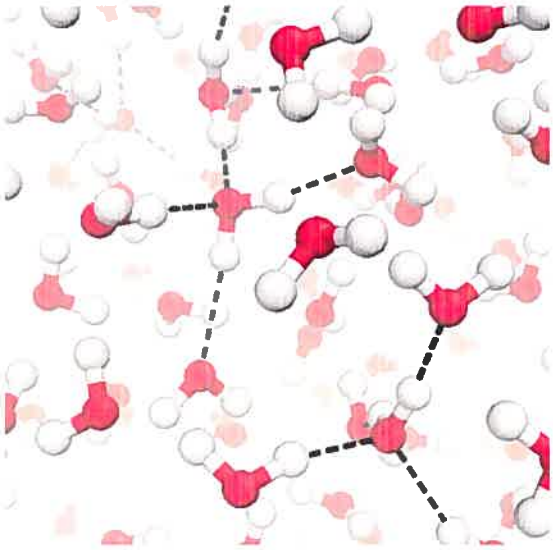

FIG. 1.2 – Illustration de la structure tétraédrique des liaisons hydrogène de la molécule d'eau en phase liquide. Les seules liaisons hydrogène représentées sont celles des molécules en possédant quatre.

amine ou le soufre d'un groupe sulfhydryle (−SH). L'oxygène et l'azote sont les deux éléments organiques les plus électronégatifs, de sorte que leurs groupes donneurs sont fortement polaires ($-O^{\delta-}H^{\delta+}$ et $>N^{\delta-}H^{\delta+}$, par exemple). L'orbitale non-liante forme quant à elle un dipôle pouvant s'étendre sur tout le groupe ($O^{\delta-}=C^{\delta+}<$, par exemple). Mentionnons que, malgré qu'il n'ait aucune fonction essentielle en biologie, le fluor peut également servir de donneur et d'accepteur.

La liaison hydrogène est une interaction dipôle–dipôle et, à ce titre, est relativement faible. Selon l'électronégativité des atomes donneur et accepteur, elle possède une énergie de formation de 2 à 10 kcal/mol, soit environ 10 fois moins qu'une liaison covalente. Par exemple, l'énergie d'une liaison hydrogène optimale entre deux molécules d'eau est environ 5 kcal/mol. Cette énergie est comparable à l'énergie des fluctuations thermiques à température ambiante et ne représente pas une barrière à la diffusion rapide d'une molécule d'eau dans le liquide. Plus exactement, il existe plusieurs mécanismes de basse énergie par lesquels une molécule d'eau du liquide se réoriente et se déplace, en brisant des liaisons hydrogène et en en formant de nouvelles [3, 4]. De fait, à température ambiante, l'eau est un liquide

de faible viscosité, comparable à celle de l'éthanol, une molécule plus grosse mais formant moins de liaisons hydrogène, et à peine le double de celle du benzène, une molécule significativement plus grosse mais ne formant pas de liaisons hydrogène.

Une molécule qui ne peut former aucune liaison hydrogène, comme l'éthane (CH_3CH_3), est de façon générale quasiment insoluble dans l'eau. Les faibles interactions *dispersives* entre l'éthane et les molécules d'eau l'entourant ne peuvent pas compenser le coût énergétique associé à la rupture de nombreuses liaisons hydrogène. Une molécule peut néanmoins posséder suffisamment de groupes polaires (susceptibles de former des liaisons hydrogène) pour être soluble dans l'eau malgré des groupes non-polaires. L'éthanol (CH_3CH_2OH), par exemple, qui possède à la fois un partie hydrophobe (CH_3CH_2-) et un groupe hydrophile ($-OH$), est parfaitement miscible avec l'eau.

C'est l'effet hydrophobe qui est responsable de la stabilité des membranes lipidiques. Les lipides sont des molécules *amphiphiles* (ou *amphipathiques*) allongées, dont une extrémité est polaire et l'autre est uniformément non-polaire. Les plus polaires—particulièrement les phospholipides, dont l'extrémité contient un groupe phosphate chargé—se solubilisent dans l'eau en formant des structures micellaires ou membranaires qui exposent leurs « têtes » hydrophiles mais protègent leurs « queues » hydrophobes. Cette séparation de phase, spontanée et stable pour des longues molécules fortement amphipathiques, est en fait le phénomène abouti analogue à la micro-agrégation métastable observée pour l'éthanol autour d'une concentration critique [5]. Cette micro-agrégation est d'autant plus marquée que la chaîne alcane est longue : plus marquée pour le propanol-1 ($CH_3CH_2CH_2OH$) que pour l'ethanol, et pour le butanol-1 ($CH_3CH_2CH_2CH_2OH$) que pour le propanol-1.

Pour autant qu'elle soit stable, la phase hydrophobe d'une membrane ou d'une micelle peut accueillir toute partie hydrophobe d'une molécule, dans la mesure où l'insertion ne force pas trop les limites de son élasticité. D'ailleurs, les protéines membranaires, qui jouent un rôle de médiation entre l'extérieur et l'intérieur d'une cellule ou d'un compartiment cellulaire, ont toutes des côtés hydrophobes d'une hauteur correspondant à l'épaisseur de la membrane.

Sans entrer dans les détails, ajoutons que la possibilité de former des liaisons hydrogène exerce un effet déterminant sur la structure des macromolécules. Les protéines présentent souvent une structure secondaire très caractéristique appelée « hélice α », qui est stabilisée par de nombreuses liaisons hydrogène entre les groupes amine et carbonyle du squelette peptidique (voir la figure 1.3). La stabilité d'une hélice α dépend de la polarité de ses chaînes latérales [6]. Par exemple, la polyalanine, dont les chaînes latérales—des groupes méthyle ($-CH_3$)—sont non-polaires, ne forme pas d'hélice α dans l'eau. À l'opposé, la polyasparagine, dont les chaînes latérales sont fortement polaires, adopte une solide structure en hélice α. Ce raisonnement n'est bien sûr valide que pour une hélice α isolée, puisque plusieurs hélices à la surface hydrophobe peuvent être stabilisées par agglomération. On sait d'ailleurs que les parties transmembranaires des protéines sont majoritairement constituées d'hélices α.

1.2 Biochimie des ions libres

Plusieurs types d'ions inorganiques participent aux processus biologiques. Les plus importants, qu'on retrouve en concentrations élevées dans tous les organismes vivants, sont le sodium (Na^+), le potassium (K^+), le chlore (Cl^-), le magnésium (Mg^{2+}) et le calcium (Ca^{2+}). Dans une cellule typique, ces concentrations vont de l'ordre de 0.1 M pour le potassium à 10^{-7} M pour le calcium. Dans le milieu extracellulaire, de même que dans certaines organites des cellules eucaryotes, la concentration de calcium est plutôt de l'ordre de 10^{-3} M, de sorte que du calcium doit être continuellement pompé à l'extérieur. Rappelons que 1 M (« un molaire ») représente une concentration de 1 mol/ℓ et qu'une solution ionique de 1 M dans des conditions de température et de pression normales contient environ une paire d'ions pour 55 molécules d'eau. Ces éléments sont très solubles dans l'eau et existent naturellement sous forme d'ions atomiques. On peut ajouter à cette courte liste le phosphore et le soufre, présents en solution sous forme de monohydrogénophosphate (HPO_4^{2-}) et de tétrasulfate (SO_4^{2-}), et fixés à des molécules biologiques sous

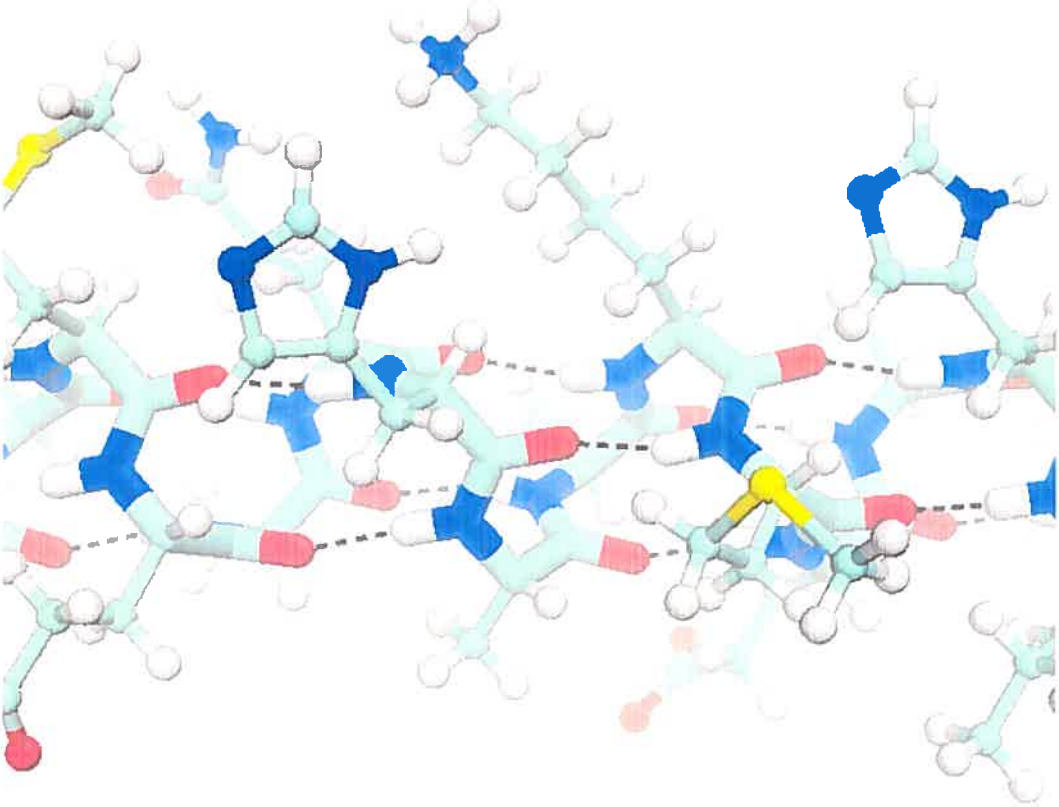

FIG. 1.3 – Illustration de la structure d'une hélice α. Les liaisons hydrogène sont montrées en lignes tiretées et le squelette peptidique est mis en évidence avec des liens de plus grand diamètre.

diverses formes de phosphates ($-PO_4^{2-}$) et de trisulfates ($=SO_3^{2-}$). S'ajoutent à ces ions « constitutifs » des ions « traces », quasiment inexistants en solution mais néanmoins essentiels au fonctionnement de plusieurs enzymes. Les plus courants sont le manganèse (Mn), le fer (Fe), le cobalt (Co), le nickel (Ni), le cuivre (Cu) et le zinc (Zn). Ces éléments, des métaux de transition, possèdent plusieurs états d'oxidation et servent de centres catalytiques à de nombreuses réactions chimiques [7].

La solubilité d'un ion dans l'eau est un facteur-clé pour en comprendre la fonction. Pour le sodium, le potassium et le chlore, l'eau crée une coordinence presque idéale : la couche de molécules d'eau en contact direct avec l'ion se réorganise de façon à profiter pleinement de la forte interaction entre la charge de l'ion et le dipole de chaque molécule d'eau. Cette coordinence ne peut être reproduite à la surface d'une protéine que par des structures moléculaires très spécifiques (comme le filtre de sélectivité des canaux potassiques [8]), de sorte que les ions se retrouvent la plupart du temps libres dans la solution. Ainsi, le sodium, le potassium et le chlore ont pour principales fonctions de contrôler l'équilibre osmotique et de diffuser rapidement un signal électrolytique [7]. Ils sont présents dans la cellule en concentrations suffisamment élevées pour établir de part et d'autre d'une membrane cellulaire un potentiel électrostatique capable d'actionner une impulsion nerveuse. Ils servent aussi à neutraliser des polyélectrolytes et à contrôler leur interaction. Par exemple, chaque paire de bases d'une molécule d'ADN possède une charge nette de $-1|e|$ neutralisée à 70% par la condensation d'une « atmosphère » d'ions positifs.

Le calcium et le magnésium existent sous forme d'ions Ca^{2+} et Mg^{2+} à l'état naturel (dans l'eau de mer, par exemple), mais peuvent être mieux coordonnés par des groupes phosphates ($-PO_4^{2-}$) ou carboxylates ($-CO_2^-$) associés à des groupes carbonyle ($\rangle C{=}O$), de sorte qu'ils se lient préférablement à des protéines [7]. Le plus polyvalent des deux est l'atome de calcium, qui peut être coordonné par une plus grande variété de groupes et selon des directions plus irrégulières [7, 9]. Il existe d'ailleurs plusieurs types de sites de liaison très sélectifs au calcium—notamment le motif « main EF » (en anglais, *EF-hand*). Par opposition, le magnésium est préféra-

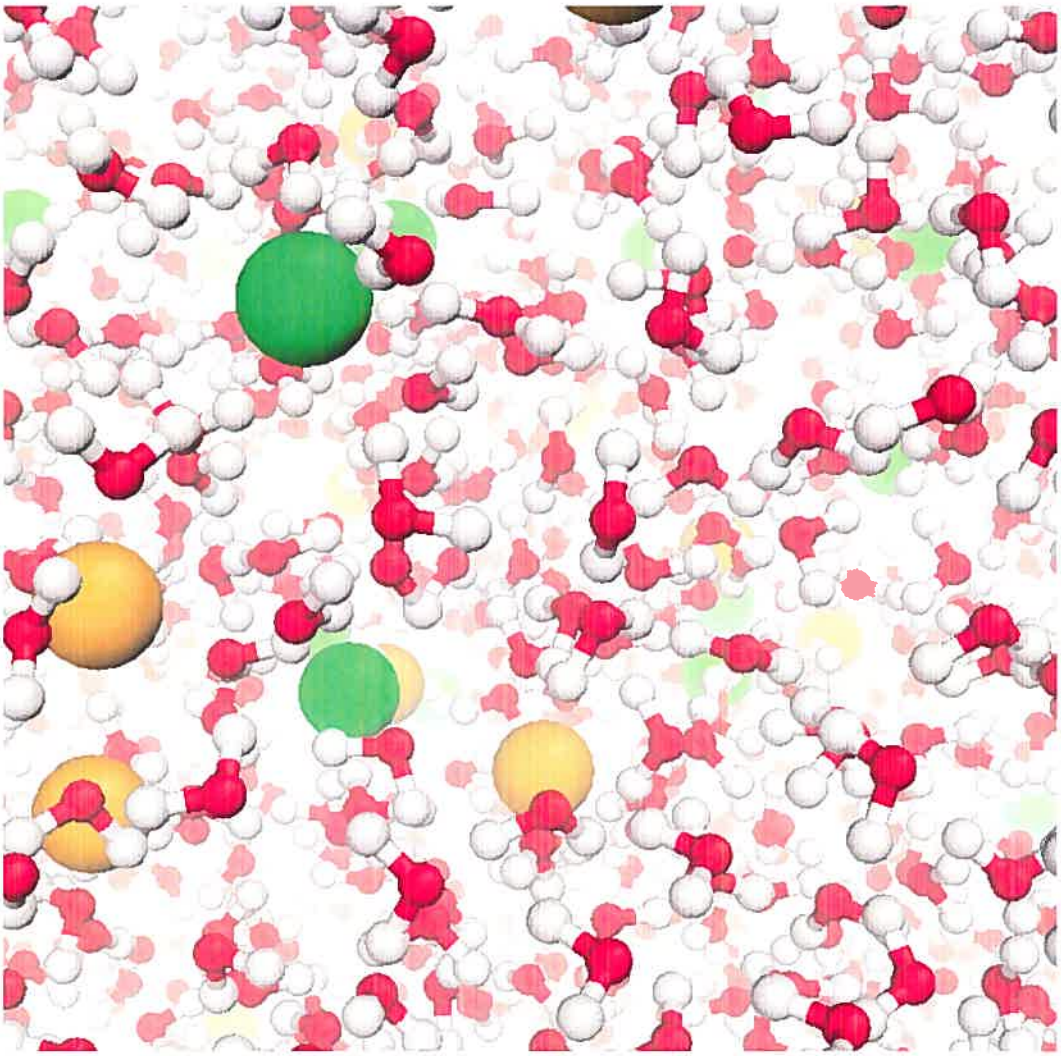

FIG. 1.4 – Illustration d'une solution de KCl à une concentration de 1 M. Les atomes de potassium sont orangés et les atomes de chlore sont verts. À cette concentration, il y a une paire K^+–Cl^- pour environ 55 molécules d'eau.

blement coordonné par six groupes (dont un groupe carboxylate), qui doivent être arrangés régulièrement. Parce qu'ils peuvent lier des acides aminés séparés dans la séquence, tous les deux jouent un important rôle structural en stabilisant certaines structures tertiaires de protéines. La plus grande partie reste néanmoins libre en solution, disponible notamment pour se fixer au site d'une protéine et y induire un changement structural. Par exemple, la contraction musculaire est déclenchée par des concentrations élevées de calcium en solution, qui se lie aux sites « main EF » de la *troponine C*. Le calcium ainsi lié déforme le complexe *troponine–tropomyosine* et permet à la *myosine* d'interagir avec les filaments d'*actine* et d'initier le cycle de contraction [1, page 308].

Les métaux de transition sont pratiquement insolubles et beaucoup mieux coordonnés par des structures moléculaires spécifiques que par des molécules d'eau. Dans leur coordinence naturelle, les énergies de liaison de ces éléments sont comparables à celles d'un lien covalent ordinaire, et les organismes vivants doivent employer des mécanismes complexes pour les en extraire. Par exemple, le fer, essentiel au transport de l'oxygène, se retrouve à l'état naturel sous forme d'hydroxide de fer trivalent ($Fe(OH)_3$), et doit passer par une série de réactions chimiques pour être employé sous forme divalente (Fe^{2+}) dans une structure en couronne extrêmement spécifique : la *porphyrine*.

En résumé, le mélange d'eau, d'ions inorganiques et de petites molécules solubles dans lequel baignent les biomolécules possède de nombreuses fonctions essentielles : il écrante les interactions non-spécifiques entre les protéines et force une cohésion hydrophobe entre certains de leurs éléments de structure, il permet à toute une classe d'ions « libres » (principalement le sodium, le potassium, le chlore, le calcium et le magnésium) de participer à des processus électrophysiologiques ou structuraux, et il assure un milieu fluide dans lequel une molécule peut facilement se déplacer. Du point de vue de la biophysique théorique, il est capital de comprendre les propriétés physiques sous-jacentes à ces fonctions.

1.3 Principes de la mécanique moléculaire

Un autre aspect frappant de la figure 1.1 (page 2) est qu'elle présente une vision moléculaire d'un organisme *complet*, et suggère qu'il est possible en principe d'avoir une compréhension à l'échelle atomique du fonctionnement global d'une bactérie (*Escherichia coli*, dans ce cas). De ce point de vue réductionniste, l'intérieur d'une bactérie n'est ni plus ni moins qu'un ensemble de molécules complexes en solution, dont la structure et les mouvements sont gouvernés par les lois de la physique. En connaissant suffisamment bien la composition et la structure d'un sous-système fonctionnant en relative indépendance du reste de l'organisme, et en disposant des ressources de calcul nécessaires pour y appliquer les lois de la physique à l'échelle atomique, on pourrait analyser en détail le déroulement d'un processus biologique et le rôle que chaque composante y joue.

1.3.1 Équation de Schrödinger

En principe, l'évolution d'un système décrit par l'hamiltonien $\mathscr{H}$ est régie par l'équation de Schrödinger dépendant du temps :

$$i\frac{\partial \Psi}{\partial t} = \mathscr{H}(t)\Psi. \tag{1.1}$$

(Notons qu'on utilise les unités atomiques, pour lesquelles $\hbar = 1$, $m_e = 1$ et $|e| = 1$.) Pour un système de N atomes, constitué de N noyaux se partageant n électrons, on a un hamiltonien de la forme suivante [10] :

$$\begin{aligned}\mathscr{H}(t) \;=\; & -\sum_{i}^{n}\frac{1}{2}\nabla_i^2 - \sum_{I}^{N}\frac{1}{2M_I}\nabla_I^2 + \sum_{i}^{n}\sum_{j>i}^{n}\frac{1}{r_{ij}} - \sum_{i}^{n}\sum_{J}^{N}\frac{Z_J}{r_{iJ}} \\ & + \sum_{I}^{N}\sum_{J>I}^{N}\frac{Z_I Z_J}{r_{IJ}} + \mathscr{E}(t).\end{aligned} \tag{1.2}$$

Les deux premières sommations représentent les énergies cinétiques des électrons et des noyaux. Les paramètres $\{M\}$ sont les masses des noyaux (en unités de

masse électronique). Les trois doubles sommations représentent, dans l'ordre, les interactions coulombiennes électron–électron, électron–noyau et noyau–noyau. Au hamiltonien *moléculaire* proprement dit est ajouté un terme $\mathscr{E}(t)$, qui représente l'effet de l'environnement. Faisant abstraction des variables de spin, l'état du système peut être décrit par la fonction d'onde $\Psi(\{\mathbf{r}\}, \{\mathbf{R}\}, t)$, où $\{\mathbf{r}\}$ représente les positions des électrons et $\{\mathbf{R}\}$ représente les positions des noyaux.

À l'aide des équations (1.1) et (1.2), on pourrait par exemple calculer le résultat de la collision à basse énergie de deux atomes d'argon. Pour des atomes dans l'état fondamental sur une « trajectoire » de collision, l'équation de Schrödinger fournirait les probabilités de transition de l'état fondamental vers des états excités, et d'un état excité vers un autre. L'intégration de cette information détaillée permettrait de prédire les sections efficaces de collision. Malgré la relative simplicité du système, la solution directe des équations (1.1) et (1.2) est un problème extrêmement difficile : chaque atome d'argon contient 18 électrons et ses états excités ont des configurations complexes. L'article de revue de Deumens *et al.* [11] est un bon point de départ sur la question. L'équation (1.1) n'a été solutionnée directement, par des méthodes numériques basées sur des théories perturbatives, que pour des systèmes de collision plus simples contenant au plus une dizaine d'électrons : une collision entre un atome d'hélium et un atome de néon, par exemple [12]. En fait, les plus récents calculs de la section efficace d'une collision de deux atomes d'argon [13] escamotent complètement l'aspect dynamique du problème en utilisant une fonction potentielle empirique et une approximation adiabatique pour le calcul des phases de diffusion. Il se trouve malgré tout que les résultats de ce calcul approximatif sont comparables aux sections efficaces expérimentales jusqu'à des énergies de collision de plusieurs keV.

L'objectif de la mécanique moléculaire est de remplacer la difficile équation du mouvement de Schrödinger par une équation simplifiée capable de décrire approximativement la mécanique de plusieurs atomes à l'aide de lois de force simplifiées appliquées à un nombre réduit de degrés de liberté.

1.3.2 Approximation de Born–Oppenheimer

L'approximation de Born–Oppenheimer [14] tire avantage du fait que les noyaux sont beaucoup plus massifs que les électrons, et donc que leur mouvement est beaucoup plus lent que celui des électrons. En d'autres termes, on peut imaginer que les électrons se déplacent dans un champ nucléaire quasiment statique et, à l'inverse, que les noyaux se déplacent dans un champ électronique moyen. Le problème quantique complet est ainsi divisé en un problème « nucléaire » dépendant du temps :

$$i\frac{\partial \Psi_{\mathrm{n}}}{\partial t} = \mathscr{H}_{\mathrm{n}}(t)\Psi_{\mathrm{n}}, \tag{1.3}$$

et un problème « électronique » dont la dépendance au temps est implicite, correspondant à chaque instant à l'équation de Schrödinger indépendante du temps :

$$\mathscr{H}_{\mathrm{e}}(\{\mathbf{R}\})\Psi_{\mathrm{e}} = E_{\mathrm{e}}(\{\mathbf{R}\})\Psi_{\mathrm{e}}. \tag{1.4}$$

L'hamiltonien électronique est

$$\mathscr{H}_{\mathrm{e}}(\{\mathbf{R}\}) = -\sum_{i}^{n}\frac{1}{2}\nabla_i^2 + \sum_{i}^{n}\sum_{j>i}^{n}\frac{1}{r_{ij}} - \sum_{i}^{n}\sum_{J}^{N}\frac{Z_J}{r_{iJ}}. \tag{1.5}$$

Il dépend du temps par l'intermédiaire des positions des noyaux : $r_{iJ} = |\mathbf{r}_i - \mathbf{R}_J(t)|$. En traitant les coordonnées nucléaires $\{\mathbf{R}\}$ de l'équation (1.4) comme des *paramètres*, on fait l'approximation que l'opérateur nucléaire ∇_I^2 a un effet négligeable sur la fonction électronique Ψ_{e}. L'hamiltonien nucléaire, quant à lui, est

$$\mathscr{H}_{\mathrm{n}}(t) = -\sum_{I}^{N}\frac{1}{2M_I}\nabla_I^2 + E_{\mathrm{e}}(\{\mathbf{R}\}) + \sum_{I}^{N}\sum_{J>I}^{N}\frac{Z_I Z_J}{r_{IJ}} + \mathscr{E}(t). \tag{1.6}$$

En utilisant la fonction E_{e} plutôt que l'opérateur $\mathscr{H}_{\mathrm{e}}$, on fait l'approximation que l'opérateur électronique ∇_i^2 est sans effet sur la fonction nucléaire Ψ_{n}. Par consé-

quent, on se trouve à considérer des solutions de la forme suivante :

$$\Psi(\{\mathbf{r}\},\{\mathbf{R}\},t) = \Psi_{\mathrm{n}}(\{\mathbf{R}\},t)\,\Psi_{\mathrm{e}}(\{\mathbf{r}\},\{\mathbf{R}\}). \tag{1.7}$$

Le problème nucléaire fait intervenir la *fonction potentielle de Born–Oppenheimer*

$$U_{\mathrm{BO}}(\{\mathbf{R}\}) = E_{\mathrm{e}}(\{\mathbf{R}\}) + \sum_{I}^{N}\sum_{J>I}^{N}\frac{Z_I Z_J}{r_{IJ}} \tag{1.8}$$

qui, plutôt que d'être calculée, pourrait en principe être définie par une fonction empirique des coordonnées nucléaires.

C'est une approche semblable qui avait été choisie par Phelps *et al.* [13] pour le calcul de la section efficace de collision de deux atomes d'argon, mentionné plus haut. La fonction potentielle utilisée pour cette étude est compliquée et manifestement empirique :

$$U_{\mathrm{BO}}(r) = \begin{cases} 0.3359\left[113212\,e^{-9.0005x-2.6027x^2}\right. & \\ \qquad \left. - F(x)\left(\dfrac{1.0997}{x^6}+\dfrac{0.54512}{x^8}+\dfrac{0.39279}{x^{10}}\right)\right] & \text{pour} \quad r>5, \\ 61.6\,\dfrac{e^{-(r/5.5)^8}}{1+(r/5)^4}\,\dfrac{1}{r^{9/2}} & \text{pour} \quad 3\le r\le 5, \\ 90.25\,e^{r/0.79}\,\dfrac{1}{r^{3/2}} & \text{pour} \quad r<3. \end{cases} \tag{1.9}$$

avec $x \equiv r/7.107$ et $F(x) = e^{-(1.004/x-1)^2}$ pour $x < 1.004$ et $F(x) = 1$ pour $x \geq 1.004$. (Tous les nombres sont en unités atomiques.) L'interaction est répulsive pour $r < 6.3$ environ, c'est-à-dire pour $r < 3.3$ Å. Le grand soin apporté à la forme de l'interaction pour des courtes distances témoigne de l'importance d'avoir une interaction précise jusqu'à des énergies de collision de plusieurs keV (1 keV = 23060.35 kcal/mol). Pour des collisions atomiques à des énergies « thermiques » (de l'ordre de $k_{\mathrm{B}}T = 0.596159$ kcal/mol $\simeq 0.026$ eV), la forme de la fonction U_{BO} peut être considérablement plus simple.

1.4 Modèles d'interactions atomiques

1.4.1 Liens covalents

Les interactions les plus fortes entre deux atomes sont les liens covalents. L'énergie de dissociation d'un lien covalent est beaucoup plus grande que $k_{\mathrm{B}}T$, et il est courant de faire l'approximation que deux atomes partageant un lien covalent sont liés harmoniquement. La distance entre les atomes liés I et J est notée ℓ_{IJ}, et l'énergie potentielle de vibration a la forme suivante :

$$U_\ell(\{\mathbf{r}\}) = \sum_{\ell(I,J)} k^\ell_{IJ}(\ell_{IJ} - \ell^\circ_{IJ})^2. \tag{1.10}$$

La constante k^ℓ_{IJ} représente la force du lien covalent. Ce modèle ne fournit pas la possibilité de briser un lien, et ne représente pas non plus le caractère anharmonique de la vibration. En réalité, un lien covalent est de plus en plus facile à étirer et peut éventuellement se briser. Une forme plus réaliste est le potentiel de Morse, $U(r) = D[1 - e^{-a(r-r^\circ)}]^2$, qui est anharmonique et possède une énergie de dissociation D. Cela dit, comme l'énergie nécessaire pour briser un lien covalent est environ deux ordres de grandeur plus grande que l'énergie des fluctuations thermiques, l'approximation harmonique est généralement valide. La très grande majorité des processus biochimiques est en fait gouvernée par des interactions moléculaires d'énergies plus faibles. Le potentiel U_ℓ, correctement calibré, peut servir à décrire les vibrations quasi-classiques des liaisons entre deux atomes « lourds » — c'est-à-dire, autres que des atomes d'hydrogène. Les vibrations des atomes d'hydrogène, quant à elles, sont suffisamment rapides que leur influence sur la dynamique des atomes lourds est inessentielle. On impose donc habituellement une distance fixe ℓ°_{IJ} entre chaque atome d'hydrogène et son atome lourd.

Pour un atome formant plusieurs liens covalents, on utilise des termes d'énergie supplémentaires qui modélisent les déformations de sa coordinence. L'angle entre deux liens (I, J) et (J, K), appelé θ_{IJK}, est maintenu proche de sa valeur à

l'équilibre θ°_{IJK} par un potentiel harmonique, et on ajoute à la fonction potentielle

$$U_{\theta}(\{\mathbf{r}\}) = \sum_{\theta(I,J,K)} k^{\theta}_{IJK}(\theta_{IJK} - \theta^{\circ}_{IJK})^2. \tag{1.11}$$

Un atome central hybridé sp^3 (le carbone d'un groupe méthyle $-CH_3$, par exemple) possède une coordinence tétraédrique et les angles à l'équilibre sont proches de 109.47°. Un atome hybridé sp^2 (le carbone d'un groupe méthylène $=CH_2$, par exemple) possède une coordinence trigonale et les angles à l'équilibre sont proches de 120° (entre les deux hydrogènes, et entre chaque hydrogène et l'atome participant à la liaison double). Finalement, un atome hybridé sp (le carbone d'un groupe méthylyne $\equiv CH$, ou celui du gaz carbonique CO_2, par exemple) possède une coordinence linéaire et l'angle à l'équilibre est 180°.

Deux atomes hybridés sp^3 formant une liaison—deux groupes méthyle formant une molécule d'éthane (CH_3CH_3), par exemple—sont plus stables sous une conformation *décalée* que sous une conformation *éclipsée*. La barrière d'énergie associée à la torsion peut être décrite en termes des angles dièdre ϕ_{IJKL}, correspondant aux angles entre les plans (I, J, K) et (J, K, L) d'une séquence de liaisons (I, J), (J, K) et (K, L) (voir la figure 1.5). Pour l'éthane, les indices J et K correspondent aux deux atomes de carbone et les indices I et L correspondent respectivement aux atomes d'hydrogène d'un groupe et de l'autre, et la barrière d'énergie varie approximativement comme $1 + \cos 3\phi$. Le facteur 3 tient compte de la symétrie du groupe méthyle. Pour l'éthylène ($CH_2{=}CH_2$), la conformation la plus stable est la conformation éclipsée, et l'énergie de torsion varie comme $1 + \cos(2\phi + 180°)$. Pour des liens quelconques, on écrira la combinaison suivante

$$U_{\phi}(\{\mathbf{r}\}) = \sum_{\phi(I,J,K,L)} \sum_{n} k^{\phi}_{IJKL,n}[1 + \cos(n\phi_{IJKL} - \delta_{IJKL,n})], \tag{1.12}$$

avec les multiplicités $n = 1$, 2, 3, 4 et 6. En plus des termes $n = 2$ et $n = 3$, les termes $n = 1$ permettent de contrôler les stabilités relatives des conformations *cis* et *trans*, et les termes $n = 4$ et $n = 6$ servent à représenter des défauts de

symétrie pour des molécules cycliques. Les phases δ sont 0 ou 180°, selon que les conformations les plus stables sont décalées ou éclipsées.

On utilise également des forces de rappel fonctions de l'angle dièdre « impropre » ω_{IJKL}, défini comme l'angle entre les plans passant par les triplets d'atomes (I, J, K) et (J, K, L) (voir la figure 1.5) :

$$U_\omega(\{\mathbf{r}\}) = \sum_{\omega(I,J,K,L)} k^\omega_{IJKL}(\omega_{IJKL} - \omega^\circ_{IJKL})^2. \tag{1.13}$$

Cet angle dièdre est nul si l'atome central I, commun aux trois liens, est dans le plan des trois atomes liés J, K et L. La valeur d'équilibre ω° permet donc de contrôler la planarité de la coordinence. Ces termes sont utiles notamment pour contrôler les mouvements hors du plan des atomes d'oxygène et d'hydrogène du squelette peptidique et des atomes d'hydrogène des cycles *imidazole* (voir la figure 1.5). Également utiles pour ajuster plus finement les fréquences des modes de déformation sont des termes de correction de type « Urey–Bradley », qui ajoutent des forces de rappel harmoniques entre les atomes I et K d'un angle (I, J, K) :

$$U_{\mathrm{UB}}(\{\mathbf{r}\}) = \sum_{\mathrm{UB}(I,J,K)} k^{\mathrm{UB}}_{IJK}(r_{IK} - r^\circ_{IK})^2. \tag{1.14}$$

1.4.2 Interactions moléculaires

Interactions électrostatiques

Lorsque deux atomes forment un lien covalent, l'atome le plus électronégatif s'accapare d'une plus grande part des électrons de valence et présente une charge nette négative δ_-. Conséquemment, l'atome le moins électronégatif présente une charge nette positive δ_+, et l'ensemble forme un dipôle d'autant plus grand que la différence d'électronégativité entre les deux atomes est grande. À des distances d'au moins quelques rayons atomiques, l'interaction entre deux tels groupes polaires correspond à l'interaction coulombienne entre les charges partielles de chacun des

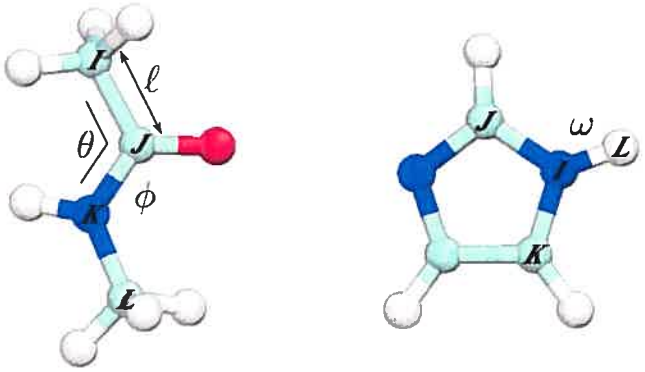

FIG. 1.5 – Schémas moléculaires définissant un lien $\ell(I,J)$, un angle $\theta(I,J,K)$, un angle dièdre $\phi(I,J,K,L)$ et un angle dièdre impropre $\omega(I,J,K,L)$. La molécule de gauche est le N-méthylacétamide, représentant le groupe amide du squelette peptidique. Celle de droite est l'imidazole, le groupe fonctionnel de l'acide aminé *histidine* (voir la figure 1.3).

quatre atomes :

$$U_{\text{élec}}(\{\mathbf{r}\}) = \frac{\delta_+\delta'_+}{|\mathbf{r}_+ - \mathbf{r}'_+|} + \frac{\delta_+\delta'_-}{|\mathbf{r}_+ - \mathbf{r}'_-|} + \frac{\delta_-\delta'_+}{|\mathbf{r}_- - \mathbf{r}'_+|} + \frac{\delta_-\delta'_-}{|\mathbf{r}_- - \mathbf{r}'_-|}. \tag{1.15}$$

Chaque terme est proportionnel à $1/r$ (où r est la distance entre les deux groupes) mais la somme est proportionnelle à $1/r^3$ et dépend fortement de l'orientation relative des deux groupes. On exclut délibérément les termes intramoléculaires $\delta_+\delta_-/|\mathbf{r}_+ - \mathbf{r}_-|$ et $\delta'_+\delta'_-/|\mathbf{r}'_+ - \mathbf{r}'_-|$ de l'équation (1.15) parce que, dans leur version *écrantée*, qui tient compte de la saturation de l'interaction à courte distance créée par l'étalement des charges partielles, ils peuvent être implicitement pris en compte par le terme U_ℓ de l'équation (1.10). De même on peut exclure le terme équivalent entre des atomes séparés par deux liaisons covalentes, puisqu'il peut être absorbé dans U_θ et U_{UB} [des équations (1.11) et (1.14)]. L'energie $U_{\text{élec}}$ modélise notamment la formation des liaisons hydrogène : elles se forment quand deux groupes polaires présentent l'un à l'autre des charges partielles de signes opposés et sont suffisamment colinéaires pour que la répulsion des charges de mêmes signes soit minimale.

Clairement, si on attribue une charge partielle à chaque atome, l'interaction électrostatique est simplement la somme des interactions de Coulomb entre tous

les atomes non-liés. Entre un groupe polaire (avec des charges partielles δ_+ et δ_-) et un cation monovalent (de charge $+1$), on a

$$U_{\text{élec}}(\{\mathbf{r}\}) = \frac{\delta_+}{|\mathbf{r}_+ - \mathbf{r}'|} + \frac{\delta_-}{|\mathbf{r}_- - \mathbf{r}'|}. \tag{1.16}$$

La somme est proportionnelle à $1/r^2$ et dépend de l'orientation du groupe polaire par rapport à l'ion. (Plus généralement, les interactions entre multipôles contribuent à l'énergie dans l'ordre suivant : charge–charge en $1/r$, charge–dipôle en $1/r^2$, charge–quadrupôle et dipôle–dipôle en $1/r^3$, charge–octopôle et dipôle–quadrupôle en $1/r^4$, et ainsi de suite.)

Énergie d'échange-répulsion-pénétration

Deux molécules chimiquement stables sont soumises à une répulsion à courte portée créée par le principe d'exclusion de Pauli (incluant la possibilité d'échange d'électrons d'une molécule à l'autre) et par la réduction de l'attraction électrostatique due à l'interpénétration des densités électroniques. Le principe d'exclusion de Pauli pénalise le recouvrement d'orbitales électroniques en contraignant les électrons de même spin à occuper des niveaux d'énergie différents. L'énergie de répulsion augmente exponentiellement avec le rapprochement des molécules. Pour des distances suffisamment courtes, elle domine toute attraction électrostatique et empêche véritablement les molécules de se superposer. Pour des collisions moléculaires « dures », elle n'a plus une dépendance exponentielle et doit être formulée comme un développement en puissances de $1/r$ [comme l'équation (1.9) en témoigne].

Énergie d'induction

L'interaction d'un groupe polaire et d'un ion, décrite par l'équation (1.16), est stabilisée par une série de termes d'*induction*. L'ion produit un champ électrique qui déforme la densité électronique du groupe polaire et y induit un moment dipolaire additionnel avec lequel l'ion interagit au même titre qu'avec le moment dipolaire permanent. Pour garder le raisonnement simple, supposons que l'ion est dans l'axe

du groupe polaire. Dans ce cas, le dipôle induit est parallèle au dipôle permanent et on pourrait représenter l'induction par une modification des charges partielles par une charge induite δ_{ind}. L'interaction électrostatique stabilisée s'écrirait

$$U_{\text{élec}}(\{\mathbf{r}\}) = \frac{\delta_+ + \delta_{\text{ind}}}{|\mathbf{r}_+ - \mathbf{r}'|} + \frac{\delta_- - \delta_{\text{ind}}}{|\mathbf{r}_- - \mathbf{r}'|}. \tag{1.17}$$

La charge induite δ_{ind} est liée au dipôle induit $\boldsymbol{\mu}_{\text{ind}}$ et à la longueur ℓ du lien polaire : $\delta_{\text{ind}} \sim \mu_{\text{ind}}/\ell$. Le dipôle induit, à son tour, est proportionnel à la polarisabilité α du groupe et au champ électrique $\mathbf{E}$ produit par l'ion. Ce champ décroit en $1/r^2$, de sorte que le terme d'induction décroit en $1/r^4$. L'équation (1.15) doit être revisitée de la même manière, pour tenir compte de la polarisabilité de chaque groupe. La charge induite δ_{ind} est proportionnelle à la polarisabilité du groupe et au champ électrique créé par le dipôle total de l'autre groupe (qui varie en $1/r^3$), de sorte que l'interaction est stabilisée par un terme proportionnel à $1/r^6$. (Plus généralement, les termes d'induction contribuent dans l'ordre suivant : charge–dipôle induit en $1/r^4$, dipôle–dipôle induit et charge–quadrupôle induit en $1/r^6$, quadrupôle–dipôle induit, dipôle–quadrupôle induit et charge–octopôle induit en $1/r^8$, et ainsi de suite.)

Les fonctions potentielles actuellement en usage pour les simulations de biomolécules sont *non-polarisables*, c'est-à-dire qu'elles incorporent les termes d'induction de manière effective. Pour une fonction potentielle non-polarisable, les charges partielles totales, représentées par $\delta_\pm \pm \delta_{\text{ind}}$ dans l'équation (1.17), sont fixées à une valeur effective qui représente l'effet d'induction d'un environnement de solvatation standard. Une fonction potentielle *polarisable*, à l'inverse, donne une forme explicite au terme d'induction, qu'il soit représenté par des incréments de charge δ_{ind} ou de véritables dipôles induits $\boldsymbol{\mu}_{\text{ind}}$. La notion de fonction potentielle polarisable est centrale à ce travail, et nous en expliquons plus longuement les principes à la section 1.4.4.

Énergie de dispersion

Il existe une interaction similaire, appelée *dispersion de London*, même pour des groupes dont les charges partielles δ sont nulles ou négligeables. Cette force provient des fluctuations quantiques dans la densité électronique d'une molécule, qui induisent des dipôles dans les molécules voisines et créent une contribution nette à l'attraction. Tout comme l'énergie d'induction, l'énergie de London est proportionnelle à la polarisabilité α de chaque groupe et décroit en $1/r^6$. On écrit

$$U_{\mathrm{L}}(\{\mathbf{r}\}) = -\frac{E_{\mathrm{I}}\alpha\alpha'}{|\mathbf{r}-\mathbf{r}'|^6}. \tag{1.18}$$

Le coefficient E_{I} est de l'ordre de l'énergie d'ionisation des atomes. Cette interaction des fluctuations dipolaires avait été incorporée dans l'équation (1.9) avec des contributions semblables pour les moments multipolaires supérieurs, en $1/r^8$ et $1/r^{10}$.

Considérons deux atomes isolés, représentés par les hamiltoniens quantiques $\hat{H}_1$ et $\hat{H}_2$ dont les états propres sont $\{\psi_{10}, \psi_{11}, \psi_{12}, \cdots\}$ et $\{\psi_{20}, \psi_{21}, \psi_{22}, \cdots\}$. Dans leurs états fondamentaux $|\psi_{10}\rangle$ et $|\psi_{20}\rangle$, ces atomes possèdent des dipôles moyens nuls. L'état fondamental du système combiné $\hat{H}_1 + \hat{H}_2$ est $|\psi_{10}, \psi_{20}\rangle = |\psi_{10}\rangle \otimes |\psi_{20}\rangle$. Si les atomes sont couplés par une interaction $\hat{W}$, le nouvel hamiltonien est $\hat{H}_1 + \hat{H}_2 + \hat{W}$, l'état couplé est $|\Psi\rangle$ et l'énergie d'interaction est

$$U = \langle\Psi|\hat{H}_1 + \hat{H}_2 + \hat{W}|\Psi\rangle - \langle\psi_{10}|\hat{H}_1|\psi_{10}\rangle - \langle\psi_{20}|\hat{H}_2|\psi_{20}\rangle. \tag{1.19}$$

Au premier ordre de la théorie des perturbations, $|\Psi\rangle$ correspond à $|\psi_{10}, \psi_{20}\rangle + |\Delta\rangle$, où

$$|\Delta\rangle = \sum_{n_1, n_2 > 0} \frac{\langle\psi_{1n_1}, \psi_{2n_2}|\hat{W}|\psi_{10}, \psi_{20}\rangle}{\varepsilon_{10} + \varepsilon_{20} - \varepsilon_{1n_1} - \varepsilon_{2n_2}} |\psi_{1n_1}, \psi_{2n_2}\rangle. \tag{1.20}$$

(ε_{in_i} est l'énergie de l'état propre $|\psi_{in_i}\rangle$.) Cette correction permet d'écrire l'énergie d'interaction jusqu'au deuxième ordre :

$$U = \langle\psi_{10}, \psi_{20}|\hat{W}|\psi_{10}, \psi_{20}\rangle + \langle\psi_{10}, \psi_{20}|\hat{W}|\Delta\rangle + \mathcal{O}(\hat{W}^3). \tag{1.21}$$

Dans le cas qui nous intéresse, $\hat{W}$ correspond à une interaction coulombienne, exprimée en terme des opérateurs de densité de charge $\hat{\rho}_1$ et $\hat{\rho}_2$:

$$\hat{W} = \int d^3r_1 d^3r_2 \, \hat{\rho}_1(\mathbf{r}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \hat{\rho}_2(\mathbf{r}_2). \tag{1.22}$$

Puisque les atomes sont neutres, le terme dominant est l'interaction dipôle-dipôle, et on peut approximer $\hat{W}$ en terme des opérateurs dipolaires $\hat{\boldsymbol{\mu}}_1$ et $\hat{\boldsymbol{\mu}}_2$:

$$\hat{W} \simeq \frac{1}{r^3} \left[\hat{\boldsymbol{\mu}}_1 \cdot \hat{\boldsymbol{\mu}}_2 - 3 \left(\hat{\boldsymbol{\mu}}_1 \cdot \hat{\mathbf{r}} \right) \left(\hat{\boldsymbol{\mu}}_2 \cdot \hat{\mathbf{r}} \right) \right]. \tag{1.23}$$

r est la distance interatomique et $\hat{\mathbf{r}}$ est le directeur de l'axe passant par les deux atomes. Le premier terme de l'équation (1.21) correspond à l'énergie d'interaction « classique » entre les dipôles moléculaires permanents—nulle dans ce cas, parce que les dipôles permanents sont nuls. Le deuxième terme de l'équation (1.21) correspond à l'énergie de dispersion dipôle–dipôle. Il varie en $1/r^6$ et dépend des polarisabilités des atomes, définies comme

$$\boldsymbol{\alpha}_i = \sum_{n_i > 0} \frac{\langle \psi_{i0} | \hat{\boldsymbol{\mu}}_i | \psi_{in_i} \rangle \langle \psi_{in_i} | \hat{\boldsymbol{\mu}}_i | \psi_{i0} \rangle + \text{c.c.}}{\varepsilon_{in_i} - \varepsilon_{i0}}. \tag{1.24}$$

Ce terme de dispersion en $1/r^6$ est le premier d'une série : dispersion dipôle–dipôle en $1/r^6$, dispersion dipôle–quadrupôle en $1/r^8$, dispersions dipôle–octopôle et quadrupôle–quadrupôle en $1/r^{10}$, et ainsi de suite.

Hiérarchie des interactions

On pourrait établir une hiérarchie approximative des affinités moléculaires. Mis à part les liens covalents, les interactions moléculaires les plus fortes sont les interactions ioniques (entre un groupe phosphate négatif et un ion magnésium, par exemple) et les interactions charge–dipole (entre un ion potassium et l'oxygène d'un groupe carbonyle, par exemple). Pour des groupes neutres, les interactions les plus fortes sont les liaisons hydrogène et les interactions dipôle–dipôle en général, suivies des interactions de London. Ces différents ordres de grandeur font que, dans

un système possédant des composantes de polarités diverses, chacune tend à se séparer des autres : les groupes chargés vont préférablement se lier entre eux, puis les groupes polaires entre eux, en enfin les groupes non-polaires entre eux. En réalité, des groupes de polarités différentes peuvent être maintenus ensemble par des liens covalents, de sorte que la tendance générale à la ségrégation est souvent résolue par des formes plus complexes—qui font d'ailleurs la variété et la complexité de la vie. Il en ressort aussi qu'autant les interactions fortes que les plus faibles ont un rôle à jouer dans la structure et le fonctionnement des systèmes biologiques. Aucune ne peut être véritablement négligée.

1.4.3 Fonction potentielle non-polarisable

La fonction potentielle non-polarisable PARAM27 [15], couramment utilisée pour la simulation de protéines et d'acide nucléiques, est définie comme suit :

$$\begin{aligned} U_{\text{PARAM27}}(\{\mathbf{r}\}) &= U_\ell(\{\mathbf{r}\}) + U_\theta(\{\mathbf{r}\}) + U_\phi(\{\mathbf{r}\}) \\ &\quad + U_\omega(\{\mathbf{r}\}) + U_{\text{UB}}(\{\mathbf{r}\}) \\ &\quad + \sum_I \sum_{J<I}{}' \frac{q_I q_J}{r_{IJ}} \\ &\quad + \sum_I \sum_{J<I}{}' 4\epsilon_{IJ} \left[\left(\frac{\sigma_{IJ}}{r_{IJ}} \right)^{12} + \left(\frac{\sigma_{IJ}}{r_{IJ}} \right)^{6} \right]. \end{aligned} \tag{1.25}$$

La notation « prime » (Σ') indique que la sommation sur les paires (I, J) exclue les paires 1–2 (atomes liés), les paires 1–3 (atomes liés à un atome commun) et les paires 1–4 (atomes liés à deux atomes eux-mêmes liés). Ces interactions sont prises en charge par les termes « liés ». Les charges partielles atomiques $\{q\}$ incluent les effets moyens d'induction. Les termes en $1/r^{12}$ et $1/r^6$ forment le potentiel de Lennard-Jones, qui décrit la répulsion dure des atomes à courte portée comme une divergence en $1/r^{12}$ et l'attraction de London, décroissant en $1/r^6$. Le terme en $1/r^{12}$ n'a pas de signification physique particulière, et a été choisi à l'origine pour simplifier le calcul analytique des coefficients du viriel des gaz monoatomiques. Les paramètres ϵ_{IJ} et σ_{IJ}, propres à chaque paire, sont formés à partir de la règle de

combinaison de Lorentz–Berthelot [16]

$$\epsilon_{IJ} = \sqrt{\epsilon_I \epsilon_J} \qquad \text{et} \qquad \sigma_{IJ} = \tfrac{1}{2}(\sigma_I + \sigma_J). \tag{1.26}$$

Cette règle de combinaison n'a pas non plus de signification particulière. Les valeurs des paramètres q_I, ϵ_I et σ_I sont ajustées pour chaque *type d'atome*, identifiant l'élément chimique de l'atome, son état d'hybridation et la nature de son environnement chimique immédiat. Le champ de force PARAM27 distingue par exemple plusieurs types de carbones : le carbone d'un carbonyle, le carbone d'un méthyle, le carbone d'un cycle aromatique, etc. De la même façon, les nombreux paramètres k des interactions « liées » sont ajustés pour chaque *type de lien*, correspondant aux types des atomes impliqués. En tout, plusieurs centaines de paramètres sont définis, décrivant les acides aminés, les bases des acides nucléiques, les principaux lipides et quelques petites molécules organiques importantes—dont l'eau, évidemment.

Les champs de force non-polarisables, comme le champ de force PARAM27 [15], considèrent chaque atome comme un objet sans structure interne, décrit uniquement par sa position dans l'espace et par la topologie de ses liaisons covalentes. D'un point de vue électrostatique, chaque atome est représenté par une charge partielle q et, dans un champ électrique $\mathbf{E}$, subit une force de Lorentz $q\mathbf{E}$. La charge partielle est fixe, ajustée à une valeur empirique qui représente l'effet moyen d'un environnement de solvatation standard.

1.4.4 Polarisation explicite

Un champ de force polarisable fournit un modèle d'induction moléculaire, c'est-à-dire qu'il permet de calculer le dipôle moléculaire induit par un champ électrique uniforme $\mathbf{E}$ d'intensité moyenne. Cette induction est définie par le tenseur de polarisabilité moléculaire $\boldsymbol{\alpha}$, qui se déduit de mesures de l'indice de réfraction ou s'obtient par des calculs quantiques *ab initio* :

$$\boldsymbol{\mu}_{\text{ind}} = \boldsymbol{\alpha} \cdot \mathbf{E}. \tag{1.27}$$

Un modèle d'induction devra produire des dipôles moléculaires induits comparables à cette prescription expérimentale. Il est possible de décrire l'induction par des transferts de charge atomique $\{\delta_{\text{ind}}\}$ d'une extrémité à l'autre de la molécule [comme il l'avait été conçu à l'équation (1.17)], auquel cas le dipôle induit sera

$$\boldsymbol{\mu}_{\text{ind}} = \sum_I \delta_{\text{ind},I} \mathbf{r}_I. \tag{1.28}$$

Il est aussi possible de décrire l'induction par des dipôles atomiques induits $\{\boldsymbol{\mu}_{\text{ind}}\}$, auquel cas on a simplement

$$\boldsymbol{\mu}_{\text{ind}} = \sum_I \boldsymbol{\mu}_{\text{ind},I}. \tag{1.29}$$

Prenons l'exemple simple de la molécule d'eau. Une molécule d'eau isolée possède un tenseur de polarisabilité légèrement anisotrope : $\alpha_{xx} = 1.525$ Å^3, $\alpha_{yy} = 1.440$ Å^3 et $\alpha_{zz} = 1.378$ Å^3, selon les calculs *ab initio* de Spelsberg et Meyer [17]. L'axe x correspond à l'axe hydrogène–hydrogène, l'axe y correspond à l'axe de symétrie de la molécule et l'axe z correspond à l'axe perpendiculaire au plan de la molécule. Un modèle polarisable simple consisterait à définir un site polarisable au centre de la molécule—le site de l'atome d'oxygène, par exemple—portant un dipôle induit suivant la formule

$$\boldsymbol{\mu}_{\text{ind}} = \alpha \mathbf{E}. \tag{1.30}$$

En choisissant α égal à la moyenne des trois composantes du tenseur $\boldsymbol{\alpha}$, soit 1.448 Å^3, le modèle reproduit la contribution isotrope du dipôle induit. On pourrait également calculer des transferts de charge en terme des composantes du champ électrique, de chaque atome d'hydrogène vers l'atome d'oxygène :

$$\delta_{\text{ind,H1}} = a_y E_y + a_x E_x \tag{1.31}$$

$$\delta_{\text{ind,H2}} = a_y E_y - a_x E_x \tag{1.32}$$

$$\delta_{\text{ind,O}} = -\delta_{\text{ind,H1}} - \delta_{\text{ind,H2}} \tag{1.33}$$

En ajustant correctement les paramètres a_x et a_y, on peut produire les valeurs désirées pour les composantes α_{xx} et α_{yy}. Par contre, la composante α_{zz} sera strictement nulle, puisqu'aucun transfert de charge n'est réalisé selon l'axe perpendiculaire au plan de la molécule.

Chaque atome possède une structure électronique interne qui dépend de son environnement électrostatique et qui, à son tour, affecte l'environnement des atomes avoisinants. Considérons par exemple l'interaction de trois atomes polarisables représentés par le modèle de l'équation (1.30). L'interaction entre deux atomes est affectée par la position du troisième, et l'ensemble des interactions doit être traité de manière cohérente. Les dipôles doivent vérifier les équations d'induction suivantes :

$$\boldsymbol{\mu}_{\mathrm{ind},I} = \alpha_I \left[-\sum_{J \neq I} q_J \boldsymbol{\nabla}_I \left(\frac{1}{r_{IJ}} \right) + \sum_{J \neq I} \boldsymbol{\mu}_{\mathrm{ind},J} \cdot \boldsymbol{\nabla}_I \boldsymbol{\nabla}_I \left(\frac{1}{r_{IJ}} \right) \right] . \tag{1.34}$$

La quantité entre crochets est le champ électrique au site I produit par les charges partielles et les dipôles induits des autres atomes. Pour l'exemple choisi, les indices I et J vont de 1 à 3, mais il est clair que cette formule est valable pour un nombre quelconque de sites polarisables.

Pour deux sites polarisables identiques dans un champ $\mathbf{E}$, chaque dipôle est induit par la somme du champ $\mathbf{E}$ et du champ produit par le dipôle induit voisin. On a pour chacun :

$$\boldsymbol{\mu}_{\mathrm{ind}} = \alpha \left[\mathbf{E} + \boldsymbol{\mu}_{\mathrm{ind}} \cdot \boldsymbol{\nabla}\boldsymbol{\nabla} \left(\frac{1}{r} \right) \right] . \tag{1.35}$$

Selon la direction du champ, $\hat{\mathbf{e}} = \mathbf{E}/|\mathbf{E}|$, on a

$$\begin{aligned} \mu_{\mathrm{ind}} &= \alpha \left[E + \mu_{\mathrm{ind}} \hat{\mathbf{e}} \cdot \boldsymbol{\nabla}\boldsymbol{\nabla} \left(\frac{1}{r} \right) \cdot \hat{\mathbf{e}} \right] \\ &\equiv \alpha \left(E + \mu_{\mathrm{ind}} \gamma \right) . \end{aligned} \tag{1.36}$$

Suivant cette définition, le dipôle induit s'écrit $\mu_{\mathrm{ind}} = \alpha E/(1 - \alpha\gamma)$. Sachant que $\boldsymbol{\nabla}\boldsymbol{\nabla}(1/r) = (3\hat{\mathbf{r}}\hat{\mathbf{r}} - \mathbf{1})/r^3$, on a $\gamma = 2/r^3$ pour $\mathbf{E}$ dans la direction de l'axe passant par les deux sites polarisables (c'est-à-dire, pour $\mathbf{e} \cdot \hat{\mathbf{r}} = 1$) et $\gamma = -1/r^3$ pour $\mathbf{E}$

dans la direction perpendiculaire ($\mathbf{e} \cdot \hat{\mathbf{r}} = 0$). Un champ d'une magnitude donnée crée donc des dipôles induits plus grands s'il est appliqué selon l'axe des sites que perpendiculairement à l'axe. Le tenseur de polarisabilité de la paire considérée comme une molécule possède les composantes suivantes :

$$\alpha_{\parallel} = \frac{2\alpha}{1 - 2\alpha/r^3} \tag{1.37}$$

$$\alpha_{\perp} = \frac{2\alpha}{1 + \alpha/r^3}. \tag{1.38}$$

Par un effet coopératif, deux polarisabilités atomiques isotropes forment une polarisabilité moléculaire anisotrope dont le rapport d'anisotropie peut être ajusté à l'aide de la distance r—ou d'un paramètre analogue représentant la force de l'interaction entre les dipôles voisins.

Qu'il s'agisse de dipôles induits $\{\boldsymbol{\mu}_{\text{ind}}\}$ ou de transferts de charge $\{\delta_{\text{ind}}\}$, un modèle polarisable introduit nécessairement des degrés de liberté supplémentaires servant à décrire la polarisation. Par analogie avec les coordonnées atomiques, ces variables sont soumises à des « forces » et doivent être ajustées selon leur « dynamique » propre. Dans l'esprit de l'approximation de Born–Oppenheimer, on peut considérer que les dipôles induits obéissent à l'équation (1.34) à tout instant de la dynamique des noyaux—c'est-à-dire, que la dynamique des électrons est couplée adiabatiquement à celle de noyaux. Le système d'equations (1.34) peut être traduit sous forme matricielle et solutionné en recourrant à une inversion de matrice. Il peut aussi être solutionné de façon itérative. Dans ce cas, les degrés de liberté de polarisation évoluent suivant une trajectoire classique ne produisant pas les fluctuations à l'origine des forces de London, et il est nécessaire d'incorporer les termes de dispersion dominants à l'aide d'un potentiel interatomique en $1/r^6$.

1.5 Objectifs de ce travail

Après avoir esquissé un vaste panorama allant de la biologie moléculaire à la physique atomique, il convient de dire que ce travail a des objectifs relativement

modestes. Nous étudierons seulement un petit sous-ensemble des nombreux types d'interactions moléculaires importantes en biologie. Et pour étudier ces quelques interactions, nous n'utiliserons qu'un petit nombre de molécules « prototypes ». C'est non sans un certain humour que nous avons présenté la figure 1.1 (à la page 2), car elle illustre en fait tout ce que nous n'étudierons pas et laisse invisible ce que nous allons directement étudier : l'eau, les ions inorganiques et les petites molécules organiques.

Notre objectif principal est de développer une forme fonctionnelle pour un modèle polarisable applicable à la simulation de processus biologiques impliquant des ions « libres » et des macromolécules. Elle doit être suffisamment précise pour représenter fidèlement les affinités relatives des différentes espèces chimiques, et suffisamment simple pour que le calcul des forces pour un système de plusieurs dizaines de milliers d'atomes puisse être répété *plusieurs millions de fois en quelques semaines* sur un ordinateur de la technologie actuelle. Avec un million de calculs des forces atomiques, la dynamique moléculaire « tout-atomique » peut étudier un processus pendant plusieurs nanosecondes, une échelle de temps caractéristique des mouvements internes des petites protéines [2, page 132]—certains sont cent fois plus rapides, d'autres cent fois plus lents. On s'intéresse d'abord aux ions dits « libres » parce qu'ils participent à des processus cinétiques rapides, étudiables à l'aide de la dynamique moléculaire. Par exemple, le temps de résidence caractéristique d'une molécule d'eau dans la couche de solvatation immédiate d'un ion potassium est de l'ordre de la nanoseconde [18].

Pour le modèle polarisable proposé, nous cherchons à supporter la double hypothèse suivante :

1. Le gain en précision apporté par un modèle polarisable permet une avancée de nature *qualitative* dans l'étude de plusieurs systèmes.
2. Le modèle de polarisation atomique peut être décrit de sorte que le coût numérique du calcul des forces atomiques est comparable à celui d'un modèle non-polarisable.

Des hypothèses semblables ont déjà été établies par d'autres chercheurs—nous en

discuterons. L'hypothèse 2 est démontrée concrètement au chapitre 2. Des versions réduites de l'hypothèse 1 sont démontrées au chapitre 4 pour l'eau, au chapitre 5 pour l'éthanol en solution et au chapitre 6 pour les ions en solution. Pour chaque système, nous cherchons à comprendre en quoi le modèle polarisable constitue une amélioration qualitative et, surtout, à quelles questions essentielles de chimie physique il permet de répondre.

Bibliographie

[1] Jan Koolman et Klaus-Heinrich Röhm. Color Atlas of Biochemistry. Thieme, Stuttgart, 1996.

[2] Kensal E. van Holde, W. Curtis Johnson et P. Shing Ho. Principles of Physical Biochemistry. Prentice Hall, Upper Saddle River, New Jersey, 1998.

[3] Francesco Sciortino, Alfons Geiger et H. Eugene Stanley. Effect of defects on molecular mobility in liquid water. *Nature*, 354(6350) : 218–221, 1991.

[4] Frank N. Keutsch et Richard J. Saykally. Water clusters : Untangling the mysteries of liquid, on molecule at a time. *Proc. Natl. Acad. Sci. U.S.A.*, 98(19) : 10533–10540, 2001.

[5] T. Sato, A. Chiba et R. Nozaki. Dynamical aspects of mixing schemes in ethanol-water mixtures in terms of the excess partial molar activation free energy, enthalpy, and entropy of the dielectric relaxation process. *J. Chem. Phys.*, 110(5) : 2508–2521, 1999.

[6] K. T. O'Neil et W. F. DeGrado. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*, 250(4981) : 646–651, 1990.

[7] J. J. R. Fraústo da Silva et R. J. P Williams. The Biological Chemistry of the Elements : The Inorganic Chemistry of Life. Clarendon, Oxford, 1991.

[8] Declan A. Doyle, João Morais Cabral, Richard A. Pfuetzner, Anling Kuo, Jacqueline M. Gulbis, Steven L. Cohen, Brian T. Chait et Roderick MacKinnon. The Structure of the Potassium Channel : Molecular Basis of K^+ Conduction and Selectivity. *Science*, 280(5360) : 69–77, 1998.

[9] Todor Dudev et Carmay Lim. Principles Governing Mg, Ca, and Zn Binding and Selectivity in Proteins. *Chem. Rev.*, 103(3) : 773–788, 2003.

[10] Attila Szabo et Neil S. Ostlund. Modern Quantum Chemistry : Introduction to Advanced Electronic Structure Theory. Dover Publications, Inc., first corrected édition, 1996.

[11] E. Deumens, A. Diz, R. Longo et Y. Öhrn. Time-dependent theoretical treatments of the dynamics of electrons and nuclei in molecular systems. *Rev. Mod. Phys.*, 66(3) : 917–983, 1994.

[12] R. Cabrera-Trujillo, J. R. Sabin, Y. Öhrn et E. Deumens. Direct diffential-cross-section calculations for ion-atom and atom-atom collisions in the keV range. *Phys. Rev. A*, 61(032719), 2000.

[13] A. V. Phelps, Chris H. Greene et J. P. Burke, Jr.. Collision cross sections from argon atoms with argon atoms for energies from 0.01 eV to 10 keV. *J. Phys. B : At. Mol. Opt. Phys.*, 33(16) : 2965–2981, 2000.

[14] M. Born et J. R. Oppenheimer. Zur Quantenteorie der Molekeln. *Ann. Phys. (Leipzig)*, 84 : 457–484, 1927.

[15] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., 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, M. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin et M. Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, 102(18) : 3586–3616, 1998.

[16] M. P. Allen et D. J. Tildesley. Computer Simulation of Liquids. Clarendon Press, Oxford, 1987.

[17] Dirk Spelsberg et Wilfried Meyer. Ab initio dynamic multipole polarizabilities and hyperpolarizabilities of H_2O and the long-range interaction coefficients for its dimer. *J. Chem. Phys.*, 108(4) : 1532–1543, 1998.

[18] L. Helm et A. E. Merbach. Water exchange on metal ions : experiments and simulations. *Coordination Chemistry Reviews*, 187 : 151–181, 1999.

Chapitre 2

Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm

Tiré de :
Guillaume Lamoureux et Benoît Roux. *J. Chem. Phys.*, 119(6) : 3025–3039, 2003.

Abstract

A simple treatment for incorporating induced polarization in computer simulations is formulated on the basis of the classical Drude oscillator model. In this model, electronic induction is represented by the displacement of a charge-carrying massless particle attached to a polarizable atom under the influence of the local electric field. The traditional self-consistent field (SCF) regime of induced polarization is reproduced if these auxiliary particles are allowed to relax instantaneously to their local energy minima for any given fixed configuration of the atoms in the system. In practice, such treatment is computationally prohibitive for generating molecular dynamics trajectories because the electric field must be recalculated several times iteratively to satisfy the SCF condition and it is important to seek a more efficient way to simulate the classical Drude oscillator model. It is demonstrated that a

close approximation to the SCF regime can be simulated efficiently by considering the dynamics of an extended Lagrangian in which a small mass is attributed to the auxiliary particles and the amplitude of their oscillations away from the local energy minimum is controlled with a low-temperature thermostat. A simulation algorithm in this modified two-temperature isobaric–isothermal ensemble is developed. The algorithm is tested and illustrated using a rigid three-site water model with one additional Drude particle attached to the oxygen which is closely related to the polarizable SPC model of Ahlström *et al.* (*Mol. Phys.*, 68: 563, 1989). The tests with the extended Lagrangian show that stable and accurate molecular dynamics trajectories for large integration time steps (1 or 2 fs) can be generated and that liquid properties equivalent to SCF molecular dynamics can be reproduced at a fraction of the computational cost.

2.1 Introduction

Most computer simulations of biomolecular systems are generated using simple potential functions which do not account for many-body induced polarization effects explicitly. Induced polarization, which arises from a perturbation of the electronic structure of the molecular species in response to external electric fields, is typically incorporated in these simulations in an average effective way. For example, widely used potential functions such as AMBER/parm94 [1], CHARMM27 [2], GROMOS [3], and OPLS [4] force fields are based on pairwise additive electrostatic interactions with fixed effective atomic partial charges. Typically, the magnitude of the atomic partial charges has been adjusted to account for the average induction arising in a typical environment. Despite the apparent simplification, such potential functions have been remarkably successful in modeling complex molecular systems over the last two decades (e.g., see Ref. [5] and references therein). For example, the properties of liquid water can be represented quite realistically without introducing induced polarizability explicitly if the dipole moment of the water molecule is enhanced compared to its gas phase value of 1.85 D, e.g., 2.35 D

for TIP3P [6], 2.27 D for SPC [7], 2.18 D for TIP4P [6], and 2.29 D for TIP5P [8]. Furthermore, such effective nonpolarizable potential functions are also able to reproduce the solvation free energy of small cations quite well [9, 10] although there are some difficulties with anions [11, 12, 13, 14]. Nonetheless, these potential functions are clearly limited in their ability to model complex biological systems such as membrane ion channels [15], and despite their usefulness there are growing concerns about their shortcomings in general [16].

In recent years, there have been increasing efforts dedicated to develop models that account for induced polarization more accurately. Most of the current models for incorporating polarizability into a potential function belong to one of three classes [16, 17]: point dipole induction models, charge transfer models (or fluctuating charge models), and classical Drude oscillator models (or shell models). The main drawback of any treatment of induced polarization comes from the computationally expensive convergence of the self-consistent field (SCF) condition, which must be satisfied at each molecular dynamics time step to determine the magnitude of the induction and the atomic forces. The SCF condition follows directly from the Born–Oppenheimer approximation, which assumes that the (quantum) electronic degrees of freedom must relax instantaneously to their ground state for every configuration of the nuclei. It is possible, however, to avoid the iterative SCF procedure altogether and to reduce the computational cost of induced polarization by treating the induction "variables" as part of the (classical) dynamical variables in an extended Lagrangian representation [18, 19]. Although such computational schemes are efficient and very attractive, the significance of the results obtained from simulations in which the induction "variables" are allowed to undergo unphysical classical thermal fluctuations is not completely clear.

The traditional point dipole induction model for liquids has a long history which goes back to Kirkwood and Onsager [20, 21] and it is, arguably, one of the most natural way to incorporate polarizability in molecular mechanical force fields (e.g., see Refs. [22, 23, 11, 18, 24, 25]). In this representation, atoms are assigned fixed atomic charges and inducible dipoles, which are then self-consistently adjusted for

any given configuration of the atoms in the systems. Such a simple model can correctly capture the polarizability in a variety of systems, e.g., organic molecules [26, 27, 28], amino acids [29], and ions in water [25]. The main drawback of this approach are the inconveniences arising from the (cumbersome) vectorial character of the point dipole and the significant modifications of the molecular mechanical force field that it requires in standard biomolecular simulation programs. Notably, the constant-pressure algorithm and the fast particle-mesh-Ewald (PME) summation procedure [30, 31] have to be extended and generalized to include charge–dipole and dipole–dipole interactions. Despite these technical difficulties, an initial version of an atomic point dipole model has been recently implemented for the AMBER/parm02 polarizable force field [32].

Charge transfer models treat the atomic partial charges as quantities allowed to vary according to a self-consistent atomic electronegativity equalization scheme [33, 34, 35]. Such models are attractive compared to point dipole induction models because they require no particular overhead in the computation of the atomic forces, since simple charge–charge Coulomb laws are maintained. Charge transfer models can, however, be seriously limited in their ability to represent the different components of the polarizability tensor of some molecular species accurately. For example, a linear molecule can only be polarized along its axis, and a planar molecule has a zero transverse polarizability. Charge transfer models were shown to be successful for liquid water [19], but unable to represent a bifurcated hydrogen bond or the transverse polarization of an aromatic ring without the addition of inducible dipoles [28]. As pointed out by Chelli *et al.* [36], atom-based charge transfer models may also considerably overestimate the polarizability of biopolymers by allowing charge to flow between remote atoms at too small an energy cost. For large molecules, it is necessary to impose further constraints to limit charge transfer to neighboring atoms [37, 28, 36].

The Drude oscillator model, which was originally proposed by Paul Drude in 1900 as a simple way to describe the dispersive properties of materials [38], represents electronic induction by introducing a massless charged particle, attached

to each polarizable atom by a harmonic spring. The position of these "auxiliary" particles is then adjusted self-consistently to their local energy minima for any given configuration of the atoms in the system. A quantum version of the model has been used in early applications to describe the dipole–dipole dispersion interactions [39, 40, 41]. A semiclassical version of the model was used more recently to describe molecular interactions [42], and electron binding [43]. The classical version of the model has been quite useful in statistical mechanical studies of dense systems, e.g., ionic crystals [44, 45, 46, 47, 48, 49], simple liquids of polarizable particles [50, 51, 52, 53], liquid water [54, 55, 56, 57, 58], and the hydration of small ions [12, 59]. One particularly attractive aspect of the Drude oscillator model is that it preserves the simple charge–charge Coulomb electrostatic interaction. Its implementation in standard biomolecular simulation programs is, therefore, relatively straightforward (including constant-pressure algorithm and PME electrostatics). Despite this technical advantage, Drude oscillators have not been as widely used as the point dipole or charge transfer models, probably because of the difficulties designing efficient computational schemes to solve the fast motion of the auxiliary particles in molecular dynamics simulations [58]. One may note that, in contrast with the point dipole [18] and charge transfer models [60, 19], extended dynamics algorithms have not been implemented to simulate Drude oscillator models and that applications to liquids have usually relied on a direct SCF iterative scheme.

In this paper, we formulate a simple polarizable model based on Drude oscillators and explore its theoretical features. In particular, the consequences of allowing the Drude particles to undergo classical thermal fluctuations are clarified. Furthermore, we present an efficient numerical integration procedure based on an extended Lagrangian technique for generating accurate molecular dynamics trajectories. The accuracy and efficiency of the simulation algorithm is illustrated by using a Drude oscillator version of the polarizable SPC (PSPC) model of Ahlström *et al.* [61]. It is shown that the trajectories generated by the extended dynamics algorithm are nearly equivalent to those generated using a rigorous SCF procedure.

The paper is concluded with a summary of the main points and an overview of future work.

2.2 Molecular dynamics of the polarizable atoms

2.2.1 Drude oscillator model

To model the polarizability α of a given atom with partial charge q, a mobile Drude particle carrying a charge q_{D} is introduced. The charge of the atom is replaced by $q - q_{\mathrm{D}}$ to preserve the net charge of the atom–Drude pair. The Drude particle is harmonically bound to the atomic particle with a force constant k_{D}, which is assumed to be the same for all atoms without any loss of generality. In the absence of a field, the Drude particle oscillates around the position of the atom, $\mathbf{r}$, and the atom appears on average as a point charge q. In a uniform field $\mathbf{E}$, the Drude particle oscillates around a displaced position $\mathbf{r} + \mathbf{d}$ with $\mathbf{d} = q_{\mathrm{D}}\mathbf{E}/k_{\mathrm{D}}$, and the average induced atomic dipole is $\boldsymbol{\mu} = q_{\mathrm{D}}^2\mathbf{E}/k_{\mathrm{D}}$. It follows that the isotropic atomic polarizability has a simple expression

$$\alpha = \frac{q_{\mathrm{D}}^2}{k_{\mathrm{D}}}. \tag{2.1}$$

For a given α, the force constant k_{D} can be chosen such that the displacement $\mathbf{d}$ of the Drude particle remains much smaller than any interatomic distance, and that the resulting induced dipole $\boldsymbol{\mu}$ is almost equivalent to a point dipole. Within this point-dipole approximation, the only relevant parameter is the combination $q_{\mathrm{D}}^2/k_{\mathrm{D}}$.

More generally, adding auxiliary Drude particles to a system of atoms yields the polarizable force field

$$U(\{\mathbf{r}\},\{\mathbf{d}\}) = U_{\mathrm{self}}(\{\mathbf{d}\}) + U_{\mathrm{bond}}(\{\mathbf{r}\}) + U_{\mathrm{elec}}(\{\mathbf{r}\},\{\mathbf{d}\}) + U_{\mathrm{LJ}}(\{\mathbf{r}\}), \tag{2.2}$$

where $U_{\text{self}}(\{\mathbf{d}\})$ represents the atom–Drude harmonic bonds (the $\frac{1}{2}kd^2$ terms), $U_{\text{bond}}(\{\mathbf{r}\})$ is the intramolecular energy contribution from the bond lengths, angles, and dihedrals, $U_{\text{elec}}(\{\mathbf{r}\},\{\mathbf{d}\})$ represents all Coulombic interactions (atom–atom, atom–Drude, and Drude–Drude), and $U_{\text{LJ}}(\{\mathbf{r}\})$ is a Lennard-Jones "12–6" nonpolar contribution.

2.2.2 Self-consistent field regime

To implement the self-consistent field (SCF) condition corresponding to the Born–Oppenheimer approximation, the forces acting on the nuclei must be computed after the Drude particles have relaxed to minimize the total potential energy. For a given nuclear configuration of N polarizable atoms, $\{\mathbf{r}\}$, the relaxed Drude particle positions $\{\mathbf{r}+\mathbf{d}^{\text{SCF}}\}$ are found by solving

$$\frac{\partial U}{\partial \mathbf{d}_i} = \mathbf{0}, \tag{2.3}$$

where U is the force field of Eq. (2.2) and where index i runs from 1 to N. U_{bond} and U_{LJ} are independent of the $\mathbf{d}$'s, and

$$\frac{\partial U_{\text{self}}}{\partial \mathbf{d}_i} + \frac{\partial U_{\text{elec}}}{\partial \mathbf{d}_i} = \mathbf{0}. \tag{2.4}$$

These equations define the force equilibria on the Drude particles

$$k_{\text{D}}\mathbf{d}_i - q_{\text{D},i}\mathbf{E}_i = \mathbf{0}, \tag{2.5}$$

where $\mathbf{E}_i$ is the total electric field in $\mathbf{r}_i+\mathbf{d}_i$, arising from the fixed charges as well as all the induced dipoles (modeled with Drude oscillators). In the point-dipole approximation, these conditions can be rewritten as the familiar self-consistent equations for induced polarization [62]

$$\boldsymbol{\mu}_i^{\text{SCF}} = \alpha_i \left[-\sum_{j\neq i} q_j \boldsymbol{\nabla}_i \left(\frac{1}{r_{ij}}\right) + \sum_{j\neq i} \boldsymbol{\mu}_j^{\text{SCF}} \cdot \boldsymbol{\nabla}_i \boldsymbol{\nabla}_i \left(\frac{1}{r_{ij}}\right) \right]. \tag{2.6}$$

(See Appendix 2.A for details.) For atomic positions $\{\mathbf{r}\}$, the relaxed displacements produce the potential

$$U^{\mathrm{SCF}}(\{\mathbf{r}\}) \equiv U(\{\mathbf{r}\}, \{\mathbf{d}^{\mathrm{SCF}}\}), \tag{2.7}$$

and the atomic motions in the SCF regime are described by

$$m_i \ddot{\mathbf{r}}_i = -\frac{\partial U^{\mathrm{SCF}}}{\partial \mathbf{r}_i}. \tag{2.8}$$

Integrating Eq. (2.8) with molecular dynamics requires that the positions of the Drude particles be set at their energy minimum at every integration time step, by solving Eq. (2.5). This simple simulation method has been widely used in molecular dynamics [63, 47, 49, 57] as well as in Monte Carlo simulations [54, 56, 64]. Nonetheless, the SCF procedure is limited and computationally expensive, because any nonconverged energy minimization introduces systematic drag forces on the physical atoms that considerably affect energy conservation and the stability of the temperature [47, 49, 58].

2.2.3 Thermalized Drude oscillators

A simple alternative to SCF is to extend the Lagrangian of the system to consider dipoles as additional dynamical degrees of freedom. This extended dynamics approach was proposed by Sprik and Klein [60] for inducible dipoles described as fluctuating charges placed on a small tetrahedron following rigidly each water molecule. A similar extended system was proposed by van Belle *et al.* for inducible point dipoles [65, 66]. In the present system, the additional degrees of freedom are the positions of the moving Drude particles. All Drude particles are attributed a small mass m_{D}, taken from the atomic masses $\{m_i\}$, and the motion of atoms and Drude particles (at positions $\{\mathbf{r}_i\}$ and $\{\mathbf{r}_{\mathrm{D},i} \equiv \mathbf{r}_i + \mathbf{d}_i\}$) is simulated on an equal dynamical footing

$$(m_i - m_{\mathrm{D}})\ddot{\mathbf{r}}_i = -\frac{\partial U}{\partial \mathbf{r}_i}, \tag{2.9}$$

$$m_{\mathrm{D}}\ddot{\mathbf{r}}_{\mathrm{D},i} = -\frac{\partial U}{\partial \mathbf{r}_{\mathrm{D},i}}. \tag{2.10}$$

The motion of Drude particles is expected to be decoupled from the atomic motion if m_{D} is sufficiently small. The obvious drawback is that a small m_{D} requires a small integration time step [67, 48, 55]. For a single Drude oscillator, a significant speedup recovery can be attained by using a multi-time-step integration approach [12, 59], but this advantage is lost for a dense system of polarizable atoms, because the long-range $1/r^3$ dipole–dipole interactions includes high-frequency oscillations and have to be integrated using the shortest time step. Provided a relaxed initial configuration obeying Eq. (2.5), the Drude oscillators will follow a quasiadiabatic dynamics. However, even if m_{D} is very small, the Drude particles will eventually reach a thermodynamic equilibrium with the rest of the system. Therefore, simulation approaches relying solely on the kinetic decoupling of the Drude oscillators to maintain a Born–Oppenheimer regime are inappropriate for long simulation runs. Taking advantage of the long thermalization time, it is possible to remain close to the SCF energy surface by periodically resetting the positions of the Drude oscillators to their energy minimum [12]. But, doing so make the simulation irreversible.

From this point of view, it is of interest to examine the consequences of full thermalization of the classical Drude oscillators on the properties of the system. This is particularly important given the fact that any classical fluctuations of the Drude oscillators are *a priori* unphysical according to the Born–Oppenheimer approximation upon which electronic induction models are based. One way to address this question is to determine the statistical probability of the nuclear configuration $\{\mathbf{r}\}$ under the influence of thermally fluctuating Drude oscillators. Because such system is concerned only with classical degrees of freedom (nuclei and Drude oscillators), this task is trivial. The probability of any nuclear configuration is proportional to the Boltzmann factor of an effective potential of mean force, $e^{-\beta U^{\mathrm{eff}}}$, which is

rigorously obtained by integrating out the displacements of the Drude oscillators

$$e^{-\beta U^{\text{eff}}(\{\mathbf{r}\})} = \frac{\int d\{\mathbf{d}\}\, e^{-\beta U(\{\mathbf{r}\},\{\mathbf{d}\})}}{\int d\{\mathbf{d}\}\, e^{-\beta U_{\text{self}}(\{\mathbf{d}\})}}. \tag{2.11}$$

Decomposing the displacements $\{\mathbf{d}\}$ in terms of the SCF displacements $\{\mathbf{d}^{\text{SCF}}\}$ and the thermal fluctuations $\{\delta\mathbf{d}\}$, the potential energy U can be expanded in powers of $\{\delta\mathbf{d}\}$

$$\begin{aligned} U(\{\mathbf{r}\},\{\mathbf{d}\}) &= U^{\text{SCF}}(\{\mathbf{r}\}) + \sum_i \delta\mathbf{d}_i \cdot \left.\frac{\partial U}{\partial \mathbf{d}_i}\right|_{\text{SCF}} \\ &\quad + \tfrac{1}{2}\sum_{ij} \delta\mathbf{d}_i \cdot \left.\frac{\partial^2 U}{\partial \mathbf{d}_i \partial \mathbf{d}_j}\right|_{\text{SCF}} \cdot \delta\mathbf{d}_j + \cdots. \end{aligned} \tag{2.12}$$

In the point-dipole approximation, only the first- and second-order terms are nonzero. The SCF contribution to the interaction can be factored out, and the first-order terms average to zero

$$e^{-\beta U^{\text{eff}}(\{\mathbf{r}\})} = e^{-\beta U^{\text{SCF}}(\{\mathbf{r}\})} \times \frac{\int d\{\delta\mathbf{d}\}\, e^{-\frac{1}{2}\beta \sum_{ij} \delta\mathbf{d}_i \cdot \frac{\partial^2 U}{\partial \mathbf{d}_i \partial \mathbf{d}_j} \cdot \delta\mathbf{d}_j}}{\int d\{\delta\mathbf{d}\}\, e^{-\beta U_{\text{self}}(\{\delta\mathbf{d}\})}}. \tag{2.13}$$

The remaining integral can be evaluated as a moment expansion. Truncating the expansion to two-body interactions yields

$$U^{\text{eff}}(\{\mathbf{r}\}) = U^{\text{SCF}}(\{\mathbf{r}\}) - \tfrac{3}{2} k_{\text{B}} T \sum_{ij} \frac{\alpha_i \alpha_j}{r_{ij}^6} + \cdots. \tag{2.14}$$

(See Appendix 2.B for the details of the calculation.) In addition to the static induction effects included in U^{SCF}, the thermalized ("hot") Drude oscillators give rise to a $1/r^6$, temperature-dependent, attractive term. This $\frac{3}{2} k_{\text{B}} T \alpha^2 / r^6$ term is the classical thermodynamic equivalent of the London quantum dispersive attraction $E_{\text{I}} \alpha^2 / r^6$ [39]. It corresponds to a small perturbation to the London forces, because $k_{\text{B}} T$ is at least two orders of magnitude smaller than the typical ionization energy

E_I. The smaller the temperature of the dipole motion is, the closer the effective potential is to the SCF potential. Equation (2.14) is independent of m_D, the mass of the oscillators. This is an indication that the argument it makes concerning a system of thermalized Drude oscillators applies to the thermodynamic properties but not to the dynamic properties.

2.2.4 Low-temperature Drude oscillators

To approximately reproduce the dynamics equivalent to the SCF regime of Eq. (2.8), we use two Nosé–Hoover thermostats [68]: a first one to keep the atoms at room temperature T and a second one to reduce the thermal fluctuations of the Drude particles by imposing a temperature $T_\star \ll T$. The idea of cooling the polarization degrees of freedom with a separate thermostat was carefully studied by Sprik [69], who showed that, for cold dipoles, both the equilibrium and diffusion properties are independent of the value of the dipole inertia parameter (the analog of m_D), as long as it is sufficiently small. For Drude oscillators, the temperature $T_\star$ should be small enough to leave almost no kinetic energy to the atom–Drude vibrations, yet large enough to allow the Drude particles to readjust to the room-temperature motion of the atoms.

The second thermostat is coupled to the motion of the Drude particles relative to their nuclei, $\{\dot{\mathbf{d}}\}$ (not to their absolute motion $\{\dot{\mathbf{r}}_\mathrm{D}\}$). Denoting $\mathbf{R}_i$ the center of mass of each $(\mathbf{r}_i, \mathbf{r}_{\mathrm{D},i})$ pair, m_i the total mass of the pair (as before), and $m'_i = m_\mathrm{D}(1 - m_\mathrm{D}/m_i)$ the reduced mass, the equations of motion are

$$m_i\ddot{\mathbf{R}}_i = \mathbf{F}_{\mathbf{R},i} - m_i\dot{\mathbf{R}}_i\dot{\eta}, \tag{2.15}$$

$$m'_i\ddot{\mathbf{d}}_i = \mathbf{F}_{\mathbf{d},i} - m'_i\dot{\mathbf{d}}_i\dot{\eta}_\star, \tag{2.16}$$

$$Q\ddot{\eta} = \sum_j m_j\dot{R}_j^2 - N_\mathrm{f}k_\mathrm{B}T, \tag{2.17}$$

$$Q_\star\ddot{\eta}_\star = \sum_j m'_j\dot{d}_j^2 - N_{\mathrm{f}\star}k_\mathrm{B}T_\star. \tag{2.18}$$

Indices i and j run from 1 to N, the total number of atoms. Because not all atoms

have to be polarizable, the total number of Drude particles, N_{D}, may be less than N. If a given atom i bears no Drude oscillator, $\mathbf{R}_i$ corresponds to $\mathbf{r}_i$, m'_i is zero, and the corresponding Eq. (2.16) is ignored. N_{f} is the number of degrees of freedom associated with the atomic motion, accounting for distance constraints imposed by SHAKE [70], and $N_{\mathrm{f}\star} \equiv 3N_{\mathrm{D}}$ is the number of degrees of freedom associated with the motion of the Drude oscillators. Q and $Q_\star$ are the inertia factors of the Nosé–Hoover thermostats. They are related to characteristic time scales of the motion, τ and $\tau_\star$, via $Q \equiv N_{\mathrm{f}} k_{\mathrm{B}} T \tau^2$ and $Q_\star \equiv N_{\mathrm{f}\star} k_{\mathrm{B}} T_\star \tau_\star^2$. The "velocities" $\dot{\eta}$ and $\dot{\eta}_\star$ are acting as friction coefficients, that is, as scaling exponents on the velocities $\{\dot{\mathbf{R}}\}$ and $\{\dot{\mathbf{d}}\}$. They are driven by "forces" G and $G_\star$—the right-hand sides of Eqs. (2.17) and (2.18), respectively—opposing any deviation from the target temperatures T and $T_\star$. $\mathbf{F}_{\mathbf{R},i} = -\partial U/\partial \mathbf{R}_i$ and $\mathbf{F}_{\mathbf{d},i} = -\partial U/\partial \mathbf{d}_i$ are the forces on the centers of mass and on the displacements. Using the chain rule, they can be written in terms of the actual forces on the particles

$$\mathbf{F}_{\mathbf{R},i} = -\frac{\partial U}{\partial \mathbf{r}_i} - \frac{\partial U}{\partial \mathbf{r}_{\mathrm{D},i}}, \tag{2.19}$$

$$\mathbf{F}_{\mathbf{d},i} = -\left(1 - \frac{m_{\mathrm{D}}}{m_i}\right) \frac{\partial U}{\partial \mathbf{r}_{\mathrm{D},i}} + \left(\frac{m_{\mathrm{D}}}{m_i}\right) \frac{\partial U}{\partial \mathbf{r}_i}. \tag{2.20}$$

The equations of motion (2.15) and (2.16) can be formulated in terms of the real coordinates of the particles, using the inverse transformation

$$\mathbf{r}_i = \mathbf{R}_i - \left(\frac{m_{\mathrm{D}}}{m_i}\right) \mathbf{d}_i, \tag{2.21}$$

$$\mathbf{r}_{\mathrm{D},i} = \mathbf{R}_i + \left(1 - \frac{m_{\mathrm{D}}}{m_i}\right) \mathbf{d}_i. \tag{2.22}$$

The system (2.15) to (2.18) conserves the extended energy

$$E_{\mathrm{NH}} = E + \tfrac{1}{2} Q \dot{\eta}^2 + \tfrac{1}{2} Q_\star \dot{\eta}_\star^2 + N_{\mathrm{f}} k_{\mathrm{B}} T \eta + N_{\mathrm{f}\star} k_{\mathrm{B}} T_\star \eta_\star, \tag{2.23}$$

where E is the sum of kinetic and potential energies of the particles. The variables η and $\eta_\star$ correspond to specific energies in each of the heat baths, from an arbitrary

reference level. Steady heat flows are expected from the Drude oscillators to the cold bath $T_\star$ and from the room temperature bath T to the atoms, that is, $\langle \dot{\eta}_\star \rangle > 0$ and $\langle \dot{\eta} \rangle < 0$. This heat transfer is small if the Drude particles and the atoms have weakly coupled kinetics (e.g., if the power spectra of their velocity autocorrelation functions do not significantly overlap).

Designing an efficient scheme to integrate Eqs. (2.15) to (2.18) requires some care. The procedure should conserve energy and simulate the appropriate thermodynamic ensemble as accurately as possible, despite the fact that heat is steadily flowing from the subsystem at room temperature T to the subsystem at $T_\star$. Following Martyna *et al.* [71], the complete evolution operator over a time step Δt, formally expressed in terms of the total Liouville operator as $e^{i\mathcal{L}\Delta t}$, is decomposed in terms of a "velocity-Verlet" propagation, denoted $e^{i\mathcal{L}_{\mathrm{VV}}\Delta t}$, and a multi-step "Nosé–Hoover" propagation, each step denoted $e^{i\mathcal{L}_{\mathrm{NH}}\Delta t/2n_\mathrm{c}}$

$$e^{i\mathcal{L}\Delta t} \simeq \left(e^{i\mathcal{L}_{\mathrm{NH}}\Delta t/2n_\mathrm{c}}\right)^{n_\mathrm{c}} e^{i\mathcal{L}_{\mathrm{VV}}\Delta t} \left(e^{i\mathcal{L}_{\mathrm{NH}}\Delta t/2n_\mathrm{c}}\right)^{n_\mathrm{c}}. \tag{2.24}$$

This integration procedure is explained in detail in Ref. [71]. The modified version developed for the current model is presented in Appendix 2.C.1.

Because the atomic forces $\{\mathbf{F}\}$ are involved only in the "velocity-Verlet" propagation, each step of the "Nosé–Hoover" propagation is an $O(N)$ calculation and thus n_c can be large without significantly affecting the efficiency. Test simulations show that, for any useful time step Δt and Drude mass m_D (e.g., in the range $\Delta t = 1.0$ fs and $m_\mathrm{D} = 0.5$ amu), the "one-step" ($n_\mathrm{c} = 1$) integrator is not sufficiently accurate. Due to systematic integration errors on the variables $\dot{\eta}$ and $\dot{\eta}_\star$, the temperatures are poorly controlled. Moreover, these errors are accumulating in η and $\eta_\star$, and the extended energy of Eq. (2.23) is drifting. With $n_\mathrm{c} > 1$, the systematic errors on $\dot{\eta}$ and $\dot{\eta}_\star$ are reduced, and the accuracy of the algorithm is significantly improved at a small computational cost.

2.2.5 Constant-pressure dynamics

The modified Andersen–Hoover barostat of Martyna *et al.* [72] was implemented to allow fluctuations in the volume of the system. The system of Eqs. (2.15) to (2.18) is further extended to include the barostat variable ε, regulating the internal pressure of the system through isotropic scaling of the simulation cell. Because of this scaling, the evolution of the positions $\{\mathbf{R}, \mathbf{d}\}$ is not directly related to the conjugate velocities $\{\mathbf{v_R}, \mathbf{v_d}\}$. The equations of motion are

$$\dot{\mathbf{R}}_i = \mathbf{v}_{\mathbf{R},i} + \mathbf{R}_i\dot{\varepsilon}, \tag{2.25}$$

$$m_i\dot{\mathbf{v}}_{\mathbf{R},i} = \mathbf{F}_{\mathbf{R},i} - m_i\mathbf{v}_{\mathbf{R},i}(\dot{\eta} + a\dot{\varepsilon}), \tag{2.26}$$

$$\dot{\mathbf{d}}_i = \mathbf{v}_{\mathbf{d},i} + \mathbf{d}_i\dot{\varepsilon}, \tag{2.27}$$

$$m'_i\dot{\mathbf{v}}_{\mathbf{d},i} = \mathbf{F}_{\mathbf{d},i} - m'_i\mathbf{v}_{\mathbf{d},i}(\dot{\eta}_\star + a\dot{\varepsilon}), \tag{2.28}$$

$$Q\ddot{\eta} = \sum_j m_j v_{\mathbf{R},j}^2 - N_\mathrm{f} k_\mathrm{B} T, \tag{2.29}$$

$$Q_\star\ddot{\eta}_\star = \sum_j m'_j v_{\mathbf{d},j}^2 - N_{\mathrm{f}\star} k_\mathrm{B} T_\star, \tag{2.30}$$

$$W\ddot{\varepsilon} = a\sum_j \left(m_j v_{\mathbf{R},j}^2 + m'_j v_{\mathbf{d},j}^2\right) + \sum_j (\mathbf{R}_j \cdot \mathbf{F}_{\mathbf{R},j} + \mathbf{d}_j \cdot \mathbf{F}_{\mathbf{d},j}) - 3PV. \tag{2.31}$$

Implicit in Eq.(2.25) is the relationship between ε and V, the volume of the simulation cell

$$\dot{V} = 3V\dot{\varepsilon}. \tag{2.32}$$

For simplicity, the N_D variables $\{\mathbf{d}\}$ are scaled along with the N atomic positions, even though they represent dipoles having almost no thermal fluctuations. Not doing so would require the virial contribution to the pressure, $\mathbf{R} \cdot \mathbf{F_R} + \mathbf{d} \cdot \mathbf{F_d}$, to be replaced by $\mathbf{R} \cdot (\mathbf{F_R} + \mathbf{F_d})$. The right-hand side of Eq. (2.31) is the driving force of the barostat, correcting for any deviation from the target pressure P. W is the inertia factor of the barostat, related to a time scale τ_ε via $W \equiv N_{\mathrm{f}\varepsilon} k_\mathrm{B} T \tau_\varepsilon^2$. $N_{\mathrm{f}\varepsilon}$ is the number of degrees of freedom effectively contributing to the pressure, that

is, $N_\mathrm{f} + N_{\mathrm{f}\star}$ if $T_\star$ equals T, but only N_f if $T_\star$ is much lower than T. Because the center of mass of the whole system is unaffected by the position scaling, three of the $N_{\mathrm{f}\varepsilon}$ degrees of freedom that are explicitly scaled are independent of the volume, and the kinetic contribution to the internal pressure is corrected by a coefficient $a = 1 + 3/N_{\mathrm{f}\varepsilon}$.

The equations of motion (2.25) to (2.31) conserve the extended energy

$$E_\mathrm{AH} = E_\mathrm{NH} + \tfrac{1}{2} W \dot{\varepsilon}^2 + PV, \tag{2.33}$$

where E_NH is defined in Eq. (2.23). They can be integrated using an extension of the method described in the previous section [73]. The details are given in Appendix 2.C.2.

2.2.6 Avoiding the "flying ice cube" problem

Even though the trajectory generated by integrating Eqs. (2.25) to (2.31) rigorously conserves the total momentum, a spurious translation of the center of mass of the system may arise after some time due to an accumulation of small numerical error, giving rise to the "flying ice cube" artifact [74]. All water molecules are gradually losing most of their relative thermal motion to develop a large coherent translation. This peculiar phenomenon is consistent with Eqs. (2.25) to (2.31), because thermostats are insensitive to how kinetic energy is partitioned among the degrees of freedom. When it is not explicitly enforced by "massive thermostatting" [75, 76, 77], energy equipartition in molecular dynamics comes from the physical coupling between the different modes of motion. Typically, an excess of kinetic energy in one mode is rapidly distributed among the other modes as a result of the tight coupling. But, an excess of kinetic energy in an uncoupled mode, such as a global translation in periodic boundary conditions, is not redistributed. This could also occur for a global rotation, if the boundary conditions have circular symmetry. In principle, a system with zero total momentum should remain so, since all internal forces have an equal and opposite reaction and no external forces are applied.

However, a coherent translation may appear if action and reaction forces show small systematic numerical errors [78], or if temperature is kept constant by velocity rescaling [74, 78]. Numerical imprecision in calculating the long range electrostatic interactions using PME may also contribute to damaging the conservation of the total momentum.

In the present case, with two competing thermostats, the origin of the "flying ice cube" artifact is clear. From Eqs. (2.26) and (2.28), the center-of-mass motion is

$$M_{\text{tot}}\dot{\mathbf{V}}_{\text{cm}} = -M_{\text{tot}}\mathbf{V}_{\text{cm}}(\dot{\eta} + a\dot{\varepsilon}), \tag{2.34}$$

where M_{tot} is the total mass of the system. Because $\langle\dot{\eta}\rangle < 0$ and $\langle\dot{\varepsilon}\rangle = 0$, any nonzero velocity will be systematically amplified, yielding a significant numerical instability. Considering a random force created by numerical errors, $\mathbf{F}_{\text{err}}$, and writing $\dot{\eta} = \langle\dot{\eta}\rangle + \delta\dot{\eta}$, the actual equation of motion of the center of mass is

$$M_{\text{tot}}\dot{\mathbf{V}}_{\text{cm}} \approx \mathbf{F}_{\text{err}} - M_{\text{tot}}\mathbf{V}_{\text{cm}}(\langle\dot{\eta}\rangle + \delta\dot{\eta} + a\dot{\varepsilon}). \tag{2.35}$$

Assuming that $\mathbf{F}_{\text{err}}$ and $\delta\dot{\eta} + a\dot{\varepsilon}$ are uncorrelated random variables, $\mathbf{V}_{\text{cm}}$ has an unstable behavior around $\mathbf{V}_{\text{cm}} = \mathbf{0}$ as long as $\langle\dot{\eta}\rangle$ is negative, that is, as long as the "flying ice cube" steady regime is not attained. Even if the center of mass is initially immobile, the numerical errors on the forces will create velocity fluctuations that will be slowly amplified by an exponential factor $e^{|\langle\dot{\eta}\rangle|t}$. To explicitly avoid this problem, the velocities $\{\mathbf{v}_{\mathbf{R}}\}$ in Eqs. (2.26) and (2.29) are replaced by $\{\mathbf{v}'_{\mathbf{R}}\}$, relative to the center-of-mass velocity $\mathbf{V}_{\text{cm}}$, and the center-of-mass motion is coupled to a heat sink. Velocities $\{\mathbf{v}_{\mathbf{d}}\}$ are unchanged, because they already are relative to $\mathbf{V}_{\text{cm}}$. The modified system is

$$\dot{\mathbf{R}}_i = \mathbf{v}_{\mathbf{R},i} + \mathbf{R}_i\dot{\varepsilon}, \tag{2.36}$$

$$m_i\dot{\mathbf{v}}'_{\mathbf{R},i} = \mathbf{F}_{\mathbf{R},i} - m_i\mathbf{v}'_{\mathbf{R},i}(\dot{\eta} + a\dot{\varepsilon}), \tag{2.37}$$

$$\dot{\mathbf{d}}_i = \mathbf{v}_{\mathbf{d},i} + \mathbf{d}_i\dot{\varepsilon}, \tag{2.38}$$

$$m'_i\dot{\mathbf{v}}_{\mathbf{d},i} = \mathbf{F}_{\mathbf{d},i} - m'_i\mathbf{v}_{\mathbf{d},i}(\dot{\eta}_\star + a\dot{\varepsilon}), \tag{2.39}$$

$$M_{\mathrm{tot}}\dot{\mathbf{V}}_{\mathrm{cm}} = -M_{\mathrm{tot}}\mathbf{V}_{\mathrm{cm}}\gamma, \tag{2.40}$$

$$Q\ddot{\eta} = \sum_j m_j v'^2_{\mathbf{R},j} - (N_{\mathrm{f}} - 3)k_{\mathrm{B}}T, \tag{2.41}$$

$$Q_\star\ddot{\eta}_\star = \sum_j m'_j v^2_{\mathbf{d},j} - N_{\mathrm{f}\star}k_{\mathrm{B}}T_\star, \tag{2.42}$$

$$W\ddot{\varepsilon} = a\sum_j \left(m_j v^2_{\mathbf{R},j} + m'_j v^2_{\mathbf{d},j}\right) + \sum_j (\mathbf{R}_j \cdot \mathbf{F}_{\mathbf{R},j} + \mathbf{d}_j \cdot \mathbf{F}_{\mathbf{d},j}) - 3PV, \tag{2.43}$$

where γ is the friction coefficient of the heat sink. The room-temperature thermostat is coupled to $N_{\mathrm{f}} - 3$ degrees of freedom instead of N_{f}. The friction coefficient γ is chosen large enough to damp the center-of-mass motion down to a negligible contribution. The heat sink constitutes a further extension of the system and the energy E_{AH} [see Eq. (2.33)], that does not include the work $-\int dt\, M_{\mathrm{tot}}V^2_{\mathrm{cm}}\gamma$ exerted on the center of mass, is not exactly conserved anymore. Equations (2.36) to (2.43) generate a modified two-temperature isobaric–isothermal ensemble that we denote $NP(T, T_\star)$, where the second temperature refers to the auxiliary polarization variables.

In contrast to the system of Eqs. (2.25) to (2.31), the modified system of Eqs. (2.36) to (2.43) is irreversible. Alternative, fully reversible treatments could be devised. For example, one could replace the dissipative term in Eq. (2.40) by a chain of Nosé–Hoover thermostats [75]. As it will be shown in Sec. 2.4.1, this is unimportant here because the total momentum is "almost" perfectly conserved by the numerical integration, and the introduction of the linear damping does not alter the precision of the simulation. In practice, the nonzero friction γ produces a very small work (less than 0.000 02 kcal/mol/ps on average). Therefore, the dissipative term has essentially no impact on the accuracy of the trajectory or on the conservation of the total energy.

2.3 Computational details

The constant-temperature and constant-pressure molecular dynamics procedures described above have been implemented in the biomolecular program CHARMM [79]. To illustrate and test the present algorithm, a cubic system of 250 water molecules represented by the polarizable SPC (PSPC) model [61] was considered. The PSPC model is based on the original SPC (simple point charge) model [7]. It has the same tetrahedral-like geometry, with $\ell_{OH} = 1.0$ Å and $\theta_{HOH} = 109.47°$. It has a single polarizable site located on the oxygen atom, and hydrogen atoms are not polarizable. Both the molecular dipole and molecular polarizability correspond to the experimental gas-phase values, with μ_0 of 1.885 27 D and α of 1.44 Å^3. The nonpolar interactions are modeled using the usual Lennard-Jones 6–12 potential between each oxygen–oxygen pair. All the parameters of the model are given in Table 2.I. Our own version of the PSPC model reproduces the polarizability by coupling a Drude particle to the oxygen site, with a harmonic spring of force constant k_D of 1000 kcal/mol/Å^2 and a positive charge $q_D = \sqrt{\alpha k_D}$, that is, 2.082 41 e. The total potential energy is

$$\begin{aligned} U(\{\mathbf{r}\}) &= \sum_i \tfrac{1}{2} k_D |\mathbf{r}_{O,i} - \mathbf{r}_{D,i}|^2 + \sum_{i,s} \sum_{j>i,t} \frac{q_{s,i} q_{t,j}}{|\mathbf{r}_{s,i} - \mathbf{r}_{t,j}|} \\ &\quad + \sum_i \sum_{j>i} U_{LJ}(|\mathbf{r}_{O,i} - \mathbf{r}_{O,j}|). \end{aligned} \tag{2.44}$$

where the labels s and t represent the four sites O, D, H_1, and H_2.

Molecular dynamics simulations at constant pressure were performed according to the three following procedures:

1. *Traditional SCF regime.* At every time step, a Newton–Raphson energy minimization is performed to keep the massless Drude particles at their minimum-energy positions until the root-mean-square force on the oscillators is under 10^{-6} kcal/mol/Å. The nuclear motion is produced by Eqs. (2.36) to (2.43) in the special case where $\{m'_i = 0\}$, $\{\mathbf{F}_{d,i} = 0\}$, $P = 1$ atm and $T_\star = 0$.

Table 2.I: Parameters and dimer properties of the SPC (Ref. [7]) and PSPC (Ref. [61]) models. The components of the traceless quadrupole Q are defined with respect to the center of mass of the molecule, using the $\angle$HOH bisector as x axis, the H_1H_2 direction as y axis, and the perpendicular direction as z axis.

	Expt	SPC[a]	PSPC[b]
ℓ_{OH} (Å)	0.9572[c]	1.0000	1.000 0
θ_{HOH} (°)	104.52[c]	109.47	109.47
q_O (e)		−0.82	−2.751 41
q_H (e)		0.41	0.334 50
q_D (e)			2.082 41
k_D (kcal/mol/Å^2)			1000
ϵ (kcal/mol)		0.1554	0.129 49
σ (Å)		3.1655	3.263 00
μ_0 (D)	1.85	2.274	1.855 27
Q_{xx} (DÅ)	−0.134[d]	−0.2938	−0.239 7
Q_{yy} (DÅ)	2.626[d]	2.1162	1.726 5
Q_{zz} (DÅ)	−2.493[d]	−1.8224	−1.486 8
α (Å^3)	1.44		1.44
U_d (kcal/mol)	−5.4[e]	−6.61	−4.29
d_d (Å)	2.98[e]	2.75	2.86
ϕ_d (°)	58 ± 6[e]	22	19
μ_d (D)	2.643[e]	3.622	3.571

[a] From Ref. [7].

[b] From Ref. [61]. The parameters q_D and k_D are specific to the Drude oscillator model and are not in the original reference.

[c] From Ref. [80].

[d] From Ref. [81].

[e] Dimer properties from Ref. [82]: binding energy U_d, oxygen–oxygen distance d_d, acceptor angle ϕ_d, and total dipole μ_d.

2. *Cold Drude oscillators regime.* Trajectories are generated in the $NP(T, T_\star)$ ensemble using Eqs. (2.36) to (2.43) with $P = 1$ atm, $T = 298.15$ K and $T_\star = 1$ K.

3. *Thermalized Drude oscillators regime, $NP(T,T)$ ensemble.* Trajectories are generated in the $NP(T = T_\star)$ ensemble using Eqs (2.36) to (2.43) with $P = 1$ atm and $T = T_\star = 298.15$ K.

Each molecule has 6 degrees of freedom for the nuclei and 3 for the Drude oscillator, thus $N_\mathrm{f} = 6N_\mathrm{mol}$ and $N_{\mathrm{f}\star} = 3N_\mathrm{mol}$. Inertia parameters Q and W are set using both τ and τ_ε equal to 0.1 ps, and using $N_{\mathrm{f}\varepsilon} = N_\mathrm{f}$ for the SCF and the cold Drude oscillators simulations, but $N_{\mathrm{f}\varepsilon} = N_\mathrm{f} + N_{\mathrm{f}\star}$ for the thermalized Drude oscillators simulations. The residual center-of-mass motion is dissipated using $\gamma = 10$ ps^{-1}, even for the SCF simulations, where the "flying ice cube" problem is very unlikely. The extended dynamics simulations are done for various m_D values: 0.8, 0.4, 0.2, and 0.1 amu. The period of free oscillation of the Drude particles is $2\pi\sqrt{m_\mathrm{D}/k_\mathrm{D}}$, corresponding to 8.69 fs for a mass of 0.8 amu and decreasing as $\sqrt{m_\mathrm{D}}$ for other masses. The characteristic response time of the second thermostat, $\tau_\star$, should be short enough to follow this fast motion. For the cold Drude oscillators simulations, $\tau_\star$ is set to 7.07 fs for $m_\mathrm{D} = 0.8$ amu, 5.00 fs for $m_\mathrm{D} = 0.4$ amu, 3.54 fs for $m_\mathrm{D} = 0.2$ amu, and 2.50 fs for $m_\mathrm{D} = 0.1$ amu. For the thermalized Drude oscillators simulations, five-time-longer and five-time-shorter time scales are tested as well: 35.36, 7.07, and 1.41 fs for $m_\mathrm{D} = 0.8$ amu, and so on for other m_D values.

The motion was integrated with time steps Δt of 0.5, 1.0, and 2.0 fs, using the multi-step thermostat integration with $n_\mathrm{c} = 100$. Drude particles having a mass of 0.4 amu are oscillating at a frequency $\omega = \sqrt{k_\mathrm{D}/m_\mathrm{D}}$, corresponding to about 1 fs^{-1}. Given that $\omega\Delta t$ should be small compared to 1 in order for the motion of the oscillator to be precisely integrated [67], none of these time steps would be recommendable for conventional molecular dynamics. The internal water geometry was constrained using the SHAKE/Roll and RATTLE/Roll procedures [71]. The

electrostatic interactions were calculated with the particle-mesh Ewald summation [83] and a 9 Å cutoff for the Lennard-Jones (LJ) interaction.

Each box was simulated for 150 ps and its properties were extracted from the last 100 ps. The following properties are accumulated: the average potential energy Δu, the average molecular volume $\langle v \rangle$, the pair distribution functions $g_{\mathrm{OO}}(r)$, $g_{\mathrm{OH}}(r)$, and $g_{\mathrm{HH}}(r)$, and the average molecular dipole $\langle \mu \rangle$. The average net gain of potential energy Δu upon formation of the dense system is

$$\Delta u = [\langle u_{\mathrm{liq}} \rangle - \langle u_{\mathrm{gas}} \rangle] = \left[\langle u_{\mathrm{liq}} \rangle - \tfrac{3}{2} k_{\mathrm{B}} T_{\star}\right] \tag{2.45}$$

and is related to the vaporization enthalpy Δh

$$\Delta h = k_{\mathrm{B}} T - \langle \Delta u \rangle. \tag{2.46}$$

It should be noted that, even though the water molecules themselves may be rigid, the thermal contribution from the Drude oscillators to the total potential energy, $\frac{3}{2} k_{\mathrm{B}} T_{\star}$, must be rigorously subtracted in an extended Lagrangian simulation (though this number is small if $T_{\star} \approx 1$). The self-diffusion constant D and the static dielectric constant ϵ were computed from the last 100 ps of ten independent simulations of 150 ps, using the expressions [84, 85]

$$D = \lim_{t \to \infty} \frac{1}{6t} \left\langle \frac{1}{N} \sum_{i=1}^{N} |\mathbf{r}_{\mathrm{O},i}(t) - \mathbf{r}_{\mathrm{O},i}(0)|^2 \right\rangle \tag{2.47}$$

and

$$\epsilon = \epsilon_{\infty} + \frac{4\pi}{3\langle V \rangle k_{\mathrm{B}} T} \left(\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2 \right), \tag{2.48}$$

where$\langle V \rangle$ is the average volume of the box and $\mathbf{M}$ is the total dipole moment of the box. For thermalized Drude oscillators, $\epsilon_{\infty} = 1$, because all electrostatic response is included in $\langle \mathbf{M}^2 \rangle$. In the SCF regime, or whenever $T_{\star}$ is very low, ϵ_{∞} can be

estimated from the Clausius–Mossotti equation

$$\frac{\epsilon_\infty - 1}{\epsilon_\infty + 2} = \frac{4\pi}{3}\frac{\alpha}{v}. \tag{2.49}$$

For $\alpha = 1.44$ Å^3, we get $\epsilon_\infty = 1.75$ if $v = 30$ Å^3 (the experimental molecular volume) and $\epsilon_\infty = 1.70$ if $v = 32$ Å^3 (the molecular volume of the PSPC model, as it will be shown later). The uncertainty on D and ϵ was estimated from the spread over the ten simulations.

2.4 Results and discussion

2.4.1 Numerical precision and stability

The stability and accuracy of the trajectories generated according to the SCF procedure is first examined. The result from those trajectories is then used as reference to assess the correctness of the trajectories generated with extended dynamics simulations with "cold" and thermalized ("hot") Drude oscillators. The trajectories generated with the SCF procedure are stable for all three time-step values (0.5, 1.0, and 2.0 fs). For $\Delta t = 0.5$ and 1.0 fs, the total extended energy E_{AH} of Eq. (2.33) has no noticeable drift (see Fig. 2.1 for $\Delta t = 1.0$ fs). For $\Delta t = 2.0$ fs, E_{AH} is very slowly drifting down, at a rate of -0.0009 kcal/mol/ps. This excellent energy conservation reflects the fact that the massless Drude particles follow adiabatically according to the motion of the atoms. The simulations are stable because the very high-frequency motion of the light Drude particles is effectively not present in the SCF trajectories. The convergence of the SCF procedure is very closely related to the number of iterations of the Newton–Raphson energy minimization. For each molecular dynamics integration time step, the SCF procedure requires about 16 iterations to converge to the desired criterion ($F_{\text{RMS}} < 10^{-6}$ kcal/mol/Å). This procedure yields a precise SCF trajectory but is computationally expensive because each time step requires 16 evaluations of the atomic forces. It may be tempting to reduce the computational cost of SCF by simply using a less stringent con-

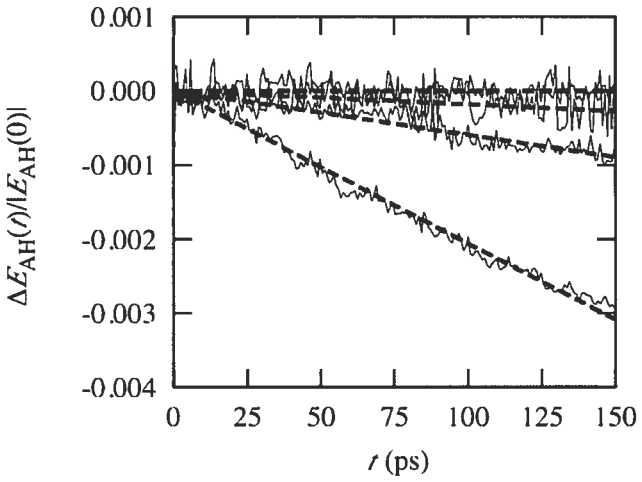

Figure 2.1: Relative deviation of the extended energy for $T_\star = 1$ K and $T_\star = 0$ K (SCF) simulations (with a 1.0 fs time step). A larger m_D creates a larger downward energy drift: -0.0338 kcal/mol/ps for $m_D = 0.8$ amu, -0.0098 kcal/mol/ps for $m_D = 0.4$ amu, -0.0031 kcal/mol/ps for $m_D = 0.2$ amu, and no noticeable drift for $m_D = 0$ amu (SCF).

vergence criteria. But, whether the properties of the system are affected by an imprecise SCF scheme is unclear. Test simulations using a convergence criterion of $F_{RMS} < 10^{-4}$ kcal/mol/Å give the same liquid properties with 12 iterations, but exhibit a somewhat faster downward drift in E_{AH} (-0.0215 kcal/mol/ps for $\Delta t = 2.0$ fs). Test simulations using $F_{RMS} < 10^{-2}$ kcal/mol/Å require only 7 iterations every step, but exhibit an unacceptable energy drift of -6.53 kcal/mol/ps. This gives rise to spurious dragging forces that lower the self-diffusion constant and distort the liquid structure. In a recent study, Yu *et al.* [58] have used a very tolerant iteration scheme that requires only 2 or 3 iterations for every time step. Our own test simulations indicate that propagating a trajectory with a convergence criteria that requires only 3 iterations, that is, $F_{RMS} < 1$ kcal/mol/Å, yields a liquid that is significantly more dense and a self-diffusion constant that is about four times smaller than the true SCF value. Our tests indicate that iterating at least until $F_{RMS} < 10^{-4}$ kcal/mol/Å guarantees that the Drude particles have a truly SCF motion. We use $F_{RMS} < 10^{-6}$ kcal/mol/Å as a precaution.

Extended dynamics simulations, which treat the Drude particles as simple clas-

sical degrees of freedom, require only a single evaluation of the forces per time step and are, thus, much more efficient than the slow SCF procedure. Overall, simulations generated with cold Drude oscillators ($T_\star = 1$ K) appear to be more stable and to conserve energy better than simulations with hot Drude oscillators ($T_\star = 298.15$ K). Using a time step of 0.5 fs with either $T_\star = 1$ K or $T_\star = 298.15$ K gives stable trajectories for all four values of the mass assigned to the Drude particles. But when a time step of 1.0 fs is used, simulations with $T_\star = 298.15$ K are unstable for $m_\mathrm{D} = 0.1$ and 0.2 amu, whereas simulations with $T_\star = 1$ K are unstable only for $m_\mathrm{D} = 0.1$ amu. Using a time step of 2.0 fs, simulations with $T_\star = 1$ K are stable for $m_\mathrm{D} = 0.8$ amu, whereas no simulations with $T_\star = 298.15$ K are stable. For stable simulations, the energy E_AH is well conserved for $T_\star = 1$ K (see Fig. 2.1 for $\Delta t = 1.0$ fs). The energy drift is slow for $m_\mathrm{D} = 0.8$ amu and becomes even slower as m_D is decreased. In contrast, the energy conservation is much poorer for $T_\star = 298.15$ K. And, although it is improved by using a five-time shorter relaxation time for the $T_\star$ thermostat (that is, $\tau_\star = 1.41$ fs for $m_\mathrm{D} = 0.8$ amu, and so on for other m_D values), the energy drift is about an order of magnitude larger for $T_\star = 298.15$ K than it is for $T_\star = 1$ K. With $m_\mathrm{D} = 0.8$ amu, using $\tau_\star = 35.36$ or even 7.07 fs creates spurious temperature oscillations. These are clear indications that, for a given Drude oscillator mass and given time step, the trajectories are more precisely integrated with $T_\star = 1$ K than with $T_\star = 298.15$ K. The kinetic temperatures,

$$T(t) \equiv \sum_j m_j v_{\mathbf{R},j}'^2 / (N_\mathrm{f} - 3) k_\mathrm{B} \tag{2.50}$$

and

$$T_\star(t) \equiv \sum_j m_j' v_{\mathbf{d},j}^2 / N_{\mathrm{f}\star} k_\mathrm{B}, \tag{2.51}$$

are correctly regulated and $\langle T(t) \rangle = 298.15$ K and $\langle T_\star(t) \rangle = 1$ K within statistical error. As shown in Fig. 2.2, the temperature $T(t)$ exhibits canonical fluctuations and is distributed according to a Gaussian with a standard deviation $\sigma = 10.9$ K, as expected for a finite system in the canonical ensemble [86] [$\sigma^2 = 2\langle T \rangle^2 / (N_\mathrm{f} - 3)$, with $N_\mathrm{f} = 6 \times 250$]. In contrast, fluctuations in the temperature $T_\star$ are not dis-

tributed according to a Gaussian. As seen in Fig. 2.2, the distribution is asymmetric and skewed toward the low temperatures. This skewness reflects the small but steady nonequilibrium heat flow in the system. It is more pronounced for a larger value of m_{D}, that is, when the kinetic coupling between the Drude and the nuclear degrees of freedom is stronger. Nonetheless, this behavior does not appear to have any consequences on the molecular motion, which obeys the canonical distribution at the correct temperature.

Although the total momentum is rigorously conserved by Eqs. (2.25) to (2.31), a spurious translation of the center of mass of the system does arise slowly after some time due to an accumulation of small numerical errors. Preliminary tests show that the center of mass can develop a net translation of about 0.03 Å/ps on average after a trajectory of 100 ps (data not shown). From Eq. (2.35), the time scale for the onset of the "flying ice cube" artifact can be estimated from $1/|\langle\dot{\eta}\rangle|$, giving about 4 ns for $m_{\mathrm{D}} = 0.8$ amu, 19 ns for $m_{\mathrm{D}} = 0.4$ amu, and longer for $m_{\mathrm{D}} = 0.2$ and 0.0 amu. These time scales are sufficiently long to allow Eqs. (2.25) to (2.31) to be used directly for simulations where rigorously reversible dynamics is required and avoid the "flying ice cube" artifact. Nevertheless, although the errors are very small for the current simulations of a simple bulk liquid, the problem may become much more acute in the case of large and complex biomolecular systems. To completely avoid any acceleration of the center of mass, the system was propagated using Eqs. (2.36) to (2.43), in which a small irreversible damping was applied to the global translation of the system. Because the original equations conserve the total momentum almost perfectly, the amount of energy dissipated by the irreversible damping is nearly negligible (less than 0.000 02 kcal/mol/ps on average).

To illustrate more directly the correspondence between the cold Drude oscillators simulations and the SCF simulations, the time-course of the oscillators of three randomly picked water molecules are compared in Fig. 2.3. Using the same initial coordinates and velocities for the nuclei, both simulation methods produce nearly identical trajectories for the first 0.3 ps, after which the chaotic nature of the motion in the liquid takes over and the trajectories diverge rapidly from one another.

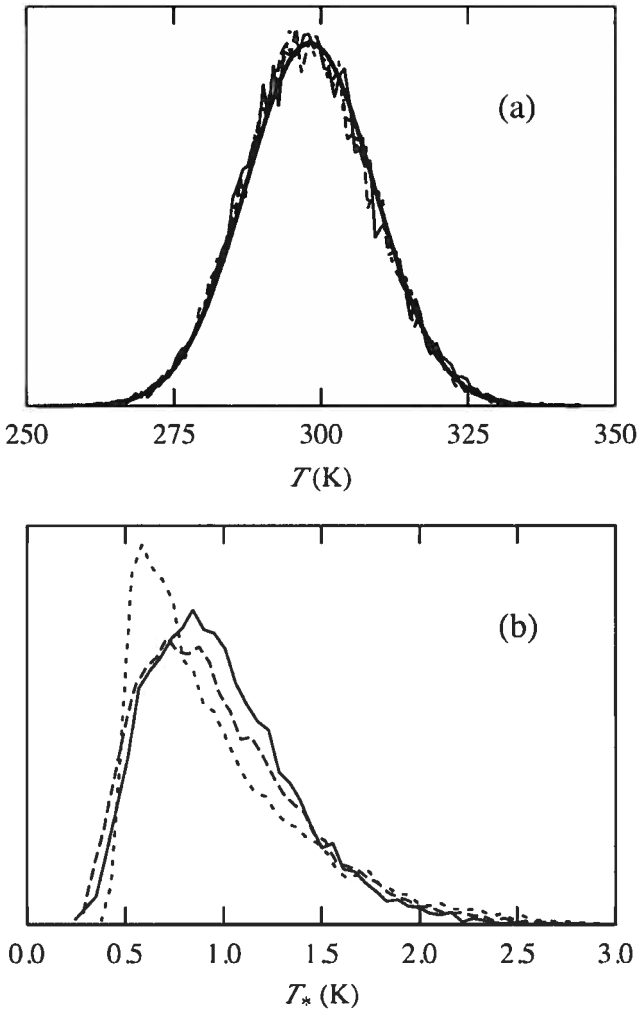

Figure 2.2: Temperature distributions (a) for T and (b) for $T_\star$ for $m_{\rm D} = 0.2$ (solid lines), $m_{\rm D} = 0.4$ (dashed lines), and $m_{\rm D} = 0.8$ (dotted lines). The smooth solid line in (a) is the expected temperature distribution in the canonical ensemble for 1497 degrees of freedom.

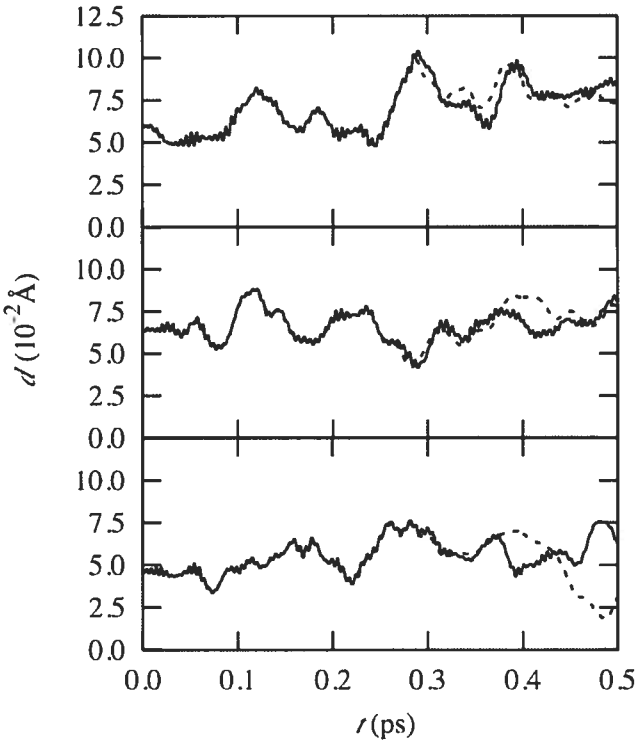

Figure 2.3: Comparison of the trajectories of three Drude particles for $m_{\mathrm{D}} = 0.4$ amu and $T_\star = 1$ K (solid lines), and SCF regime (dashed lines).

This time scale is comparable to the decay time of the velocity autocorrelation function, suggesting that dynamical properties should be simulated precisely.

As a side issue, one may note that the kinetic coupling between the high-frequency Drude oscillators and the remaining degrees of freedom in the system is very weak. Thermalization of the motion of the Drude oscillators can be achieved from this weak coupling only, but it is significantly accelerated with a coupling to a separate, fast-reacting thermostat. In the simulations with thermalized Drude oscillators, the second thermostat at $T_\star = 298.15$ K allows a rapid equipartition of the energy, whereas a few tens of picoseconds were needed in the simulations of van Belle *et al.* based on an extended dynamics scheme with point dipoles [66]. In those simulations a dipole inertia factor of 0.5 amu/e^2, roughly equivalent to a mass of $m_{\mathrm{D}} = 2.168$ amu, was used. The thermalization time grows as the inertia of the dipoles is decreased. For instance, our own test simulations without a separate

thermostat on the Drude oscillators show that oscillators with $m_D = 0.4$ amu require as much as 150 ps to fully thermalize. This is a very long simulation time during which the thermodynamic properties are unreliable.

2.4.2 Average properties of the liquid

Table 2.II summarizes the average properties of the liquid obtained with the various simulation procedures. In the following, the SCF simulations are used as a reference to assess the accuracy of the extended dynamics trajectories. It is observed that the simulations with the cold Drude oscillators give consistent results for all the values of the mass m_D that were tested. For a given time step, Δu, $\langle v \rangle$, $\langle \mu \rangle$, and the $g(r)$ distribution functions (see Fig. 2.4) are all within the statistical errors. Most importantly, they are very close to the properties calculated from the trajectories generated according to the SCF procedure. The time step seem to have a very small systematic effect on the properties, but the effect is probably not specific to the motion of the Drude oscillators because it is independent of m_D and it is noticeable for the SCF values as well. The self-diffusion constant D and the dielectric constant ϵ are close to the SCF values as well—within the estimated uncertainties. Figure 2.5 shows that the histograms of the induced dipole for the simulations generated with cold Drude oscillators superimpose to the SCF results and are independent of the mass attributed to the Drude particles.

It is clear that the simulations generated with thermalized Drude oscillators exhibit some systematic deviations relative to those generated with SCF or with the cold Drude oscillators. In particular, there is a slight increase in the liquid density and of the average molecular dipole. This is caused by the presence of the small additional cohesion interaction arising from (classical) dipole–dipole correlations, given to lowest order term in Eq. (2.14). Furthermore, Fig. 2.5 shows that the $T_\star = 298.15$ K simulations exhibit a spurious enhancement of the dipole fluctuations if the mass of the Drude particle becomes too small. This observation is in accord with previous analysis which shows that the relative error on the mean-squared displacement of an oscillator of frequency ω simulated with a

Table 2.II: Liquid properties for the PSPC water model for cold, thermalized, and SCF Drude oscillators, compared with SPC and experimental properties. Dashes indicate unstable simulations.

$T_\star$ (K)	Δt (fs)	m_D (amu)	Δu (kcal/mol)	$\langle v \rangle$ (Å^3)	$\langle \mu \rangle$ (D)	D (10^{-5}cm^2/s)	ϵ
1.00	0.5	0.8	−7.94	32.13	2.714	3.98 ± 0.06	195 ± 30
		0.4	−7.94	32.13	2.714	3.91 ± 0.09	175 ± 30
		0.2	−7.95	32.11	2.715	3.87 ± 0.07	195 ± 30
		0.1	−7.95	32.12	2.715	3.93 ± 0.06	205 ± 30
	1.0	0.8	−7.94	32.13	2.714	3.93 ± 0.06	225 ± 30
		0.4	−7.94	32.12	2.714	3.97 ± 0.08	185 ± 30
		0.2	−7.94	32.13	2.714	3.96 ± 0.10	170 ± 30
		0.1	—				
	2.0	0.8	−7.92	32.14	2.712	4.03 ± 0.07	210 ± 30
		0.4	—				
		0.2	—				
		0.1	—				
298.15	0.5	0.8	−7.95	31.88	2.744		
		0.4	−7.93	31.90	2.745		
		0.2	−7.88	31.61	2.752		
		0.1	−7.74	31.13	2.761		
	1.0	0.8	−7.86	31.56	2.750		
		0.4	−7.71	31.03	2.760		
		0.2	—				
		0.1	—				
SCF	1.0	0.0	−7.94	32.14	2.714	3.86 ± 0.05	190 ± 30
	2.0	0.0	−7.92	32.16	2.712	3.98 ± 0.06	195 ± 30
Ahlström *et al.* [61][a]			−9.08	29.92	2.9	2.0 ± 0.2	
van Belle *et al.* [65][a]			−9.11	29.92	2.8	2.4 ± 0.2	
SPC[b]			−9.79	31.2	2.274	3.91 ± 0.04	73 ± 10
Experiment			−9.92[c]	30.0[d]		2.3[e]	78.4[f]

[a] Constant-volume simulations.

[b] Liquid properties computed from 20 simulations of 250 molecules in the isothermal–isobaric ensemble, using particle-mesh Ewald and a 9 Å Lennard-Jones cutoff.

[c] From the experimental vaporization enthalpy $\Delta h = 10.52$ kcal/mol (Ref. [87]) using $\Delta h = k_B T - \Delta u$.

[d] Corresponding to 0.997 g/cm^3.

[e] From Ref. [88].

[f] From Ref. [89].

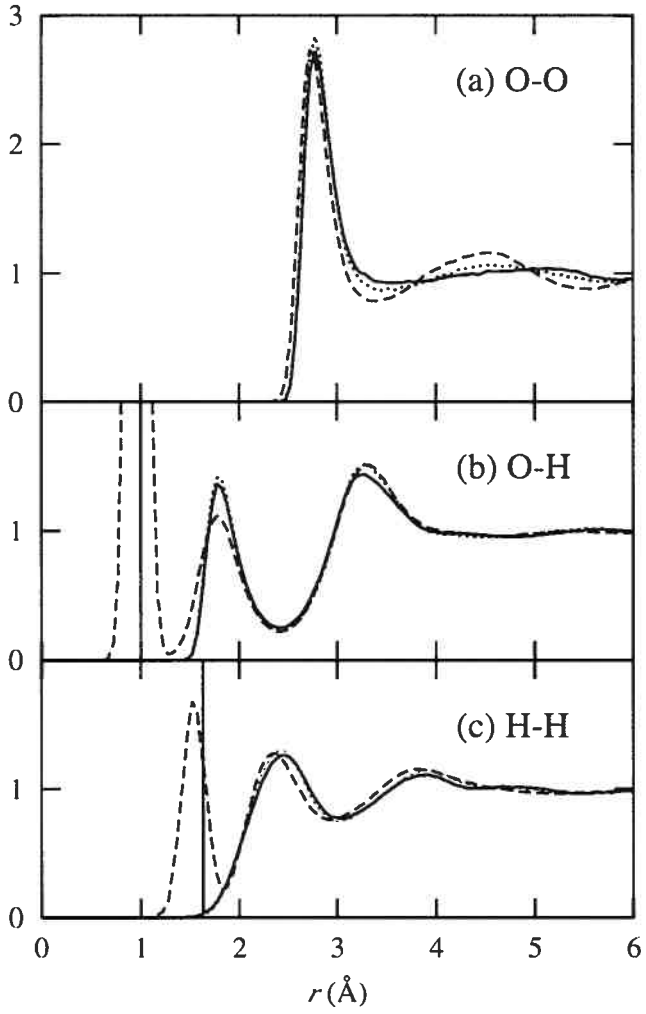

Figure 2.4: (a) Oxygen–oxygen; (b) oxygen–hydrogen; and (b) hydrogen–hydrogen radial distribution functions for PSPC model (solid lines), compared to SPC distributions (dotted lines) and experimental distributions from Soper *et al.* (Ref. [90]) (dashed lines). Only the PSPC distributions obtained with $T_\star = 0$ K with $\Delta t = 2.0$ fs are shown, but $T_\star = 1$ K gives indistinguishable curves.

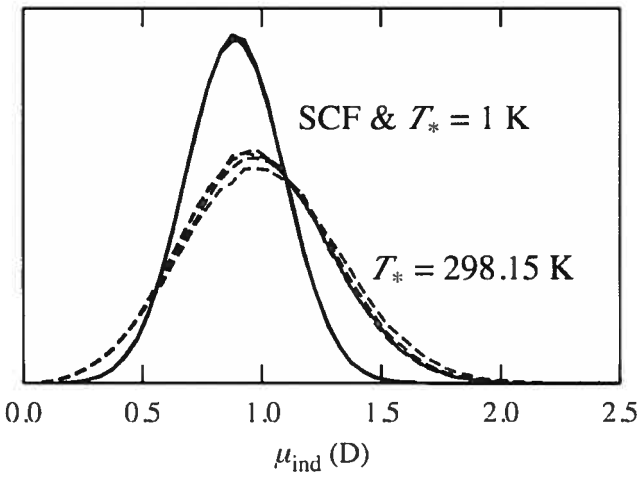

Figure 2.5: Induced dipole histograms for SCF simulations with $\Delta t = 1.0$ fs (solid line); $T_\star = 1$ K simulations with $\Delta t = 1.0$ fs and $m_D = 0.8$, 0.4 and 0.2 amu (three curves in solid lines, indistinguishable from the SCF curve); and $T_\star = 298.15$ K simulations with $\Delta t = 0.5$ fs (dashed lines). Dashed histograms get lower and wider as m_D decreases from 0.8, 0.4, 0.2, to 0.1 amu. It shows that, for room-temperature Drude oscillators, both average-induced dipole and the dipole fluctuations have unphysical dependences on m_D. For $T_\star = 1$ K simulations, the induced dipoles are rarely larger than 1.5 D, corresponding to an actual Drude displacement of about 0.15 Å.

velocity-Verlet algorithm varies as $\omega^2 \Delta t^2$ [67]. As a consequence of the inaccurate integration of the high-frequency motion of the Drude oscillators, the average equilibrium thermodynamics properties now exhibit an unphysical dependence upon the choice of mass attributed to the Drude particles. As m_D decreases, $\langle \mu \rangle$ slightly increases, $\langle v \rangle$ decreases, and Δu is becoming less negative. In contrast, the results from the cold Drude oscillators simulations do not vary as m_D is decreased. This strongly indicates that generating trajectories with cold Drude oscillators should be the method of choice to simulate such models.

It is of interest to compare the present results with those of previous simulations and with experimental data. The PSPC model was previously simulated using a relatively tolerant SCF procedure by Ahlström *et al.* [61], and using an extended dynamics algorithm by van Belle *et al.* [65]. The main results from those simulations are included at the bottom of Table 2.II. The results of Ahlström *et al.* and van Belle *et al* are relatively uniform, but show significant discrepancies with the present results, particularly concerning the enthalpy, the average density, and the self-diffusion constant. It is unlikely that these discrepancies are due to the fact that the original point dipole of the PSPC model was replaced by a Drude oscillator, because for the simulations at $T_\star = 298.15$ K the Drude particles are almost never displaced by more than 0.2 Å (see Fig. 2.5). Even at this separation, the point dipole remains a good approximation. The discrepancies are most probably due to the fact that the present work uses a constant-pressure algorithm, whereas the previous simulations were performed at a constant volume corresponding to the experimental density of water. In addition, a PME summation was used here instead of a screening and cutoff scheme [61, 65]. To further compare with the results of Ahlström *et al.* and van Belle *et al.*, additional $T_\star = 298.15$ K simulations were done at constant volume [using Eqs. (2.36) to (2.43) with an infinite W and assuming a constant density of 1 g/cm^3]. At constant volume, with $m_D = 1.0$ amu and using the Ewald summation, we get $\Delta u = -7.41$ kcal/mol. The remaining discrepancy with the original results (-9.08 ± 0.07 [61] and -9.11 ± 0.05 kcal/mol [65]) is caused by the differences in the treatment of the electrostatic interactions.

Additional test simulations using an electrostatic screening and cutoff scheme instead of PME reproduces the previous results for the enthalpy [61, 65]. The present constant-pressure PME simulations show that the true liquid density of the PSPC model is, in fact, too low and that the average potential energy Δu is about 2 kcal/mol off compared to experiments. Lastly, the present simulations show that the dielectric constant is on the order of 200. This high dielectric constant is consistent with the average molecular dipole of 2.71 D observed in the simulations. According to a trend observed for a number of simple water models similar to PSPC [91, 92, 93, 94], the average molecular dipole in the liquid should be around 2.4 to 2.6 D to reproduce the correct dielectric constant of 78.4 [89]. The present analysis is not meant to be a criticism of the previous efforts to design and simulate the PSPC model more than a decade ago [61, 65]. Nonetheless, it highlights the general importance of avoiding truncated electrostatics with cutoff schemes in simulations of polar liquids, and the difficulties in designing and parametrizing a simple computational model that reproduces the properties of liquid water accurately.

2.5 Summary

We have presented a general polarizable model based on Drude oscillators and demonstrated that the SCF motion of the Drude oscillators can be efficiently and accurately simulated with an extended molecular dynamics simulation procedure keeping the polarization variables at low temperature. This is done by using two separate thermostats for the "atomic" and "dipolar" motions of the polarizable atoms. An algorithm to generate molecular dynamics trajectories in this modified two-temperature isobaric–isothermal $NP(T, T_\star)$ ensemble was described. Tests demonstrated that this is the most advantageous approach for molecular dynamics simulations of polarizable model based on classical Drude oscillators. The traditional SCF procedure, in which the Drude particles are kept at the minimum of their local energy surface, is at least one order of magnitude more expensive com-

putationally than this algorithm. An alternative extended dynamical approach, in which the Drude particles are allowed to equilibrate thermally with the nuclei at room temperature, was found to yield some liquid properties inaccurately and was also less stable numerically.

The present results show that simulations in the $NP(T,T_\star)$ ensemble have a clear advantage over simulations over $NP(T,0)$ and $NP(T,T)$ simulations. With separate atomic and dipolar thermostatting, the particles are kept in a regime where the stiff Drude oscillator forces counterbalance the induction forces and leave a slowly varying net force that can be integrated using a reasonably large time step. A 2.0 fs time step can safely be used with $m_{\mathrm{D}} = 0.8$ amu and retain accurate thermodynamic as well as dynamical properties. Molecular dynamics with cold Drude oscillators ($T_\star = 1$ K) is more than an order of magnitude more efficient than the SCF regime but produces nearly equivalent liquid properties. The modified isobaric–isothermal $NP(T,T_\star)$ method is more efficient than molecular dynamics with thermalized Drude oscillators, that requires a smaller time step to integrate the motion of the hot Drude oscillators with the same accuracy. It is preferable as well, because the thermal fluctuations of the Drude oscillators have a small but undesirable cohesive effect on the liquid.

The Drude oscillator model has several advantages over a point-dipole model. It gives an intuitive physical picture in terms of a displacement of the electronic distribution. The auxiliary particle can inherit any of the properties of a model atom: dipole screening could be implicitly incorporated using charge screening for the Drude particles and the nuclei, and dispersion and repulsion effects could be extended to the dipole using a shallow Lennard-Jones potential on the Drude particles, thus allowing mechanical polarizability [17]. Representing a dipole as two point charges allows any pre-existing, charge-based, implementation of the particle-mesh Ewald summation [83] to be used. Last, because the dipole is not point-like, the model is able to represent an actual charge delocalization without the need of additional nonatomic sites.

As for any explicitly polarizable model, the drawback is that the system has

additional degrees of freedom, thereby increasing the computational cost of the simulation. For pure water, having additional auxiliary particles only on the oxygen atoms makes a standard $O(N \log N)$ computation of the atomic forces about 40% more expensive for a three-site model like the one used in this work, and about 30% for a four-site model. Therefore, trajectories generated in the two-temperatures isobaric–isothermal $NP(T, T_\star)$ ensemble are essentially equivalent to those generated according to the traditional SCF procedure at a fraction of the cost.

In the near future, efforts will be dedicated to the parametrization of a polarizable water model based on classical Drude oscillators that correctly reproduces all the essential properties of the liquid phase while improving the accuracy of the microscopic water–water interactions compared to the non-polarizable models. Because polarizability is essential for a better understanding of the affinity and coordination of small charged ligands to protein or nucleic acid molecules [15], the following logical step will be to get similar polarizable models to improve the microscopic representation of ion–water and ion–protein interactions. The present scheme will be the basis for an explicitly polarizable force field for molecular dynamics of proteins and nucleic acids.

Acknowledgements

Helpful discussions with Alexander D. MacKerell Jr., Bruce J. Berne, and Mark E. Tuckerman are gratefully acknowledged. G. L. is grateful to Alain Caillé for his support. This work was supported by Grant No. 0110847 from the National Science Foundation.

2.A Self-consistent equations for induced polarization

For unscreened Coulomb interactions

$$U_{\text{elec}} = \sum_i \sum_{j>i} \left[\frac{(q_i - q_{\text{D},i})(q_j - q_{\text{D},j})}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{(q_i - q_{\text{D},i})q_{\text{D},j}}{|\mathbf{r}_i - \mathbf{r}_j - \mathbf{d}_j|} + \frac{q_{\text{D},i}(q_j - q_{\text{D},j})}{|\mathbf{r}_i + \mathbf{d}_i - \mathbf{r}_j|} + \frac{q_{\text{D},i}q_{\text{D},j}}{|\mathbf{r}_i + \mathbf{d}_i - \mathbf{r}_j - \mathbf{d}_j|} \right]. \tag{2.52}$$

Assuming that the $\mathbf{d}$'s are small, all terms can be expanded into charge–charge, charge–dipole, and dipole–dipole contributions

$$U_{\text{elec}} \simeq \sum_i \sum_{j>i} \left[\frac{q_i q_j}{r_{ij}} + (q_{\text{D},i}\mathbf{d}_i q_j - q_i q_{\text{D},j}\mathbf{d}_j) \cdot \boldsymbol{\nabla}_i \left(\frac{1}{r_{ij}} \right) - q_{\text{D},i}\mathbf{d}_i q_{\text{D},j}\mathbf{d}_j : \boldsymbol{\nabla}_i \boldsymbol{\nabla}_i \left(\frac{1}{r_{ij}} \right) \right]. \tag{2.53}$$

All gradients have been converted to $\boldsymbol{\nabla}_i \equiv \partial/\partial \mathbf{r}_i$. Condition (2.4) of the text now reads

$$k_{\text{D}}\mathbf{d}_k = \sum_{j \neq k} \left[-q_{\text{D},k} q_j \boldsymbol{\nabla}_k \left(\frac{1}{r_{kj}} \right) + q_{\text{D},k} q_{\text{D},j} \mathbf{d}_j \cdot \boldsymbol{\nabla}_k \boldsymbol{\nabla}_k \left(\frac{1}{r_{kj}} \right) \right]. \tag{2.54}$$

With $\boldsymbol{\mu}_k = q_{\text{D},k}\mathbf{d}_k$ and $\alpha_k = q_{\text{D},k}^2/k_{\text{D}}$, these are Eqs (2.6) of the text.

2.B Effective field for thermalized Drude oscillators

With $\boldsymbol{\mu}_i = q_{\text{D},i}\mathbf{d}_i$, and because only U_{self} and U_{elec} depend on the $\mathbf{d}$'s

$$\delta\mathbf{d}_i \cdot \frac{\partial^2 U}{\partial \mathbf{d}_i \partial \mathbf{d}_j} \cdot \delta\mathbf{d}_j = \delta\boldsymbol{\mu}_i \cdot \left(\frac{\partial^2 U_{\text{self}}}{\partial \boldsymbol{\mu}_i \partial \boldsymbol{\mu}_j} + \frac{\partial^2 U_{\text{elec}}}{\partial \boldsymbol{\mu}_i \partial \boldsymbol{\mu}_j} \right) \cdot \delta\boldsymbol{\mu}_j. \tag{2.55}$$

With $\alpha_i = q_{\mathrm{D},i}^2/k_{\mathrm{D}}$

$$\begin{aligned} \frac{\partial^2 U_{\mathrm{self}}}{\partial \boldsymbol{\mu}_i \partial \boldsymbol{\mu}_j} &= \delta_{ij} \frac{\mathbf{1}}{\alpha_i} \\ &\equiv \mathbf{A}_{ij}^{-1}, \end{aligned} \tag{2.56}$$

where $\mathbf{1}$ is the 3×3 identity matrix. From Eq. (2.53) in the point dipole approximation,

$$\begin{aligned} \frac{\partial^2 U_{\mathrm{elec}}}{\partial \boldsymbol{\mu}_i \partial \boldsymbol{\mu}_j} &= -\boldsymbol{\nabla}_i \boldsymbol{\nabla}_i \left(\frac{1}{r_{ij}} \right) \\ &\equiv -\mathbf{T}_{ij}. \end{aligned} \tag{2.57}$$

Integrating over point-dipole fluctuations $\{\delta\boldsymbol{\mu}\}$ instead of displacement fluctuations $\{\delta\mathbf{d}\}$, the correction factor on the right-hand-side of Eq. (2.13) is

$$\frac{\int d\{\delta\boldsymbol{\mu}\} \prod_{ij} e^{-\frac{1}{2}\beta \delta\boldsymbol{\mu}_i \cdot \left(\mathbf{A}_{ij}^{-1} - \mathbf{T}_{ij}\right) \cdot \delta\boldsymbol{\mu}_j}}{\int d\{\delta\boldsymbol{\mu}\} \prod_{ij} e^{-\frac{1}{2}\beta \delta\boldsymbol{\mu}_i \cdot \mathbf{A}_{ij}^{-1} \cdot \delta\boldsymbol{\mu}_j}} \equiv \left\langle \prod_{ij} e^{\frac{1}{2}\beta \delta\boldsymbol{\mu}_i \cdot \mathbf{T}_{ij} \cdot \delta\boldsymbol{\mu}_j} \right\rangle, \tag{2.58}$$

where $\langle \ldots \rangle$ is an average over the thermal fluctuations of noninteracting dipoles. The average is expanded as a moment series

$$\begin{aligned} \left\langle \prod_{ij} e^{\frac{1}{2}\beta \delta\boldsymbol{\mu}_i \cdot \mathbf{T}_{ij} \cdot \delta\boldsymbol{\mu}_j} \right\rangle &= 1 + \frac{\beta}{2} \sum_{ij} \mathbf{T}_{ij} \cdot \langle \delta\boldsymbol{\mu}_i \delta\boldsymbol{\mu}_j \rangle \\ &\quad + \tfrac{1}{2} \left(\frac{\beta}{2} \right)^2 \sum_{ijkl} \mathbf{T}_{ij} \mathbf{T}_{kl} \cdot \langle \delta\boldsymbol{\mu}_i \delta\boldsymbol{\mu}_j \delta\boldsymbol{\mu}_k \delta\boldsymbol{\mu}_l \rangle + \cdots. \end{aligned} \tag{2.59}$$

The two-point correlation functions are explicitly integrated

$$\begin{aligned} \langle \delta\boldsymbol{\mu}_i \delta\boldsymbol{\mu}_j \rangle &= \frac{\int d\{\delta\boldsymbol{\mu}\} \prod_{kl} \delta\boldsymbol{\mu}_i \delta\boldsymbol{\mu}_j e^{-\frac{1}{2}\beta \delta\boldsymbol{\mu}_k \cdot \mathbf{A}_{kl}^{-1} \cdot \delta\boldsymbol{\mu}_l}}{\int d\{\delta\boldsymbol{\mu}\} \prod_{kl} e^{-\frac{1}{2}\beta \delta\boldsymbol{\mu}_k \cdot \mathbf{A}_{kl}^{-1} \cdot \delta\boldsymbol{\mu}_l}} \\ &= \frac{1}{\beta} \mathbf{A}_{ij}. \end{aligned} \tag{2.60}$$

From Wick's theorem (e.g., see Ref. [95]), the four-point correlation functions are written as sums of two-point correlation functions

$$\langle \delta\boldsymbol{\mu}_i \delta\boldsymbol{\mu}_j \delta\boldsymbol{\mu}_k \delta\boldsymbol{\mu}_l \rangle = \frac{1}{\beta^2} \left(\mathbf{A}_{ij}\mathbf{A}_{kl} + \mathbf{A}_{ik}\mathbf{A}_{jl} + \mathbf{A}_{il}\mathbf{A}_{jk} \right). \tag{2.61}$$

Because $\mathbf{T}_{ij} = \mathbf{0}$ for $i = j$ and $\mathbf{A}_{ij} = \mathbf{0}$ for $i \neq j$, the first term of Eq. (2.59) is

$$\begin{aligned} \frac{\beta}{2} \sum_{ij} \mathbf{T}_{ij} \cdot \langle \delta\boldsymbol{\mu}_i \delta\boldsymbol{\mu}_j \rangle &= \tfrac{1}{2} \sum_{ij} \mathbf{T}_{ij} \cdot \mathbf{A}_{ij} \\ &= 0, \end{aligned} \tag{2.62}$$

and the second term is

$$\begin{aligned} &\tfrac{1}{2} \left(\frac{\beta}{2} \right)^2 \sum_{ijkl} \mathbf{T}_{ij}\mathbf{T}_{kl} \cdot \langle \delta\boldsymbol{\mu}_i \delta\boldsymbol{\mu}_j \delta\boldsymbol{\mu}_k \delta\boldsymbol{\mu}_l \rangle \\ &\qquad = \tfrac{1}{8} \sum_{ijkl} \mathbf{T}_{ij}\mathbf{T}_{kl} \cdot \left(\mathbf{A}_{ij}\mathbf{A}_{kl} + \mathbf{A}_{ik}\mathbf{A}_{jl} + \mathbf{A}_{il}\mathbf{A}_{jk} \right) \\ &\qquad = \tfrac{1}{4} \sum_{ijkl} \mathbf{T}_{ij}\mathbf{T}_{kl} \cdot \mathbf{A}_{ik}\mathbf{A}_{jl}. \end{aligned} \tag{2.63}$$

Because $\mathbf{T}_{ij} = (3\hat{\mathbf{r}}_{ij}\hat{\mathbf{r}}_{ij} - \mathbf{1})/r_{ij}^3$,

$$\tfrac{1}{4} \sum_{ijkl} \mathbf{T}_{ij}\mathbf{T}_{kl} \cdot \mathbf{A}_{ik}\mathbf{A}_{jl} = \tfrac{3}{2} \sum_{ij} \frac{\alpha_i \alpha_j}{r_{ij}^6} \tag{2.64}$$

and

$$\left\langle \prod_{ij} e^{\frac{1}{2}\beta \delta\boldsymbol{\mu}_i \cdot \mathbf{T}_{ij} \cdot \delta\boldsymbol{\mu}_j} \right\rangle = 1 + \tfrac{3}{2} \sum_{ij} \frac{\alpha_i \alpha_j}{r_{ij}^6} + \cdots. \tag{2.65}$$

Back to expression (2.13), the two-body correction to the self-consistent field is

$$U^{\text{eff}}(\{\mathbf{r}\}) = U^{\text{SCF}}(\{\mathbf{r}\}) - \tfrac{3}{2} k_{\text{B}} T \sum_{ij} \frac{\alpha_i \alpha_j}{r_{ij}^6} + \cdots. \tag{2.66}$$

2.C Numerical integrators

2.C.1 Nosé–Hoover thermostats

The Liouville operator corresponding to Eqs. (2.15) to (2.18) is

$$i\mathcal{L} = i\mathcal{L}_r + i\mathcal{L}_v + i\mathcal{L}_\eta + i\mathcal{L}_{\dot{\eta}}, \tag{2.67}$$

where

$$i\mathcal{L}_r = \sum_i \left(\mathbf{v}_{\mathbf{R},i} \cdot \frac{\partial}{\partial \mathbf{R}_i} + \mathbf{v}_{\mathbf{d},i} \cdot \frac{\partial}{\partial \mathbf{d}_i} \right), \tag{2.68}$$

$$i\mathcal{L}_v = \sum_i \left(\frac{\mathbf{F}_{\mathbf{R},i}}{m_i} \cdot \frac{\partial}{\partial \mathbf{v}_{\mathbf{R},i}} + \frac{\mathbf{F}_{\mathbf{d},i}}{m'_i} \cdot \frac{\partial}{\partial \mathbf{v}_{\mathbf{d},i}} \right), \tag{2.69}$$

$$i\mathcal{L}_\eta = \dot{\eta}\frac{\partial}{\partial \eta} + \dot{\eta}_\star \frac{\partial}{\partial \eta_\star} - \sum_i \left(\eta \mathbf{v}_{\mathbf{R},i} \cdot \frac{\partial}{\partial \mathbf{v}_{\mathbf{R},i}} + \eta_\star \mathbf{v}_{\mathbf{d},i} \cdot \frac{\partial}{\partial \mathbf{v}_{\mathbf{d},i}} \right), \tag{2.70}$$

$$i\mathcal{L}_{\dot{\eta}} = \frac{G}{Q}\frac{\partial}{\partial \dot{\eta}} + \frac{G_\star}{Q_\star}\frac{\partial}{\partial \dot{\eta}_\star}. \tag{2.71}$$

Variables $\{\mathbf{v}_\mathbf{R}, \mathbf{v}_\mathbf{d}\}$ are the conjugate velocities of $\{\mathbf{R}, \mathbf{d}\}$.

The velocity-Verlet propagation [see Eq. (2.24)] is

$$e^{i\mathcal{L}_\mathrm{v}\Delta t} \simeq e^{i\mathcal{L}_v \Delta t/2}\, e^{i\mathcal{L}_r \Delta t}\, e^{i\mathcal{L}_v \Delta t/2}. \tag{2.72}$$

Applied to the left of a phase space vector $\mathbf{X}$, this sequence of three differential operators translates into a step of the usual velocity-Verlet algorithm [96]. Over a time interval $\tau = \Delta t/2n_\mathrm{c}$, the Nosé–Hoover propagation [see Eq. (2.24)] is

$$e^{i\mathcal{L}_\mathrm{NH}\tau} \simeq e^{i\mathcal{L}_{\dot{\eta}}\tau/2}\, e^{i\mathcal{L}_\eta \tau}\, e^{i\mathcal{L}_{\dot{\eta}}\tau/2}. \tag{2.73}$$

Applied to $\mathbf{X}$, this sequence can also be translated into a simple algorithm [71].

The overall propagation operator $e^{i\mathcal{L}\Delta t}$ can be written as pseudocode, to be

read from right to left

$$\underbrace{e^{i\mathcal{L}_{\dot\eta}\tau/2}\,G\,e^{i\mathcal{L}_{\eta}\tau}\,e^{i\mathcal{L}_{\dot\eta}\tau/2}\,G}_{\text{repeat } n_c \text{ times}}\,\left(R_v\,e^{i\mathcal{L}_v\Delta t/2}\,F\,R_r\,e^{i\mathcal{L}_r\Delta t}\,e^{i\mathcal{L}_v\Delta t/2}\right)\,\underbrace{e^{i\mathcal{L}_{\dot\eta}\tau/2}\,G\,e^{i\mathcal{L}_{\eta}\tau}\,e^{i\mathcal{L}_{\dot\eta}\tau/2}\,G}_{\text{repeat } n_c \text{ times}} \tag{2.74}$$

R_r and R_v stand for the SHAKE [70] and RATTLE [97] operations on the positions and on the velocities. F stands for the computation of the forces $\{\mathbf{F}\}$, after the update of the positions. G stands for the computation of "forces" G and $G_\star$, before every occurrence of $i\mathcal{L}_{\dot\eta}$.

2.C.2 Andersen–Hoover barostat

The Liouville operator for Eqs. (2.25) to (2.31) is

$$i\mathcal{L} = i\mathcal{L}_r^\varepsilon + i\mathcal{L}_v^\varepsilon + i\mathcal{L}_\eta + i\mathcal{L}_{\dot\eta} + i\mathcal{L}_\varepsilon + i\mathcal{L}_{\dot\varepsilon}, \tag{2.75}$$

where $i\mathcal{L}_r^\varepsilon$ and $i\mathcal{L}_v^\varepsilon$ are extensions of $i\mathcal{L}_r$ and $i\mathcal{L}_v$

$$i\mathcal{L}_r^\varepsilon = \sum_i \left[(\mathbf{v}_{\mathbf{R},i} + \dot\varepsilon\mathbf{R}_i)\cdot\frac{\partial}{\partial\mathbf{R}_i} + (\mathbf{v}_{\mathbf{d},i} + \dot\varepsilon\mathbf{d}_i)\cdot\frac{\partial}{\partial\mathbf{d}_i}\right], \tag{2.76}$$

$$i\mathcal{L}_v^\varepsilon = \sum_i \left[\left(\frac{\mathbf{F}_{\mathbf{R},i}}{m_i} - a\dot\varepsilon\mathbf{R}_i\right)\cdot\frac{\partial}{\partial\mathbf{v}_{\mathbf{R},i}} + \left(\frac{\mathbf{F}_{\mathbf{d},i}}{m'_i} - a\dot\varepsilon\mathbf{d}_i\right)\cdot\frac{\partial}{\partial\mathbf{v}_{\mathbf{d},i}}\right], \tag{2.77}$$

and where

$$i\mathcal{L}_\varepsilon = \dot\varepsilon\frac{\partial}{\partial\varepsilon}, \tag{2.78}$$

$$i\mathcal{L}_{\dot\varepsilon} = \frac{G_\varepsilon}{W}\frac{\partial}{\partial\dot\varepsilon}. \tag{2.79}$$

G_ε corresponds to the right-hand side of Eq. (2.31).

We use a yet unpublished decomposition of the Andersen–Hoover barostat propagator proposed by Tuckerman [73], where the barostat is treated along with the particles, but the thermostats are integrated separately, at the beginning and the

end of each time step. The propagator can be written

$$e^{i\mathcal{L}\Delta t} \simeq \left(e^{i\mathcal{L}_{\mathrm{NH}}\Delta t/2n_c}\right)^{n_c} e^{i\mathcal{L}^{\varepsilon}_{\mathrm{vV}}\Delta t} \left(e^{i\mathcal{L}_{\mathrm{NH}}\Delta t/2n_c}\right)^{n_c}. \tag{2.80}$$

The operator $e^{i\mathcal{L}_{\mathrm{NH}}\tau}$ is propagating the η and $\eta_\star$ thermostats as in Eq. (2.73). The operator $e^{i\mathcal{L}^{\varepsilon}_{\mathrm{vV}}\Delta t}$ is a modified velocity-Verlet propagator that also propagates the barostat ε:

$$e^{i\mathcal{L}^{\varepsilon}_{\mathrm{vV}}\Delta t} \simeq e^{i\mathcal{L}_{\dot{\varepsilon}}\Delta t/2}\, e^{i\mathcal{L}^{\varepsilon}_{v}\Delta t/2}\, e^{i\mathcal{L}^{\varepsilon}_{r}\Delta t}\, e^{i\mathcal{L}_{\varepsilon}\Delta t}\, e^{i\mathcal{L}^{\varepsilon}_{v}\Delta t/2}\, e^{i\mathcal{L}_{\dot{\varepsilon}}\Delta t/2}. \tag{2.81}$$

Applied to $\mathbf{X}$, this sequence can be translated into a simple algorithm.

Constraints on the bond lengths are enforced with the SHAKE/Roll and RATTLE/Roll procedures of Martyna *et al.*[71]. The $e^{i\mathcal{L}^{\varepsilon}_{\mathrm{vV}}\Delta t}$ propagator, applied from the left, is written in the following pseudocode:

$$\left(\underbrace{R^{\varepsilon}_{v}\, e^{i\mathcal{L}_{\dot{\varepsilon}}\Delta t/2}\, G_{\varepsilon}\, e^{i\mathcal{L}_{v}\Delta t/2}\, r}_{\text{iterate until consistent}}\; s\, F\; \underbrace{R^{\varepsilon}_{r}\, e^{i\mathcal{L}_{r}\Delta t}\, e^{i\mathcal{L}_{\varepsilon}\Delta t}\, e^{i\mathcal{L}_{v}\Delta t/2}\, e^{i\mathcal{L}_{\dot{\varepsilon}}\Delta t/2}\, G_{\varepsilon}\, r}_{\text{iterate until consistent}}\; s \right). \tag{2.82}$$

F stands for the computation of the forces $\{\mathbf{F}\}$ and G_ε stands for the computation of "force" G_ε. R^{ε}_{r} and R^{ε}_{v} stand for the SHAKE/Roll and RATTLE/Roll operations, and the braces contain the operations to iterate to self-consistency. s and r stand for "store a copy of the phase vector $\mathbf{X}$" and "recall the copy of $\mathbf{X}$". The $e^{i\mathcal{L}_{\mathrm{NH}}\tau}$ propagator is as described in Appendix 2.C.1.

References

[1] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, 117(19) : 5179–5197, 1995.

[2] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., 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, M. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, 102(18) : 3586–3616, 1998.

[3] W. F. van Gunsteren, X. Daura, and A. E. Mark. The GROMOS force field. In P. von Ragué Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, III, and P. R. Schreiner (editors) : Encyclopedia of Computational Chemistry, volume 2, pages 1211–1216. Wiley & Sons, New York, 1998.

[4] William L. Jorgensen and Julian Tirado-Rives. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.*, 110(6) : 1657–1666, 1988.

[5] Martin Karplus. Special Issue on Molecular Dynamics Simulations of Biomolecules. *Acc. Chem. Res.*, 35 : 321–323, 2002.

[6] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2) : 926–935, 1983.

[7] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. Interaction models for water in relation to protein hydration. In B. Pullman (editor) : Intermolecular Forces, pages 331–342. Reidel, Dordrecht, 1981.

[8] Michael W. Mahoney and William L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, 112(20) : 8910–8922, 2000.

[9] Johan Åqvist. Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *J. Phys. Chem.*, 94 : 8021–8024, 1990.

[10] Dmitrii Beglov and Benoît Roux. Finite representation of an infinite bulk system : Solvent boundary potential for computer simulations. *J. Chem. Phys.*, 100(12) : 9050–9063, 1994.

[11] Terry P. Lybrand and Peter A. Kollman. Water-water and water-ion potential functions including terms for many body effects. *J. Chem. Phys.*, 83(6) : 2923–2933, 1985.

[12] Steven J. Stuart and Bruce J. Berne. Effects of Polarizability on the Hydration of the Chloride Ion. *J. Phys. Chem.*, 100(29) : 11934–11943, 1996.

[13] Douglas J. Tobias, Pavel Jungwirth, and Michele Parrinello. Surface solvation of halogen anions in water clusters : An *ab initio* molecular dynamics study of the $Cl^-(H_2O)_6$ complex. *J. Chem. Phys.*, 114(16) : 7036–7044, 2001.

[14] Liem X. Dang. Computational Study of Ion Binding to the Liquid Interface of Water. *J. Phys. Chem. B*, 106(40) : 10388–10394, 2002.

[15] Benoît Roux and Simon Bernèche. On the Potential Functions used in Molecular Dynamics Simulations of Ion Channels. *Biophys. J.*, 82(3) : 1681–1684, march 2002.

[16] Thomas A. Halgren and Wolfgang Damm. Polarizable force fields. *Curr. Opin. Struct. Biol.*, 11 : 236–242, 2001.

[17] Steven W. Rick and Steven J. Stuart. Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. In Kenny B. Lipkowitz and Donald B. Boyd (editors) : Reviews in Computational Chemistry, volume 18, pages 89–146. Wiley-VCH, Hoboken, NJ, 2002.

[18] Daniel van Belle, Ignace Couplet, Martine Prevost, and Shoshana J. Wodak. Calculations of Electrostatic Properties in Proteins. *Journal of Molecular Biology*, 198 : 721–735, 1987.

[19] Steven W. Rick, Steven J. Stuart, and Bruce J. Berne. Dynamical fluctuating charge force field : Application to liquid water. *J. Chem. Phys.*, 101(7) : 6141–6156, 1994.

[20] Lars Onsager. Electric Moments of Molecules in Liquids. *J. Am. Chem. Soc.*, 58 : 1486–1493, 1936.

[21] J. G. Kirkwood. The dielectric polarization of polar liquids. *J. Chem. Phys.*, 7 : 911–919, 1939.

[22] E. L. Pollock and B. J. Alder. Effective Field of a Dipole in Polarizable Fluids. *Phys. Rev. Lett.*, 39(5) : 299–302, 1977.

[23] P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn. Cooperative effects in simulated water. *Nature*, 282 : 459–464, 1979.

[24] Piotr Cieplak and Peter Kollman. A new water potentiel including polarization : Application to gas-phase, liquid, and crystal properties of water. *J. Chem. Phys.*, 92(11) : 6755–6760, 1990.

[25] P. Cieplak and P. A. Kollman. Monte Carlo simulations of aqueous solutions of Li^+ and Na^+ using many-body potentials. Coordination numbers, ion solvation enthalpies, and the relative free energy of solvation. *J. Chem. Phys.*, 92 : 6761–6767, 1990.

[26] B. T. Thole. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.*, 59 : 341–350, 1981.

[27] Piet Th. van Duijnen and Marcel Swart. Molecular and Atomic Polarizabilities : Thole's Model Revisited. *J. Phys. Chem. A*, 102(14) : 2399–2407, 1998.

[28] Harry A. Stern, George A. Kaminski, Jay L. Banks, Ruhong Zhou, Bruce J. Berne, and Richard A. Friesner. Fluctuating Charge, Polarizable Dipole, and Combined Models : Parameterization from ab Initio Quantum Chemistry. *J. Phys. Chem. B*, 103(22) : 4730–4737, 1999.

[29] Julia M. Goodfellow. Cooperative effects in water-biomolecule crystal systems. *Proc. Natl. Acad. Sci. U.S.A.*, 79 : 4977–4979, 1982.

[30] Zhong-Hui Duan and Robert Krasny. An Ewald summation based multipole method. *J. Chem. Phys.*, 113(9) : 3492–3495, 2000.

[31] Abdulnour Toukmaji, Celeste Nagui, John Board, and Tom Darden. Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. *J. Chem. Phys.*, 113(24) : 10913–10927, 2000.

[32] Piotr Cieplak, James Caldwell, and Peter 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(10) : 1048–1057, 2001.

[33] Roman F. Nalewajski. A Study of Electronegativity Equalization. *J. Phys. Chem.*, 89(13) : 2831–2837, 1985.

[34] Wilfried J. Mortier, Swapan K. Ghosh, and S. Shankar. Electronegativity Equalization Method for the Calculation of Atomic Charges in Molecules. *J. Am. Chem. Soc.*, 108(15) : 4315–4320, 1986.

[35] Anthony K. Rappé and William A. Goddard, III. Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.*, 95(8) : 3358–3363, 1991.

[36] Riccardo Chelli, Piero Procacci, Roberto Righini, and Salvatore Califano. Electrical response in chemical potential equalization schemes. *J. Chem. Phys.*, 111(18) : 8569–8575, 1999.

[37] Jay L. Banks, George A. Kaminski, Ruhong Zhou, Daniel T. Mainz, Bruce J. Berne, and Richard A. Friesner. Parametrizing a polarizable force field from *ab initio* data. I. The fluctuating point charge model. *J. Chem. Phys.*, 110(2) : 741–754, 1999.

[38] Paul Drude. The Theory of Optics. Longmans, Green, and Co., New York, 1902. Translation by C. Riborg Mann and Robert A. Millikan.

[39] F. London. The General Theory of Molecular Forces. *Trans. Faraday Soc.*, 33 : 8–26, 1937.

[40] W. L. Bade. Drude-Model Calculation of Dispersion Forces. I. General Theory. *J. Chem. Phys.*, 27(6) : 1280–1284, 1957.

[41] W. L. Bade and John G. Kirkwood. Drude-Model Calculation of Dispersion Forces. II. The Linear Lattice. *J. Chem. Phys.*, 27(6) : 1284–1288, 1957.

[42] A. T. Amos. Bond Properties Using a Modern Version of the Drude Model. *Int. J. Quantum Chem.*, 60(1) : 67–74, 1996.

[43] F. Wang and K. D. Jordan. Application of a Drude model to the binding of excess electrons to water clusters. *J. Chem. Phys.*, 116(16) : 6973–6981, 2002.

[44] B. G. Dick and A. W. Overhauser. Theory of the Dielectric Constants of Alkali Halide Crystals. *Phys. Rev.*, 112 : 90, 1958.

[45] J. E. Hanlon and A. W. Lawson. Effective Ionic Charge in Alkali Halides. *Phys. Rev.*, 113 : 472–478, 1959.

[46] G. Jacucci, I. R. McDonald, and K. Singer. Introduction of Shell-Model of Ionic Polarizability into Molecular-Dynamics Calculations. *Phys. Lett. A*, 50(2) : 141–143, 1974.

[47] M. J. L Sangster and M. Dixon. Interionic potentials in alkali halides and their use in simulations of the molten salts. *Adv. Phys.*, 25(3) : 247–342, 1976.

[48] P. J. Mitchell and David Fincham. Shell model simulations by adiabatic dynamics. *J. Phys. Cond. Mat.*, 5 : 1031–1038, 1993.

[49] P. J. D. Lindan and M. J. Gillan. Shell-model molecular dynamics simulation of superionic conduction in CaF_2. *J. Phys. Cond. Mat.*, 5 : 1019–1030, 1993.

[50] J. S. Høye and G. Stell. Dielectric theory for polar molecules with fluctuating polarizability. *J. Chem. Phys.*, 73(1) : 461–468, 1980.

[51] Lawrence R. Pratt. Effective field of a dipole in non-polar polarizable fluids. *Mol. Phys.*, 40(2) : 347–360, 1980.

[52] Jianshu Cao and B. J. Berne. Theory of polarizable liquid crystals : Optical birefringence. *J. Chem. Phys.*, 99(3) : 2213–2220, 1993.

[53] F. Lado. Molecular theory of a charged particle in a polarizable nonpolar liquid. *J. Chem. Phys.*, 106(11) : 4707–4713, 1997.

[54] H. Saint-Martin, C. Medina-Llanos, and I. Ortega-Blake. Nonadditivity in an analytical intermolecular potential : The water-water interaction. *J. Chem. Phys.*, 93(9) : 6448–6452, 1990.

[55] N. H. de Leeuw and S. C. Parker. Molecular-dynamics simulation of MgO surfaces in liquid water using a shell-model potential for water. *Phys. Rev. B*, 58(20) : 13901–13908, 1998.

[56] Humberto Saint-Martin, Jorge Hernández-Cobos, Margarita I. Bernal-Uruchurtu, Iván Ortega-Blake, and Herman J. C. Berendsen. A mobile charge densities in harmonic oscillators (MCDHO) molecular model for numerical simulations : The water-water interaction. *J. Chem. Phys.*, 113(24) : 10899–10912, 2000.

[57] Paul J. van Maaren and David van der Spoel. Molecular Dynamics Simulations of Water with Novel Shell-Model Potentials. *J. Phys. Chem. B*, 105(13) : 2618–2626, 2001.

[58] Haibo Yu, Tomas Hansson, and Wilfred F. van Gunsteren. Development of a simple, self-consistent polarizable model for liquid water. *J. Chem. Phys.*, 118(1) : 221–234, 2003.

[59] Steven J. Stuart and B. J. Berne. Surface Curvature Effects in the Aqueous Ionic Solvation of the Chloride Ion. *J. Phys. Chem. A*, 103(49) : 10300–10307, 1999.

[60] Michiel Sprik and Michael L. Klein. A polarizable model for water using distributed charge sites. *J. Chem. Phys.*, 89(12) : 7556–7560, 1988.

[61] Peter Ahlström, Anders Wallqvist, Sven Engström, and Bo Jönsson. A molecular dynamics study of polarizable water. *Mol. Phys.*, 68(3) : 563–581, 1989.

[62] C. J. F. Böttcher. Theory of electric polarization. Volume 1 : Dielectrics in static fields. Elsevier Scientific Publishing Company, Amsterdam, second edition, 1973.

[63] G. Jacucci, I. R. McDonald, and A. Rahman. Effects of polarization on equilibrium and dynamics properties of ionic systems. *Phys. Rev. A*, 13(4) : 1581–1592, 1976.

[64] Michael W. Mahoney and William L. Jorgensen. Rapid estimation of electronic degrees of freedom in Monte Carlo calculations for polarizable models of liquid water. *J. Chem. Phys.*, 114(21) : 9337–9349, 2001.

[65] Daniel van Belle, Matheus Froeyen, Guy Lippens, and Shoshana J. Wodak. Molecular dynamics simulation of polarizable water by an extended Lagrangian method. *Mol. Phys.*, 77(2) : 239–255, 1992.

[66] D. Van Belle and S. J. Wodak. Extended Lagrangian formalism applied to temperature control and electronic polarization effects in molecular dynamics simulations. *Comput. Phys. Comm.*, 91 : 253–262, 1995.

[67] R. W. Pastor, B. R. Brooks, and A. Szabo. An Analysis of the Accuracy of Langevin and Molecular-Dynamics Algorithms. *Mol. Phys.*, 65(6) : 1409–1419, 1988.

[68] William G. Hoover. Canonical dynamics : Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3) : 1695–1697, 1985.

[69] Michiel Sprik. Computer Simulation of the Dynamics of Induced Polarization Fluctuations in Water. *J. Phys. Chem.*, 95(6) : 2283–2291, 1991.

[70] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of Cartesian equations of motion of a system with constraints – Molecular dynamics of *n*-alkanes. *J. Comput. Phys.*, 23(3) : 327–341, 1977.

[71] Glenn J. Martyna, Mark E. Tuckerman, Douglas J. Tobias, and Michael L. Klein. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.*, 87(5) : 1117–1157, 1996.

[72] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, 101(5) : 4177–4189, 1994.

[73] Mark E. Tuckerman. private communication. 2001.

[74] Stephen C. Harvey, Robert K.-Z. Tan, and Thomas E. Cheatham, III. The Flying Ice Cube : Velocity Rescaling in Molecular Dynamics Leads to Violation of Energy Equipartition. *J. Comput. Chem.*, 19(7) : 726–740, 1998.

[75] Glenn J. Martyna, Michael L. Klein, and Mark Tuckerman. Nosé-Hoover chains : The canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97(4) : 2635–2643, 1992.

[76] Mark E. Tuckerman, Bruce J. Berne, Glenn J. Martyna, and Michael L. Klein. Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals. *J. Chem. Phys.*, 99(4) : 2796–2808, 1993.

[77] Douglas J. Tobias, Glenn J. Martyna, and Michael L. Klein. Molecular Dynamics Simulations of a Protein in the Canonical Ensemble. *J. Phys. Chem.*, 97(49) : 12959–12966, 1993.

[78] See-Wing Chiu, Michael Clark, Shankar Subramaniam, and Eric Jakobsson. Collective Motion Artifacts Arising in Long-Duration Molecular Dynamics Simulations. *J. Comput. Chem.*, 21(2) : 121–131, 2000.

[79] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM : A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4 : 187–217, 1983.

[80] Felix Franks (editor). Water : A Comprehensive Treatise, volume 1 : The Physics and Physical Chemistry of Water. Plenum Press, New York, 1972.

[81] J. Verhoeven and A. Dymanus. Magnetic Properties and Molecular Quadrupole Tensor of Water Molecule by Beam-Maser Zeeman Spectroscopy. *J. Chem. Phys.*, 52 : 3222, 1970.

[82] Thomas R. Dyke, Kenneth M. Mack, and J. S. Muenter. The structure of water dimer from molecular beam electric resonance spectroscopy. *J. Chem. Phys.*, 66(2) : 498–510, 1977.

[83] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald : An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12) : 10089–10092, 1993.

[84] M. P. Allen and D. J. Tildesley. Computer Simulation of Liquids. Clarendon Press, Oxford, 1987.

[85] M. Neumann and O. Steinhauser. Computer Simulation and the Dielectric Constant of Polarizable Polar Systems. *Chem. Phys. Lett.*, 106(6) : 563–569, 1984.

[86] Brad Lee Holian and Arthur F. Voter. Thermostatted molecular dynamics : How to avoid the Toda demon hidden in Nosé-Hoover dynamics. *Phys. Rev. E*, 52(3) : 2338–2347, 1995.

[87] Gábor Jancsó and W. Alexander van Hook. Condensed Phase Isotope Effects (Especially Vapor Pressure Isotope Effects). *Chem. Rev.*, 74(6) : 689–750, 1974.

[88] K. Krynicki, C. D. Green, and D. W. Sawyer. Pressure and Temperature-Dependence of Self-Diffusion in Water. *Discuss. Faraday Soc.*, 66 : 199–208, 1978.

[89] D. P. Fernandez, Y. Mulev, A. R. H. Goodwin, and J. M. H. L. Sengers. A Database for the Static Dielectric-Constant of Water and Steam. *J. Phys. Chem. Ref. Data*, 24(1) : 33–69, 1995.

[90] A. K. Soper, F. Bruni, and M. A. Ricci. Site-site pair correlation functions of water from 25 to 400°C : Revised analysis of new and old diffraction data. *J. Chem. Phys.*, 106(1) : 247–254, 1997.

[91] Michiel Sprik. Hydrogen bonding and the static dielectric constant in liquid water. *J. Chem. Phys.*, 95(9) : 6762–6769, 1991.

[92] Jean-Christophe Soetens, Marília T. C. Martins Costa, and Claude Millot. Static dielectric constant of the polarizable NCC water model. *Mol. Phys.*, 94(3) : 577–579, 1998.

[93] Anders Wallqvist and Raymond D. Mountain. Molecular Models of Water : Derivation and Description, volume 13 of Reviews in Computational Chemistry, chapter 4, pages 183–247. Wiley-VCH, New York, 1999.

[94] L. Delle Site, A. Alavi, and R. M. Lynden-Bell. The electrostatic properties of water molecules in condensed phases : an *ab initio* study. *Mol. Phys.*, 96(11) : 1683–1693, 1999.

[95] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski. Methods of Quantum Field Theory in Statistical Physics. Dover Publications, Inc., New York, 1975.

[96] W.C. Swope, H.C. Andersen, P.H. Berens, and K.R. Wilson. A computer-simulation method for the calculation of equilibrium-constants for the formation of physical clusters of molecules – Application to small water clusters. *J. Chem. Phys.*, 76(1) : 637–649, 1982.

[97] H. C. Andersen. Rattle : A "velocity" version of the Shake algorithm for molecular dynamics calculations. *J. Comput. Phys.*, 52(1) : 24–34, 1983.

Chapitre 3

Intégration numérique des équations du mouvement

L'annexe C du chapitre 2 présente une bonne quantité de détails sur l'intégration numérique des équations du mouvement mais ne fournit pas de vue d'ensemble sur le sujet. Ce chapitre présente de façon plus systématique et plus abordable les idées essentielles derrière les méthodes modernes d'intégration numérique des équations de la dynamique moléculaire.

3.1 Le problème de base

La dynamique d'un ensemble de N particules est habituellement décrite par les équations de Newton :

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i. \tag{3.1}$$

Sauf dans certains cas particuliers concernant peu la biologie, il s'agit d'un ensemble d'équations différentielles non-intégrables analytiquement. Bien qu'on ne puisse pas obtenir une forme analytique des trajectoires $\{\mathbf{r}(t)\}$ à partir des positions initiales $\{\mathbf{r}(0)\}$ et des vitesses initiales $\{\dot{\mathbf{r}}(0)\}$, on peut produire une série de positions $\{\mathbf{r}(\delta t)$, $\mathbf{r}(2\delta t)$, $\mathbf{r}(3\delta t)$, $\ldots\}$ qui obéissent à des équations aux différences finies équivalentes aux équations de Newton lorsque $\delta t \to 0$. Par exemple, on pourrait générer des

trajectoires discrètes à partir des équations

$$m_i \frac{\mathbf{r}_i(t+\delta t) - 2\mathbf{r}_i(t) + \mathbf{r}_i(t-\delta t)}{\delta t^2} = \mathbf{F}_i(t). \tag{3.2}$$

On calculerait alors $\mathbf{r}_i(t+\delta t)$ à partir de $\mathbf{r}_i(t)$ et de $\mathbf{r}_i(t-\delta t)$:

$$\mathbf{r}_i(t+\delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\delta t) + \delta t^2 \frac{\mathbf{F}_i(t)}{m_i}. \tag{3.3}$$

Cette dernière équation correspond en fait à la formule de Verlet [1], utilisée pour la première fois en 1967 pour l'intégration numérique du mouvement d'un grand nombre de particules. La première simulation de dynamique moléculaire a été réalisée par Rahman en 1964 [2], et elle utilisait un algorithme de prédiction-correction.

Évidemment, la même approche peut être tentée à partir des équations de Hamilton (exprimées ici en coordonnées cartésiennes) :

$$\dot{\mathbf{r}}_i = \frac{\partial H}{\partial \mathbf{p}_i}, \tag{3.4}$$

$$\dot{\mathbf{p}}_i = -\frac{\partial H}{\partial \mathbf{r}_i}. \tag{3.5}$$

Pour un hamiltonien de la forme

$$H = U(\{\mathbf{r}\}) + \sum_i \frac{p_i^2}{2m_i}, \tag{3.6}$$

les équations de Hamilton (3.4) et (3.5) sont équivalentes aux équations de Newton (3.1), avec $\mathbf{F}_i = -\partial U/\partial \mathbf{r}_i$. On pourrait par exemple générer les trajectoires discrètes avec les équations aux différences finies suivantes :

$$\frac{\mathbf{r}_i(t+\delta t/2) - \mathbf{r}_i(t-\delta t/2)}{\delta t} = \mathbf{v}_i(t), \tag{3.7}$$

$$\frac{\mathbf{v}_i(t+\delta t/2) - \mathbf{v}_i(t-\delta t/2)}{\delta t} = \frac{\mathbf{F}_i(t)}{m_i}. \tag{3.8}$$

En remplaçant t par $t+\delta t/2$ dans la première équation, on peut calculer $\mathbf{r}_i(t+\delta t)$

à partir de $\mathbf{r}_i(t)$ et de $\mathbf{v}_i(t+\delta t/2)$, et $\mathbf{v}_i(t+\delta t/2)$ à partir de $\mathbf{v}_i(t-\delta t/2)$:

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \delta t \mathbf{v}_i(t+\delta t/2), \tag{3.9}$$

$$\mathbf{v}_i(t+\delta t/2) = \mathbf{v}_i(t-\delta t/2) + \delta t \frac{\mathbf{F}_i(t)}{m_i}. \tag{3.10}$$

Ces équations correspondent aux formules *leap-frog* [3], elles-aussi très couramment utilisées pour la dynamique moléculaire.

Le problème de base de la dynamique moléculaire consiste à trouver un ensemble d'équations aux différences finies qui produise des trajectoires discrètes $\{\mathbf{r}(0), \mathbf{r}(\delta t), \mathbf{r}(2\delta t), \ldots\}$ *proches* des trajectoires exactes $\{\mathbf{r}(t)\}$. Comme le mouvement d'un grand nombre de particules couplées est généralement chaotique, il est plus juste de dire qu'on cherche à produire des trajectoires discrètes *similaires* aux trajectoires exactes : si petit soit δt, il viendra un temps où les trajectoires discrètes divergeront des trajectoires exactes.

3.2 L'oscillateur harmonique

Dans la section précédente, nous avons sciemment veillé à ce que les équations aux différences finies (3.2), (3.7) et (3.8) soient réversibles dans le temps. Avec l'exemple habituel de l'oscillateur harmonique, nous montrerons maintenant que cette précaution nous assure que les équations conservent l'énergie totale du système.

L'hamiltonien d'un oscillateur harmonique unidimensionnel est $H = \frac{1}{2}kx^2 + \frac{1}{2}p^2/m$, les équations de Hamilton sont $\dot{x} = p/m$ et $\dot{p} = -kx$, et l'équation de Newton correspondante est $m\ddot{x} = -kx$. Pour les conditions initiales $x(0) = 0$ et $v(0) = v_0$, la trajectoire exacte est

$$x(t) = \frac{v_0}{\omega}\sin(\omega t), \tag{3.11}$$

avec $\omega \equiv \sqrt{k/m}$. La période de l'oscillation est $\tau = 2\pi/\omega$ et son amplitude est $A = v_0/\omega$. L'énergie totale de l'oscillateur est $\frac{1}{2}mv_0^2$. Une méthode d'intégration

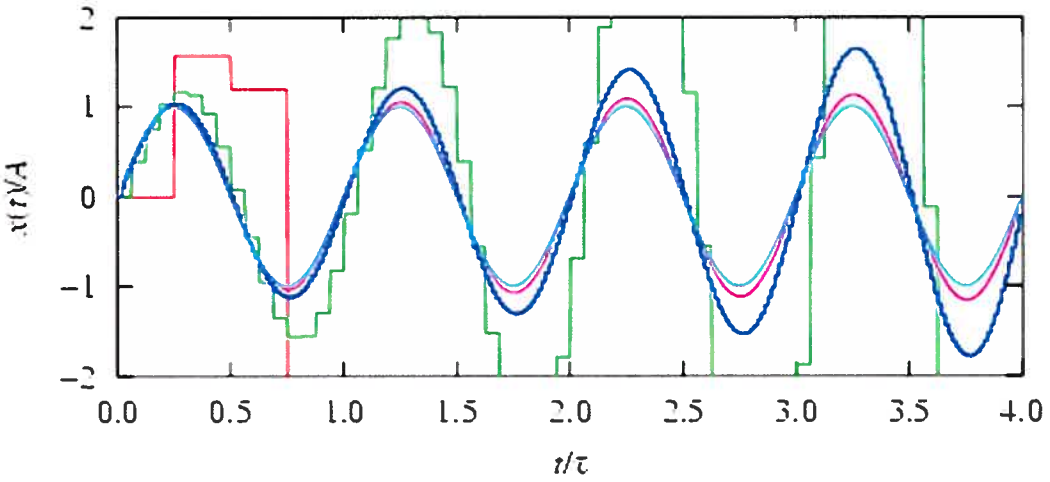

FIG. 3.1 – Trajectoires discrètes générées avec l'algorithme « naif » des équations (3.12) et (3.13), pour des pas d'intégration δt de plus en plus petits : $\delta t = \tau/4$ en rouge, $\tau/16$ en vert, $\tau/64$ en bleu marine et $\tau/256$ en rose. La courbe cyan est la trajectoire exacte.

naïve consisterait à utiliser les équations du mouvement uniformément accéléré :

$$x(\delta t) = x(0) + v(0)\delta t + \tfrac{1}{2}\frac{F(0)}{m}\delta t^2, \tag{3.12}$$

$$v(\delta t) = v(0) + \frac{F(0)}{m}\delta t. \tag{3.13}$$

La Figure 3.1 présente les trajectoires obtenues avec cette approche pour des pas d'intégration δt de longueurs variées. Il ressort clairement que, malgré que des pas très courts produisent une trajectoire proche de la trajectoire exacte, l'amplitude de l'oscillation va toujours en augmentant. L'énergie totale du système n'est visiblement pas conservée.

À l'opposé, on utilise les formules *velocity-Verlet* [4] :

$$v(\delta t/2) = v(0) + \tfrac{1}{2}\frac{F(0)}{m}\delta t, \tag{3.14}$$

$$x(\delta t) = x(0) + v(\delta t/2)\delta t, \tag{3.15}$$

$$v(\delta t) = v(\delta t/2) + \tfrac{1}{2}\frac{F(\delta t)}{m}\delta t. \tag{3.16}$$

Ces équations sont réversibles dans le temps et produisent des trajectoires stables

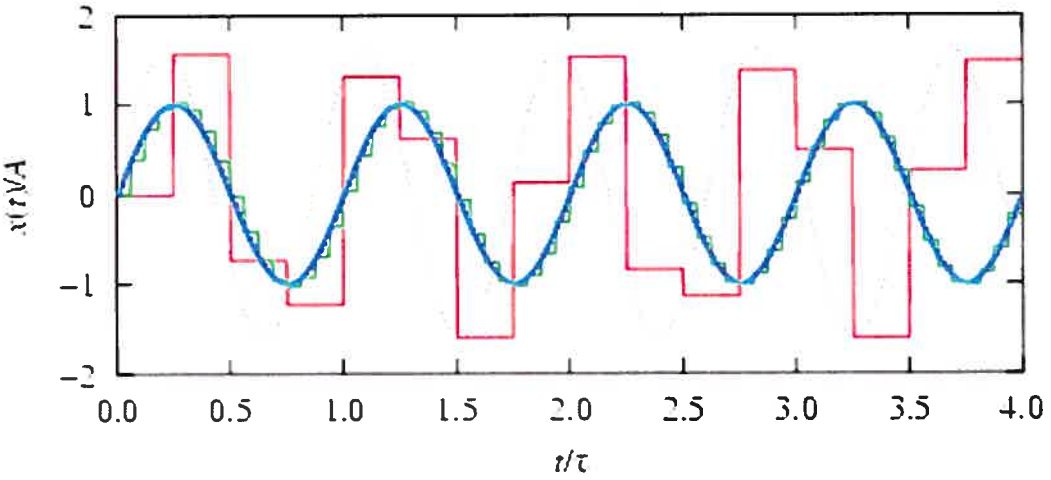

FIG. 3.2 – Trajectoires discrètes générées avec l'algorithme *velocity-Verlet*, pour des pas d'intégration de plus en plus petits. La courbe cyan est la trajectoire exacte. La sinusoïde en ligne tiretée est ajustée sur la trajectoire produite avec $\delta t = \tau/4$ (en rouge).

même pour des pas d'intégration δt moins d'un ordre de grandeur plus courts que la période naturelle de l'oscillateur (voir Figure 3.2, notamment la trajectoire en rouge). Plus intéressant encore est le fait qu'on peut faire passer presque exactement par les points de la trajectoire discrète une trajectoire continue similaire à la trajectoire exacte : une sinusoïde de fréquence et d'amplitude légèrement différentes. En d'autres termes, malgré qu'elles utilisent un pas d'intégration fini, les formules *velocity-Verlet* intègrent précisément un hamiltonien *modifié* $H'(\delta t)$ similaire au hamiltonien original H [5]. Et la fréquence modifiée $\omega'(\delta t)$ se rapproche rapidement de la valeur originale ω lorsque le pas d'intégration raccourcit.

3.3 Formalisme de l'opérateur d'évolution

Le formalisme de l'opérateur d'évolution a été appliqué à la dynamique moléculaire par Tuckerman, Berne et Martyna [6], sur la base de l'idée que l'évolution temporelle d'un système peut être formellement représentée par un opérateur de propagation dans le temps (voir par exemple les sections 9–4 et 9–5 de la référence [7]) et que, selon le théorème de Trotter [8, 9, 10, 11], cet opérateur peut

être décomposé avec une précision arbitrairement grande. La présentation que nous en faisons ici est légèrement différente de celle de Tuckerman *et al.*, et insiste davantage sur les notions d'*équation différentielle modifiée* [12, 5] et de *hamiltonien modifié* [13, 14].

3.3.1 Intégration d'un hamiltonien

Les équations de Hamilton peuvent être résolues de façon formelle en introduisant l'opérateur de Liouville iL (ou *liouvillien*), défini comme le crochet de Poisson avec l'hamiltonien H, fonction des variables canoniques $\{q\}$ et $\{p\}$:

$$iL \equiv \{\cdots, H\} = \sum_i \left(\frac{\partial H}{\partial p_i} \frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial}{\partial p_i} \right). \tag{3.17}$$

Rappelons que le crochet de Poisson, l'analogue classique du commutateur quantique, est défini comme

$$\{A, B\} \equiv \sum_i \left(\frac{\partial A}{\partial q_i} \frac{\partial B}{\partial p_i} - \frac{\partial B}{\partial q_i} \frac{\partial A}{\partial p_i} \right). \tag{3.18}$$

Représentant l'état du système par le vecteur $\mathbf{\Gamma} \equiv (\{q\}, \{p\})$ appartenant à l'espace des phases, les équations de Hamilton (3.4) et (3.5) s'écrivent alors sous la forme compacte suivante :

$$\dot{\mathbf{\Gamma}} = iL\mathbf{\Gamma}. \tag{3.19}$$

On peut formellement écrire la solution des équations du mouvement

$$\mathbf{\Gamma}(t) = e^{iLt}\mathbf{\Gamma}(0). \tag{3.20}$$

L'opérateur différentiel e^{iLt} correspond à un propagateur dans le temps, et représente simplement la série de Taylor suivante :

$$\mathbf{\Gamma}(t) = \mathbf{\Gamma}(0) + t \left. \frac{\partial \mathbf{\Gamma}}{\partial t} \right|_0 + \frac{t^2}{2!} \left. \frac{\partial^2 \mathbf{\Gamma}}{\partial t^2} \right|_0 + \frac{t^3}{3!} \left. \frac{\partial^3 \mathbf{\Gamma}}{\partial t^3} \right|_0 + \cdots. \tag{3.21}$$

Comme Goldstein le fait remarquer au dernier paragraphe de la section 9–5 de son *Classical Mechanics* [7], l'équation (3.20) ne représente pas un véritable progrès pour solutionner les équations du mouvement d'un système de façon analytique. Sa forme exponentielle n'est à toutes fins pratiques qu'une réécriture de la triviale équation (3.21). Elle présente néanmoins un point de vue intéressant sur l'opération formelle de faire avancer un système dans le temps. Il est possible de décomposer l'intégration de 0 à t en une série de n *pas* de longueur $\delta t = t/n$:

$$e^{iLt} = \left(e^{iL\delta t}\right)^n . \tag{3.22}$$

L'opération de e^{iLt} sur un vecteur d'état $\Gamma(0)$ peut donc en principe être résolue pas à pas, mais exige d'écrire une expression calculable pour l'opération de $e^{iL\delta t}$. Évidemment, il n'est pas plus facile de trouver une formule *exacte* pour $e^{iL\delta t}$ qu'il ne l'est pour e^{iLt}. Par contre, une formule *approximative* sera plus précise décomposée en n pas qu'utilisée globalement. Autrement dit, on peut obtenir une erreur ϵ arbitrairement petite en décomposant une formule approximative $f(t)$ en un nombre de pas n arbitrairement grand :

$$e^{iLt} = [f(\delta t)]^n + \epsilon. \tag{3.23}$$

La forme retenue pour l'expression $f(\delta t)$ aura d'importantes conséquences pratiques : on se souviendra que l'approche naïve des équations (3.12) et (3.13) et l'approche de *velocity-Verlet* des équations (3.14) à (3.16) sont toutes les deux valides pour n suffisamment grand, mais que la forme de la trajectoire produite par l'algorithme *velocity-Verlet* est similaire à celle de la trajectoire exacte pour des pas δt relativement longs.

L'opérateur iL est divisé en deux parties iL_q et iL_p, avec

$$iL_q = \sum_i \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i} \qquad \text{et} \qquad iL_p = -\sum_i \frac{\partial H}{\partial q_i}\frac{\partial}{\partial p_i}. \tag{3.24}$$

Malgré que les opérateurs iL_q et iL_p ne commutent pas, il est possible de factoriser

approximativement le propagateur $e^{iL\delta t}$ au moyen du développement de Baker–Campbell–Hausdorff. Pour deux variables non-commutatives A et B, les premiers termes de ce développement sont

$$e^{A\delta t}e^{B\delta t} = \exp\left\{\delta t\left[A+B+\tfrac{1}{2}\delta t[A,B]+\tfrac{1}{12}\delta t^2[(A-B),[A,B]]+\cdots\right]\right\}. \quad (3.25)$$

En interchangeant A et B, on observe une intéressante alternance de symétrie :

$$e^{B\delta t}e^{A\delta t} = \exp\left\{\delta t\left[A+B-\tfrac{1}{2}\delta t[A,B]+\tfrac{1}{12}\delta t^2[(A-B),[A,B]]-\cdots\right]\right\}. \quad (3.26)$$

On peut regrouper chacun des termes du développement en une partie symétrique F_s qui ne change pas de signe sous permutation de A et B, et une partie antisymétrique F_a qui change de signe :

$$F_\mathrm{s}(\delta t) \equiv A+B+\tfrac{1}{12}\delta t^2[(A-B),[A,B]]+\cdots, \quad (3.27)$$

$$F_\mathrm{a}(\delta t) \equiv \tfrac{1}{2}\delta t[A,B]+\cdots. \quad (3.28)$$

On peut alors écrire le produit des équations (3.25) et (3.26) sous la forme

$$e^{B\delta t/2}e^{A\delta t}e^{B\delta t/2} = e^{(F'_\mathrm{s}-F'_\mathrm{a})\delta t/2}e^{(F'_\mathrm{s}+F'_\mathrm{a})\delta t/2}. \quad (3.29)$$

Pour que le produit représente un seul pas d'intégration, δt a été remplacé par $\delta t/2$. Pour raccourcir la notation, on a défini $F'_\mathrm{s} \equiv F_\mathrm{s}(\delta t/2)$ et $F'_\mathrm{a} \equiv F_\mathrm{a}(\delta t/2)$. En appliquant la formule de Baker–Campbell–Hausdorff une seconde fois, au terme de droite de l'équation précédente, on obtient

$$e^{B\delta t/2}e^{A\delta t}e^{B\delta t/2} = \exp\left\{\delta t\left[F'_\mathrm{s}+\tfrac{1}{2}\delta t[F'_\mathrm{s},F'_\mathrm{a}]+\tfrac{1}{12}\delta t^2[[F'_\mathrm{s},F'_\mathrm{a}],F'_\mathrm{a}]+\cdots\right]\right\}. \quad (3.30)$$

Si A représente iL_q et B représente iL_p, on a directement

$$e^{iL_p\delta t/2}e^{iL_q\delta t}e^{iL_p\delta t/2} = \exp\left\{\delta t\left[iL+iL_2(\delta t^2)+\mathcal{O}(\delta t^4)\right]\right\}, \quad (3.31)$$

avec

$$\begin{aligned} iL_2(\delta t^2) &= \tfrac{1}{48}\delta t^2[(iL_q - iL_p),[iL_q, iL_p]] + \tfrac{1}{16}\delta t^2[iL,[iL_q, iL_p]] \\ &= \tfrac{1}{24}\delta t^2[(2iL_q + iL_p),[iL_q, iL_p]]. \end{aligned} \tag{3.32}$$

En d'autres termes, la séquence de trois opérateurs du membre gauche de l'équation (3.31) correspond à un propagateur $e^{iL'\delta t}$, où iL' correspond au liouvillien original iL modifié par un terme iL_2 dépendant du carré du pas d'intégration, puis par un terme $\mathcal{O}(\delta t^4)$. On remarque aussi que, de par sa construction à symétrie bilatérale, ce propagateur est réversible dans le temps :

$$\Gamma(\delta t) = e^{iL_p\delta t/2}e^{iL_q\delta t}e^{iL_p\delta t/2}\Gamma(0), \tag{3.33}$$

$$e^{-iL_p\delta t/2}e^{-iL_q\delta t}e^{-iL_p\delta t/2}\Gamma(\delta t) = \Gamma(0). \tag{3.34}$$

Pour un hamiltonien de la forme (3.6), l'opérateur de Liouville s'écrit

$$iL = \sum_i \left(\dot{\mathbf{r}}_i \cdot \frac{\partial}{\partial \mathbf{r}_i} + \frac{\mathbf{F}_i}{m_i} \cdot \frac{\partial}{\partial \mathbf{v}_i}\right) = iL_r + iL_v, \tag{3.35}$$

avec

$$iL_r = \sum_i \dot{\mathbf{r}}_i \cdot \frac{\partial}{\partial \mathbf{r}_i} \qquad \text{et} \qquad iL_v = \sum_i \frac{\mathbf{F}_i}{m_i} \cdot \frac{\partial}{\partial \mathbf{v}_i}. \tag{3.36}$$

Sachant que

$$e^{a(\partial/\partial x)}x = \sum_{n=0}^{\infty} \frac{a^n}{n!}\frac{\partial^n}{\partial x^n}x = x + a, \tag{3.37}$$

on peut calculer l'effet de la séquence d'opérations $e^{iL_v\delta t/2}e^{iL_r\delta t}e^{iL_v\delta t/2}$ sur le vecteur d'état $\mathbf{\Gamma} = (\{\mathbf{r}\},\{\mathbf{v}\})$ par étapes. L'effet de l'opérateur $e^{iL_v\delta t/2}$ sur le vecteur $(\{\mathbf{r}(0)\},\{\mathbf{v}(0)\})$ est

$$e^{iL_v\delta t/2}\begin{bmatrix} \{\mathbf{r}(0)\} \\ \{\mathbf{v}(0)\} \end{bmatrix} = \begin{bmatrix} \{\mathbf{r}(0)\} \\ \{\mathbf{v}(0) + \tfrac{1}{2}\delta t\mathbf{F}(0)/m\} \end{bmatrix} \equiv \begin{bmatrix} \{\mathbf{r}(0)\} \\ \{\mathbf{v}(\delta t/2)\} \end{bmatrix}. \tag{3.38}$$

L'effet de l'opérateur $e^{iL_r\delta t}$ sur le vecteur $(\{\mathbf{r}(0)\}, \{\mathbf{v}(\delta t/2)\})$ est

$$e^{iL_r\delta t}\begin{bmatrix} \{\mathbf{r}(0)\} \\ \{\mathbf{v}(\delta t/2)\} \end{bmatrix} = \begin{bmatrix} \{\mathbf{r}(0) + \delta t\mathbf{v}(\delta t/2)\} \\ \{\mathbf{v}(\delta t/2)\} \end{bmatrix} \equiv \begin{bmatrix} \{\mathbf{r}(\delta t)\} \\ \{\mathbf{v}(\delta t/2)\} \end{bmatrix}. \tag{3.39}$$

Finalement, la seconde application de l'opérateur $e^{iL_v\delta t/2}$ donne

$$e^{iL_v\delta t/2}\begin{bmatrix} \{\mathbf{r}(\delta t)\} \\ \{\mathbf{v}(\delta t/2)\} \end{bmatrix} = \begin{bmatrix} \{\mathbf{r}(\delta t)\} \\ \{\mathbf{v}(\delta t/2) + \frac{1}{2}\delta t\mathbf{F}(\delta t)/m\} \end{bmatrix} \equiv \begin{bmatrix} \{\mathbf{r}(\delta t)\} \\ \{\mathbf{v}(\delta t)\} \end{bmatrix}. \tag{3.40}$$

Les forces $\{\mathbf{F}(\delta t)\}$ auront été calculées avec les nouvelles positions $\{\mathbf{r}(\delta t)\}$. On reconnaît les trois opérations de l'algorithme *velocity-Verlet*, présenté aux équations (3.14) à (3.16).

3.3.2 Hamiltonien modifié

L'opérateur iL_2 de l'équation (3.32) possède toutes les propriétés d'un crochet de Poisson, et il existe un hamiltonien $H_2(\delta t^2)$ tel que $iL_2 = \{\cdots, H_2\}$. Le crochet de Poisson vérifie notamment l'identité de Jacobi :

$$\{A, \{B, C\}\} + \{C, \{A, B\}\} + \{B, \{C, A\}\} = 0, \tag{3.41}$$

où A, B et C sont des variables non-commutatives quelconques. En utilisant la notation $D_A \equiv \{\cdots, A\}$, on peut réexprimer cette identité sous une forme pour laquelle la troisième variable (C) est implicite :

$$[D_A, D_B] = D_{\{B,A\}}. \tag{3.42}$$

Le liouvillien original peut s'écrire $iL = D_H$ et, pour un hamiltonien de la forme séparable $H = U(\{q\}) + K(\{p\})$, on peut aussi écrire $iL_q = D_K$ et $iL_p = D_U$. L'opérateur iL_2 peut être reformulé de la même façon :

$$iL_2 = \frac{1}{24}\delta t^2[(2D_K + D_U), [D_K, D_U]]$$

$$= \frac{1}{24}\delta t^2 [D_{2K+U}, D_{\{U,K\}}]$$
$$= \frac{1}{24}\delta t^2 D_{\{\{U,K\},2K+U\}}. \tag{3.43}$$

Par comparaison directe, on a

$$H_2 = \frac{1}{24}\delta t^2 \{\{U,K\}, 2K+U\}$$
$$= \frac{1}{24}\delta t^2 \sum_{i,j} \left(2\frac{\partial K}{\partial p_i}\frac{\partial^2 U}{\partial q_i \partial q_j}\frac{\partial K}{\partial p_j} - \frac{\partial U}{\partial q_i}\frac{\partial^2 K}{\partial p_i \partial p_j}\frac{\partial U}{\partial q_j} \right). \tag{3.44}$$

Comme les termes du développement de Baker–Campbell–Hausdorff sont calculables à partir d'algorithmes symboliques relativement simples [15, 16], le développement du hamiltonien modifié peut être continué à un ordre quelconque avec un programme de calcul symbolique tel que MAPLE, MATHEMATICA ou MATLAB. On peut ainsi analyser la fiabilité de tout algorithme d'intégration en calculant les termes successifs du développement du hamiltonien modifié $H' = H + H_1(\delta t) + H_2(\delta t^2) + H_3(\delta t^3) + \cdots$. À ce sujet, nous renvoyons le lecteur aux références [12, 17, 5, 13, 14].

On peut revisiter l'exemple de l'oscillateur harmonique présenté à la section 3.2 pour comprendre les déviations systématiques de la solution numérique obtenue avec l'algorithme *velocity-Verlet* (voir la figure 3.2). Pour l'oscillateur harmonique, l'hamiltonien est $H = \frac{1}{2}kx^2 + \frac{1}{2}p^2/m$ et, suivant l'équation (3.44), le premier terme du développement du hamiltonien modifié est

$$H_2(\delta t) = \tfrac{1}{2}\left(-\tfrac{1}{12}\delta t^2 k/m\right) kx^2 + \tfrac{1}{2}\left(\tfrac{1}{6}\delta t^2 k/m\right) p^2/m. \tag{3.45}$$

La somme $H + H_2$ correspond au hamiltonien d'un oscillateur harmonique modifié $H' = \frac{1}{2}k'x^2 + \frac{1}{2}p^2/m'$, dont la fréquence est

$$\omega'(\delta t) = \sqrt{k'/m'} = \omega\sqrt{\left(1 - \tfrac{1}{12}\delta t^2\omega^2\right)\left(1 + \tfrac{1}{6}\delta t^2\omega^2\right)}. \tag{3.46}$$

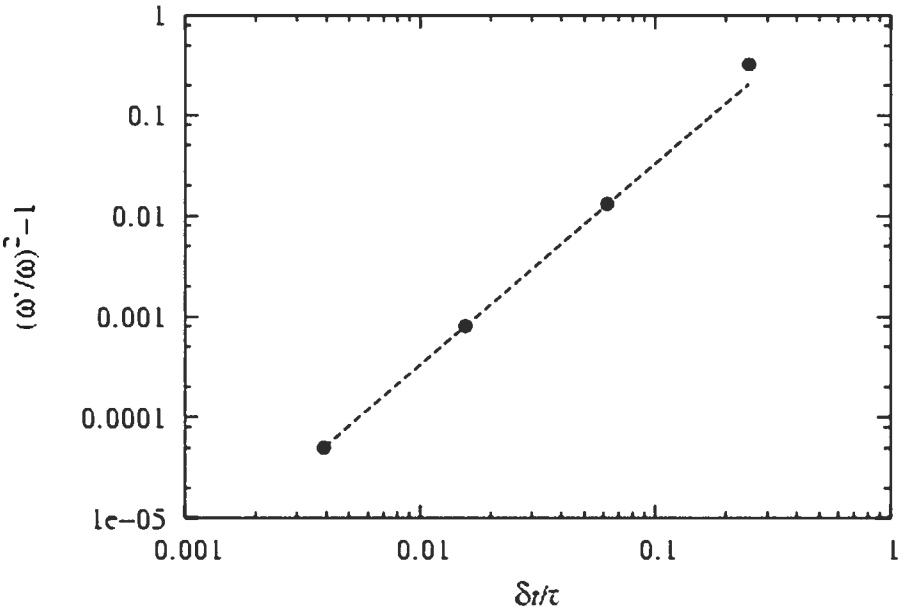

FIG. 3.3 – Fréquences modifiées $\omega'(\delta t)$ pour l'algorithme *velocity-Verlet*. On présente en fait la combinaison $(\omega'/\omega)^2 - 1$. La droite tiretée correspond à $\frac{1}{12}\delta t^2\omega^2$, le premier terme de la correction H_2.

À l'ordre δt^2, la différence relative des carrés des fréquences est

$$(\omega'/\omega)^2 - 1 = \tfrac{1}{12}\delta t^2\omega^2. \tag{3.47}$$

C'est cette fonction qui est présentée à la figure 3.3. Les fréquences modifiées $\omega'(\delta t)$ sont obtenues par l'ajustement d'une sinusoïde à chaque trajectoire discrète de la figure 3.2. On constate que le terme H_2 explique l'écart par rapport à la fréquence véritable pour des pas d'intégration aussi élevés que un dixième de la période naturelle d'oscillation. Pour des pas d'intégration plus longs, il faudrait considérer les termes H_4, H_6 et suivants.

On comprend le problème rencontré par l'approche d'intégration naïve à la section 3.2 en retournant à l'équation (3.25) et en y remplaçant A par iL_q et B par iL_p :

$$e^{iL_q\delta t}e^{iL_p\delta t} = \exp\left\{\delta t\left[iL + iL_1(\delta t) + \mathcal{O}(\delta t^2)\right]\right\}, \tag{3.48}$$

avec

$$iL_1 = \tfrac{1}{2}\delta t[iL_q, iL_p]. \tag{3.49}$$

L'hamiltonien correspondant au générateur iL_1 est

$$\begin{aligned} H_1 &= \tfrac{1}{2}\delta t\{U, K\} \\ &= \tfrac{1}{2}\delta t \sum_i \frac{\partial U}{\partial q_i}\frac{\partial K}{\partial p_i}. \end{aligned} \tag{3.50}$$

Pour l'oscillateur harmonique unidimensionnel intégré avec une telle formule, on a

$$H_1 = \tfrac{1}{2}\delta t(k/m)xp. \tag{3.51}$$

L'hamiltonien modifié $H + H_1$ possède donc une source d'énergie, sous forme d'une friction *négative*, qui fait croitre indéfiniment l'amplitude d'oscillation. Malgré que l'opérateur $e^{iL_x\delta t}e^{iL_v\delta t}$ ne traduise pas exactement l'approche naïve des équations (3.12) et (3.13), on observe un comportement semblable pour leur solution numérique (voir la figure 3.1). [Pour reproduire exactement les équations (3.12) et (3.13), il faudrait appliquer la *somme* de facteurs $e^{iL_x\delta t}e^{iL_v\delta t/2} + e^{iL_v\delta t/2} - 1$, opérateur qui ne se prête pas facilement à l'analyse.]

3.3.3 Stabilité et précision

L'efficacité de l'algorithme *velocity-Verlet*—de tout algorithme de construction semblable, en fait—est ultimement limitée par un phénomène de résonance [18, 19] qui déstabilise les trajectoires discrètes $\{\mathbf{r}(0), \mathbf{r}(\delta t), \mathbf{r}(2\delta t), \ldots\}$ si δt est supérieur à une valeur critique.

Analysons la stabilité de l'intégrateur *velocity-Verlet* en utilisant encore une fois l'exemple de l'oscillateur harmonique. Pour un oscillateur harmonique unidimensionnel de fréquence naturelle $\omega = \sqrt{k/m}$, une application de l'algorithme *velocity-Verlet* correspond à la formule de récurrence suivante [18] (écrite sous

forme matricielle) :

$$\begin{aligned}\begin{bmatrix} \omega x(t+\delta t) \\ v(t+\delta t) \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ -\frac{1}{2}\omega\delta t & 1 \end{bmatrix} \begin{bmatrix} 1 & \omega\delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2}\omega\delta t & 1 \end{bmatrix} \begin{bmatrix} \omega x(t) \\ v(t) \end{bmatrix} \\ &= \begin{bmatrix} 1-\frac{1}{2}\omega^2\delta t^2 & \omega\delta t \\ -\omega\delta t\left(1-\frac{1}{4}\omega^2\delta t^2\right) & 1-\frac{1}{2}\omega^2\delta t^2 \end{bmatrix} \begin{bmatrix} \omega x(t) \\ v(t) \end{bmatrix} \end{aligned} \tag{3.52}$$

Il s'agit d'un cas particulier des transformations des équations (3.14) à (3.16), simplifié du fait que $F = -kx$. Les trois étapes élémentaires de l'algorithme sont composées en une seule application linéaire, que nous appellerons S. Pour que tous les éléments de matrice soient sans unités, nous avons converti la position x en unités de vitesse, sous la forme ωx. Le système possède une dynamique stable sous itération de l'application linéaire S si les valeurs propres λ de celle-ci ont une amplitude inférieure à 1. Les deux valeurs propres sont les racines du polynôme caractéristique $\det(S - \lambda\mathbf{1})$:

$$\lambda = 1 - \tfrac{1}{2}\omega^2\delta t^2 \pm i\tfrac{1}{4}\omega\delta t\sqrt{1+\tfrac{1}{4}\omega^2\delta t^2}. \tag{3.53}$$

On montre facilement que la condition de stabilité $|\lambda| < 1$ est vérifiée lorsque $\omega\delta t < 2$. On pourrait donc en principe utiliser un pas d'intégration aussi long que τ/π, la période naturelle d'oscillation divisée par π. Par comparaison, la courbe rouge de la figure 3.2 correspond à un pas d'intégration $\delta t = \tau/4$, légèrement plus court que la limite de stabilité τ/π. La courbe verte, considérablement plus précise que la rouge, correspond à un pas d'intégration $\tau/16$. Pour des simulations de processus de nature *diffusive* (« diffusif » par opposition à « ballistique »), la prescription d'usage est de choisir un pas d'intégration vingt fois plus court que la période d'oscillation du mouvement le plus rapide du système [20, chapitre 3, section V–B].

Suivant une méthode récursive imaginée indépendamment par Yoshida [21] et Suzuki [22], on peut réduire indéfiniment l'erreur sur le développement d'un pas d'intégration $e^{(A+B)\delta t}$ en composant des produits d'opérateurs non-commutatifs de

complexité croissante :

$$
\begin{aligned}
e^{(A+B)\delta t} &= 1+\mathcal{O}(\delta t) && (3.54)\\
&= \underbrace{e^{B\delta t/2}e^{A\delta t}e^{B\delta t/2}}_{P_3(\delta t)}+\mathcal{O}(\delta t^3) && (3.55)\\
&= \underbrace{P_3(w_3\delta t)P_3[(1-2w_3)\delta t]P_3(w_3\delta t)}_{P_5(\delta t)}+\mathcal{O}(\delta t^5) && (3.56)\\
&= \underbrace{P_5(w_5\delta t)P_5[(1-2w_5)\delta t]P_5(w_5\delta t)}_{P_7(\delta t)}+\mathcal{O}(\delta t^7), && (3.57)
\end{aligned}
$$

où $w_3 = 1/(2-\sqrt[3]{2})$, $w_5 = 1/(2-\sqrt[5]{2})$, et ainsi de suite. Cette méthode tire avantage des propriétés de récursion du développement de Baker–Campbell–Hausdorff, et réduit le problème de trouver un propagateur P_{2n+1} correct à l'ordre $2n+1$ à celui de trouver le paramètre w_{2n+1} qui annulle exactement l'erreur $\mathcal{O}(\delta t^{2n-1})$. Comme les propagateurs sont réversibles par construction, l'erreur $\mathcal{O}(\delta t^{2n})$ est toujours nulle et seule subsiste une erreur $\mathcal{O}(\delta t^{2n+1})$.

Un algorithme d'intégration d'ordre élevé est utile lorsque le coût du calcul des opérateurs A et B augmente extrêmement rapidement avec le nombre de degrés de liberté (pour des calculs de dynamique quantique, par exemple) ou lorsqu'une très grande précision est requise (pour des calculs astronomiques, par exemple). Il permet d'obtenir avec un pas d'intégration plus long une précision comparable à celle d'un algorithme d'ordre plus bas. Pour la dynamique moléculaire classique de N particules, dont les forces atomiques peuvent être obtenues par un calcul de complexité $\mathcal{O}(N \log N)$, et pour laquelle la précision des trajectoires n'est pas critique, le gain en efficacité d'un algorithme d'ordre $2n+1$ est réduit par l'augmentation géométrique de la complexité de chaque pas d'intégration.

3.3.4 Symétrie symplectique

L'algorithme *velocity-Verlet*, en dépit du fait qu'il ne solutionne pas exactement le mouvement du hamiltonien H, préserve la structure *symplectique* du flot dans l'espace des phases, dont la propriété essentielle est de conserver le volume d'un

élément de l'espace des phases. La structure symplectique est directement liée à la symétrie des équations de Hamilton, qui peuvent être écrites à l'aide de la matrice symplectique J :

$$\dot{\mathbf{\Gamma}} = J\frac{\partial H}{\partial \mathbf{\Gamma}}. \tag{3.58}$$

La matrice J est simplement

$$J = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{1} \\ \hline -\mathbf{1} & \mathbf{0} \end{array}\right], \tag{3.59}$$

où $\mathbf{1}$ est la matrice identité et $\mathbf{0}$ est la matrice nulle. Une application linéaire S qui transforme le vecteur des phases $\mathbf{\Gamma}$ en un vecteur $\mathbf{\Gamma}' = S\mathbf{\Gamma}$ est dite *symplectique* si l'équation du mouvement pour $\mathbf{\Gamma}'$ conserve sa structure symplectique :

$$\dot{\mathbf{\Gamma}}' = J\frac{\partial H}{\partial \mathbf{\Gamma}'}. \tag{3.60}$$

Cette condition est assurée si S possède la propriété suivante :

$$S^{\mathrm{T}} J S = J. \tag{3.61}$$

Montrons que l'opérateur de propagation correspondant à l'algorithme *velocity-Verlet* est symplectique. Pour un problème unidimensionnel, un pas de l'algorithme *velocity-Verlet* correspond à l'application de la matrice suivante [23] :

$$\begin{aligned} P_{\mathrm{vV}} &= \begin{bmatrix} 1 & 0 \\ -\frac{1}{2}\delta t\lambda'/m & 1 \end{bmatrix} \begin{bmatrix} 1 & \delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{2}\delta t\lambda/m & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1-\frac{1}{2}\delta t^2\lambda/m & \delta t \\ -\frac{1}{2}\delta t^2(\lambda+\lambda')/m + \frac{1}{4}\delta t^4\lambda\lambda'/m^2 & 1-\frac{1}{2}\delta t^2\lambda'/m \end{bmatrix}. \end{aligned} \tag{3.62}$$

Sans perte de généralité, on a défini λ tel que $F = -\lambda x$, et λ' tel que $F' = -\lambda' x'$. Cette matrice est analogue à celle de l'équation (3.52), mis à part qu'elle s'applique à un vecteur (x, v) plutôt qu'à un vecteur $(\omega x, v)$. On vérifie aisément

que $P_{\mathrm{vV}}^{\mathrm{T}} J P_{\mathrm{vV}} = J$ et donc que l'intégrateur *velocity-Verlet* est symplectique. Nous sommes donc assuré que l'application répétée de P_{vV} produit un flot dans l'espace des phases qui vérifie le théorème de Liouville [7, section 9–8], indépendamment du flot créé par les réelles équations du mouvement. (Il va sans dire que la composition de plusieurs transformations symplectiques est symplectique elle-aussi.) À l'opposé, l'opérateur correspondant à l'intégration naïve décrite par les équations (3.12) et (3.13) n'est pas symplectique.

Sous une transformation de variables continue $\mathbf{T} = \mathbf{T}(\mathbf{\Gamma})$ quelconque, les équations du mouvement deviennent

$$\dot{\mathbf{T}} = J_{\mathbf{T}} \frac{\partial H}{\partial \mathbf{T}}, \tag{3.63}$$

avec

$$(J_{\mathbf{T}})_{ij} = \sum_{kl} \frac{\partial T_i}{\partial \Gamma_k} J_{kl} \frac{\partial T_j}{\partial \Gamma_l}. \tag{3.64}$$

La matrice $J_{\mathbf{T}}$ vérifie l'équivalent de l'identité de Jacobi [24] :

$$\sum_n \left[(J_{\mathbf{T}})_{in} \frac{\partial (J_{\mathbf{T}})_{jk}}{\partial T_n} + (J_{\mathbf{T}})_{kn} \frac{\partial (J_{\mathbf{T}})_{ij}}{\partial T_n} + (J_{\mathbf{T}})_{jn} \frac{\partial (J_{\mathbf{T}})_{ki}}{\partial T_n} \right] = 0. \tag{3.65}$$

Même dans ce cas plus général, la transformation $\mathbf{T}$ est symplectique—c'est-à-dire qu'elle ne change pas la forme des équations du mouvement—si la matrice $J_{\mathbf{T}}$ est égale à la matrice symplectique J. Naturellement, cette condition revient à l'équation (3.61) pour une transformation linéaire. Faisons aussi remarquer qu'une transformation symplectique est aussi *canonique* [7, section 9–1], puisque les équations du mouvement transformée peuvent évidemment encore être écrites sous une forme hamiltonienne.

La matrice symplectique J, qui définit les équations de Hamilton (3.58), constitue aussi une définition du crochet de Poisson :

$$\{A, B\} = \frac{\partial A}{\partial \mathbf{\Gamma}} J \frac{\partial B}{\partial \mathbf{\Gamma}}. \tag{3.66}$$

Le fait que la matrice J soit antisymétrique (c'est-à-dire, que $J^{\mathrm{T}} = -J$) correspond à une importante propriété du crochet de Poisson : $\{B, A\} = -\{A, B\}$ et en particulier $\{A, A\} = 0$. À cause de cette antisymétrie, la dynamique définie par l'équation (3.58) conserve la fonction H :

$$\frac{dH}{dt} = \frac{\partial H}{\partial \mathbf{\Gamma}}\dot{\mathbf{\Gamma}} = \frac{\partial H}{\partial \mathbf{\Gamma}} J \frac{\partial H}{\partial \mathbf{\Gamma}} = 0. \tag{3.67}$$

H, on le sait, représente l'énergie totale du système. Le système défini par un hamiltonien $H(\mathbf{\Gamma})$, indépendant du temps et possédant une structure symplectique représentée par la matrice J, est simplement dit *hamiltonien*. Insistons sur le fait que dH/dt demeure nul si J est remplacé par une matrice antisymétrique A quelconque. Des équations du mouvement $\dot{\mathbf{\Gamma}} = A(\partial H/\partial \mathbf{\Gamma})$, à défaut d'être *hamiltoniennes*, auraient au moins le mérite de conserver l'énergie totale H.

3.3.5 Intégration d'un système non-hamiltonien

Le formalisme symplectique peut être étendu à l'intégration de systèmes non-hamiltoniens, dont le crochet de Poisson n'est pas défini par la matrice J. C'est notamment le cas des systèmes où les atomes sont couplés à un réservoir de chaleur agissant comme thermostat et simulant l'ensemble canonique (NVT) plutôt que l'ensemble microcanonique (NVE). Prenons l'exemple simple du thermostat de Nosé–Hoover [25], dont les équations du mouvement sont

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i - m_i \dot{\mathbf{r}}_i \dot{\eta}, \tag{3.68}$$

$$Q\ddot{\eta} = \sum_j m_j \dot{r}_j^2 - N_{\mathrm{f}} k_{\mathrm{B}} T. \tag{3.69}$$

La signification des variables et paramètres est la même qu'au chapitre précédent. L'énergie totale du système inclut la contribution du réservoir de chaleur :

$$H = U(\{\mathbf{r}\}) + N_{\mathrm{f}} k_{\mathrm{B}} T \eta + \sum_i \frac{p_i^2}{2m_i} + \frac{p_\eta^2}{2Q}. \tag{3.70}$$

Ce n'est que tout récemment [26] qu'il a été prouvé que ces équations généraient l'ensemble canonique (NVT). La difficulté de la preuve venait du fait que les équations du mouvement ne sont pas des équations de Hamilton—au sens où elles ne peuvent pas être écrites sous la forme (3.58)—et donc que le flot dans l'espace des phases obéit à une géométrie plus compliquée [27].

Suivant Sergi et Ferrario [28], on peut écrire les équations du mouvement sous forme matricielle :

$$\begin{bmatrix} \{\dot{\mathbf{r}}\} \\ \dot{\eta} \\ \{\dot{\mathbf{p}}\} \\ \dot{p}_\eta \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & \{-\mathbf{p}\} \\ 0 & -1 & \{\mathbf{p}\} & 0 \end{bmatrix} \begin{bmatrix} \{\partial U/\partial \mathbf{r}\} \\ N_\mathrm{f} k_\mathrm{B} T \\ \{\mathbf{p}/m\} \\ p_\eta/Q \end{bmatrix} \tag{3.71}$$

Cette équation correspond à une forme généralisée, non-hamiltonienne, de l'équation (3.58) :

$$\dot{\mathbf{\Gamma}} = \mathcal{J} \frac{\partial H}{\partial \mathbf{\Gamma}}. \tag{3.72}$$

La matrice $\mathcal{J}$ est antisymétrique, ce qui indique que la fonction H de l'équation (3.70) est bel et bien une constante du mouvement. Par contre, la matrice $\mathcal{J}$ ne vérifie pas l'équation (3.65). On ne peut donc pas lui associer de véritable crochet de Poisson, et le système n'est pas hamiltonien à proprement parler [29]. On peut néanmoins définir un crochet de Poisson *non-hamiltonien* (identifié par un indice $\mathcal{J}$)

$$\{A, B\}_\mathcal{J} \equiv \frac{\partial A}{\partial \mathbf{\Gamma}} \mathcal{J} \frac{\partial B}{\partial \mathbf{\Gamma}}. \tag{3.73}$$

Comme $\mathcal{J}$ est antisymétrique mais non-hamiltonienne, le crochet possède la propriété d'antisymétrie, $\{B, A\}_\mathcal{J} = -\{A, B\}_\mathcal{J}$ mais falsifie l'identité de Jacobi :

$$\{A, \{B, C\}_\mathcal{J}\}_\mathcal{J} + \{C, \{A, B\}_\mathcal{J}\}_\mathcal{J} + \{B, \{C, A\}_\mathcal{J}\}_\mathcal{J} \neq 0. \tag{3.74}$$

Ce crochet généralisé peut à son tour servir à définir un opérateur de Liouville

non-hamiltonien [29] (que l'on notera $i\mathcal{L}$ au lieu de iL)

$$i\mathcal{L} \equiv \{\cdots, H\}_{\mathcal{J}} = -\{H, \cdots\}_{\mathcal{J}} = -\frac{\partial H}{\partial \mathbf{\Gamma}} \mathcal{J} \frac{\partial}{\partial \mathbf{\Gamma}}. \tag{3.75}$$

Formellement, la trajectoire du système peut être obtenue par l'opération du propagateur $e^{i\mathcal{L}}$.

Quelles propriétés devrait posséder un propagateur P pour intégrer « correctement » le système défini par l'équation (3.72) ? Pour que l'énergie étendue du système soit conservée, il faut qu'il soit réversible dans le temps. Il faudrait aussi qu'il n'affecte pas la forme des équations du mouvement, c'est-à-dire que

$$P^{-1} \mathcal{J} P = \mathcal{J}. \tag{3.76}$$

Pour les équations de Nosé–Hoover (3.71), l'opérateur de Liouville non-hamiltonien est

$$\begin{aligned} i\mathcal{L} &= \sum_j \frac{\mathbf{p}_j}{m_j} \cdot \frac{\partial}{\partial \mathbf{r}_j} + \frac{p_\eta}{Q} \frac{\partial}{\partial \eta} + \sum_j \left(\mathbf{F}_j - \frac{p_\eta}{Q} \mathbf{p}_j \right) \cdot \frac{\partial}{\partial \mathbf{p}_j} + G \frac{\partial}{\partial p_\eta} \\ &= \underbrace{\sum_j \mathbf{v}_j \cdot \frac{\partial}{\partial \mathbf{r}_j}}_{i\mathcal{L}_r} + \underbrace{\sum_j \frac{\mathbf{F}_j}{m_j} \cdot \frac{\partial}{\partial \mathbf{v}_j}}_{i\mathcal{L}_v} + \underbrace{\dot{\eta} \frac{\partial}{\partial \eta} - \sum_j \dot{\eta} \mathbf{v}_j \cdot \frac{\partial}{\partial \mathbf{v}_j}}_{i\mathcal{L}_\eta} + \underbrace{\frac{G}{Q} \frac{\partial}{\partial \dot{\eta}}}_{i\mathcal{L}_{\dot{\eta}}}, \end{aligned} \tag{3.77}$$

avec

$$G \equiv \sum_j \frac{p_i^2}{m_i} - N_{\mathrm{f}} k_{\mathrm{B}} T. \tag{3.78}$$

Dans l'esprit de l'algorithme *velocity-Verlet*, on factorise l'opérateur $e^{iL\delta t}$ de la façon suivante [30] :

$$e^{(i\mathcal{L}_\eta + i\mathcal{L}_{\dot{\eta}})\delta t/2} e^{i\mathcal{L}_v \delta t/2} e^{i\mathcal{L}_r \delta t} e^{i\mathcal{L}_v \delta t/2} e^{(i\mathcal{L}_\eta + i\mathcal{L}_{\dot{\eta}})\delta t/2} = \exp\left\{ \delta t [i\mathcal{L} + \mathcal{O}(\delta t^2)] \right\}. \tag{3.79}$$

Les trois opérateurs du centre sont ceux de l'équation (3.31) et agissent de la même façon [voir les équations (3.38), (3.39) et (3.40)]. L'opérateur $e^{(i\mathcal{L}_\eta + i\mathcal{L}_{\dot{\eta}})\delta t/2}$ au début

et à la fin de chaque pas d'intégration est factorisé ainsi :

$$e^{i\mathcal{L}_\eta \delta t} e^{i\mathcal{L}_\eta \delta t} e^{i\mathcal{L}_\eta \delta t} = \exp\left\{\tfrac{1}{2}\delta t[i\mathcal{L}_\eta + i\mathcal{L}_{\dot\eta} + \mathcal{O}(\delta t^2)]\right\}. \tag{3.80}$$

Comme il ne dépend pas des forces atomiques, il pourrait aussi être factorisé à un ordre plus élevé (comme il est fait au chapitre 2) :

$$\left(e^{i\mathcal{L}_{\dot\eta} \delta t/2n_c} e^{i\mathcal{L}_\eta \delta t/n_c} e^{i\mathcal{L}_{\dot\eta} \delta t/2n_c}\right)^{n_c} = \exp\left\{\tfrac{1}{2}\delta t[i\mathcal{L}_\eta + i\mathcal{L}_{\dot\eta} + \mathcal{O}(\delta t^2/n_c^2)]\right\}. \tag{3.81}$$

Nous reviendrons sur ce genre de développement à la section suivante. Sachant que $e^{a(\partial/\partial x)}x = x + a$ [voir l'équation (3.37)] et que

$$e^{ax(\partial/\partial x)}x = \sum_{n=0}^{\infty} \frac{a^n}{n!}\left(x\frac{\partial}{\partial x}\right)^n x = e^a x, \tag{3.82}$$

les effets des neuf opérations sur un vecteur des phases Γ sont, dans l'ordre,

$$\begin{aligned}
e^{i\mathcal{L}_{\dot\eta}\delta t/4} &: \dot\eta \leftarrow \dot\eta + \tfrac{1}{4}\delta t G(\{\mathbf{v}\})/Q, && (3.83)\\
e^{i\mathcal{L}_\eta \delta t/2} &: \mathbf{v}_i \leftarrow \mathbf{v}_i\, e^{-\dot\eta \delta t/2}, && (3.84)\\
&\quad \eta \leftarrow \eta + \tfrac{1}{2}\delta t \dot\eta, && (3.85)\\
e^{i\mathcal{L}_{\dot\eta}\delta t/4} &: \dot\eta \leftarrow \dot\eta + \tfrac{1}{4}\delta t G(\{\mathbf{v}\})/Q,\\
e^{i\mathcal{L}_v \delta t/2} &: \mathbf{v}_i \leftarrow \mathbf{v}_i + \tfrac{1}{2}\delta t \mathbf{F}_i(\{\mathbf{r}\})/m_i, && (3.86)\\
e^{i\mathcal{L}_r \delta t} &: \mathbf{r}_i \leftarrow \mathbf{r}_i + \delta t \mathbf{v}_i, && (3.87)\\
e^{i\mathcal{L}_v \delta t/2} &: \mathbf{v}_i \leftarrow \mathbf{v}_i + \tfrac{1}{2}\delta t \mathbf{F}_i(\{\mathbf{r}\})/m_i,\\
e^{i\mathcal{L}_{\dot\eta}\delta t/4} &: \dot\eta \leftarrow \dot\eta + \tfrac{1}{4}\delta t G(\{\mathbf{v}\})/Q,\\
e^{i\mathcal{L}_\eta \delta t/2} &: \mathbf{v}_i \leftarrow \mathbf{v}_i\, e^{-\dot\eta \delta t/2},\\
&\quad \eta \leftarrow \eta + \tfrac{1}{2}\delta t \dot\eta,\\
e^{i\mathcal{L}_{\dot\eta}\delta t/4} &: \dot\eta \leftarrow \dot\eta + \tfrac{1}{4}\delta t G(\{\mathbf{v}\})/Q.
\end{aligned}$$

Il est possible de vérifier que la composition de ces neuf opérations, traduite sous

forme matricielle, vérifie l'équation (3.76). C'est donc une « bonne » méthode d'intégration des équations du mouvement (3.71).

3.3.6 Intégration multi-échelles

L'avantage conceptuel du formalisme de l'opérateur d'évolution devient évident lorsqu'on cherche à générer un algorithme d'intégration *multi-échelles*. Imaginons que l'opérateur de Liouville iL possède une partie *lente* iL', qui peut être intégrée à l'aide d'un pas $\delta t'$, et une partie *rapide* iL'', qui doit être intégrée à l'aide d'un pas plus court, $\delta t'' = \delta t'/m$. Plutôt que de propager l'entièreté du système à l'aide du plus court pas d'intégration, on n'utilise le pas $\delta t''$ que pour la partie rapide du liouvillien :

$$\left(e^{iL\delta t''}\right)^m \simeq e^{iL'\delta t'/2}\left(e^{iL''\delta t''}\right)^m e^{iL'\delta t'/2}. \tag{3.88}$$

Cette décomposition est particulièrement avantageuse si iL' est cher à calculer mais que iL'' ne l'est pas.

Pour la dynamique moléculaire, l'idée a été développée de manière formelle par Tuckerman, Berne et Martyna [6], et perfectionnée pour une variété de situations dynamiques : particules massives au milieu de particules légères [31], interactions rapides à courte portée et lentes à longue portée [32], molécules possédant des modes de vibrations rapides [33], particules très chaudes dans un bain de particules froides [34]. C'est une méthode puissante, mais qui doit être appliquée avec discernement, vu que les opérateurs iL' et iL'' ne sont jamais exactement découplables (voir notamment les recommandations de la référence [35], et les subtilités soulevées par la séparation des forces électrostatiques à longue portée [36, 37, 38]). Elle présente aussi des problèmes de résonance [39] plus subtils que ceux discutés à la section 3.3.3 dans le cas simple de l'intégrateur *velocity-Verlet*.

Bibliographie

[1] Loup Verlet. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, 159(1) : 98–103, 1967.

[2] A. Rahman. Correlations in the Motion of Atoms in Liquid Argon. *Phys. Rev.*, 136(2A) : A405–A411, 1964.

[3] M. P. Allen et D. J. Tildesley. Computer Simulation of Liquids. Clarendon Press, Oxford, 1987.

[4] W.C. Swope, H.C. Andersen, P.H. Berens et K.R. Wilson. A computer-simulation method for the calculation of equilibrium-constants for the formation of physical clusters of molecules – Application to small water clusters. *J. Chem. Phys.*, 76(1) : 637–649, 1982.

[5] Robert D. Skeel, Guihua Zhang et Tamar Schlick. A family of symplectic integrators : Stability, accuracy, and molecular dynamics applications. *SIAM J. Sci. Comput.*, 18(1) : 203, 1997.

[6] M. Tuckerman, B. J. Berne et G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97(3) : 1990–2001, 1992.

[7] Herbert Goldstein. Classical mechanics. Addison-Wesley, Reading, Massachusetts, 2^{e} édition, 1980.

[8] H. F. Trotter. On the Product of Semi-Groups of Operators. *Proc. Am. Math. Soc.*, 10(4) : 545–551, 1959.

[9] Hans De Raedt et Bart De Raedt. Application of the generalized Trotter formula. *Phys. Rev. A*, 28(6) : 3575, 1983.

[10] Minoru Takahashi et Masatoshi Imada. Monte Carlo Calculation of Quantum Systems. II. Higher Order Correction. *J. Phys. Soc. Jpn*, 53(11) : 3765–3769, 1984.

[11] Michael Creutz et Andreas Gocksch. Higher-order hybrid Monte Carlo algorithms. *Phys. Rev. Lett.*, 63(1–3) : 9–12, 1989.

[12] R. F. Warming et B. J. Hyett. Modified equation approach to stability and accuracy analysis of finite-difference methods. *J. Comput. Phys.*, 14(2) : 159–179, 1974.

[13] Robert D. Skeel et David J. Hardy. Practical Construction of Modified Hamiltonians. *SIAM J. Sci. Comput.*, 23(4) : 1172–1188, 2001.

[14] C. M. Field et F. W. Nijhoff. A note on modified Hamiltonians for numerical integrations admitting an exact invariant. *Nonlinearity*, 16(5) : 1673–1683, 2003.

[15] H. Munthe-Kaas et B. Owren. Computations in a Free Lie Algebra. *Phil. Trans. Royal Soc. A*, 357 : 957–981, 1999.

[16] Matthias W. Reinsch. A simple expression for the terms in the Baker-Campbell-Hausdorff series. *J. Math. Phys.*, 41(4) : 2434–2442, 2000.

[17] R. W. Pastor, B. R. Brooks et A. Szabo. An Analysis of the Accuracy of Langevin and Molecular-Dynamics Algorithms. *Mol. Phys.*, 65(6) : 1409–1419, 1988.

[18] Tamar Schlick, Margaret Mandziuk, Robert D. Skeel et K. Srinivas. Nonlinear Resonance Artifacts in Molecular Dynamics Simulations. *J. Comput. Phys.*, 140(1) : 1–29, 1998.

[19] Qun Ma, Jesús A. Izaguirre et Robert D. Skeel. Verlet-I/R-RESPA/Impulse is Limited by Nonlinear Instabilities. *SIAM J. Sci. Comp.*, 24(6) : 1951–1973, 2003.

[20] Oren M. Becker, Alexander D. MacKerell, Jr., Benoît Roux et Masakatsu Watanabe (rédacteurs). Computational Biochemistry and Biophysics. Marcel Dekker, Inc., New York, 2001.

[21] Haruo Yoshida. Construction of higher order symplectic integrators. *Phys. Lett. A*, 150(5–7) : 262–268, 1990.

[22] Masuo Suzuki. General theory of fractal path integrals with applications to many-body theories and statistical physics. *J. Math. Phys.*, 32(2) : 400–407, 1991.

[23] Paul F. Batcho et Tamar Schlick. Special stability advantages of position-Verlet over velocity-Verlet in multiple-time step integration. *J. Chem. Phys.*, 115(9) : 4019–4029, 2001.

[24] P. J. Morrison. Hamiltonian description of the ideal fluid. *Rev. Mod. Phys.*, 70(2) : 467–521, 1998.

[25] William G. Hoover. Canonical dynamics : Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3) : 1695–1697, 1985.

[26] Mark E. Tuckerman, Yi Liu, Giovanni Ciccotti et Glenn J. Martyna. Non-Hamiltonian molecular dynamics : Generalizing Hamiltonian phase space principles to non-Hamiltonian systems. *J. Chem. Phys.*, 115(4) : 1678–1702, 2001.

[27] M. E. Tuckerman, C. J. Mundy et G. J. Martyna. On the classical statistical mechanics of non-Hamiltonian systems. *Europhys. Lett.*, 45(2) : 149–155, 1999.

[28] Alessandro Sergi et Mauro Ferrario. Non-Hamiltonian equations of motion with a conserved energy. *Phys. Rev. E*, 64 : 056125, 2001.

[29] Alessandro Sergi. Non-Hamiltonian equilibrium statistical mechanics. *Phys. Rev. E*, 67 : 021101, 2003.

[30] Glenn J. Martyna, Mark E. Tuckerman, Douglas J. Tobias et Michael L. Klein. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.*, 87(5) : 1117–1157, 1996.

[31] Mark E. Tuckerman, Bruce J. Berne et Angelo Rossi. Molecular dynamics algorithm for multiple time scales : Systems with disparate masses. *J. Chem. Phys.*, 94(2) : 1465–1469, 1991.

[32] Mark E. Tuckerman, Bruce J. Berne et Glenn J. Martyna. Molecular dynamics algorithm for multiple time scales : Systems with long range forces. *J. Chem. Phys.*, 94(10) : 6811–6815, 1991.

[33] Masakatsu Watanabe et Martin Karplus. Dynamics of molecules with internal degrees of freedom by multiple time-steps methods. *J. Chem. Phys.*, 99(10) : 8063–8074, 1993.

[34] Ruhong Zhou, Steven J. Stuart et B. J. Berne. Molecular dynamics for nonequilibrium systems in which there are a small number of very hot particles in a cold bath : Reference system propagator methods. *J. Chem. Phys.*, 105(1) : 235–239, 1996.

[35] Steven J. Stuart, Ruhong Zhou et B. J. Berne. Molecular dynamics with multiple time scales : The selection of efficient reference system propagators. *J. Chem. Phys.*, 105(4) : 1426–1436, 1996.

[36] Piero Procacci, Massimo Marchi et Glenn J. Martyna. Electrostatic calculations and multiple time scales in molecular dynamics simulation of flexible molecular systems. *J. Chem. Phys.*, 108(21) : 8799–8803, 1998.

[37] Paul F. Batcho, David A. Case et Tamar Schlick. Optimized particle-mesh Ewald/multiple-time step integration for molecular dynamics simulations. *J. Chem. Phys.*, 115(9) : 4003–4018, 2001.

[38] Xiaoliang Qian et Tamar Schlick. Efficient multiple-time-step integrators with distance-based force splitting for particle-mesh-Ewald molecular dynamics simulations. *J. Chem. Phys.*, 116(14) : 5971–5983, 2002.

[39] Jeffrey J. Biesiadecki et Robert D. Skeel. Dangers of Multiple Time Step Methods. *J. Comput. Phys.*, 109 : 318–328, 1993.

Chapitre 4

A simple polarizable model of water based on classical Drude oscillators

Tiré de :
Guillaume Lamoureux, Alexander D. MacKerell, Jr. et Benoît Roux. *J. Chem. Phys.*, 119(10) : 5185–5197, 2003.

Abstract

A simple polarizable water model is developed and optimized for molecular dynamics simulations of the liquid phase under ambient conditions. The permanent charge distribution of the water molecule is represented by three point charges: two hydrogen sites and one additional M site positioned along the HOH bisector. Electronic induction is represented by introducing a classical charged Drude particle attached to the oxygen by a harmonic spring. The oxygen site carries an equal and opposite charge, and is the center of an intermolecular Lennard-Jones interaction. The HOH gas-phase experimental geometry is maintained rigidly and the dipole of the isolated molecule is 1.85 D, in accord with experiment. The model is simulated by considering the dynamics of an extended Lagrangian in which a small mass is attributed to the Drude particles. It is parametrized to reproduce the salient properties of liquid water under ambient conditions. The optimal model, refered to as SWM4-DP for "simple water model with four sites and Drude polarizability," yields

a vaporization enthalpy of 10.52 kcal/mol, a molecular volume of 29.93 Å^3, a static dielectric constant of 79 ± 5, a self-diffusion constant of $(2.30 \pm 0.04) \times 10^{-5}$ cm^2/s, and an air/water surface tension of 66.9 ± 0.9 dyn/cm, all in excellent accord with experiments. The energy of the water dimer is -5.18 kcal/mol, in good accord with estimates from experiments and high level *ab initio* calculations. The polarizability of the optimal model is 1.04 Å^3, which is smaller than the experimental value of 1.44 Å^3 in the gas phase. It is likely that such a reduced molecular polarizability, which is essential to reproduce the properties of the liquid, arises from the energy cost of overlapping electronic clouds in the condensed phase due to Pauli's exclusion principle opposing induction.

4.1 Introduction

For meaningful theoretical studies of biomolecular systems, it is necessary to have potential energy functions providing a realistic and accurate representation of both the microscopic interactions and thermodynamic properties. This presents a difficult challenge, particularly when charged or highly polar species are involved (see, e.g., Ref. [1]), where electronic polarization is expected to have a significant role in both structure and energetic properties. Current biomolecular potential functions typically account for many-body polarization effects in an average way using an effective parametrization of the atomic partial charges [2]. Because of this approximation, the optimal parametrization is the result of a compromise between an accurate representation of the microscopic energies and bulk solvation properties. Such potential functions can yield meaningful results of semiquantitative accuracy, but are not uniformly reliable under all the solvation conditions relevant to biomolecules. Accordingly, improvements in the representation of biomolecules are anticipated if nonadditive many-body polarization is explicitly taken into account. At the present time, computational chemists and theoreticians are actively pursuing the development of a new generation of force fields for computational studies of biological systems that will include induced polarization [3, 4]. However,

much more work is needed before such potential functions are ready to be used in simulations of heterogenous biological systems.

Any effort to develop a force field for biomolecular systems must start with a model for water. Water is ubiquitous in biological systems, and a high quality water force field is essential for meaningful simulation studies of biological systems. Many of the models of water currently used in biomolecular simulations are based on fixed effective partial charges which were adjusted to yield accurate bulk liquid properties [5, 6]. Although such models, which incorporate the average influence of electronic induction in an isotropic liquid, have been and remain exceedingly useful, there are some well-justified concerns about their ability to represent the properties of liquid in inhomogeneous environments.

Electronic polarization of water is highly sensitive to its environment. In the gas phase, an isolated water molecule has a dipole moment of 1.85 D [7]. However, the average molecular dipole is 2.1 D in the water dimer [8], increases in larger water clusters [8], reaching, in the condensed phase, a value between 2.4–2.6 D, as suggested from classical molecular dynamics simulations of the dielectric properties [9, 10, 11], and 2.95 D, as obtained from *ab initio* molecular dynamics [12, 13] and from analysis of experimental data [14, 15]. This bulk value is close to the maximum dipole any charged or polar molecule can induce in a water molecule. For instance, adding a sodium ion or a dimethyl phosphate anion to bulk water does not significantly add to the induction effect [16].

Polarizable water models better reproduce the molecular dipoles in contexts where the hydrogen bonds network of room-temperature bulk water is either partially destroyed or significantly perturbed [8, 17], but, most importantly, they better reproduce the interaction energies. While reproducing the liquid-vapor enthalpy as well as the nonpolarizable models, they have dimer energies closer to -5.0 kcal/mol (the consensus *ab initio* value [18]): -4.69 kcal/mol for the POL3 model [19], -4.51 kcal/mol for the TIP4P-FQ model [20], -5.33 kcal/mol for the model of Burnham *et al.* [17], and -5.00 kcal/mol for the MCDHO model [21]. They adequately reproduce the total binding energy of the water trimer

[19, 22, 17, 21], although the many-body energy contribution itself may be underestimated.

Polarizability appears essential to accommodate the local disruption of the hydrogen bond network created by anions such as chloride [23, 24, 25, 26, 27] or fluoride [28, 29, 30, 31, 32], or to reproduce the polarization effects of small multivalent cations on the first hydration shell [33, 34]. And although explicit polarizability does not appear to have any significant effect on the reorganization of water molecules at flat liquid–hydrophobic [35], liquid–vapor [36, 19], or liquid–metal [37, 38] interfaces, it may play an decisive role for the specific water–water interactions near small nonpolar moieties [39, 40]. In short, polarizability is essential to get from the same water model accurate energetics in the vicinity of highly polar moieties (such as carbonyl groups), small ions (such as sodium or chloride), as well as in anisotropic nonpolar environments.

Although a polarizable water model based on a simple potential may not capture the delicate energy balance needed to describe all phase transformations accurately from ice to water vapor, it should considerably help in increasing the accuracy of biomolecular simulations, which are largely generated around ambient conditions of pressure and temperature. It is important to continue to search for simple models of water that are accurate and computationally inexpensive, and that are adapted to investigate problems in large complex biomolecular systems, which require very long simulation times. That said, it is important to keep in mind that even an *exact* microscopic potential function might not automatically lead to accurate thermodynamic properties in classical simulations, which neglect the quantum nature of the nuclei. For example, there are clear, albeit small, differences between classical and quantum simulations of the bulk liquid based on a given water model [41]. For this reason, parametrization of the potential function is unavoidable to reproduce macroscopic properties accurately.

The goal of the present paper is to present a simple model for water and determine its optimal parametrization. In this model, a water molecule is represented as a rigid object imposing the experimental gas phase molecular geometry, with

four interaction sites: the oxygen, carrying the molecular polarizability but no net charge, the two hydrogens, and an additional site located along the HOH bisector. An isotropic polarizability is introduced by adding a charged mobile auxiliary particle attached to the oxygen by a harmonic spring: a classical Drude oscillator [42].

Many polarizable models of comparable complexity exist in the literature. Some of the most recent (and most successful) are the RPOL model of Dang [43], the TIP4P-FQ model of Rick *et al.* [20], the BSV model of Brodholt *et al.* [44], the "polarizable point-charge" (PPC) model of Svishchev *et al.* [45], the POL model of Dang and Chang [19], the MCDHO model of Saint-Martin *et al.* [21], the shell water (SW) models of van Maaren and van der Spoel [46], the POL5 models of Stern *et al.* [47], and the "charge-on-spring" (COS) models of Yu *et al.* [48]. Most of these models significantly overestimate the static dielectric constant ϵ of liquid water. Typically, ϵ is higher that 100, compared to the expected value of 78.4 [49]. Such an incorrect electrostatic shielding is a important shortcoming for a model to be used as a solvent for biomolecules. Only TIP4P-FQ and PPC models have ϵ close to 80. But they have electrostatic representations that may be considered limiting: their out-of-plane molecular polarizability components are zero [20, 45], whereas the molecular polarizability of water is almost isotropic [50].

The model presented in this work is adjusted to reproduce the potential energy, density, self-diffusion constant, and static dielectric constant of bulk liquid water under normal temperature and pressure conditions, using the double-thermostat molecular dynamics algorithm described previously [51]. Simply by reproducing these four bulk properties, we find a consistent microscopic picture: good dimer energy and good hydrogen-bonding geometry, as well as consistent energies for water clusters.

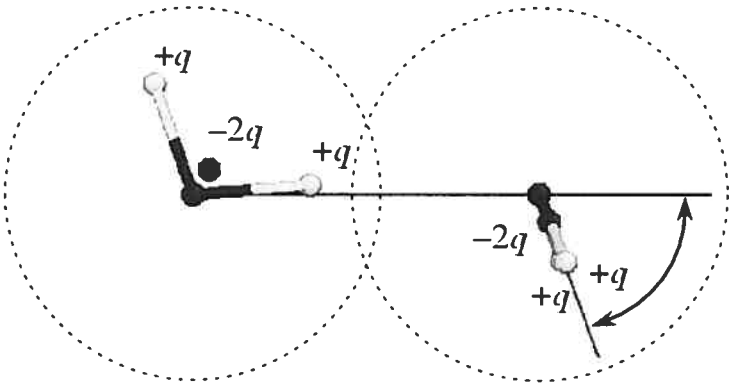

Figure 4.1: Illustration of the SWM4-DP water dimer. The molecular structures are shown as ball-and-stick models, with separate M sites. The Drude particles are too close to the oxygen sites to be shown. The arrow defines the acceptor angle θ_A.

4.2 Theory and methods

4.2.1 Model and computational details

The model considered here is closely related to the TIP4P model [5]. The water molecule is kept in a fixed geometry, with four interaction sites: the oxygen and hydrogen atoms and an additional massless site, called "M," located at a fixed distance ℓ_{OM} along the bisector of the HOH angle (see Fig. 4.1). The oxygen and hydrogen atoms are kept in the experimental geometry of the water molecule in the gas phase with a OH bond length of 0.9572 Å and a HOH angle of 104.52° [52]. A mobile auxiliary particle, a classical Drude oscillator [42], is attached by a harmonic spring of force constant k_D to the oxygen site to represent induced polarizability [51]. The Drude particle carries a positive charge q_D while the oxygen carries a charge $-q_D$. During dynamics, the Drude particle stays near the minimum energy position, $\mathbf{r}_O + q_D\mathbf{E}/k_D$, where $\mathbf{E}$ is the field arising from the surrounding molecules. The displacement of the Drude particle gives rise to an induced dipole, $q_D^2\mathbf{E}/k_D$. Correspondingly, the molecular polarizability α is equal to q_D^2/k_D [51]. The permanent dipole of the isolated molecule in the gas phase is set by the charge carried by the hydrogen atoms and the massless interaction site M; the charge on the hydrogens is $q_H = q$ and the the charge on M is $q_M = -2q$.

Unshielded $1/r$ Coulombic potentials are assumed between all the intermolecular pairs of interaction sites s (O, H1, H2, M, and D) and the core repulsion and nonpolar interactions are represented using a Lennard-Jones "12-6" potential between each oxygen–oxygen (OO) pair. The total potential energy for a system of water molecules is

$$\begin{aligned} U_{\text{tot}}(\{\mathbf{r}\}) &= \sum_i \tfrac{1}{2} k_{\text{D}} |\mathbf{r}_{\text{O},i} - \mathbf{r}_{\text{D},i}|^2 + \sum_{i<j} \sum_{s,s'} \frac{q_s q_{s'}}{|\mathbf{r}_{s,i} - \mathbf{r}_{s',j}|} \\ &\quad + \sum_{i<j} 4\epsilon_{\text{OO}} \left[\left(\frac{\sigma_{\text{OO}}}{|\mathbf{r}_{\text{O},i} - \mathbf{r}_{\text{O},j}|} \right)^{12} - \left(\frac{\sigma_{\text{OO}}}{|\mathbf{r}_{\text{O},i} - \mathbf{r}_{\text{O},j}|} \right)^{6} \right] \end{aligned} \tag{4.1}$$

The model is very simple and has only a few parameters which need to be optimized: the length ℓ_{OM} between the oxygen and the M site, the charge q, the spring constant k_{D} and the charge of the Drude particle, q_{D}, and the coefficients of the Lennard-Jones potential ϵ_{OO} and σ_{OO}. In fact, the model is highly constrained and there are only four independent parameters: ℓ_{OM}, α, ϵ_{OO} and σ_{OO}. All the other parameters can be deduced directly. The magnitude of the charge q on the M site is constrained so that the total permanent dipole for an isolated molecule, $\mu_0 = 2q[\ell_{\text{OH}} \cos(\frac{1}{2}\theta_{\text{HOH}}) - \ell_{\text{OM}}]$, matches the experimental gas phase value of 1.85 D [7]. Furthermore, as long as the spring constant k_{D} is sufficiently large, the point-dipole approximation is valid and only the combination $q_{\text{D}}^2/k_{\text{D}}$, corresponding to the polarizability α, is a free parameter. Within this approximation, the sign of the charge carried by the Drude particle is irrelevant, and we chose arbitrarily q_{D} to be positive. We use a spring constant of 1000 kcal/mol/Å^2, which insures that the point-dipole limit is valid [51]. The four parameters of the model are optimized to match a number of properties as accurately as possible according to the procedure described in the next section.

4.2.2 Parametrization strategy

We seek to optimize the parameters of the model to reproduce the vaporization enthalpy, density, static dielectric constant, and self-diffusion constant in the liq-

uid phase at ambient conditions of pressure and temperature. The ability of the model to reproduce the radial distribution functions from experiments [53, 54] is monitored, as a final criteria, though it is not used directly in the optimization procedure. The following parameters have to be adjusted: ℓ_{OM}, α, ϵ_{OO}, and σ_{OO}. In fact, for given values of ℓ_{OM} and α, it is possible to adjust the Lennard-Jones parameters ϵ_{OO} and σ_{OO} to get specific values of the interaction energy U_{dimer} and the oxygen–oxygen equilibrium distance d_{OO} of the optimized water dimer. In seeking to characterize different models in terms of the average properties of the bulk liquid, we found it more practical to use those quantities directly, instead of the Lennard-Jones parameters ϵ_{OO} and σ_{OO}. The main advantage is that they are more directly related to the vaporization enthalpy and density of the liquid than the Lennard-Jones parameters, thus making the parameter search easier to interpret. Since there is a one-to-one correspondence between $(\epsilon_{OO}, \sigma_{OO})$ and (U_{dimer}, d_{OO}) in the range of interest, this choice does not limit the number of possible models.

The overall parametrization strategy is the following. First, a large number of models with different values of ℓ_{OM}, α, U_{dimer}, and d_{OO} were generated in order to explore the parameter space and determine the region yielding reasonable values for the average molecular volume $\langle v \rangle$, interaction energy Δu, self-diffusion constant D, and static dielectric constant ϵ. In the following stage, a "grid" search is set up to explore the properties of the model by varying the four parameters systematically. A $3 \times 3 \times 4 \times 3$ grid was used (3 values for ℓ_{OM}, α, and d_{OO}, and 4 values for U_{dimer}), corresponding to a total of 108 simulations. Then, a quadratic polynomial response-surface model $\mathbf{P}$ was fitted to the results of the 108 simulations to provide an interpolation formula predicting the liquid properties for any parametrization within the grid without having to actually perform the simulation. From the polarizability α, the ℓ_{OM} distance, and the energy U_{dimer} and oxygen–oxygen distance d_{OO} for the water dimer, the response function $\mathbf{P}$ provides predicted values for Δu, $\langle v \rangle$, height of the first peak in the oxygen–oxygen radial distribution function, $g_{OO}^{(1)}$,

and average dipole $\langle\mu\rangle$:

$$\{\Delta\hat{u}, \langle\hat{v}\rangle, \hat{g}^{(1)}_{\mathrm{OO}}, \langle\hat{\mu}\rangle\} = \mathbf{P}(\alpha, \ell_{\mathrm{OM}}, U_{\mathrm{dimer}}, d_{\mathrm{OO}}). \tag{4.2}$$

(The hat is to distinguish the predicted values from the actual values one would get from performing the simulation.) The coefficients of the response function are determined using a least-squares fit. There are 15 coefficients for each property for a total number of 60 coefficients. Such a polynomial response function can be formally inverted ($\mathbf{P}^{-1}$) to get the parameters that produce some chosen values for $\hat{g}^{(1)}_{\mathrm{OO}}$ and $\langle\hat{\mu}\rangle$ while satisfying the conditions $\langle\hat{v}\rangle = 29.94$ Å^3 (corresponding to a density of 0.0334 molecules/Å^3) and $\Delta\hat{u} = -9.92$ kcal/mol. In the final stage, the parameters were adjusted to reproduce the experimental values for Δu and $\langle v\rangle$, and to obtain good agreement for D and ϵ.

4.2.3 Computational details

All simulations were performed by considering the dynamics of an extended Lagrangian in which a small mass m_{D} and kinetic energy is attributed to the Drude particles. The amplitude of their oscillations away from the local energy minimum is controlled with a low-temperature thermostat acting in the local center-of-mass reference frame of each oxygen–Drude pair [51]. The Drude particle is attached to the oxygen with a harmonic spring with a force constant k_{D} of 1000 kcal/mol/Å^2. Its mass is set to 0.4 amu and the mass of the oxygen is set to 15.5994 amu, such that the total mass of the oxygen–Drude pair is equal to 15.9994 amu (the correct mass of an oxygen atom). To ensure that the time course of the induced dipoles stays close to the self-consistent-field (SCF) solution, a Nosé–Hoover thermostat at a temperature $T_\star = 1$ K is applied to the relative motion of each oxygen–Drude pair (in their local center-of-mass reference frame). The relaxation time of the Drude oscillator thermostat at temperature $T_\star$ is 0.005 ps (5 fs). It was previously shown that the trajectories generated according to this procedure are very close to those generated by the SCF regime of induced polarization [51]. To control the

global thermalization of the system, a second Nosé–Hoover thermostat at room temperature T is applied to the center of mass of the oxygen–Drude pairs as well as the hydrogens atoms. A modified Andersen–Hoover barostat is used to maintain the system at constant pressure P of 1 atm [55, 51]. A relaxation time of 0.1 ps is assigned to the barostat, which, combined with the global thermostat, maintains the system at ambient conditions of pressure and temperature. The water molecules are constrained to the experimental geometry using the SHAKE/Roll and RATTLE/Roll procedures [56]. At each dynamical step, the total force acting on the massless M sites is redistributed onto the oxygen and hydrogen atoms and their position is geometrically reconstructed after each move. The trajectory is propagated with a 1.0 fs time step, using a multistep integration procedure for the thermostat variables [51]. This modified two-temperature isobaric–isothermal ensemble, or $NP(T, T_\star)$ ensemble, is a variant of the isobaric–isothermal equations of Martyna *et al.* [55] and has been implemented in the biomolecular program CHARMM [57]. See Ref. [51] for further detail.

Systems of 250 water molecules were simulated with cubic periodic boundary conditions. The electrostatic interactions were treated without truncation using particle-mesh Ewald (PME) summation [58] (with $\kappa = 0.33$ for the charge screening and fourth-order splines for the mesh interpolations) and a 15 Å cutoff for the Lennard-Jones interactions. Some of the exploratory simulations were performed using smaller boxes (125 molecules) and smaller cutoffs (9 Å) for the Lennard-Jones interactions. Each box was simulated for 150 ps and the average liquid properties were extracted from the last 100 ps. For each simulation, the average molecular volume $\langle v \rangle$ and the vaporization enthalpy Δh are calculated and compared with the experimental values. The correct target values are 29.94 Å^3 for the molecular volume (corresponding to 0.997 g/cm^3) and 10.52 kcal/mol for the vaporization enthalpy [59]. Under the constant pressure simulation conditions, the average molecular volume $\langle v \rangle$ is extracted directly from the average volume $\langle V \rangle$ of the periodic simulation box. The vaporization enthalpy Δh is calculated from the

average net gain of potential energy Δu upon formation of the dense system

$$\begin{aligned} \Delta h &= k_{\mathrm{B}}T - \Delta u \\ &= k_{\mathrm{B}}T - \left(\langle u\rangle_{\mathrm{liq}} - \langle u\rangle_{\mathrm{gas}}\right) \\ &= k_{\mathrm{B}}T - \left(\langle u\rangle_{\mathrm{liq}} - \tfrac{3}{2}k_{\mathrm{B}}T_{\star}\right). \end{aligned} \tag{4.3}$$

It should be noted that, even though the water molecules themselves are rigid, the thermal contribution from the Drude oscillators to the total potential energy, $\frac{3}{2}k_{\mathrm{B}}T_{\star}$, must be subtracted in an extended Lagrangian simulation (even though this correction is small for $T_{\star} = 1$ K). Practically, Δu rather than Δh is considered in the analysis. The average structure of the liquid was characterized by calculating the oxygen–oxygen, oxygen–hydrogen, and hydrogen–hydrogen radial distribution functions, $g_{\mathrm{OO}}(r)$, $g_{\mathrm{OH}}(r)$, and $g_{\mathrm{HH}}(r)$, and compared with experiments [53, 54]. The dielectric constant of the liquid ϵ was calculated using [60],

$$\epsilon = \epsilon_{\infty} + \frac{4\pi}{3\langle V\rangle k_{\mathrm{B}}T}\left(\langle \mathbf{M}^2\rangle - \langle \mathbf{M}\rangle^2\right), \tag{4.4}$$

where $\mathbf{M}$ is the total dipole moment of the box and ϵ_{∞} is estimated from the Clausius–Mossotti equation (see, e.g., Ref. [61]):

$$\frac{\epsilon_{\infty} - 1}{\epsilon_{\infty} + 2} = \frac{4\pi}{3}\frac{\alpha}{v}. \tag{4.5}$$

The convergence of ϵ is monitored by plotting its value as the time series, $\mathbf{M}(t)$, is accumulated (see Fig. 4.2). It is also useful to monitor the average molecular dipole $\langle\mu\rangle$ because it is strongly correlated with the dielectric constant even though its convergence is an order of magnitude faster than Eq. (4.4). The Debye relaxation time τ_{D} was extracted from the first 40 ps of the exponential decay of the dipole time autocorrelation function [62],

$$\Phi(t) = \frac{\langle \mathbf{M}(t)\cdot\mathbf{M}(0)\rangle}{\langle \mathbf{M}^2\rangle}. \tag{4.6}$$

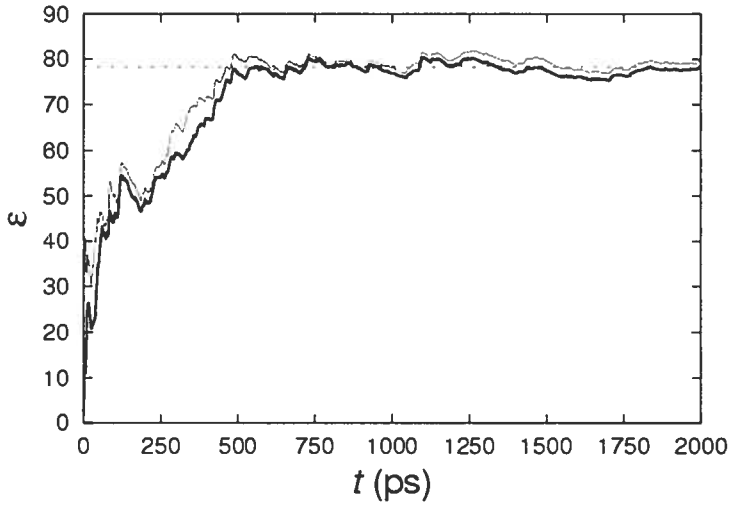

Figure 4.2: Convergence of the dielectric constant as $\mathbf{M}(t)$ is accumulated [see equation (4.4)]. $\mathbf{M}(t)$ is the concatenation of the 20 100 ps time series. The thin line corresponds to Eq. (4.4) assuming $\langle \mathbf{M} \rangle = 0$ by symmetry. The fine horizontal line shows the experimental value $\epsilon = 78.4$ (Ref. [49]).

The self-diffusion constant D was calculated from the mean squared displacement [63],

$$D = \lim_{t \to \infty} \frac{1}{6t} \left\langle \frac{1}{N} \sum_{i=1}^{N} \left[\mathbf{r}_{\mathrm{O},i}(t) - \mathbf{r}_{\mathrm{O},i}(0) \right]^2 \right\rangle . \tag{4.7}$$

The self-diffusion constant D and static dielectric constant ϵ are computed from 10 independent simulations of 150 ps, using the same initial conformation but assigning different random initial velocities. The uncertainties on both ϵ and D are estimated from the spread of the values for the 10 independent simulations. The properties of the final model are further converged using ten additional independent simulations.

As an additional validation of the optimal model, interfacial properties were characterized by calculating the air/water surface tension γ, and the interfacial electrostatic potential $\Delta\phi$. A system of 250 water molecules was simulated in a slab geometry in a 19.56 Å $\times$ 19.56 Å $\times$ 58.68 Å tetragonal periodic cell. The surface tension γ is computed from the pressure increase perpendicular to the interface [64],

$$\gamma = \tfrac{1}{2} L_z \left[\langle P_{zz} \rangle - \tfrac{1}{2} (\langle P_{xx} \rangle + \langle P_{yy} \rangle) \right] , \tag{4.8}$$

where $L_z = 58.68$ Å. The volume of the system was kept constant and the pressure tensor was monitored. The interfacial potential was calculated from the average electrostatic potential along the z-axis, $\phi(z)$, obtained by integrating the charge density $\rho(z)$:

$$\phi(z) = -\frac{2\pi}{\epsilon_0} \int_{-\infty}^{z} dz' \left[\int_{-\infty}^{z'} dz'' \, \rho(z'') - \int_{z'}^{\infty} dz'' \, \rho(z'') \right]. \tag{4.9}$$

Five independent simulations were equilibrated for 100 ps and data was accumulated for an additional 1000 ps. Last, to assess how the optimal model reproduces the energetics and the geometry of hydrogen bonding, we examined the binding energies of small water clusters (trimer, tetramer, pentamer, and hexamer).

4.3 Results and discussion

4.3.1 Optimal parametrization

For the sake of simplicity, we chose a rigid model imposing the experimental geometry of the water molecules in the gas phase. The permanent molecular dipole is created by partial charges on the hydrogen atoms and on a nonatomic M site positioned along the HOH bisector. An isotropic polarizability is attributed exclusively to the oxygen site by a classical Drude oscillator (site D). Site D has a positive charge, matched by a negative charge on the oxygen site. The total potential energy is represented by simple radial functions as expressed in Eq. (4.1).

In several ways, the current model is exceedingly simple, and this has important consequences. In particular, the model does not account for obvious factors such as the flexibility of the water molecule and the small increase in the OH bond length and in the HOH bond angle, known to occur in the liquid [65]. Furthermore, nuclear quantum effects giving rise to zero point vibration, which can be substantial because of the light mass of the hydrogen nuclei, are completely ignored. It is therefore unrealistic to expect that the model be able to provide

a correct representation of all properties of water, going all the way from small systems (monomer, dimer, trimer) up to dense bulk phases (liquid, solid) under a variety of conditions of temperature and pressure. The optimal parametrization of the current model reflects a compromise that is aimed at the most relevant situations for biomolecular systems. Thus, where to make compromises becomes an important issue. Widely used effective potentials, with no induced polarizability, were parametrized to quantitatively match the enthalpy and density of the liquid as correctly as possible. These models often have tolerated fairly large inaccuracies in the representation of small clusters. For example, the permanent dipole of effective nonpolarizable water models deviates significantly from the dipole of an isolated water molecule in the gas phase. The experimental dipole is 1.85 D whereas it is 2.35 D for TIP3P [5], 2.27 D for SPC [6], and 2.18 D for TIP4P [5]. The dipole of these models was enhanced to yield better thermodynamic properties for the liquid phase. Similarly, these models significantly overestimate the interaction energy of the water dimer by 1 to 1.5 kcal/mol. Nonetheless, effective nonpolarizable models have displayed a remarkable ability to describe biomolecular systems under ambient conditions semiquantitatively [66]. This observation carries an important message for the development of novel polarizable force field:even though it is important to have as good a representation of the small systems as possible, it is, first and foremost, essential to reproduce the basic thermodynamic properties of the bulk phase. Therefore, it seems reasonable to accept small deviations for the small systems if this allows for reproduction of the thermodynamic properties. Furthermore, it is also important to realize that there are often uncertainties about the "correct" target values, particularly in the case of small clusters. For example, the interaction energy of the water dimer is -5.4 ± 0.7 kcal/mol from experimental measurements [67, 68], but is -5.0 ± 0.1 kcal/mol, according to high-level *ab initio* calculation [18]. Such variations are small, but can nonetheless have an important impact on the properties of the condensed phase. Therefore, models which yield a dimer energy anywhere around the estimates from *ab initio* or gas phase experiments are acceptable and should be tested.

Even though the magnitude of the polarizability α of an isolated water molecule in the gas phase is 1.44 Å^3, in the bulk phase it is likely that molecular polarization is opposed by the energetic cost arising from Pauli's exclusion principle due to the overlap of neighboring electronic charge distributions [69]. This phenomena is accounted for by a "renormalized" effective α in the liquid that is smaller than the gas phase value. Recent *ab initio* calculations on the water dimer [70] and on small water clusters [71, 72] suggest that the condensed phase molecular polarizability can be 7% to 9% less than in gas phase. In the present work, the need for a smaller polarizability to reproduce the properties of the liquid phase became apparent from a large number of simulations (several hundred) exploring many models over a wide range of the three parameters ℓ_{OM} (between 0.0 and 0.35 Å), U_{dimer} (between -5.5 and -4.5 kcal/mol), and d_{OO} (between 2.7 and 3.0 Å) while keeping α at 1.44 Å^3. From this large number of simulations, it was clear that models with the experimental gas-phase polarizability systematically yield an overestimated dielectric constant, typically in the range of 150 to 200. Furthermore, none of these models could get both the correct density and enthalpy: liquid densities close to the experimental value always resulted in a vaporization enthalpies that were too favorable by about 1 to 2 kcal/mol. The average dipole of those models is around 2.9 D, which is consistent with the overestimated dielectric constant [9]. (For similar water models having a simple electrostatic representation based on unified atomic charges and polarizabilities, a dielectric constant in the correct range requires an average dipole between 2.4 and 2.6 D [9, 10, 11].) On the basis of these observations, the molecular polarizability α was then treated as an empirical adjustable parameter in optimizing the model. Models with values of α varying between 0.6 to 1.4 Å^3 were simulated. The results clearly indicated that the value of α must be around 1.0 Å^3 to yield reasonable liquid properties.

Once the reasonable range of the four parameters was established, the properties of different models were explored by a systematic grid search procedure. A total of 108 models were simulated with all combinations of the following parameter values: $\ell_{\mathrm{OM}} = 0.23, 0.24, 0.25$ Å; $\alpha = 0.95, 1.00, 1.05$ Å^3; $U_{\mathrm{OO}} = -5.4$, -5.3,

-5.2, -5.1 kcal/mol; and $d_{OO} = 2.81$, 2.82, 2.83 Å. The polynomial response function $\mathbf{P}$ of Eq. (4.2) was fitted to the results of the simulations and numerically inverted to help determine the optimal parameters for the model. As long as the models are kept within the range of parameters that were actually explored by the simulations, the polynomial prediction turned out to be very accurate, even more reliable than a single simulation, because the random statistical errors of individual simulations are filtered out by the fitting procedure. In the final stage of refinement, the strong correlation between the average molecular dipole in the liquid and the dielectric constant, as well as the strong correlation between the height of the first peak in the oxygen–oxygen radial distribution function and the diffusion constant, were exploited to converge more rapidly towards an optimal model. Assuming that everything else remains the same, the self-diffusion constant decreases when the first peak is enhanced. Within the constraints of the current model, it is possible to have an acceptable self-diffusion constant only if the maximum of the first peak is close to 3.0. Similarly, the dielectric constant increases when the average dipole per molecule in the liquid is increasing. An acceptable dielectric constant is produced only if the average dipole is around 2.45 D, within a narrow range. For example, a model, whose parameters were chosen via $\mathbf{P}^{-1}$ to impose a maximum in the radial distribution function $g_{OO}^{(1)}$ of 3.00 Å and an average dipole $\langle\mu\rangle$ of 2.44 D, yields a diffusion constant D of 2.38×10^{-5} cm^2/s and a dielectric constant ϵ equal to 79 ± 10. A closely related model, whose parameters were chosen to impose $g_{OO}^{(1)} = 3.05$ Å yields $D = 2.17 \times 10^{-5}$ cm^2/s but $\epsilon = 71 \pm 10$. Combining this information, four more models were simulated with parameters chosen to yield $g_{OO}^{(1)} = 3.05$ Å, and an average dipole $\langle\mu\rangle$ equal to 2.445, 2.450, 2.455, and 2.460 D. The results were 2.27, 2.24, 2.29, and $(2.21 \pm 0.07) \times 10^{-5}$ cm^2/s for the diffusion constant, and 74, 79, 79, 82 ± 10 for the dielectric constant. The third (and best) model was chosen. Table 4.I shows the final set of parameters and Table 4.II the liquid properties. We call this optimized model SWM4-DP, standing for "simple water model (4-site) with Drude polarizability."

We obtain better converged properties for the SWM4-DP model by performing

Table 4.I: Parameters of the SWM4-DP water model compared to the TIP3P model. For SWM4-DP, ℓ_{OM}, $q_D = -q_O$, ϵ, and σ_{OO} were adjusted to reproduce the liquid properties.

	TIP3P[a]	SWM4-DP[b]
ℓ_{OH} (Å)	0.9572	0.9572
θ_{HOH} (°)	104.52	104.52
ℓ_{OM} (Å)		0.238 08
q_O (e)	-0.834	-1.771 85
q_M (e)		-1.107 40
q_H (e)	0.417	0.553 70
q_D (e)		1.771 85
k_D (kcal/mol/Å^2)		1000
ϵ_{OO} (kcal/mol)	0.1521	0.205 68
σ_{OO} (Å)	3.1506	3.180 30

[a] From Ref. [5].
[b] This work.

10 additional independent simulations. With a total of 20 simulations, the diffusion constant is $(2.30 \pm 0.04) \times 10^{-5}$ cm^2/s and the dielectric constant is 79 ± 5. The convergence of the dielectric constant is shown in Fig. 4.2. For the SWM4-DP model, the Clausius–Mossotti equation (4.5) gives $\epsilon_\infty = 1.52$. A more direct estimation of ϵ_∞ can be obtained by computing the dielectric contribution of the dipole fluctuations for many frozen nuclear configurations,

$$\epsilon_\infty = 1 + \frac{4\pi}{3\langle V\rangle_{(f)} k_B T_\star} \left(\langle \mathbf{M}^2 \rangle_{(f)} - \langle \mathbf{M} \rangle^2_{(f)} \right), \tag{4.10}$$

where $\langle \cdots \rangle_{(f)}$ indicates an average over induced-dipole fluctuations only. This average is obtained for Drude oscillators moving according to Langevin dynamics at temperature $T_\star$ and for all nuclei frozen in a configuration extracted from the molecular dynamics. Equation (4.10) exploits the fact that, under a linear response theory, the classical fluctuations of the induced dipoles are a direct measure of the total polarizability A of the box. This treatment is equivalent to the approach of

Table 4.II: Parameters and properties of a selection of water models. The components of the traceless quadrupole Q are defined using the MO direction as x-axis, the H_1H_2 direction as y-axis, and the perpendicular direction as z-axis.

	Expt.[a]	TIP3P[b]	RPOL[c]	TIP4P-FQ[d]	POL[e]	MCDHO[f]	SW-RIGID[g]	POL5/TZ[h]	COS/B2[i]	SWM4-DP[j]
ℓ_{OH} (Å)	0.9572	0.9572	1.0000	0.9572	0.9572	0.9590	0.9572	0.9572	1.0000	0.9572
θ_{HOH} (°)	104.52	104.52	109.50	104.52	104.52	104.83	104.52	104.52	109.47	104.52
ℓ_{OM} (Å)				0.15	0.215		0.1374			0.238 08
μ_0 (D)	1.85	2.347	2.024	1.85	1.848	1.85	1.85	1.854	2.07	1.85
α (Å^3)	1.44		1.47	1.123	1.444	1.352	1.47	1.291	0.93	1.042 52
Q_{xx} (DÅ)	−0.134	−0.080	−0.263	−0.098	−0.188	−0.24	−0.13	−0.002	−0.27	−0.2421
Q_{yy} (DÅ)	2.626	1.762	1.885	1.882	2.235	2.67	2.63	2.337	1.93	2.4068
Q_{zz} (DÅ)	−2.493	−1.681	−1.623	−1.785	−2.047	−2.44	−2.50	−2.335	−1.66	−2.1647
U_{dimer} (kcal/mol)	−5.4	−6.50	−5.5	−4.5	−4.69	−5.00	−5.24	−4.96	−5.57	−5.18
d_{OO} (Å)	2.98	2.74	2.82	2.92	2.87	2.92	2.93	2.896	2.79	2.82
θ_A (°)	58	20		27		56	56	63	51	70
μ_{dimer} (D)	2.643	3.866		3.430	2.05	2.681	2.47	2.435	3.76	2.087
Δu (kcal/mol)	−9.92	−9.82	−9.94	−9.9	−9.84	−10.40	−9.96	−9.92	−9.97	−9.927
$\langle v \rangle$ (Å^3)	30.0	29.9	30.1	30.0	30.1	29.33	30.04	30.00	30.1	29.93
$\langle \mu \rangle$ (D)		2.35	2.62	2.62	2.75	3.01	2.61	2.712	2.62	2.456
D (10^{-5}cm^2/s)	2.3	5.1	2.4 ± 0.3	1.9	2.1 ± 0.1		3.2	1.81	2.6	2.30 ± 0.04
ϵ	78.4	92 ± 5	106 ± 18	79 ± 8			97	98 ± 8	121.6	79 ± 5
τ_D (ps)	8.3		11 ± 4	8			7.3		14.9	9.4 ± 0.7
γ (dyn/cm)	72.0	52.7			92 ± 5					66.9 ± 0.9
$\Delta\phi$ (mV)		−500			−500					−540

[a] Molecular geometry from Ref. [52], μ_0 from Ref. [7], quadrupole from Ref. [73], dimer properties from Ref. [67], Δu from Ref. [59] using Eq. (4.3), D from Ref. [74], ϵ from Ref. [49], τ_D a consensus from Refs. [75, 76, 77].

[b] CHARMM version of the original TIP3P (Ref. [5]). The liquid properties were computed from 10 simulations of 125 molecules in the isothermal–isobaric ensemble, using particle-mesh Ewald and a 9 Å Lennard-Jones cutoff. The surface tension γ and the surface potential $\Delta\phi$ are from Ref. [78].

[c] Reference [43]. The diffusion constant D, the dielectric constant ϵ, and the Debye relaxation time τ_D are from Ref. [79].

[d] Reference [20]. Constant volume simulations.

[e] Reference [19]. The surface potential $\Delta\phi$ is from Reference [80].

[f] Reference [21].

[g] Reference [46].

[h] Reference [47]. The 5-site model has explicit lone pairs at ℓ_{OL}= 0.5 Å.

[i] Reference [48].

[j] This work.

Neumann and Steinhauser [60]. The temperature of the Langevin simulation is irrelevant and gives almost identical results whether $T_\star = 1$ K or 300 K. Averaging ϵ_∞ for ten uncorrelated nuclear conformations gives $\epsilon_\infty = 1.68$ for the SWM4-DP model, a value that is slightly higher than the mean-field estimate predicted by the Clausius–Mossotti equation (1.52) but lower than the experimental estimate (1.79) [61].

It is of interest to examine the sensitivity of the parameters in the context of a given model. Figure 4.3 shows some projections of the response surface $\mathbf{P}$. It is observed that the interaction energy of the water dimer, U_{dimer}, and the oxygen–oxygen equilibrium distance d_{OO} have almost opposite effects on the average molecular energy Δu and volume $\langle v \rangle$ in the liquid. Generating different models with variations only in these two quantities, which are closely related to the parameters of the Lennard-Jones potential, ϵ_{OO} and σ_{OO}, is a very inefficient way to span the $\langle v \rangle$–Δu plane. These figures show why models with inappropriate values for α and ℓ_{OM} cannot be "fixed" by adjusting U_{dimer} and d_{OO}: They would require unphysical values of U_{dimer} and d_{OO}, that would yield bad liquid structure and bad self-diffusion constant. On the other hand, α cannot be freely adjusted to get correct $\langle v \rangle$ and Δu, because it has a major influence on the average dipole in the liquid, hence on the dielectric constant (see Fig. 4.3b). From the vector diagrams of Fig. 4.3, it is clear that treating the molecular polarizability as a parameter removes some important collinearities and that the four parameters of the model provide altogether enough freedom to adjust all four target liquid properties.

The oxygen–oxygen, oxygen–hydrogen, and hydrogen–hydrogen radial distribution functions of the final model are shown in Fig. 4.4. The agreement with experimental radial distribution functions is generally excellent, which is in itself remarkable since no specific adjustments were made to reproduce this experimental data. The best models yielding good properties for the dimer and the average enthalpy and density naturally give reasonable liquid structure. The narrow shape of the first peak in the $g_{\text{OO}}(r)$ radial distribution function is probably due to the steepness of the Lennard-Jones repulsive potential and the lack of intramolecular

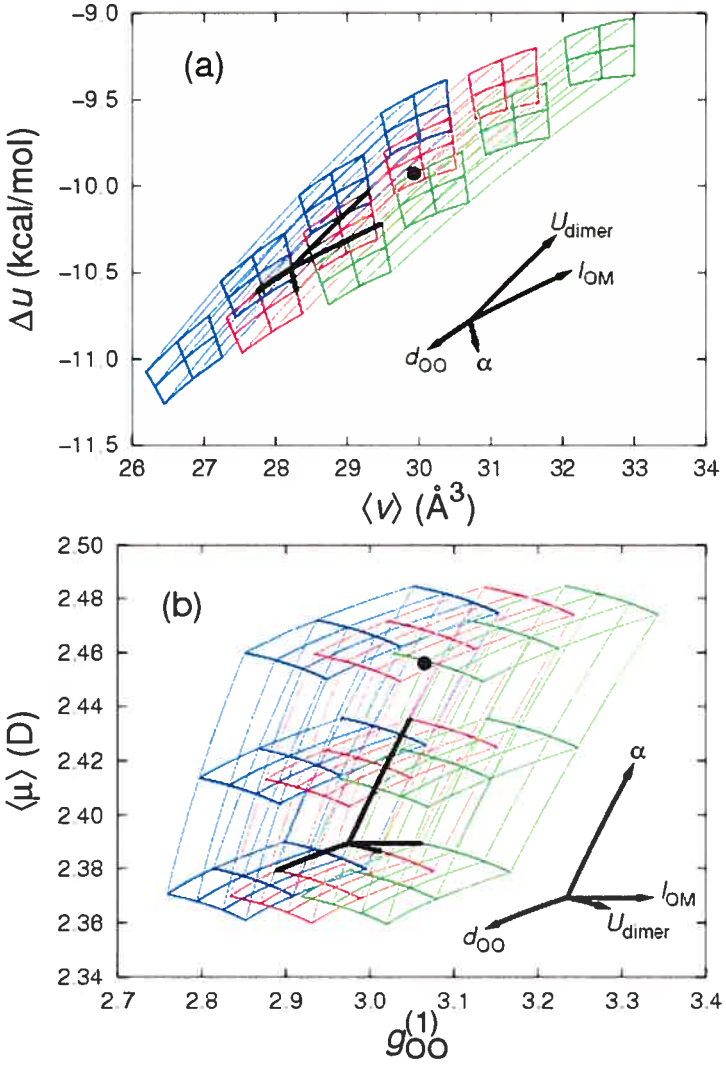

Figure 4.3: Projections of the response surface $\mathbf{P}(\ell_{\mathrm{OM}}, \alpha, U_{\mathrm{dimer}}, d_{\mathrm{OO}})$ (a) in the $\langle v \rangle$-Δu plane and (b) in the $g^{(1)}_{\mathrm{OO}}$-$\langle \mu \rangle$ plane: in blue for $\ell_{\mathrm{OM}} = 0.23$ Å, in red for $\ell_{\mathrm{OM}} = 0.24$ Å, and in green for $\ell_{\mathrm{OM}} = 0.25$ Å. Each $3 \times 4 \times 3$ grid is showing how the properties are varying according to the three remaining parameters. The arrows show how to follow the grid for each parameter: $\alpha = 0.95, 1.00, 1.05$ Å^3; $U_{\mathrm{dimer}} = -5.4, -5.3, -5.2, -5.1$ kcal/mol; and $d_{\mathrm{OO}} = 2.81, 2.82, 2.83$ Å. The black dot shows the SWM4-DP liquid properties.

flexibility. For the same reason, the first intramolecular peak of the $g_{OH}(r)$ distribution is slightly outward ($r_{OH}^{(1)} = 1.85$ Å instead of 1.78 Å). The height of the first peak in the oxygen–oxygen radial distribution function is equal to 3.065. This value is somewhat high but almost within experimental errors ($g_{OO}^{(1)} = 2.7 \pm 0.3$ from the neutron diffraction experiment [53]). The position and height of the first peak are necessary to match the experimental value of the self-diffusion constant in the liquid. It would be possible to reproduce the radial distribution function exactly by sacrificing the diffusion constant. Having $g_{OO}^{(1)}$ around 2.7 or lower gives $D > 4 \times 10^{-5}$ cm^2/s, much higher than 2.3×10^{-5} cm^2/s, the experimental value [74]. In part, the intrinsic limitations of the conventional Lennard-Jones potential appear to be responsible for the inaccuracies of the average liquid structure. The harsh $1/r^{12}$ repulsion may not allow the position of the first peak in the oxygen–oxygen radial distribution function, $r_{OO}^{(1)}$, to differ significantly from d_{OO}, the oxygen–oxygen equilibrium distance in the water dimer. In the SWM4-DP model, d_{OO} is equal to 2.82 Å and the first peak is at 2.79 Å, smaller by only 0.03 Å. In contrast, while the position of the first peak determined experimentally is 2.73 Å, the equilibrium oxygen–oxygen distance in the water dimer is around 2.91 Å based on high level *ab initio* calculations [81], perhaps even as large as 2.98 Å according to gas phase measurements [67]. This implies that there should be a shift on the order of 0.2 Å between $r_{OO}^{(1)}$ and d_{OO}. A softer repulsive potential may allow for such a difference to be reproduced. In the SWM4-DP model, both the large $r_{OO}^{(1)}$ and the relatively flat second peak of the $g_{OO}(r)$ distribution are a necessary tradeoff to reproduce the diffusion constant. Nevertheless, the coordination number, defined as

$$N_c = 4\pi\rho \int_0^{r_{min}} dr\, g_{OO}(r), \tag{4.11}$$

is 4.63 for the SWM4-DP model (using $r_{min} = 3.35$ Å), in good agreement with the recent x-ray diffraction result of $N_c = 4.7$ [82, 54].

The permanent dipole of the isolated molecule, μ_0, is set to exactly match the

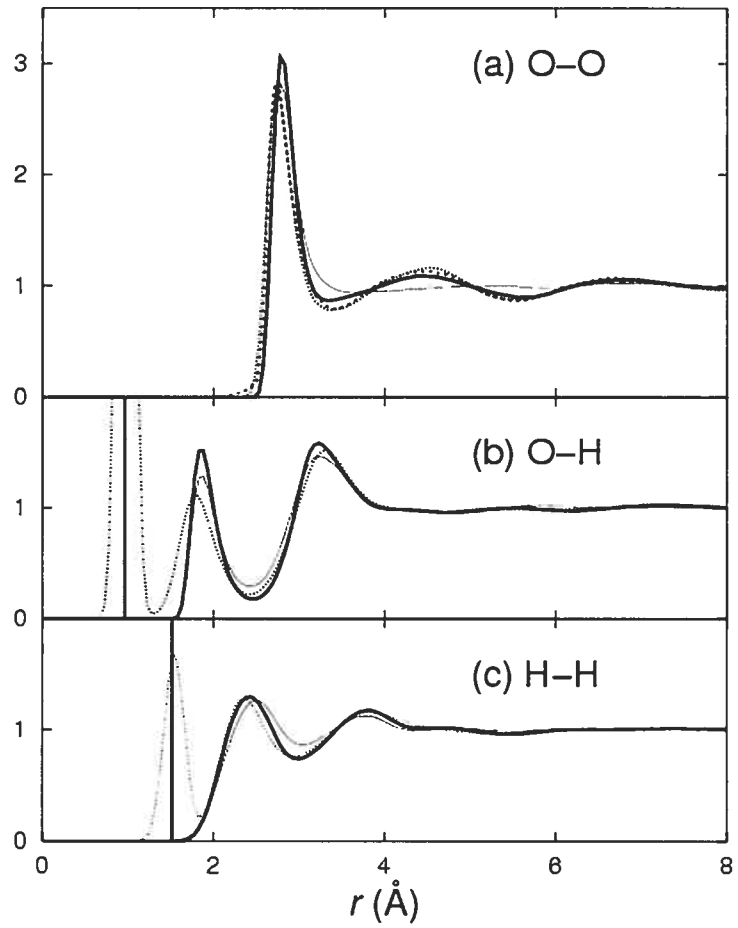

Figure 4.4: (a) Oxygen–oxygen, (b) oxygen–hydrogen, and (c) hydrogen–hydrogen radial distribution functions for SWM4-DP (thick solid lines), compared to TIP3P distributions (thin solid lines) and experimental distributions: neutron diffraction data of Soper *et al.* (Ref. [53]) for (a), (b), and (c) (dotted lines); x-ray diffraction data of Hura *et al.* (Ref. [54]) for (a) (dashed lines).

experimental gas phase value. The position of the M site is the only remaining free parameter which affects the permanent molecular quadrupole of an isolated molecule. But given the simplicity of the charge distribution, each individual component of the quadrupole cannot be reproduced exactly. Q_{xx} would be reproduced with $\ell_{\mathrm{OM}} = 0.1789$ Å, Q_{yy} with $\ell_{\mathrm{OM}} = 0.2642$ Å, and Q_{zz} with $\ell_{\mathrm{OM}} = 0.2928$ Å. As shown in Table 4.II, the optimal value of 0.23808 Å for ℓ_{OM} yields a reasonable compromise for each component of the molecular quadrupole, suggesting that a realistic molecular quadrupole is important for obtaining good liquid properties. This was also noted previously by Chialvo and Cummings [83]. Parenthetically, it may be noted that the present value of ℓ_{OM} is larger than for the TIP4P model (0.15 Å) [5]. The TIP4P-FQ model [20], a polarizable water model with charge-transfer, left the distance ℓ_{OM} unchanged from its value in the TIP4P model. For the polarizable TIP4P-FQ model, this distance is perhaps sub-optimal, since a larger distance of 0.23 yields a permanent quadrupole that is closer to the experimental value (see Table 4.II). For the SWM4-DP model, a larger ℓ_{OM} also improves the hydrogen bond geometry of the water dimer. The acceptor angle θ_{A} is 70° instead of 27° for TIP4P-FQ [20], closer to the experimental value $58 \pm 6°$ [67]. This 70° may be a little too high, as it makes the total dipole of the dimer, μ_{dimer}, smaller than the experimental value (see Table 4.II). Such an open angle reproduces the tetrahedral geometry of hydrogen bonding in water and is likely to give a correct liquid structure. Indeed, the liquid structure of SWM4-DP shows a significant improvement over that of TIP3P (see Fig. 4.4): The $g_{\mathrm{OO}}(r)$ distribution has a distinct second peak and even a third peak, and the first intramolecular peak of the $g_{\mathrm{HH}}(r)$ distribution has the correct position ($r_{\mathrm{HH}}^{(1)} = 2.43$ Å). These are two signs of a better tetrahedral structure.

The polarizability of the optimal model is 1.04 Å^3, significantly smaller than the value of 1.44 Å^3 in the gas phase. Although this observation is undoubtedly dependent on the details of the present model, it is strongly supported by a very thorough exploration of the parameter space. Figure 4.5 shows how the dielectric constant is related to the average molecular dipole. This trend was initially dis-

played by Sprik [9, 10, 11, 84], from which it was concluded that an average dipole about 2.4 to 2.6 D was needed to correctly reproduce ϵ using classical molecular dynamics simulations. The SWM4-DP model has $\langle\mu\rangle = 2.456$ D, a value smaller than that of any other polarizable model, although still within the broad range indicated by *ab initio* or density functional Car-Parrinello simulations on small water boxes: from 2.43 (Ref. [84]) to 3.0 D [12]. An unequivocal value of $\langle\mu\rangle$ is difficult to calculate from a pure quantum mechanical approach because there is no unique way to partition the electronic density among the molecules. In addition, the computational cost of the method makes it difficult to simulate a bulk liquid during a time sufficiently long to sample many representative conformations. Hybrid quantum mechanical/molecular mechanical (QM/MM) descriptions seem to yield intermediate dipole values: 2.90 D for a quantum water molecule surrounded by four classical molecules placed in a spherical dielectric cavity [85], and 2.71 D for a quantum water molecule surrounded by polarizable classical molecules [86]. Interestingly, consistent values are obtained from *ab initio* calculations using liquid conformations obtained from classical simulations of the TIP3P model [5] or the TIP5P model [87]: 2.65 (Ref. [71]) and 2.60 ± 0.14 D [88], respectively. A gas-phase-like α yields overestimated average dipole and dielectric constant. For example, the RPOL model [43, 79], with a permanent dipole of 2.024 D and a polarizability of 1.47 $\mathrm{\AA}^3$, and the SW-RIGID model [46], with $\mu_0 = 1.85$ and $\alpha = 1.47$ $\mathrm{\AA}^3$, have both high dielectric constants (see Table 4.II). In contrast, the TIP4P-FQ model [20] ($\mu_0 = 1.85$ D and $\alpha = 1.123$ $\mathrm{\AA}^3$), the PPC model [45] ($\mu_0 = 2.14$ and $\alpha = 0.56$ $\mathrm{\AA}^3$), and the present SWM4-DP model ($\mu_0 = 1.85$ D and $\alpha = 1.04$ $\mathrm{\AA}^3$) all have correct dielectric constants (see Fig. 4.5). Therefore, the fact that a reduced molecular polarizability is essential to reproduce the dielectric constant of the liquid appears to be a robust conclusion of the current effort. It is likely that this condensed-phase renormalization of the polarizability arises from the energy cost of overlapping electronic clouds due to Pauli's exclusion principle and that it should apply to other polar molecules.

The average molecular dipole $\langle\mu\rangle = 2.456$ D in a bulk liquid of SWM4-DP

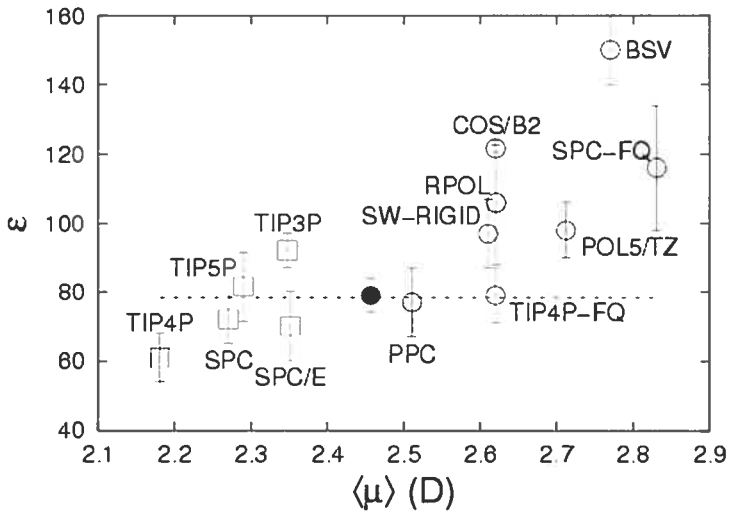

Figure 4.5: Dielectric constants for selected water models. Nonpolarizable models (squares): TIP4P (Refs. [5] and [89]), SPC (Refs. [6] and [89]), TIP5P (Ref. [87]), TIP3P (Ref. [5]) (ϵ computed by the authors), and SPC/E (Ref. [89]). Polarizable models (circles): SWM4-DP (this work, full circle), PPC (Ref. [45]), SW-RIGID (Ref. [46]), TIP4P-FQ (Ref. [20]), RPOL (Refs. [43] and [79]), COS/B2 (Ref. [48]), POL5/TZ (Ref. [47]), BSV (Ref. [44]), and SPC-FQ (Ref. [20]). Values with no reported uncertainties were assigned ± 10. The horizontal line shows the experimental value $\epsilon = 78.4$.

water molecules comprises the permanent dipole 1.85 D, plus an induced dipole of 0.606 D. With the current choice of spring constant to attach the Drude particle to the oxygen, this corresponds to a displacement of about 0.07 Å, away from the oxygen site toward the M site. Because of the symmetry of the molecule, the average induced dipole is parallel to the MO axis.

The TIP4P-FQ, SW-RIGID, and SWM4-DP models have very similar electrostatic properties (same permanent gas-phase dipole and same geometry), but have contrasting features that suggest why correct dielectric constants can be obtained for diverse average dipoles $\langle\mu\rangle$ (SWM4-DP *versus* TIP4P-FQ), and why similar average dipoles can produce different dielectric constants (TIP4P-FQ *versus* SW-RIGID). Both SWM4-DP and TIP4P-FQ have correct dielectric constants but, while the SWM4-DP model has a perfectly isotropic polarizability ($\alpha_x = \alpha_y = \alpha_z = 1.04$ Å^3), the TIP4P-FQ model (a charge-transfer model) has a very anisotropic molecular polarizability ($\alpha_x = 0.82$ Å^3, $\alpha_y = 2.55$ Å^3, $\alpha_z = 0$ Å^3) and a small molecular quadrupole compared with the experimental gas-phase (see Table 4.II). The SW-RIGID model, on the other hand, produces the save average dipole as TIP4P-FQ, but has a nonrenormalized, isotropic polarizability. Its electrostatics differs from SWM4-DP mostly because of a larger α, which explains the larger average dipole and the larger dielectric constant. Ultimately, these differences are creating specific tetrahedral arrangements [90] and specific orientational fluctuations [91] that have an impact on the value of the dielectric constant.

The Debye relaxation time of the SWM4-DP model is 9.4 ± 0.7 ps, in reasonable agreement with the experimental values 8.27 ps [75], 8.32 ps [76], and 8.40 ps [77]. This result was obtained by fitting a single exponential to an average of the 20 autocorrelation functions $\Phi(t)$ for t between 0 and 40 ps.

4.3.2 Interfacial properties

Nonpolarizable water models with effective constant dipole may be reasonably valid for representing the interactions between molecules immersed in the bulk region, where they are assumed to be uniformly polarized by their surrounding.

Because such nonpolarizable models are designed and optimized to reproduce the average properties of an isotropic bulk liquid, their ability to represent markedly anisotropic environments is expected to be limited. Such a nonisotropic environment is, perhaps, well-exemplified in the case of the vacuum-liquid interface. For this reason it is of interest to examine the properties of the vacuum-liquid interface with the SWM4-DP model.

The surface tension extracted from the five slab simulations of 1000 ps each is 66.9 ± 0.9 dyn/cm. This value compares well with the experimental surface tension of 72.0 dyn/cm and constitutes an improvement over the TIP3P model, that has a surface tension of 52.7 dyn/cm [78]. The anisotropic orientation of the water molecules at the interface gives rise to a nonzero electrostatic potential difference between the vacuum and the liquid. Figure 4.6 shows the interfacial potential across the z-axis, $z = 0$ being the position of the center of mass of the water slab. As a test charge crosses the air/liquid interface (i.e., as z goes from -13 to -6 Å), the potential profile displays a rapid drop from $\phi = 0$ to $\phi = -540$ mV. This potential drop reflects the average inward ordering of the molecular dipoles at the surface of the slab [92]. The interfacial potential $\Delta\phi$ cannot be measured directly, and experimental propositions for its magnitude—even for its sign—vary greatly (see Ref. [93] and references therein). For the present model, we get $\Delta\phi = -540$ mV, in good agreement with the nonpolarizable TIP3P and TIP4P models [5] (-500 mV [78] and -500 mV [80], respectively) and with the polarizable POL model (-500 mV) [80]. In contrast to these models, the SWM4-DP model displays a notable potential well at $z = -8$ Å (see Fig. 4.6). At this position, the interface has essentially the density of bulk water (data not shown), but the SWM4-DP molecules still display a significant orientational order. The order parameter $\langle \rho(z) \cos\theta(z)/\rho_{\text{bulk}} \rangle$ [78], where θ is the angle between the $\overrightarrow{\text{MO}}$ directors and the normal to the interface, has a significant negative value for z between -11 and -6 Å (with an extremum -0.08 at $z = -8.5$ Å). This indicates that the oxygen atoms at the interface are pointing toward the gas phase, in agreement with recent *ab initio* simulations [94]

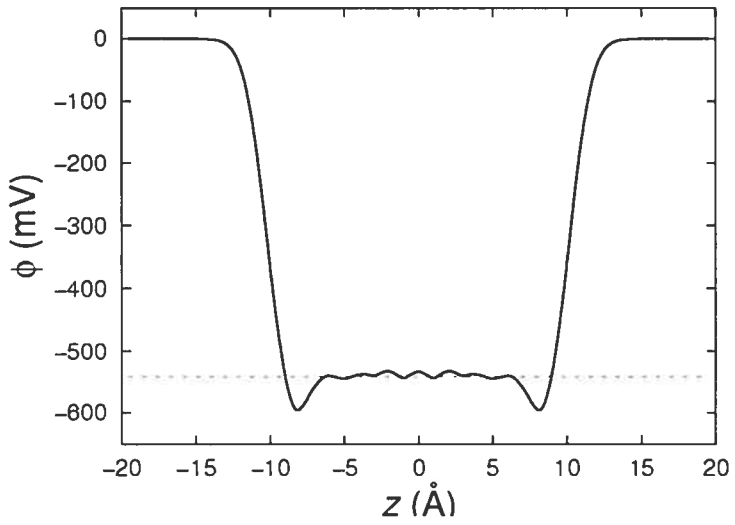

Figure 4.6: Interfacial potential for the SWM4-DP model. The potential drop $\Delta\phi$ is represented by a dashed line at -540 mV.

[After the present article was published, Kuo and Mundy [95] reported *ab initio* molecular dynamics on an air/water interface system of 216 molecules (instead of 32 for Vassilev *et al.* [94]). Their simulation shows that water molecules at the interface participate in hydrogen bonds slightly more as single proton donors than as single proton acceptors (66% versus 56%) and slightly less as double-donors than as double-acceptors (14% versus 20%). This indicates an excess of dangling bonds at the interface. It is difficult to compare those results with our own analysis of molecular orientation, because the connexion between the nature of the hydrogen bonding of the surface molecules and their orientation relative to the plane of the interface is unknown.]

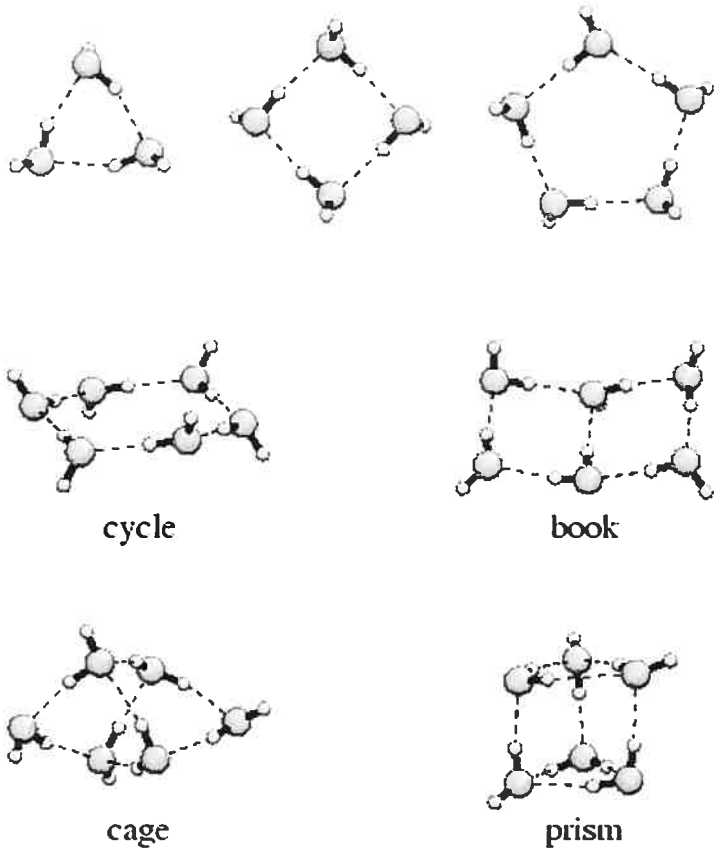

Figure 4.7: Illustration of the SWM4-DP water clusters reported in Table 4.III. For clarity, the M sites are not shown.

4.3.3 Water clusters

Although the SWM4-DP model is optimal for the bulk liquid properties, we expect the energetics of small water clusters ($N = 3$ to 6) to be significantly better than with a nonpolarizable model like TIP3P [5]. The total binding energies of SWM4-DP clusters are presented in Table 4.III for the cyclic trimer, cyclic tetramer, cyclic pentamer, and some standard hexamers: cycle, book, cage, and prism. Each or the hexamers has a definite hydrogen-bond topology but has many distinct stable conformers defined by how each water molecule participates in the hydrogen bonding. In Table 4.III and Fig. 4.7, we report only the lowest-energy SWM4-DP conformers.

The binding energies of the SWM4-DP clusters are systematically less negative

Table 4.III: Binding energies for water clusters. N_{Hb} is the number of hydrogen bonds. Energies in parenthesis refer to a conformer different from the SWM4-DP conformer.

N	Isomer	N_{Hb}	*Ab initio*[a]	TIP3P[b]	TIP4P-FQ[c]	MCDHO[d]	POL5/TZ[c]	SWM4-DP[e]
3	cycle	3	−15.8	−17.75	−12.58	−13.98	−13.42	−14.44
4	cycle	4	−27.6	−29.47	−23.64	−25.86	−25.53	−25.22
5	cycle	5	−36.3	−38.91	−32.95	−35.30	−34.11	−33.43
6	cycle	6	−44.8	−47.64	−41.37	−44.26	−41.79	−41.27
6	book	7	−45.6	−48.23	−40.15	(−43.98)	−42.46	−42.62
6	cage	8	(−45.8)	−47.45	(−39.30)	−43.69	(−41.78)	−43.26
6	prism	9	−45.9	−46.69		(−44.12)		−42.96

[a] From Ref. [96].

[b] CHARMM version of the original TIP3P (Ref. [5]). The lowest-energy TIP3P conformers are equivalent to those found for SWM4-DP, except for the trimer and tetramer being flat.

[c] From Ref. [47].

[d] From Ref. [21].

[e] This work.

than the *ab initio* results of Xantheas *et al.* [96] (MP2 theory converged to the complete basis set limit), also presented in Table 4.III. This problem seems common for polarizable water models (see Table 4.III): TIP4P-FQ [20] and POL5/TZ [47] were optimized for liquid properties, and are significantly less negative than SWM4-DP for the cage and prism isomers. The MCDHO model [21] was parametrized from *ab initio* data for the dimer and is better for small water clusters, but has inferior liquid properties. The problem is opposite for the nonpolarizable TIP3P model [5], where the enhanced charges overestimate the stability of the water clusters. For the SWM4-DP model, the deviation from *ab initio* becomes smaller—per hydrogen bond—as the water molecules are more coordinated, that is, as the cluster is more representative of bulk water. Because each molecule is accepting only one hydrogen bond, the cyclic SWM4-DP clusters have the largest discrepancies: 0.45 kcal/mol/H-bond for the trimer, 0.60 kcal/mol/H-bond for the tetramer, 0.58 kcal/mol/H-bond for the pentamer, and 0.59 kcal/mol/H-bond for the cyclic hexamer. The errors on the hexamer binding energies decrease for more compact structures: 0.43 kcal/mol/H-bond for the book isomer, 0.32 kcal/mol/H-bond for the cage isomer, and 0.33 kcal/mol/H-bond for the prism isomer. Apparently, the renormalized polarizability of the SWM4-DP model is not optimal for molecules with less than four nearest neighbors. For rigid molecules, not allowed to relax the OH bond lengths and HOH bond angles, a smaller renormalization factor would be more appropriate. Indeed, a modified version of SWM4-DP with a nonrenormalized polarizability $\alpha = 1.44$ Å^3 gives significantly lower energies for the equivalent structures: -15.56 kcal/mol for the trimer, -28.19 kcal/mol for the tetramer, -38.42 kcal/mol for the pentamer, -48.24 kcal/mol for the cyclic hexamer, -48.52 kcal/mol for the book hexamer, -47.89 kcal/mol for the cage hexamer, and -47.68 kcal/mol for the prism hexamer. A challenge would be to design a model able to capture the progressive renormalization of α in increasingly large clusters, but that is beyond the scope of the present effort.

In agreement with recent *ab initio* studies [96, 97], the SWM4-DP model designates a nonplanar hexamer structure (a cage conformer) as the global energy min-

imum. This is contrasting with other polarizable models, such as TIP4P-FQ [20], MCDHO [21], and POL5/TZ [47], that give planar global minima (see Table 4.III). This is an important quality of the SWM4-DP model because the water hexamer is the smallest cluster for which the global energy minimum has a nonplanar hydrogen-bond topology [8]. The hexamer binding energy sequence for the SWM4-DP model is also consistent with *ab initio* [96, 97]: $E_{\text{prism}} \approx E_{\text{cage}} < E_{\text{book}} < E_{\text{cycle}}$. The global SWM4-DP energy minimum corresponds to the cage structure illustrated in Fig. 4.7, and E_{cage} is actually 0.30 kcal/mol less than E_{prism}. The optimal SWM4-DP cage structure is equivalent to the reported *ab initio* cage (du){1} conformer of Losada and Leutwyler [97]. The second most stable SWM4-DP structure, at only 0.12 kcal/mol higher in energy, is equivalent to the *ab initio* cage conformer reported by Xantheas *et al.* [96] ((uu){1} according to Gregory and Clary's nomenclature [98]). The third most stable structure is the prism conformer, followed by the book structure (see Table 4.III and Fig. 4.7). Thus, although there is a small systematic underestimation of the water cluster binding energies, the relative energies are in good agreement with high-level *ab initio* data.

4.4 Conclusion

The proposed water model uses Drude polarizability as the only new ingredient. It keeps a point-charge electrostatic representation and the common Lennard-Jones potential between the oxygen atoms. This simple polarizable model provides the essential electrostatics to reproduce the properties of the liquid phase at room temperature and pressure: the gas-phase molecular dipole (1.85 D), a good molecular quadrupole, and a smaller renormalized molecular polarizability.

The model uses an effective value of 1.04 Å^3 for the molecular polarizability, smaller than the experimental gas phase value of 1.44 Å^3. Such a renormalized value, which reflects the influence of the microscopic factors opposing polarization in the condensed phase, is necessary to obtain accurate properties of the liquid. This implies that the model retains some empirical character, i.e., some parameters

are set to effective values designed to yield optimal properties for the liquid. It may be possible to design more realistic models by introducing some short range interactions between induced dipoles to account for the microscopic factors opposing polarization in the condensed phase. Nonetheless, this may not be necessary for our purpose. In the same spirit as the previous generation of nonpolarizable models with effective interactions, it can be realistically hoped that simple models with an effective renormalized polarizability can accurately simulate biomolecular systems. Nonpolarizable models, in which the permanent dipole of the isolated molecule has been enhanced to simulate the bulk phase reasonably well [5, 6], owe much of their success to the fact that an effective dipole makes sense in most relevant situations because the polarity of the environment is similar. In the current polarizable model, that aspect of reality is accurately simulated, i.e., the correct permanent dipole can be attributed to the isolated molecule and the enhanced polarization arises naturally in the bulk phase. However, the polarizability is reduced and set to an effective empirical value. Nevertheless, such models can represent a wide range of situations accurately since the microscopic factors giving rise to the reduction in the molecular polarizability of a molecular moiety (Pauli's exclusion principle and the overlap of electronic clouds) may be assumed to have a similar effect in most dense systems.

The SWM4-DP model uses point charges and Lennard-Jones potentials because they are simple and give satisfactory properties. The optimal SWM4-DP model indicates that the "12-6" Lennard-Jones potential gives rise to a hard repulsion that cannot correctly reproduce both the oxygen–oxygen distance in the dimer, d_{OO}, and the position of the first peak in the oxygen–oxygen pair correlation function, $r_{OO}^{(1)}$. The SWM4-DP model compromises on both: d_{OO} is 2.82 Å instead of 2.91 Å [81] and $r_{OO}^{(1)}$ is 2.79 Å instead of 2.73 Å [53, 54]. Moreover, to get a correct self-diffusion constant, the SWM4-DP model somewhat compromises on the liquid structure: having $r_{OO}^{(1)}$ closer to the experimental value would have improved both the position and the amplitude of the second peak of the oxygen–oxygen radial distribution, but would have yielded a less diffusive liquid. And the $g_{OO}^{(1)}$ would

have been even larger. To get a correct dielectric constant, the SWM4-DP model needs a relatively small average condensed-phase dipole and significant adjustment of the molecular polarizability. Some of these compromises could probably be alleviated by adopting a rigid molecular geometry adapted to the condensed phase, but the largest contribution is most likely due to the overall shape of the short-range and midrange interactions. For these reasons, we conclude that it might be worth considering alternative, *softer* interactions such as the Buckingham "exp-6" potential instead of Lennard-Jones potential and smeared charges instead of point charges in future models of water. In order to keep a correct self-diffusion constant along with a closer and lower first $g_{OO}(r)$ peak, such "softer" models should allow the oxygen atoms to move more freely within the first hydration shell.

The SWM4-DP model is optimal for the purpose of carrying classical molecular dynamics or Monte Carlo simulations of bulk solvent systems at room temperature and pressure. It may not yield accurate results if nonclassical effects arising from the quantum nature of the nuclei were taken into account explicitly, e.g., using simulations based on discretized Feynmann path integral [41]. Accordingly, it may need to be reparametrized if the internal geometry was allowed to be flexible. The SWM4-DP model has a simple functional form, yet it accurately reproduces the essential properties of liquid water under normal conditions. This makes it well suited for simulating the hydration of charged or highly polar molecules, including large biomolecules.

Acknowledgements

B. R. is supported by NSF Grant No. 0110847. A. D. M. is supported by NIH Grant No. GM51501. Helpful discussions with R. Friesner and B. Berne are gratefully acknowledged. G. L. is grateful to Alain Caillé for his support.

References

[1] Oren M. Becker, Alexander D. MacKerell, Jr., Benoît Roux, and Masakatsu Watanabe (editors). Computational Biochemistry and Biophysics. Marcel Dekker, Inc., New York, 2001.

[2] Alexander D. MacKerell, Jr.. Computational Biochemistry and Biophysics, chapter 2. Marcel Dekker, Inc., New York, 2001.

[3] Thomas A. Halgren and Wolfgang Damm. Polarizable force fields. *Curr. Opin. Struct. Biol.*, 11 : 236–242, 2001.

[4] Steven W. Rick and Steven J. Stuart. Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. In Kenny B. Lipkowitz and Donald B. Boyd (editors) : Reviews in Computational Chemistry, volume 18, pages 89–146. Wiley-VCH, Hoboken, NJ, 2002.

[5] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2) : 926–935, 1983.

[6] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. Interaction models for water in relation to protein hydration. In B. Pullman (editor) : Intermolecular Forces, pages 331–342. Reidel, Dordrecht, 1981.

[7] Thomas R. Dyke and J. S. Muenter. Electric Dipole-Moments of Low J States of H_2O and D_2O. *J. Chem. Phys.*, 59(6) : 3125–3127, 1973.

[8] J. K. Gregory, D. C. Clary, K. Liu, M. G. Brown, and R. J. Saykally. The Water Dipole Moment in Water Clusters. *Science*, 275 : 814–817, 1997.

[9] Michiel Sprik. Hydrogen bonding and the static dielectric constant in liquid water. *J. Chem. Phys.*, 95(9) : 6762–6769, 1991.

[10] Jean-Christophe Soetens, Marília T. C. Martins Costa, and Claude Millot. Static dielectric constant of the polarizable NCC water model. *Mol. Phys.*, 94(3) : 577–579, 1998.

[11] Anders Wallqvist and Raymond D. Mountain. Molecular Models of Water : Derivation and Description, volume 13 of Reviews in Computational Chemistry, chapter 4, pages 183–247. Wiley-VCH, New York, 1999.

[12] Pier Luigi Silvestrelli and Michele Parrinello. Water Molecule Dipole in the Gas and in the Liquid Phase. *Phys. Rev. Lett.*, 82(16) : 3308–3311, 1999.

[13] Pier Luigi Silvestrelli and Michele Parrinello. Structural, electronic, and bonding properties of liquid water from first principles. *J. Chem. Phys.*, 111(8) : 3572–3580, 1999.

[14] Y. S. Badyal, M.-L. Saboungi, D. L. Price, S. D. Shastri, D. R. Haeffner, and A. K. Soper. Electron distribution in water. *J. Chem. Phys.*, 112(21) : 9206–9208, 2000.

[15] Anna V. Gubskaya and Peter G. Kusalik. The total molecular dipole moment for liquid water. *J. Chem. Phys.*, 117(11) : 5290–5302, 2002.

[16] I-Feng Kuo and Douglas J. Tobias. Electronic Polarization and Hydration of the Dimethyl phosphate Anion : An an Initio Molecular Dynamics Study. *J. Phys. Chem. B*, 105(24) : 5827–5832, 2001.

[17] Christian J. Burnham, Jichen Li, Sotiris S. Xantheas, and Maurice 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 Ih. *J. Chem. Phys.*, 110(9) : 4566–4581, 1999.

[18] M. W. Feyereisen, D. Feller, and D. A. Dixon. Hydrogen bond energy of the water dimer. *J. Phys. Chem.*, 100(8) : 2993–2997, 1996.

[19] Liem X. Dang and Tsun-Mei Chang. Molecular dynamics study of water clusters, liquid, and liquid-vapor interface of water with many-body potentials. *J. Chem. Phys.*, 106(19) : 8149–8159, 1997.

[20] Steven W. Rick, Steven J. Stuart, and Bruce J. Berne. Dynamical fluctuating charge force field : Application to liquid water. *J. Chem. Phys.*, 101(7) : 6141–6156, 1994.

[21] Humberto Saint-Martin, Jorge Hernández-Cobos, Margarita I. Bernal-Uruchurtu, Iván Ortega-Blake, and Herman J. C. Berendsen. A mobile charge densities in harmonic oscillators (MCDHO) molecular model for numerical simulations : The water-water interaction. *J. Chem. Phys.*, 113(24) : 10899–10912, 2000.

[22] Yi-Ping Liu, Kyungsun Kim, Bruce J. Berne, Richard A. Friesner, and Steven W. Rick. Constructing *ab initio* force fields for molecular dynamics simulations. *J. Chem. Phys.*, 108(12) : 4739–4755, 1998.

[23] Terry P. Lybrand and Peter A. Kollman. Water-water and water-ion potential functions including terms for many body effects. *J. Chem. Phys.*, 83(6) : 2923–2933, 1985.

[24] Michiel Sprik, Michael L. Klein, and Kyoko Watanabe. Solvent Polarization and Hydration of the Chlorine Anion. *J. Phys. Chem.*, 94(16) : 6483–6488, 1990.

[25] Liem X. Dang and David E. Smith. Molecular dynamics simulations of aqueous ionic clusters using polarizable water. *J. Chem. Phys.*, 99(9) : 6950–6956, 1993.

[26] Steven J. Stuart and Bruce J. Berne. Effects of Polarizability on the Hydration of the Chloride Ion. *J. Phys. Chem.*, 100(29) : 11934–11943, 1996.

[27] Douglas J. Tobias, Pavel Jungwirth, and Michele Parrinello. Surface solvation of halogen anions in water clusters : An *ab initio* molecular dynamics study of the $Cl^-(H_2O)_6$ complex. *J. Chem. Phys.*, 114(16) : 7036–7044, 2001.

[28] Sotiris S. Xantheas and Liem X. Dang. Critical Study of Fluoride-Water Interactions. *J. Phys. Chem.*, 100(10) : 3989–3995, 1996.

[29] Richard A. Bryce, Mark A. Vincent, Nathaniel O. J. Malcolm, Ian H. Hillier, and Neil A. Burton. Cooperative effects in the structuring of fluoride water clusters : Ab initio hybrid quantum mechanical/molecular mechanical model incorporating polarizable fluctuating charge solvent. *J. Chem. Phys.*, 109(8) : 3077–3085, 1998.

[30] Orlando M. Cabarcos, Corey J. Weinheimer, James M. Lisy, and Sotiris S. Xantheas. Microscopic hydration of the fluoride anion. *J. Chem. Phys.*, 110(1) : 5–8, 1999.

[31] Jiwon Baik, Jongseob Kim, D. Majumdar, and Kwang S. Kim. Structures, energetics, and spectra of fluoride-water clusters $F^-(H_2O)_n$, $n = 1$–6 : *Ab initio* study. *J. Chem. Phys.*, 110(18) : 9116–9127, 1999.

[32] William H. Robertson, Eric G. Diken, Erica A. Price, Joong-Won Shin, and Mark A. Johnson. Spectroscopic Determination of the OH^- Solvation Shell in the $OH^- \cdot (H_2O)_n$ Clusters. *Science*, 299 : 1367–1372, 2003.

[33] Felice C. Lightstone, Eric Schwegler, Randolph Q. Hood, François Gygi, and Giulia Galli. A first principles molecular dynamics simulation of the hydrated magnesium ion. *Chem. Phys. Lett.*, 343 : 549–555, 2001.

[34] I. Bakó, J. Hutter, and G. Pálinkás. Car-Parrinello molecular dynamics simulation of the hydrated calcium ion. *J. Chem. Phys.*, 117(21) : 9838–9843, 2002.

[35] Anders Wallqvist. Polarizable water at a hydrophobic wall. *Chem. Phys. Lett.*, 165(5) : 437–442, 1990.

[36] Kazi A. Motakabbir and Max L. Berkowitz. Liquid-vapor interface of TIP4P water : comparison between a polarizable and a nonpolarizable model. *Chem. Phys. Lett.*, 176(1) : 61–66, 1991.

[37] Axel Kohlmeyer, Wolfgang Witschel, and Eckhard Spohr. Molecular dynamics simulations of water/metal and water/vacuum interfaces with a polarizable water model. *Chem. Phys.*, 213 : 211–216, 1996.

[38] In-Chul Yeh and Max L. Berkowitz. Effects of the polarizability and water density constraint on the structure of water near charged surfaces : Molecular dynamics simulations. *J. Chem. Phys.*, 112(23) : 10491–10485, 2000.

[39] Michael H. New and B. J. Berne. Molecular Dynamics Calculation of the Effect of Solvent Polarizability on the Hydrophobic Interaction. *J. Am. Chem. Soc.*, 117(27) : 7172–7179, 1995.

[40] Steven W. Rick and B. J. Berne. Free Energy of the Hydrophobic Interaction for Molecular Dynamics Simulations : The Effects of Solute and Solvent Polarizability. *J. Phys. Chem. B*, 101(49) : 10488–10493, 1997.

[41] Robert A. Kuharski and Peter J. Rossky. Quantum Mechanical Contributions to the Structure of Liquid Water. *Chem. Phys. Lett.*, 103(5) : 357–362, 1984.

[42] Paul Drude. The Theory of Optics. Longmans, Green, and Co., New York, 1902. Translation by C. Riborg Mann and Robert A. Millikan.

[43] Liem X. Dang. The nonadditive intermolecular potential for water revised. *J. Chem. Phys.*, 97(4) : 2659–2660, 1992.

[44] John Brodholt, Marco Sampoli, and Renzo Vallauri. Parametrizing a polarizable intermolecular potential for water. *Mol. Phys.*, 86(1) : 149–158, 1995.

[45] Igor M. Svishchev, Peter G. Kusalik, Jian Wang, and Russel J. Boyd. Polarizable point-charge model for water : Results under normal and extreme conditions. *J. Chem. Phys.*, 105(11) : 4742–4750, 1996.

[46] Paul J. van Maaren and David van der Spoel. Molecular Dynamics Simulations of Water with Novel Shell-Model Potentials. *J. Phys. Chem. B*, 105(13) : 2618–2626, 2001.

[47] Harry A. Stern, F. Rittner, B. J. Berne, and Richard A. Friesner. Combined fluctuating charge and polarizable dipole models : Application to a five-site water potential function. *J. Chem. Phys.*, 115(5) : 2237–2251, 2001.

[48] Haibo Yu, Tomas Hansson, and Wilfred F. van Gunsteren. Development of a simple, self-consistent polarizable model for liquid water. *J. Chem. Phys.*, 118(1) : 221–234, 2003.

[49] D. P. Fernandez, Y. Mulev, A. R. H. Goodwin, and J. M. H. L. Sengers. A Database for the Static Dielectric-Constant of Water and Steam. *J. Phys. Chem. Ref. Data*, 24(1) : 33–69, 1995.

[50] William F. Murphy. The Rayleigh depolarization ratio and rotational Raman spectrum of water vapor and the polarizability components for the water molecule. *J. Chem. Phys.*, 67(12) : 5877–5882, 1977.

[51] Guillaume Lamoureux and Benoît Roux. Modeling induced polarization with classical Drude oscillators : Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.*, 119(6) : 3025–3039, 2003.

[52] Felix Franks (editor). Water : A Comprehensive Treatise, volume 1 : The Physics and Physical Chemistry of Water. Plenum Press, New York, 1972.

[53] A. K. Soper, F. Bruni, and M. A. Ricci. Site-site pair correlation functions of water from 25 to 400°C : Revised analysis of new and old diffraction data. *J. Chem. Phys.*, 106(1) : 247–254, 1997.

[54] Greg Hura, Jon M. Sorenson, Robert M. Glaeser, and Teresa Head-Gordon. A high-quality x-ray scattering experiment on liquid water at ambient conditions. *J. Chem. Phys.*, 113(20) : 9140–9148, 2000.

[55] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, 101(5) : 4177–4189, 1994.

[56] Glenn J. Martyna, Mark E. Tuckerman, Douglas J. Tobias, and Michael L. Klein. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.*, 87(5) : 1117–1157, 1996.

[57] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM : A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4 : 187–217, 1983.

[58] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald : An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12) : 10089–10092, 1993.

[59] Gábor Jancsó and W. Alexander van Hook. Condensed Phase Isotope Effects (Especially Vapor Pressure Isotope Effects). *Chem. Rev.*, 74(6) : 689–750, 1974.

[60] M. Neumann and O. Steinhauser. Computer Simulation and the Dielectric Constant of Polarizable Polar Systems. *Chem. Phys. Lett.*, 106(6) : 563–569, 1984.

[61] A. D. Buckingham. A Theory of the Dielectric Polarization of Polar Substances. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.*, 238(1213) : 235–244, 1956.

[62] M. Neumann and O. Steinhauser. On the Calculation of the Frequency-Dependent Dielectric Constant in Computer Simulations. *Chem. Phys. Lett.*, 102(6) : 508–513, 1983.

[63] M. P. Allen and D. J. Tildesley. Computer Simulation of Liquids. Clarendon Press, Oxford, 1987.

[64] J. G. Kirkwood and F. P. Buff. The statistical mechanical theory of surface tension. *J. Chem. Phys.*, 17(3) : 338–343, 1949.

[65] K. Ichikawa, Y. Kameda, T. Yamaguchi, H. Wakita, and M. Misawa. Neutron-diffraction investigation of the intramolecular structure of a water molecule in the liquid-phase at high-temperatures. *Mol. Phys.*, 73(1) : 79–86, 1991.

[66] Martin Karplus. Special Issue on Molecular Dynamics Simulations of Biomolecules. *Acc. Chem. Res.*, 35 : 321–323, 2002.

[67] Thomas R. Dyke, Kenneth M. Mack, and J. S. Muenter. The structure of water dimer from molecular beam electric resonance spectroscopy. *J. Chem. Phys.*, 66(2) : 498–510, 1977.

[68] L. A. Curtiss, D. J. Frurip, and M. Blander. Studies of Molecular Association in H_2O and D_2O Vapors by Measurement of Thermal-Conductivity. *J. Chem. Phys.*, 71(6) : 2703–2711, 1979.

[69] Akihiro Morita and Shigeki Kato. An *ab initio* analysis of medium perturbation on molecular polarizabilities. *J. Chem. Phys.*, 110(24) : 11987–11998, 1999.

[70] M. in het Panhuis, P. L. A. Popelier, R. W. Munn, and J. G. Ángyán. Distributed polarizability of the water dimer : Field-induced charge transfer along the hydrogen bond. *J. Chem. Phys.*, 114(8) : 7951–7961, 2001.

[71] Yaoquan Tu and Aatto Laaksonen. The electronic properties of water molecules in water clusters and liquid water. *Chem. Phys. Lett.*, 329 : 283–288, 2000.

[72] Akihiro Morita. Water Polarizability in Condensed Phase : *Ab initio* Evaluation by Cluster Approach. *J. Comput. Chem.*, 23(15) : 1466–1471, 2002.

[73] J. Verhoeven and A. Dymanus. Magnetic Properties and Molecular Quadrupole Tensor of Water Molecule by Beam-Maser Zeeman Spectroscopy. *J. Chem. Phys.*, 52 : 3222, 1970.

[74] K. Krynicki, C. D. Green, and D. W. Sawyer. Pressure and Temperature-Dependence of Self-Diffusion in Water. *Discuss. Faraday Soc.*, 66 : 199–208, 1978.

[75] Udo Kaatze. Complex Permittivity of Water as a Function of Frequency and Temperature. *J. Chem. Eng. Data*, 34 : 371–374, 1989.

[76] J. Barthel, K. Bachhuber, R. Buchner, and H. Hetzenauer. Dielectric spectra of some common solvents in the microwave region. Water and lower alcohols. *Chem. Phys. Lett.*, 165(4) : 369–373, 1990.

[77] J. T. Kindt and C. A. Schmuttenmaer. Far-Infrared Dielectric Properties of Polar Liquids Probed by Femtosecond Terahertz Pulse Spectroscopy. *J. Phys. Chem.*, 100(24) : 10373–10379, 1996.

[78] Scott E. Feller, Richard W. Pastor, Atipat Rojnuckarin, Stephen Bogusz, and Bernard R. Brooks. Effects of Electrostatic Force Truncation on Interfacial and Transport Properties of Water. *J. Phys. Chem.*, 100(42) : 17011–17020, 1996.

[79] David E. Smith and Liem X. Dang. Computer simulations of NaCl association in polarizable water. *J. Chem. Phys.*, 100(5) : 3757–3766, 1994.

[80] Liem X. Dang and Tsun-Mei Chang. Molecular Mechanism of Ion Binding to the Liquid/Vapor Interface of Water. *J. Phys. Chem. B*, 106(2) : 235–238, 2002.

[81] Matthew P. Hodges, Anthony J. Stone, and Sotiris S. Xantheas. Contribution of Many-Body Terms to the Energy for Small Water Clusters : A Comparison of ab Initio Calculations and Accurate Model Potentials. *J. Phys. Chem. A*, 101(48) : 9163–9168, 1997.

[82] Jon M. Sorenson, Greg Hura, Robert M. Glaeser, and Teresa Head-Gordon. What can x-ray scattering tell us about the radial distribution functions of water ? *J. Chem. Phys.*, 113(20) : 9149–9161, 2000.

[83] Ariel A. Chialvo and Peter T. Cummings. Engineering a simple polarizable model for the molecular simulation of water applicable over wide ranges of state conditions. *J. Chem. Phys.*, 105(18) : 8274–8281, 1996.

[84] L. Delle Site, A. Alavi, and R. M. Lynden-Bell. The electrostatic properties of water molecules in condensed phases : an *ab initio* study. *Mol. Phys.*, 96(11) : 1683–1693, 1999.

[85] Stéphanie Chalmet and Manuel F. Ruiz-López. The reaction field of a water molecule in liquid water : Comparison of different quantum/classical models. *J. Chem. Phys.*, 115(11) : 5220–5227, 2001.

[86] Tina D. Poulsen, Peter R. Ogilby, and Kurt V. Mikkelsen. Linear response properties for solvated molecules described by a combined multiconfigurational self-consistent-field/molecular mechanics model. *J. Chem. Phys.*, 116(9) : 3730–3738, 2002.

[87] Michael W. Mahoney and William L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, 112(20) : 8910–8922, 2000.

[88] K. Coutinho, R. C. Guedes, B. J. Costa Cabral, and Sylvio Canuto. Electronic polarization of liquid water : converged Monte Carlo-quantum mechanics

results for the multipole moments. *Chem. Phys. Lett.*, 369(3–4) : 345–353, 2003.

[89] Kyoko Watanabe and Michael L. Klein. Effective pair potentials and the properties of water. *Chem. Phys.*, 131 : 157–167, 1989.

[90] Peter Höchtl, Stefan Boresch, Wolfgang Bitomsky, and Othmar Steinhauser. Rationalization of the dielectric properties of common three-site water models in terms of their force field parameters. *J. Chem. Phys.*, 109(12) : 4927–4937, 1998.

[91] Bertrand Guillot. A Reappraisal of What We Have Learnt During Three Decades of Computer Simulations on Water. *J. Mol. Liq.*, 101(1–3) : 219–260, 2002.

[92] M. C. Goh, J. M. Hicks, K. Kemnitz, G. R. Pinto, T. F. Heinz, K. B. Eisenthal, and K. Bhattacharyya. Absolute orientation of water molecules at the neat water surface. *J. Phys. Chem.*, 92(18) : 5074–5075, 1988.

[93] Maria Paluch. Electrical properties of free surface of water and aqueous solutions. *Adv. Colloid Interface Sci.*, 84 : 27–45, 2000.

[94] Peter Vassilev, Christoph Hartnig, Marc T. M. Koper, Frédéric Frechard, and Rutger A. van Santen. *Ab initio* molecular dynamics simulation of liquid water and water-vapor interface. *J. Chem. Phys.*, 115(21) : 9815–9820, 2001.

[95] I-Feng W. Kuo and Christopher J. Mundy. An ab Initio Molecular Dynamics Study of the Aqueous Liquid-Vapor Interface. *Science*, 303 : 658–660, 2004.

[96] Sotiris S. Xantheas, Christian J. Burnham, and Robert J. Harrison. Development of transferable interaction models of water. II. Accurate energetics of the first few water clusters from first principles. *J. Chem. Phys.*, 116(4) : 1493–1499, 2002.

[97] Martin Losada and Samuel Leutwyler. Water hexamer clusters : Structures, energies, and predicted mid-infrared spectra. *J. Chem. Phys.*, 117(5) : 2003–2016, 2002.

[98] Jonathon K. Gregory and David C. Clary. Theoretical Study of the Cage Water Hexamer Structure. *J. Phys. Chem. A*, 101(36) : 6813–6819, 1997.

Chapitre 5

Molecular dynamics study of hydration in ethanol–water mixtures using a polarizable force-field

Tiré de :
Sergei Yu. Noskov, Guillaume Lamoureux et Benoît Roux. *J. Phys. Chem.*, article sous presse.
(Les deux premiers auteurs doivent être considérés premiers auteurs conjoints.)

Abstract

The abnormal physico-chemical characteristics of ethanol solvation in water are commonly attributed to the phenomenon of hydrophobic hydration. To investigate the structural organization of hydrophobic hydration in water–ethanol mixtures, we use molecular dynamics simulations based on detailed atomic models. Induced polarization is incorporated into the potential function on the basis of the classical Drude oscillator model. Water–ethanol mixtures are simulated at eleven ethanol molar fractions, from 0.05 to 0.9. Although the water and ethanol models are parametrized separately to reproduce the vaporization enthalpy, static dielectric constant, and self-diffusion constant of neat liquids at ambient conditions, they also reproduce the energetic and dynamical properties of the mixtures accurately. Furthermore, the calculated dielectric constant for the various water–alcohol mix-

tures is in excellent agreement with experimental data. The simulations provide a detailed structural characterization of the mixtures. A depletion of water–water hydrogen bonding in the first hydration shell of ethanol is compensated by an enhancement in the second hydration shell. The structuring effect from the second solvation shell gives rise to a net positive hydrogen bonding excess for ethanol molar fractions up to $\simeq 0.5$. For larger molar fractions, the second hydration shell is not sufficiently populated to overcome the net H-bond depletion from the first shell.

5.1 Introduction

Solvation of polar and nonpolar moieties in liquid water is one of the key factors that governs the structural stability of proteins. Despite the great importance of hydrophobic hydration, many of its aspects are not completely understood [1, 2, 3]. From this point of view, the solvation of alcohols in liquid water provides perhaps the most elementary but non-trivial model system in which hydration of nonpolar and polar hydrogen-bonding groups are directly and intimately competing. Hydration of alcohols is commonly interpreted structurally in terms of the classical "iceberg" picture of Frank and Evans [4]. Within this framework, it has been suggested that the water molecules surrounding a nonpolar moiety are rearranging into a low-entropy cage with stronger hydrogen bonds (H-bonds). Although there is broad support for this view [5, 6, 7, 8], a different perspective is suggested by recent scattering measurements [9, 10, 11], as well as molecular dynamics simulations [12, 13, 14]. Studies combining diffraction data with computations by Soper *et al.* [15], found that there are only minor changes H-bond occurrence in the first hydration shell of hydrated alcohols, but that there is a higher solvent density ("compaction") in the second hydration shell. These results suggest that the structural complexity of alcohol–water mixtures may in fact, extend well beyond a simple picture based on the first solvation shell. Numerous experimental evidence collected from compressibility, UV and IR absorption spectra [16], dielectric relax-

ation and Raman spectra investigations [6] and neutron scattering studies [10, 17]. indicate that there is a competition between formation of solute–solute, solvent–solvent and solute–solvent clusters. The nature of these structural rearrangements is not completely understood [6, 10]. Monatomic alcohols may be considered as the simplest molecular solute simultaneously presenting hydrophobic and hydrophilic moieties. Therefore, aqueous solutions of alcohols represent a rich test-ground for investigating hydration using theoretical approaches.

Generally, computer simulations of biomolecular systems are generated using simple potential functions which do not account for many-body induced polarization effects explicitly. Induced polarization is typically incorporated in an average effective way. For example, widely used potential functions such as AMBER/parm94 [18], CHARMM27 [19], GROMOS [20], and OPLS [21] force fields are based on pairwise additive electrostatic interactions with fixed effective atomic partial charges that have been adjusted to account for the average induction arising in a typical environment. While such potential functions can be remarkably successful in modeling complex molecular systems (e.g., see Ref. [22]), a meaningful representation of the balance of hydration forces around polar and nonpolar moieties requires a careful consideration of induced polarization. A polarizable model can produce a polarization response varying gradually with the composition of the mixture.

At the present time, computational chemists and biophysicists are actively pursuing the development of a new generation of force-fields for biomolecular simulations [23, 24, 25, 26, 27]. Recently, a simple polarizable potential function for water molecule based on the classical Drude oscillator has been developed and implemented [28]. The current efforts are made within the framework of the CHARMM/PARAM-22 force-field [19], with the long-term goal of developing a complete polarizable potential function for biomolecular simulations. A logical next step after water is to parametrize more complicated systems such as monatomic alcohols [25, 24]. In addition to its obvious methodological interest, an accurate

model of ethanol is very important to address a number of fundamental questions about the role of atomic polarizability regarding hydration phenomena.

The aims of the present paper are two-fold. From a methodological perspective, we want to develop a new simple polarizable model for ethanol consistent with the existing CHARMM/PARAM-22 force-field [19, 28]. To assure the validity of the model both in the gas- and condensed-phases, we parametrize it against a wealth of experimental information, complemented by quantum-chemical data on the ethanol gas-phase dipole moment, polarizability, and multiple dimer structures. At a more fundamental level, our goal is to elucidate the structure and dynamics underlying their abnormal mixing properties, and to ascertain the significance of hydrophobic hydration in ethanol–water solutions for a wide range of concentration.

5.2 Models and Methods

The intermolecular potential for the polarizable ethanol model is similar to the functional form used previously for the SWM4-DP water model [28]. Each atom is described by a partial point-charge q and a set of Lennard-Jones parameters ϵ and σ (combined with the others following the usual Lorentz–Berthelot rules [29]). In addition, the heavy atoms are carrying an auxiliary Drude particles [28], tethered by a harmonic spring with force constant k_{D}.

In a field $\mathbf{E}$, a Drude particle carrying a charge δq shifts its equilibrium, zero-field position by $\mathbf{d} = \delta q \mathbf{E}/k_{\mathrm{D}}$, giving rise to an induced dipole $\boldsymbol{\mu} = \delta q^2 \mathbf{E}/k_{\mathrm{D}}$. Hence, the atomic polarizability created by a Drude model is $\alpha = \delta q^2/k_{\mathrm{D}}$ and, assuming a uniform $k_{\mathrm{D}} = 1000$ kcal/mol/Å^2 for all three heavy atoms of the ethanol molecule, the charges δq can be adjusted to produce appropriate α's. Within a given molecule, the induced atomic dipoles are coupled with a screened electrostatic interaction. The original scheme of Thole [30] was adapted to the present case where finite dipoles are formed of two point charges. For two interacting dipoles 1 and 2, the pair of charges δq_1 and $-\delta q_1$ interacts with the pair of charges δq_2 and $-\delta q_2$ via a screened Coulomb interaction. The interaction energy is the sum

of four terms of the form

$$\frac{\delta q_1 \delta q_2}{r_{12}} \left[1 - \left(1 - \frac{\bar{r}_{12}}{2} \right) e^{-\bar{r}_{12}} \right]. \tag{5.1}$$

The normalized distance $\bar{r}_{12}$ is defined as

$$\bar{r}_{12} = a \frac{r_{12}}{\sqrt[6]{\alpha_1 \alpha_2}}, \tag{5.2}$$

where a is a dimensionless number that is adjusted empirically. This interaction creates a cooperativity among the induced dipoles that favors a longitudinal polarization in accord with *ab initio* calculations. The molecular polarizability of ethanol would be isotropic in the absence of this intramolecular coupling.

The parameter a is set to 2.6, a value which is known to create a correct polarizability anisotropy ratio for an idealized benzene molecule, and which is similar to the various estimates of Thole [30] and van Duijnen and Swart [31]. Apart from the Thole dipole–dipole intramolecular interactions, we keep the original CHARMM/PARAM-22 intramolecular interactions [19] (bond stretch, angle bend, and torsion). The atomic charges and atomic polarizabilities (as well as the Lennard-Jones parameters for the ethanol oxygen) were adjusted to reproduce the essential properties of liquid ethanol: evaporation enthalpy, density, dielectric constant, and self-diffusion constant. All the parameters of the ethanol model are listed in Table 5.I. The oxygen atom is labeled O_E, the carbon atom bound to the oxygen is labeled C_1, and the final carbon is labeled C_2. The other parameters are kept the same as in the CHARMM/PARAM-22 force-field [19]. The proposed ethanol model reproduces fairly well the experimental value of the gas phase dipole, $\mu = 1.71$ D [32].

The evaporation enthalpy Δh is obtained from the average potential energy per molecule ($\langle u \rangle_{\text{liq}}$) in the liquid phase relative to the average potential energy of a single molecule in the gas phase ($\langle u \rangle_{\text{gas}}$). The latter is obtained from a Langevin dynamics simulation with the three induced dipoles kept in their energy minimum

Table 5.I: Parameters of the polarizable ethanol model. C_1 is the carbon atom bound to the oxygen, and C_2 is the carbon atom of the methyl group. H_1 are the two atoms of the C_1 group, and H_2 are the three atoms of the C_2 group. The parameters unchanged from the CHARMM/PARAM-22 force-field are in italic.

Atom	q (e)	α (Å^3)	ϵ (kcal/mol)	σ (Å)
O_E	−0.520351	0.977723	−0.1521	1.532345
HO_E	0.361419	—	−0.046	*0.200007*
C_1	0.020246	1.217571	−0.055	*1.937705*
H_1	0.073583	—	−0.022	*1.175986*
C_2	−0.269975	1.237576	−0.08	*1.835251*
H_2	0.087165	—	−0.022	*1.175986*

at every time of the simulation.

$$\Delta h = k_B T - \left(\langle u \rangle_{liq} - \langle u \rangle_{gas} \right). \tag{5.3}$$

The density is reported in terms of the molecular volume $v = \langle V \rangle / N$, where N is the number of molecules and $\langle V \rangle$ is the average volume of the simulation box. The dielectric constant is computed from the fluctuations of $\mathbf{M}$, the total dipole moment of the box [33, 34]:

$$\epsilon = \epsilon_\infty + \frac{4\pi}{3\langle V \rangle k_B T} \left(\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2 \right). \tag{5.4}$$

The high-frequency contribution ϵ_∞, which is not produced by inducible dipoles in the self-consistent field regime, is estimated from the Clausius–Mossotti equation [35]:

$$\frac{\epsilon_\infty - 1}{\epsilon_\infty + 1} = \frac{4\pi}{3} \frac{\alpha}{v}. \tag{5.5}$$

For the present ethanol model at its equilibrium density, $\epsilon_\infty = 1.69$. The self-diffusion constant D is computed from the mean squared displacement of the oxygen atoms. For the molecular dynamics (MD) simulations of the mixtures, this ethanol model is combined with the SWM4-DP water model of Lamoureux *et al.* [28]. The water oxygens will be referred to as O_w. All MD simulations are

performed in the NPT ensemble at room temperature and pressure ($T = 300$ K and $P = 1$ atm) using a double-thermostat scheme described previously [28]. The dipoles are maintained in the self-consistent-field induction regime using an additional low-temperature thermostat [36]. The integration time-step is 2 fs. Cubic periodic boundary conditions with particle-mesh Ewald summation algorithm [37] were used to treat long-range electrostatic interactions. Truncation of van der Waals interactions are accounted for by an effective long-range correction [38]. The primary cubic box contains 250 molecules for all simulations. The initial dimensions of the box were set up according to the experimental density of ethanol and ethanol–water solutions [32]. All ethanol–water mixtures were simulated for 3 ns. To get a better convergence, the pure ethanol system was simulated for 6 ns.

5.3 Results and Discussion

5.3.1 Gas-phase properties

The first step in validating the ethanol model is to examine its ability to accurately reproduce the geometries and energies of various hydrogen-bonded complexes from previous studies. Table 5.II summarizes the results for the present polarizable model, along with available quantum-chemical computations [39] and molecular-mechanical computations using the CHARMM/PARAM-22 force-field [19], the OPLS force-field [40, 41], and the polarizable PIPF force-field of Gao *et al.* [25] It should be noted that results from the current model are similar to previous studies using OPLS potential [40] and quantum chemistry computations [39]. The oxygen–oxygen distance is very similar to those of the OPLS potential model, which is about 0.1 Å shorter than the same distance from *ab initio* estimates.

5.3.2 Liquid ethanol

The simulated and experimental results on enthalpy of evaporation are presented in Table 5.II for different potential models. The performance of all models are

Table 5.II: Computed and experimental properties for ethanol.

	This work	PARAM-22[a]	OPLS[b]	PIPF[c]	Ab initio/Expt.
Trans–trans ethanol dimer:					
E_{int} (kcal/mol)	−7.29	−7.77	−7.04	−6.19	−4.86 – −7.96[d]
$r_{\text{O–O}}$ (Å)	2.79	2.80	2.74	2.74	2.847[d]
Gauche–trans ethanol dimer:					
E_{int} (kcal/mol)	−7.79	−7.78			−4.82 – −8.16[d]
$r_{\text{O–O}}$ (Å)	2.79	2.81			2.86[d]
Liquid state properties:					
Δh (kcal/mol)	10.19	9.68	10.23	10.09	10.15[e]
ϵ	24.1 ± 1.2	17.1 ± 0.6	16.1 ± 0.4	23.9 ± 0.4	24.3[f]
D (10^{-5} cm^2/s)	0.967	1.59	1.56		1.09[g]

[a] Ref. [19].
[b] Refs. [40, 41].
[c] Ref. [25].
[d] From Ref. [39].
[e] From Ref. [42].
[f] From Ref. [43].
[g] From Ref. [44].

comparable and agree well with experimental results. The computed molecular volume differs from experiment by only 2.7%, which is comparable or better than results from previous simulations [25, 40, 41]. Compared to nonpolarizable models [45, 41], the inclusion of polarizability is expected to improve the accuracy of the model for simulating self-diffusion coefficient and dielectric constant. The change in molecular dipole moment of ethanol transferring from the gas phase to the liquid is a good indicator of the condensed phase polarization effect. The average dipole in neat liquid ethanol is 2.38 D compared to 1.71 D in the gas phase. Although it is lower than experimental [46] or Car-Parrinello (CP) molecular dynamics [47] estimates (3.04 D and 3.1 D, respectively), this value is needed to correctly reproduces the dielectric constant of the liquid.

The structure of the liquid can be characterized by a number of pair-distribution functions (RDFs). Figure 5.1 summarizes two of the most important RDFs for ethanol, together with the running integration numbers. There is a displacement of

the first peak in the O_E–O_E and O_E–H RDFs in comparison to nonpolarizable models [40]. Experimental results indicate a first peak in the O_E–O_E RDFs at around 2.7–2.8 Å, with a coordination number of 1.8–2.0 [48, 49]. These observations are in reasonable agreement with the present results, which yield a coordination number of 1.94 (by integration of the O_E–O_E RDFs to the minimum at 3.45 Å). Hydrogen-bonding interactions in ethanol are reflected in the strong first peak in the O_E–H distribution function (Fig. 5.1). The first maximum is around 1.75 Å for the polarizable model and 1.85 Å for the CHARMM/PARAM-22 force-field. The integration of each distribution function to the position of the first minimum gives coordination number of 0.96 (up to 2.55 Å) and 0.95 (up to 2.60 Å) for polarizable and nonpolarizable potential models, respectively. It was found that the use of a long-range correction [38] to the summation of the van der Waals dispersion terms is essential for a precise simulation of the volume in NPT simulations. Similar conclusions were reached for liquid alkanes [38].

5.3.3 Water–Ethanol Mixtures

The O_E–O_w, C_1–O_w, and C_2–O_w RDFs were computed for all concentrations (see the Supplementary Material). For diluted ethanol mixtures, the water structure around ethanol is similar to the results at infinite dilution of Fidler and Rodger [12] and van Erp and Meijer [47]. The positions and shapes of all peaks are comparable, though the O_E–O_w peak is higher in the current simulations.

Dielectric properties

A demanding test for the transferability of the polarizable force-field is the accurate reproduction of the dielectric behavior of binary mixtures as a function of molar fractions of water and ethanol. The dependence of the dielectric constant on the mixture compositions, estimated from the MD simulations, is depicted on Fig. 5.2 along with experimental data [43]. The inclusion of explicit polarizability succeeds in reproducing the variation of the static dielectric constant accurately

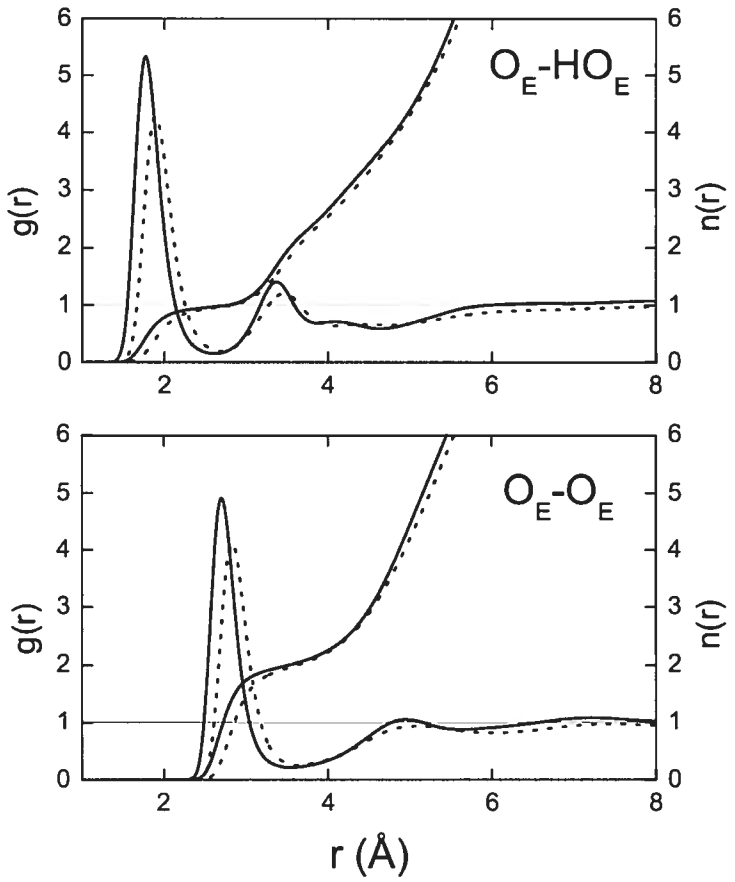

Figure 5.1: O_E–HO_E and O_E–O_E radial distribution functions with running integration numbers (solid line: this work; dotted line: CHARMM/PARAM-22 force-field).

over the full molar fraction range. In contrast, nonpolarizable models for ethanol fail to reproduce ϵ. Perhaps additional parametrization is required to the correct description of the dielectric constants by simulations with nonpolarizable models for ethanol.

The uncertainty on the dielectric constant estimates from MD simulations, especially in the 0.50–0.80 range of ethanol molar fraction, is due to the slow convergence in fluctuations of the total dipole. Figure 5.2b shows the contributions to ϵ of the water and ethanol components of the mixture. The expression $\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2$ of formula (5.4) is expanded into contributions from the fluctuations of the total dipole created by the water molecules, $\langle \mathbf{M}_\mathrm{w}^2 \rangle - \langle \mathbf{M}_\mathrm{w} \rangle^2$, from the fluctuations of the dipole created by the ethanol molecules, $\langle \mathbf{M}_\mathrm{EtOH}^2 \rangle - \langle \mathbf{M}_\mathrm{EtOH} \rangle^2$, and from the correlated fluctuations for both components, $2(\langle \mathbf{M}_\mathrm{w} \cdot \mathbf{M}_\mathrm{EtOH} \rangle - \langle \mathbf{M}_\mathrm{w} \rangle \cdot \langle \mathbf{M}_\mathrm{EtOH} \rangle)$. Interestingly, it reveals that dipole fluctuation cross-correlations are maximum in the $X_\mathrm{EtOH} = 0.2$ to 0.4 range. Those might be related to the presence of a cooperative dynamics between water and ethanol in this range of concentrations.

Upon dilution in water, the average dipole of the ethanol molecule changes smoothly from 2.38 D in pure liquid ethanol to about 2.75 D at infinite dilution in liquid water. By comparison, the average molecular dipole of water goes from 2.45 D in pure water to approximately 2.37 D in pure ethanol (see Supplementary Material, Figure 5.10). These variations signal the importance of the polarity of the surrounding medium on the molecules. Over the whole range of ethanol–water mixtures, the magnitude of the dipole of ethanol changes by 0.37 D. These changes, which affect all ethanol–ethanol as well as ethanol–water intermolecular interactions, make it necessary to account for the induced polarization explicitly in such system. The contribution of each polarizable atom to the total induced dipole varies with ethanol concentration. Surprisingly, the C_1 and C_2 atoms are contributing almost of half the total induced dipole per molecule. (see Supplementary Material). For pure ethanol, the apolar part of the molecule has an induced dipole of 0.35 D, in agreement with the CP simulations of van Erp and Meijer [47].

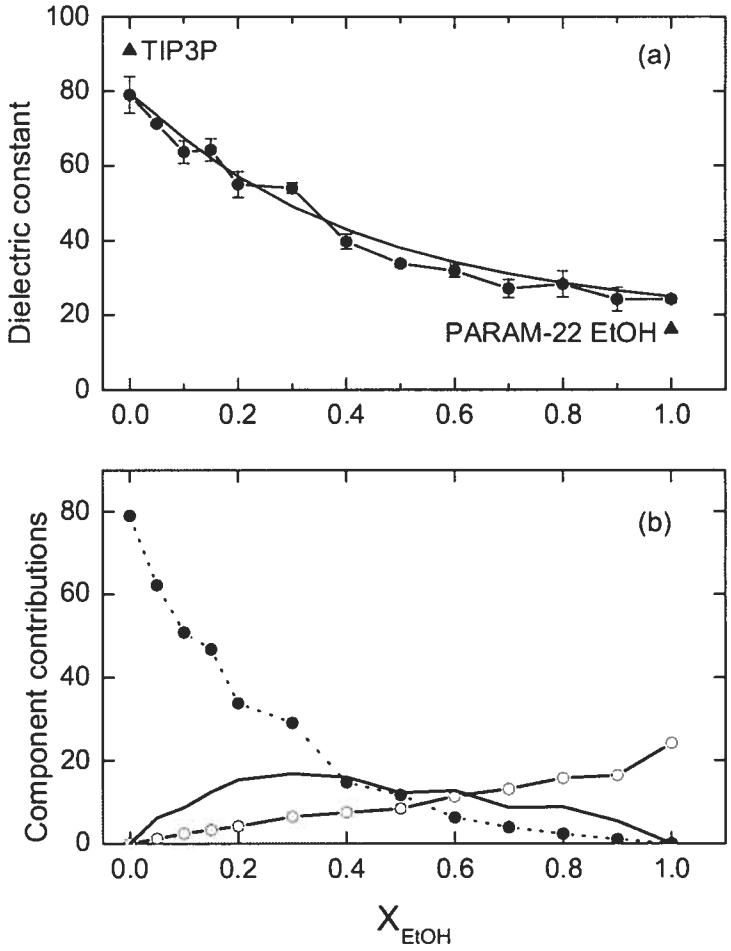

Figure 5.2: (a) Dependence of the dielectric constant on the composition of the ethanol–water mixtures (circles: this work; solid line: experiment at 20°C [43]). Results for pure TIP3P liquid [50, 28] and pure CHARMM/PARAM-22 ethanol (triangles) are shown for comparison. The results for pure SWM4-DP liquid are from Ref. [28]. (b) Contribution of the water component (full symbols), the ethanol component (empty symbols), and water–ethanol dipole cross-correlations (solid line, no symbols) to the dielectric constant of panel (a).

At a molar fraction X_{EtOH} of 0.05, it contributes as much as the polar oxygen (see Supplementary Material).

Dynamics

Many of the anomalous properties of alcohol–water mixtures result from inhomogeneous mixing at different compositions. One of the salient characteristic of the non-ideality in ethanol–water mixtures is the variation of the self-diffusion coefficients as a function of molar fraction of alcohol. Computed and experimentally measured self-diffusion coefficients for water and ethanol are depicted in Fig. 5.3. Remarkably, the present model is able to reproduce the position of the minima in the self-diffusion coefficients for water ($X_{\mathrm{EtOH}} = 0.3$) and ethanol ($X_{\mathrm{EtOH}} = 0.2$). In contrast, MD studies of ethanol–water mixtures using nonpolarizable OPLS model for ethanol and TIP4P for water [51] do not reproduce this important feature of the ethanol–water mixtures.

A number of possible explanations of the nonlinear dependence of diffusion characteristics have been proposed over the last decade [52, 53]. It was hypothesized that the large drop in the self-diffusion coefficients of water with a small increase of ethanol concentration ($X_{\mathrm{EtOH}} \simeq 0.05$–$0.10$) was caused by the existence of clathrate-like hydrates. Nuclear magnetic resonance studies [54] and molar excess enthalpy measurements [55] generally support the idea that the hydrogen bond network between water molecules surrounding the alkyl group is strengthened. However, diffraction studies do not substantiate this view in diluted water–alcohol mixtures [10, 56]. The minimum of ethanol self-diffusion around $X_{\mathrm{EtOH}} \simeq 0.2$ was associated with competitive self-association of ethanol and water by measurements of the dielectric relaxation and permittivity spectra [57]. Compressibility measurements [58] suggested that the hydrophobic hydration becomes negligible after $X_{\mathrm{EtOH}} \simeq 0.3$ (minimum in water self diffusion). Contrary to these results, recent structural studies of methanol–water solutions suggested that water still forms hydrogen bonded clusters, even at a very high alcohol concentrations [17]. Despite the numerous experiments, these questions are not resolved. It may be

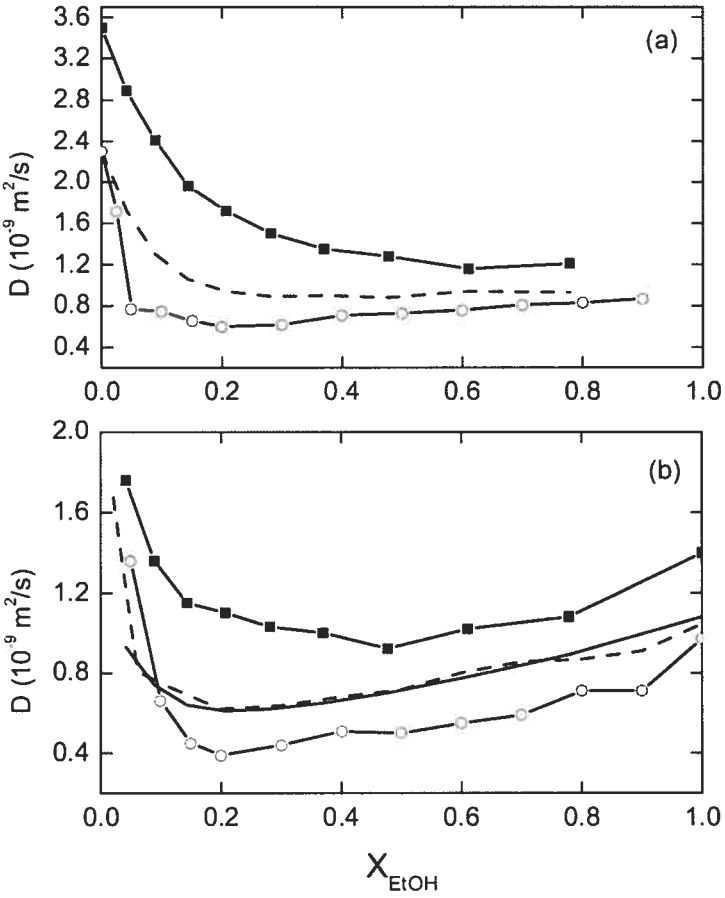

Figure 5.3: Self-diffusion coefficients for (a) water and (b) ethanol as function of molar fraction of ethanol. The empty circles are the results for the present work. Available experimental (solid [52] and dashed [53] lines, no symbols) and theoretical [51] (nonpolarizable OPLS force-field, square symbols) data are depicted for both species.

hoped that the current MD simulations offer a realistic and accurate description of the microscopic properties of ethanol–water mixtures. This provides a unique opportunity to address some of the questions about local structure effect and its influence on dynamics in water–ethanol mixtures.

5.3.4 Hydrogen-bonding analysis

Radial pair distribution functions, despite their usefulness, do not reveal all the complex structural aspects of the mixtures. To complement the RDF analysis, we analyzed the three-dimensional topology and clustering of the solute and solvent molecules as a function of ethanol concentration. For the present analysis, two molecules are considered to be H-bonded if their H$\cdots$O distance is less than 2.4 Å and if the O–H$\cdots$O angle is larger than 150°. This definition is similar to that of Ferrario *et al.* [59], of De Loof *et al.* [60], and of Luzar and Chandler [61, 62]. An energetic definition of the H-bond is not practical, as the energy of the polarizable model is not pair-wise decomposable. For each MD configuration, the list of donor–acceptor pairs matching the H-bond definition is compiled into a set of H-bonding statistics. Because there is no perfectly objective definition of a H-bond, the H-bond counts should be compared to the analogous count for a reference liquid (pure water or pure ethanol).

Average hydrogen-bonding in water–ethanol mixtures

For pure liquid water, we find an average of 3.03 H-bonds per water molecule with the ($d_{\mathrm{HO}} < 2.4$ Å)/150° geometric definition of the H-bond. This value is smaller than the coordination number of 4.63 for neighboring oxygens [28], reflecting the weakening/breakage of hydrogen bonds in tetrahedrally arranged water molecules throughout the simulation. The 3.03 value is consistent with other studies. From simulations calibrated against neutron diffraction data and a ($d_{\mathrm{OO}} < 3.5$ Å)/150° definition, Soper *et al.* [63] obtain 3.58 H-bonds per molecule. From an MD simulation of pure water using the TIP4P potential [50] and a

($d_{OO} < 3.5$ Å)/($U_{int} < -2.6$ kcal/mol) definition of the H-bond, Oleinikova *et al.* [64] obtain 3.26 H-bonds per molecule. The contrasting x-ray absorption and x-ray Raman scattering measurements of Wernet *et al.* [65], yielding 2.2 ± 0.5 H-bonds per molecule, suggest the picture of a hydrated water molecule forming two strong and two weak H-bonds. Recent *ab initio* molecular simulations by Kuo and Mundy [66] yield 2.5 H-bonds per molecule in the bulk phase, but only 1.9 H-bonds for molecules near a liquid-vapor interface [using a ($d_{OH} < 2.27$ Å)/140° definition]. From our simulation of the pure liquid ethanol, we find 1.65 H-bonds per molecule; the O_E–O_E coordination number is 1.94. Overall, these comparisons indicate that the coordination number of the oxygen–oxygen RDF is a somewhat biased estimator of the number of H-bonds.

Figure 5.4 shows the average total number of H-bonds per molecule as the ethanol concentration is increased. This total number includes contributions from both water and ethanol molecules. The top graph is showing the number of H-bonds per water molecule, broken down into contributions from water–water (W/W) and water–ethanol (W/E) H-bonds. The bottom graph is showing the same analysis for the ethanol component of the mixtures [ethanol–ethanol (E/E) and ethanol–water (E/W)]. The deviations of these curves from ideal straight lines correspond to the excess H-bonding. The positive excesses for the W/W and E/E curves and the negative excesses for the W/E and E/E curves are indicating that water is an overall better solvent than ethanol: at any ethanol fraction, both water and ethanol are preferably solvated by water.

At high-ethanol concentrations, the number of H-bonds decreases to 2.64 per water molecule. In other words, a water molecule solvated in pure ethanol loses only 14% of its H-bonding relative to pure liquid water. In comparison, the solvation of ethanol is enhanced by its dilution in water, going from 1.65 to 2.23 H-bonds per ethanol as X_{EtOH} goes from 1 to 0. For X_{EtOH} larger than 0.50, the excess of water–water H-bonds per water molecule is almost exactly compensated by the deficit of ethanol–water H-bonds per water molecule. This compensatory effect makes the total H-bonding per water appears like it is following the ideal

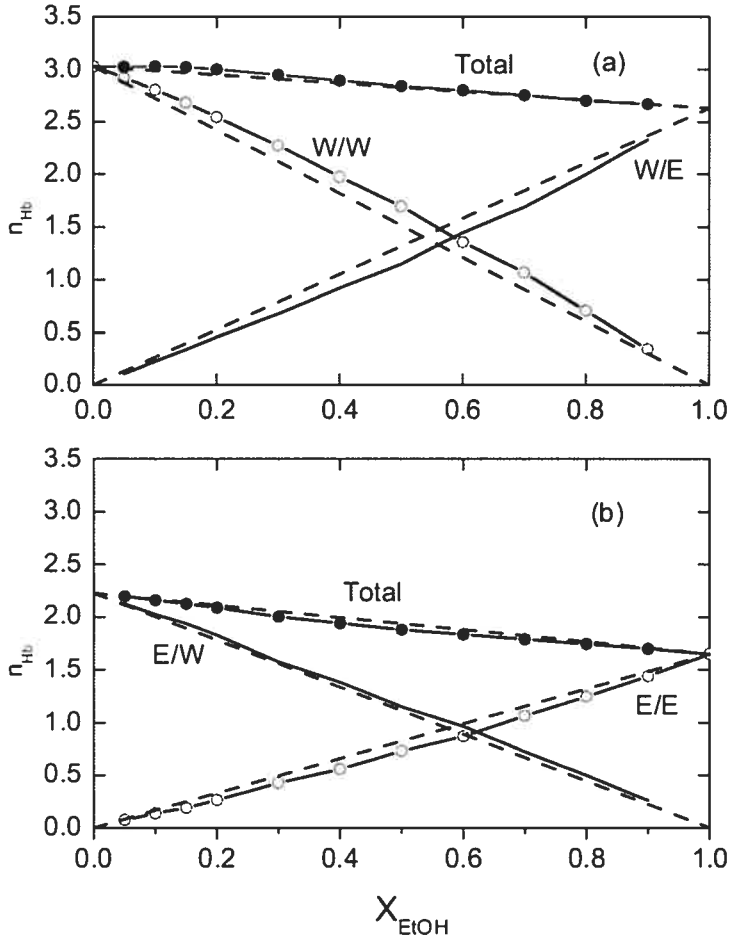

Figure 5.4: Average number of hydrogen bonds per molecule (n_{Hb}) for (a) water and (b) ethanol. In panel (a), the total n_{Hb} per water molecule is broken down into water–water (W/W) and water–ethanol (W/E) contributions. In panel (b), the total per ethanol molecule is broken down into ethanol–water (E/W) and ethanol–ethanol (E/E) contributions. The straight dotted lines represent the ideal mixing.

mixing curve. For X_{EtOH} less than 0.50 however, there is a (positive) excess in the total H-bonding of water (see Fig. 5.4a). Such a increase is the usual signature of the hydrophobic hydration [67, 58]. The increase is sufficiently large that the total number of H-bonds for water molecules is actually stronger at low-ethanol concentrations than in pure water. The number of H-bonds per water molecule is maximum around $X_{\mathrm{EtOH}} = 0.10$. Knowing that the water–water H-bonding dominates the thermodynamics at low ethanol concentration, this could be related to the maximum in excess enthalpy and entropy at $X_{\mathrm{EtOH}} \simeq 0.10$ and a maximum in excess free energy at $X_{\mathrm{EtOH}} = 0.18$–0.22 observed from dielectric relaxation measurements [5, 68, 69, 8]. What is not clear, however, is the nature of these H-bond-rich regions and how they are affected by the increase in ethanol concentration.

Structure of hydrogen-bonding around ethanol

Figures 5.5a and 5.5b show two isosurfaces for the density of water oxygen around ethanol at $X_{\mathrm{EtOH}} = 0.05$ (the lowest ethanol concentration simulated). The highest-density surface (Fig. 5.5a, in opaque yellow) shows the first coordination shell of the ethanol hydroxyl group (in the *trans* conformation). It is divided into a relatively confined region where one water molecule is an acceptor to the ethanol hydroxyl, and a wide band in which water acts as a proton donor. The lower-density surface (Fig. 5.5b, in translucent yellow)—still higher that the bulk density—shows an accumulation of water at the surface of the hydrophobic group, as well as a water-rich "cap" region where water can coordinate first-hydration-shell water molecules in both lobes of Fig. 5.5a. The strong water–ethanol H-bonds are giving rise to a visible secondary density maximum on two sides of the ring-shaped isosurface of Fig. 5.5a.

To complement this structural information, we examine the average number of H-bonds a water molecule at a position $\mathbf{r}$ relative to the closest ethanol will form with any other molecule (water or ethanol, including the reference ethanol molecule). Fig. 5.5c shows the distribution $n_{\mathrm{Hb}}(\mathbf{r})$ for a slice in the O_E–C_1–C_2

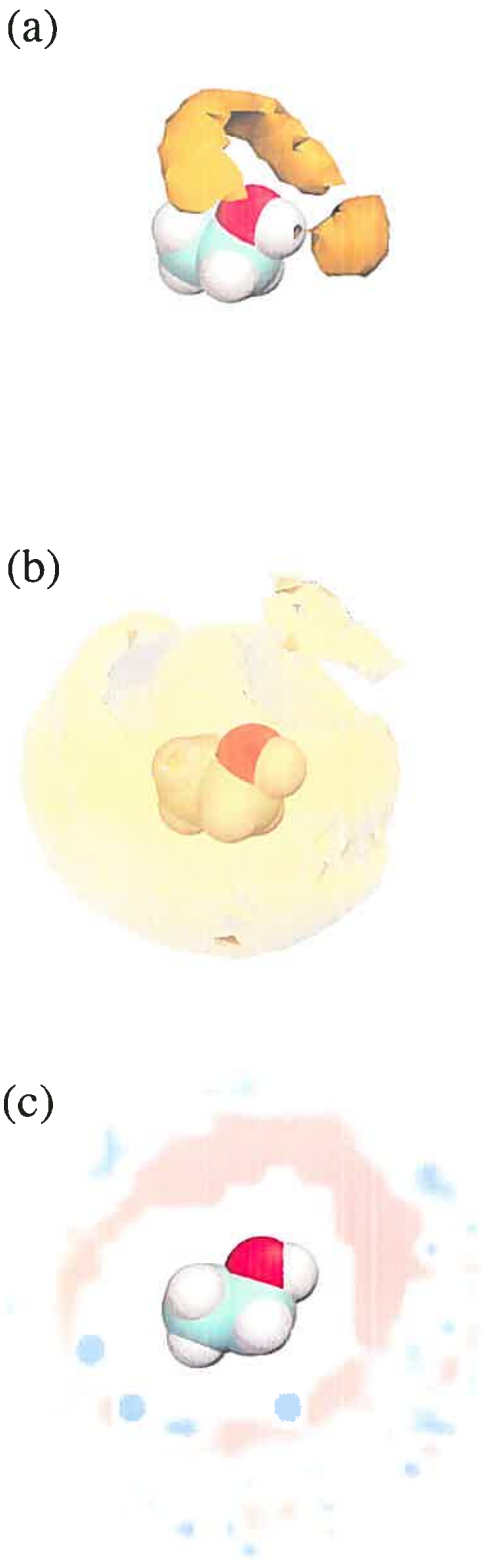

Figure 5.5: Water oxygen density around the closest ethanol molecule for $X_{\mathrm{EtOH}} = 0.05$. Two isodensity surfaces are shown: the higher-density surface of panel (a) illustrates the first hydration shell of the ethanol hydroxyl group, and the lower-density surface of panel (b) shows the overall packing of water around the hydrophobic group and the second hydration shell of the ethanol hydroxyl. (c) Hydrogen bonding excess for a water molecule at a given position with respect to the closest ethanol molecule. Blue denotes regions where the water molecule forms more H-bonds than the bulk average, and red denotes regions where it forms less H-bonds than the bulk average. This function is presented for a slice in the O_E–C_1–C_2 plane of the ethanol molecule, and up to a distance of 6 Å from the C_1 atom (after which the average value gets noisier because of a lack of sampling).

plane of the reference ethanol molecule. Because n_{Hb} is the number of hydrogen bonds *per molecule*, its value at a point $\mathbf{r}$ is not directly related to the probability of a water molecule actually being there. The regions where a water molecule makes less H-bonds than the bulk average (in red in Fig. 5.5c) are where it is the most strongly bonded to ethanol, i.e., in the first hydration shell of the ethanol hydroxyl and, to a lesser extent, anywhere within H-bonding distance. In contrast, the water molecules in a narrow region of the first coordination shell of the nonpolar group have a H-bonding above the bulk average. This seems to conjure up the classical "clathrate" or "iceberg" picture of the hydrophobic hydration [4], but overall, the net contribution to the excess water H-bonding from the first coordination shell of the nonpolar and hydroxyl groups is *negative*.

The average *positive* H-bonding excess reported in Fig. 5.4 is in fact explained by the secondary hydration structure of ethanol. There is a small but significant H-bonding excess in most of the second shell (see Fig. 5.5c). Although it is small, it involves a larger number of water molecules and creates a *positive* net contribution to the water H-bonding excess that compensates for the negative contribution from the first hydration shell. This can be quantified by integrating the contributions to the excess number of water H-bonds from concentric shells centered around the closest C_1 atom:

$$N_{\mathrm{Hb}}^{(\mathrm{xs})}(R) = \int_0^R dr \left[\rho_{\mathrm{Hb}}(r) - n_{\mathrm{Hb}}^{(\mathrm{id})}\rho_{\mathrm{O}}(r)\right]. \tag{5.6}$$

$\rho_{\mathrm{Hb}}(r)$ is the total number of H-bonds for the water molecules whose oxygen atom is in the shell of radius r and thickness dr, and $\rho_{\mathrm{O}}(r)$ is the number of such oxygen atoms. The ratio $\rho_{\mathrm{Hb}}(r)/\rho_{\mathrm{O}}(r)$ corresponds to $n_{\mathrm{Hb}}(r)$, the average number of H-bonds for a single water molecule in the shell of radius r. It is the radial average of the quantity presented in the contour plot of Fig. 5.5c. The value $n_{\mathrm{Hb}}^{(\mathrm{id})}$ is the number of H-bonds expected from an ideal, noninteracting water–ethanol mixture:

$$n_{\mathrm{Hb}}^{(\mathrm{id})} = 3.03 - X_{\mathrm{EtOH}} \times (3.03 - 2.64) \tag{5.7}$$

This number depends on the concentration, and can be read from the dashed line under the total curve of Fig. 5.4a. The excess H-bonding $n_{\mathrm{Hb}}(r)$ [corresponding to the quantities inside the square brackets of Eq. (5.6), divided by ρ] and the running integrals $N_{\mathrm{Hb}}^{(\mathrm{xs})}(R)$ are presented in Fig. 5.6 for the low ethanol concentrations. They all converge to positive values, reflecting the fact that the water H-bonding is always in excess. Consistently with Fig. 5.4a, the maximum water H-bonding excess is for $X_{\mathrm{EtOH}} = 0.15$. For $X_{\mathrm{EtOH}} = 0.05$, an excess of H-bonding can be observed as far as 8 Å from the closest ethanol.

In Fig. 5.6, the two vertical lines indicate the positions of the first maximum and first minimum in the C_1–O RDF. They are not changing significantly in this low-ethanol concentration range. This shows that, although some water molecules in the outer region of the first shell may have a positive excess H-bonding, they are not contributing sufficiently to compensate for the negative excess H-bonding in the first shell. This suggests that the excess H-bonding is not due to the water in the first hydration shell, where the classical "clathrate" structure would be found. The distance at which the positive excess H-bonding overcomes the negative excess can be read directly from the intersection of the integral (in dashed line) with the zero-abscissa horizontal line. It gets shorter as the ethanol concentration increases but, interestingly, does not quite penetrate the first hydration shell. At $X_{\mathrm{EtOH}} = 0.20$, a significant compensation is occurring in the first shell (around 4 Å), but the negative excess is still overcome in the second hydration shell (around 5–6 Å). A similar "compaction" of the first hydration shell was noticed by Dixit *et al.* [10], who also suggested that it could be explaining the entropic contribution to the hydrophobic interaction.

The analysis of Fig. 5.6 cannot be done for $X_{\mathrm{EtOH}} \geq 0.40$ because the ethanol molecules are too condensed for a secondary hydration structure to be significantly populated. Any water molecule in the second hydration shell of one ethanol molecule is also in the first shell of another. At such concentration, the excess integrals are close to zero, and never become significantly negative for higher ethanol fraction (as shown in Fig. 5.4a). This could be explained by the fact that the

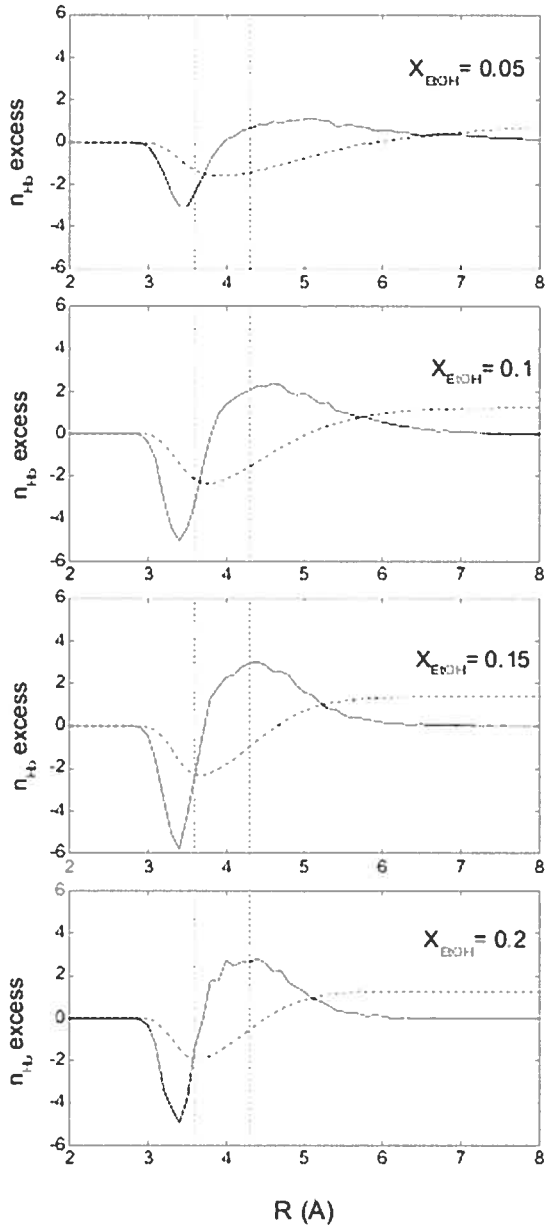

Figure 5.6: Radial distributions (solid lines) of the excess n_{Hb} for a water molecule at given distance from the nearest C_1 atom. The dashed lines are the running integrals $N_{\mathrm{Hb}}^{(\mathrm{xs})}(R)$ (see text) and represent the total H-bonding excess. The dashed vertical lines show the positions of the first peak of the C_1–O RDF and of the density minimum between the first and second solvation shells.

water clusters are robust enough to provide for themselves the contribution of the water–water H-bonding missing at high-ethanol concentration.

Hydrogen-bonded clusters

Liquid water is characterized by the existence of an extensive inter-connected network of hydrogen bonds that percolates throughout the entire phase [70]. Such a view is confirmed by the simulations based on the SWM4-DP model. Indeed, cluster analysis of the H-bond connectivity of pure water [28] reveals a H-bonding network extending over the whole liquid. It shows that less than 1% of the molecules are not connected to the bulk H-bond network. Upon the admixture of ethanol, the H-bond network of any type (water–water, water–ethanol, or ethanol–ethanol) becomes less dense, and larger disconnected clusters begin to be observed. Figure 5.7 presents the distribution of H-bonded cluster sizes (number of participating molecules).

According to Fig. 5.7a, a sufficiently high concentration of ethanol breaks the percolating water network. For $X_{\mathrm{EtOH}} \leq 0.3$, most of the water component is forming extensive clusters of sizes close to the total number of water molecules in the simulation box. For $X_{\mathrm{EtOH}} = 0.7$, most of the water is forming small clusters (significantly smaller than the total 75 water molecules in the simulation). The critical percolation point of the water network appears to be for X_{EtOH} slightly less than 0.5. At this concentration, cluster of all sizes can be observed, their distribution obeying approximately a power law [71]. From randomly generated liquid structures constrained by experimental data, as well as from standard MD simulations, Dougan *et al.* [15] found a comparable critical fraction (near 0.54) for methanol–water mixtures. Our analysis is also consistent with the results of neutron-diffraction scattering, showing non-ideal mixing of water at $X_{\mathrm{EtOH}} = 0.7$ [17].

The cluster analysis shows that, at $X_{\mathrm{EtOH}} = 0.90$, 69% of the water component is present as dispersed monomers, 22% as water dimers, 5% as water trimers, and 2% as water tetramers (see Fig. 5.7a). At $X_{\mathrm{EtOH}} = 0.70$, 29% of the water are

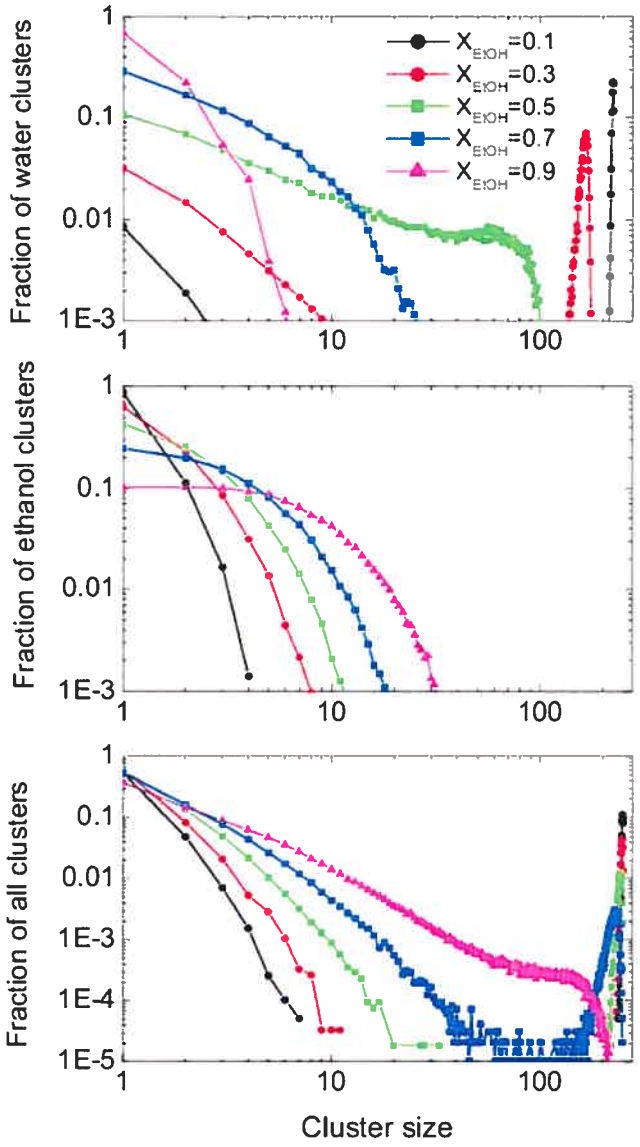

Figure 5.7: Molar contribution of H-bonded clusters in terms of their sizes for (a) pure-water clusters, (b) pure-ethanol clusters, and (c) mixed water–ethanol clusters. It is the molar contribution to the water component of the mixture for panel a, the ethanol component of the mixture for panel b, and the whole liquid for panel c.

monomers and 17% are dimers, 11% are trimers and 9% are tetramers. This is consistent with the neutron-diffraction results of Dixit *et al.* [17] for methanol–water mixtures, which show that, at a methanol fraction of 0.7, only a small fraction of the water component is monomeric.

Because of its lower H-bonding capability, small clusters of 1–5 molecules disconnected from the bulk network contain mostly ethanol. At $X_{\mathrm{EtOH}} = 0.10$, approximately 25% of all the single molecules not forming any H-bond are ethanol. Interestingly, this bias persists for clusters of mesoscopic sizes (at least at ethanol concentrations high enough for these clusters to be seen). For $X_{\mathrm{EtOH}} = 0.90$, clusters up to 20 molecules contain on average more than 90% ethanol. For $X_{\mathrm{EtOH}} = 0.70$, clusters up to 50 molecules contain on average more than 70% ethanol.

The most striking observation from Fig. 5.7b is that the ethanol component of the mixture does not percolate at any concentration point. This suggests that the mechanism of proton mobility in ethanol could be qualitatively different from the Grotthuss mechanism proposed for pure water [72], and might involve a reorganization of the H-bonding structure as described by Morrone and Tuckerman [73] for pure methanol. Alternatively, if one considers the network formed of any type of H-bond (see Fig. 5.7c), it appears that the presence of a relatively small amount of water ($X_{\mathrm{EtOH}} < 0.3$) forms a percolating two-component network from the ethanol clusters (which are by themselves non-percolating, see Fig 5.7b). This may, perhaps, be related to the maximum observed in the ethanol–water cross-contribution to the dielectric constant around the same molar fraction, as the total dielectric constant depends on the fluctuations of the totale dipole [34]. The formation of a two-component percolating network affects the H-bonding between water molecules and ethanol clusters at $X_{\mathrm{EtOH}} < 0.2$. There is optimal number of water molecules to solvate each ethanol molecule at $X_{\mathrm{EtOH}} = 0.2$. This provides an energetical and structural basis for the existence of the quasi-stable H-bonded ethanol–water clusters and suggests an explanation for the slow increase in the self-diffusion coefficients for X_{EtOH} larger than 0.2 (see Fig. 5.3). For methanol–water solutions,

Dougan *et al.* [15] observed that the self-diffusion coefficient reaches a minimum in the concentration regime; they note that this point is where both the H-bonding network of water and a network of hydrophobic packing of methanol are percolating. This mixed water–ethanol network is critically percolating at X_{EtOH} slightly less than 0.9. Although the exact point at which this percolation transition occur cannot be determined very accurately because of the finite size of the simulation box (250 molecules) and the finite simulation time, its existence is unambiguous.

5.4 Summary

A polarizable model for ethanol was parametrized to reproduce the vaporization enthalpy, static dielectric constant, and self-diffusion constant of the neat liquid at ambient conditions. It was used in combination with the polarizable SWM4-DP water model [28] to investigate the structural and dynamical origins of the hydrophobic hydration phenomenon observed in water–ethanol mixtures. Even though the ethanol model was parametrized independently of the water model, the properties of the mixtures obtained from MD simulations are in excellent agreement with experiment. The dielectric constant is correctly reproduced across the whole ethanol concentration range (see Fig. 5.2). The important features of the water and ethanol mobilities are reproduced as well: a rapid decrease in diffusivity for low ethanol concentration, an absolute mobility minimum as $X_{\mathrm{EtOH}} \sim 0.2$, and a gradual increase as the ethanol concentration gets higher (see Fig. 5.3). An important aspect of the polarizable model is that it produces a polarization response gradually with the composition of the mixture. That, of course, can be attained only with a model that accounts for induced polarization explicitly. It is not clear to what extent this feature explains the good accuracy obtained for the dielectric properties of the mixtures, but knowing how strongly the average induced dipole correlates with the dielectric constant [74], it is undoubtedly a significant factor.

The present work suggests a unifying view on alcohol hydration. The current simulations support the notion that the water–water H-bonding is enhanced

in the first hydration shell of the nonpolar group, as commonly assumed in the "iceberg" picture. But according to the simulations, such an effect does not dominate the overall structure because the water–water H-bonding is depleted in the neighborhood of the hydroxyl group. The net effect from the entire first hydration shell is a reduction rather than an excess of water H-bond, and the dominant contribution arises from the structuring of water in the second hydration shell of ethanol. The simulations also indicate the presence of water clustering at very high ethanol concentration, and reveal the existence of a transition between a fully percolating network of H-bonded water molecules at low ethanol concentration, to a non-percolating H-bonded network at high ethanol concentration.

Acknowledgements

Useful discussions with C. L. Brooks III, M. G. Kiselev, A. D. MacKerell Jr., and A. K. Soper are gratefully acknowledged. B. R. was supported by NIH grant GM-072558 and by NSF grant 0415784. S. Yu. N. was supported by the American Epilepsy Society and UCB Pharma Inc.

5.A Supplementary material

[See Figures 5.8, 5.9, and 5.10 on following pages.]

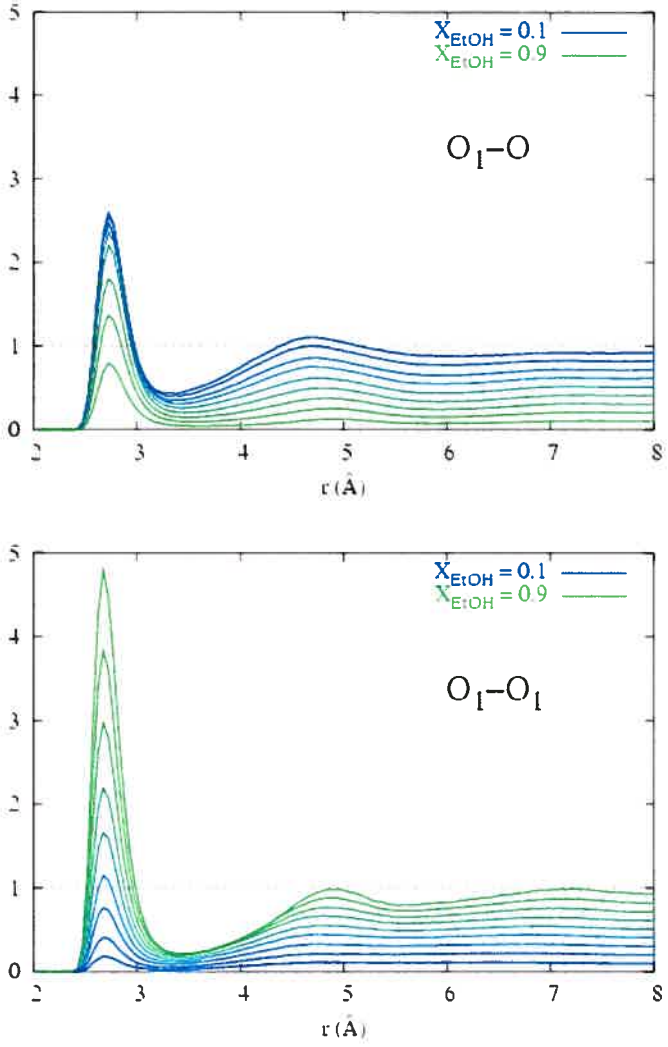

Figure 5.8: Oxygen–oxygen pair correlation functions for ethanol. The functions of the top graph are normalized according to the proportion of water, and those of the bottom graph, to the proportion of ethanol.

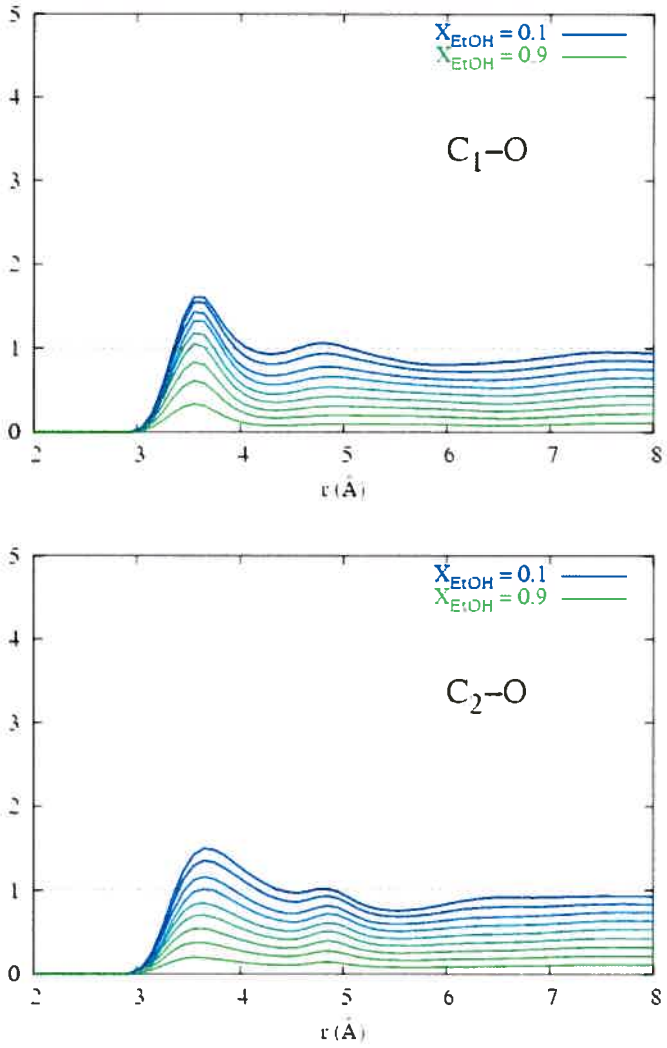

Figure 5.9: Carbon–oxygen pair correlation functions for ethanol. The functions are normalized according to the proportion of water.

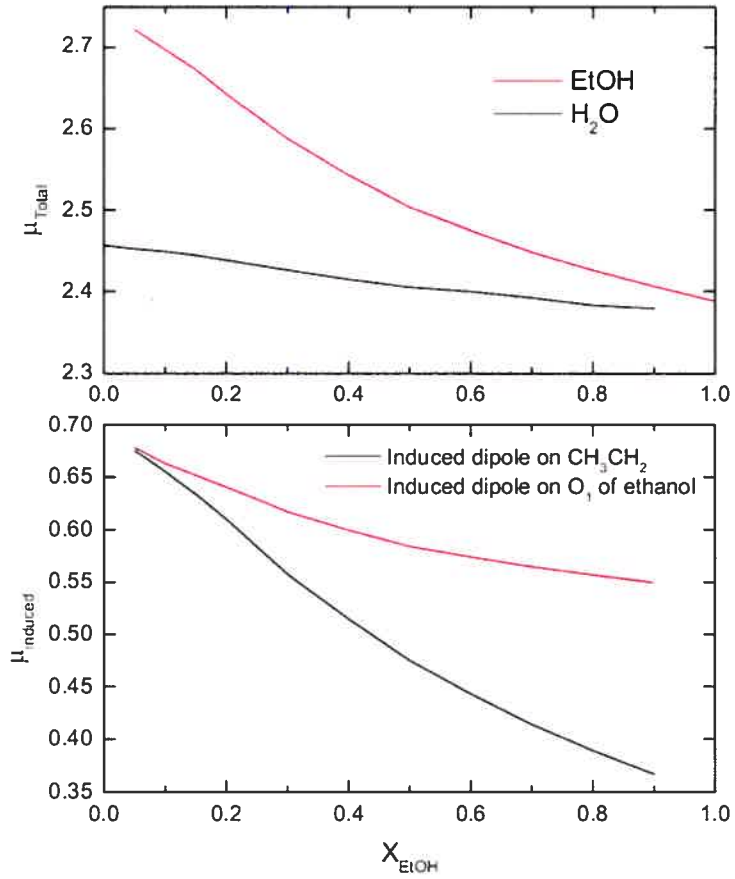

Figure 5.10: Molecular induction effects in ethanol and water as a function of the ethanol concentration. The top graph is the total molecular dipole (in Debyes) and the bottom graph is the induced dipole in the polar and apolar groups of ethanol.

References

[1] G. Hummer, S. Garde, A. E. García, and L. R. Pratt. New perspectives on hydrophobic effects. *Chem. Phys.*, 258(2–3) : 349–370, 2000.

[2] David M. Huand and David Chandler. Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. *Proc. Natl. Acad. Sci. U.S.A.*, 97(15) : 8324–8327, 2000.

[3] Noel T. Southall, Ken A. Dill, and A. D. J. Haymet. A View of the Hydrophobic Effect. *J. Phys. Chem. B*, 106(3) : 521–533, 2002.

[4] H. S. Frank and M. W. Evans. Free volume and entropy in condensed systems. 3. Entropy in binary liquid mixtures—Partial molal entropy in dilute solutions—Structure and thermodynamics in aqueous electrolytes. *J. Chem. Phys.*, 13(11) : 507–532, 1945.

[5] J. A. Boyne and Arthur Gordon Williams. Enthalpies of Mixing of Ethanol and Water at 25°C. *J. Chem. Eng. Data*, 12(3) : 318–318, 1967.

[6] K. Egashira and N. Nishi. Low-Frequency Raman Spectroscopy of Ethanol–Water Binary Solution : Evidence for Self-Association of Solute and Solvent Molecules. *J. Phys. Chem. B*, 102(21) : 4054–4057, 1998.

[7] William S. Price, Hiroyuki Ide, and Yoji Arata. Solution Dynamics in Aqueous Monohydric Alcohol Systems. *J. Phys. Chem. A*, 107(24) : 4784–4789, 2003.

[8] Yoshikata Koga, Keiko Nishikawa, and Peter Westh. "Icebergs" or No "Icebergs" in Aqueous Alcohols? : Composition-Dependent Mixing Schemes. *J. Phys. Chem. A*, 108(17) : 3873–3877, 2004.

[9] John L. Finney, Daniel T. Bowron, and Alan K. Soper. The structure of aqueous solutions of tertiary butanol. *J. Phys. Cond. Mat.*, 12(8A) : A123–A128, 2000.

[10] S. Dixit, A. K. Soper, J. L. Finney, and J. Crain. Water structure and solute association in dilute aqueous methanol. *Europhys. Lett.*, 59(3) : 377–383, 2002.

[11] J. L. Finney, D. T. Bowron, R. M. Daniel, P. A. Timmins, and M. A. Roberts. Molecular and mesoscale structures in hydrophobically driven aqueous solutions. *Biophys. Chem.*, 105(2–3) : 391–409, 2003.

[12] Fidler and Rodger. Solvation Structure around Aqueous Alcohols. *J. Phys. Chem. B*, 103(36) : 7695–7703, 1999.

[13] S. Y. Noskov, M. G. Kiselev, A. M. Kolker, and B. M. Rode. Structure of methanol-methanol associates in dilute methanol-water mixtures from molecular dynamics simulation. *J. Mol. Liq.*, 91(1–3) : 157–165, 2001.

[14] M. Kiselev and D. Ivlev. The study of hydrophobicity in water–methanol and water–*tert*-butanol mixtures. *J. Mol. Liq.*, 110(1–3) : 193–199, 2004.

[15] L. Dougan, S. P. Bates, R. Hargreaves, J. P. Fox, J. Crain, J. L. Finney, V. Réat, and A. K. Soper. Methanol–water solutions : A bi-percolating liquid mixture. *J. Chem. Phys.*, 121(13) : 6456–6462, 2004.

[16] M. D'Angelo, G. Onori, and A. Santucci. Self-association of monohydric alcohols in water : Compressibility and infrared absorption measurements. *J. Chem. Phys.*, 100(4) : 3107–3113, 1994.

[17] S. Dixit, J. Crain, W. C. K. Poon, J. L. Finner, and A. K. Soper. Molecular segregation observed in a concentration alcohol–water solution. *Nature*, 416(6883) : 829–832, 2002.

[18] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, 117(19) : 5179–5197, 1995.

[19] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., 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, M. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, 102(18) : 3586–3616, 1998.

[20] W. F. van Gunsteren, X. Daura, and A. E. Mark. The GROMOS force field. In P. von Ragué Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, III, and P. R. Schreiner (editors) : Encyclopedia of Computational Chemistry, volume 2, pages 1211–1216. Wiley & Sons, New York, 1998.

[21] William L. Jorgensen and Julian Tirado-Rives. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.*, 110(6) : 1657–1666, 1988.

[22] Martin Karplus. Special Issue on Molecular Dynamics Simulations of Biomolecules. *Acc. Chem. Res.*, 35 : 321–323, 2002.

[23] George A. Kaminski, Richard A. Friesner, and Ruhong Zhou. A computationally inexpensive modification of the point dipole electrostatic polarization model for molecular simulations. *J. Comput. Chem.*, 24(3) : 267–276, 2003.

[24] Marco Paolantoni and Branka M. Ladanyi. Polarizability anisotropy relaxation in liquid ethanol : A molecular dynamics study. *J. Chem. Phys.*, 117(8) : 3856–3873, 2002.

[25] Jiali Gao, Dariush Habibollazadeh, and Lei Shao. A Polarizable Intermolecular Potential Function for Simulation of Liquid Alcohols. *J. Phys. Chem.*, 99(44) : 16460–16467, 1995.

[26] Sandeep Patel and Charles L. Brooks, III. CHARMM fluctuating charge force field for proteins : I. Parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.*, 25(1) : 1–16, 2004.

[27] Sandeep Patel, Alexander D. MacKerell, Jr., and Charles L. Brooks, III. CHARMM fluctuating charge force field for proteins : II. *J. Comput. Chem.*, 25(12) : 1504–1514, 2004.

[28] Guillaume Lamoureux, Alexander D. MacKerell, Jr., and Benoît Roux. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.*, 119(10) : 5185–5197, 2003.

[29] M. P. Allen and D. J. Tildesley. Computer Simulation of Liquids. Clarendon Press, Oxford, 1987.

[30] B. T. Thole. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.*, 59 : 341–350, 1981.

[31] Piet Th. van Duijnen and Marcel Swart. Molecular and Atomic Polarizabilities : Thole's Model Revisited. *J. Phys. Chem. A*, 102(14) : 2399–2407, 1998.

[32] E. W. Washburn (editor). International Critical Tables of Numerical Data, Physics, Chemistry and Technology. Knovel, New York, 2003.

[33] J. G. Kirkwood. The dielectric polarization of polar liquids. *J. Chem. Phys.*, 7 : 911–919, 1939.

[34] M. Neumann and O. Steinhauser. Computer Simulation and the Dielectric Constant of Polarizable Polar Systems. *Chem. Phys. Lett.*, 106(6) : 563–569, 1984.

[35] A. D. Buckingham. A Theory of the Dielectric Polarization of Polar Substances. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.*, 238(1213) : 235–244, 1956.

[36] Guillaume Lamoureux and Benoît Roux. Modeling induced polarization with classical Drude oscillators : Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.*, 119(6) : 3025–3039, 2003.

[37] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald : An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12) : 10089–10092, 1993.

[38] Patrick Lagüe, Richard W. Pastor, and Bernard R. Brooks. Pressure-Based Long-Range Correction for Lennard-Jones Interactions in Molecular Dynamics Simulations : Application to Alkanes and Interfaces. *J. Phys. Chem. B*, 108(1) : 363–368, 2004.

[39] Robert A. Provencal, Raphael N. Casaes, Katja Roth, Joshua B. Paul, Chris N. Chapo, Richard J. Saykally, Gregory S. Tschumper, and Henry F. Schaeffer, III. Hydrogen Bonding in Alcohol Clusters : A Comparative Study by Infrared Cavity Ringdown Laser Absorption Spectroscopy. *J. Phys. Chem. A*, 104(7) : 1423–1429, 2000.

[40] William L. Jorgensen. Optimized intermolecular potential functions for liquid alcohols. *J. Phys. Chem.*, 90(7) : 1276–1284, 1986.

[41] L. Saiz, J. A. Padrò, and E. Guàrdia. Structure and Dynamics of Liquid Ethanol. *J. Phys. Chem. B*, 101(1) : 78–76, 1997.

[42] B. D. Smith and R. Srivastava. Thermodynamic Data for Pure Compounds, Part B : Halogenated Hydrocarbons and Alcohols. Elsevier, Amsterdam, 1986.

[43] Gösta Åkerlöf. Dielectric constants of some organic solvent–water mixtures at various temperatures. *J. Am. Chem. Soc.*, 54(11) : 4125–4139, 1932.

[44] S. Meckl and M. D. Zeidler. Self-diffusion measurements of ethanol and propanol. *Mol. Phys.*, 63(1) : 85–95, 1988.

[45] Leonor Saiz, Elvira Guàrdia, and Joan-Angel Padrò. Dielectric properties of liquid ethanol : A computer simulation study. *J. Chem. Phys.*, 113(7) : 2814–2821, 2000.

[46] A. L. McClellan. Tables of Experimental Dipole Moments, volume 3. Rahara Enterprises, El Cerrito, CA, 1989.

[47] Titus S. van Erp and Evert Jan Meijer. *Ab initio* molecular dynamics study of aqueous solvation ef ethanol and ethylene. *J. Chem. Phys.*, 118(19) : 8831–8840, 2003.

[48] Chris J. Benmore and Yen L. Loh. The structure of liquid ethanol : A neutron diffraction and molecular dynamics study. *J. Chem. Phys.*, 112(13) : 5877–5883, 2000.

[49] A. H. Narten and A. Habenschuss. Hydrogen bonding in liquid methanol and ethanol determined by x-ray diffraction. *J. Chem. Phys.*, 80(7) : 3387–3391, 1984.

[50] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2) : 926–935, 1983.

[51] Erik J. W. Wensink, Alex C. Hoffmann, Paul J. van Maaren, and David van der Spoel. Dynamic properties of water/alcohol mixtures studied by computer simulation. *J. Chem. Phys.*, 119(14) : 7308–7317, 2003.

[52] Allan J. Easteal, A. Vernon, J. Edge, and Lawrence A. Woolf. Effects of proton exchange on diffusion in aqueous solutions of methanol. *J. Phys. Chem.*, 89(7) : 1066–1069, 1985.

[53] Harris, Newitt, and Derlacki. Alcohol tracer diffusion, density, NMR and FTIR studies of aqueous ethanol and 2,2,2-trifluoroethanol solutions at 25°C. *J. Chem. Soc.—Faraday Trans.*, 94(14) : 1963–1970, 1998.

[54] Kazuko Mizuno, Yasue Miyashita, Yohji Shindo, and Hideo Ogawa. NMR and FT-IR Studies of Hydrogen Bonds in Ethanol–Water Mixtures. *J. Phys. Chem.*, 99(10) : 3225–3228, 1995.

[55] J. A. Larkin. Thermodynamic properties of aqueous non-electrolyte mixtures. I. Excess enthalpy for water + ethanol at 298.15 to 383.15 K. *J. Chem. Thermodyn.*, 7(2) : 137–148, 1975.

[56] J. Turner and A. K. Soper. The effect of apolar solutes on water structure : Alcohols and tetraalkylammonium ions. *J. Chem. Phys.*, 101(7) : 6116–6125, 1994.

[57] Petong, Pottel, and Kaatze. Water–Ethanol Mixtures at Different Compositions and Temperatures. A Dielectri Relaxation Study. *J. Phys. Chem. A*, 104(32) : 7420–7428, 2000.

[58] G. Onori and A. Santucci. Dynamical and structural properties of water/alcohol mixtures. *J. Mol. Liq.*, 69 : 161–181, 1996.

[59] Mauro Ferrario, Michael Haughney, Ian R. McDonald, and Michael L. Klein. Molecular-dynamics simulation of aqueous mixtures : Methanol, acetone, and ammonia. *J. Chem. Phys.*, 93(7) : 5156–5166, 1990.

[60] Hans De Loof, Lennart Nilsson, and Rudolf Rigler. Molecular dynamics simulation of galanin in aqueous and nonaqueous solution. *J. Am. Chem. Soc.*, 114(11) : 4028–4035, 1992.

[61] Alenka Luzar and David Chandler. Hydrogen-bond kinetics in liquid water. *Nature*, 379(6560) : 55–57, 1996.

[62] Alenka Luzar and David Chandler. Effect of Environment on Hydrogen Bond Dynamics in Liquid Water. *Phys. Rev. Lett.*, 76(6) : 928–931, 1996.

[63] A. K. Soper, F. Bruni, and M. A. Ricci. Site-site pair correlation functions of water from 25 to 400°C : Revised analysis of new and old diffraction data. *J. Chem. Phys.*, 106(1) : 247–254, 1997.

[64] A. Oleinikova, I. Brovchenko, A. Geiger, and B. Guillot. Percolation of water in aqueous solution and liquid–liquid immiscibility. *J. Chem. Phys.*, 117(7) : 3296–3304, 2002.

[65] Ph. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L. Å. Näslund, T. K. Hirsch, L. Ojamäe, P. Glatzel, L. G. M. Pettersson, and A. Nilsson. The Structure of the First Coordination Shell in Liquid Water. *Science*, 304 : 995–999, 2004.

[66] I-Feng W. Kuo and Christopher J. Mundy. An ab Initio Molecular Dynamics Study of the Aqueous Liquid-Vapor Interface. *Science*, 303 : 658–660, 2004.

[67] F. Franks and D. J. G Ives. Structural properties of alcohol–water mixtures. *Q. Rev. Chem. Soc.*, 20(1) : 1–44, 1966.

[68] T. Sato, A. Chiba, and R. Nozaki. Dynamical aspects of mixing schemes in ethanol-water mixtures in terms of the excess partial molar activation free energy, enthalpy, and entropy of the dielectric relaxation process. *J. Chem. Phys.*, 110(5) : 2508–2521, 1999.

[69] Takaaki Sato and Richard Buchner. Dielectric Relaxation Processes in Ethanol/Water Mixtures. *J. Phys. Chem. A*, 108(23) : 5007–5015, 2004.

[70] H. E. Stanley and J. Teixeira. Interpretation of the unusual behavior of H_2O and D_2O at low-temperatures—Tests of a percolation model. *J. Chem. Phys.*, 73 : 3404–3422, 1980.

[71] Christian D. Lorenz and Robert M. Ziff. Precise determination of the bond percolation thresholds and finite-size scaling corrections for the sc, fcc, and bcc lattices. *Phys. Rev. E*, 57(1) : 230–236, 1998.

[72] C. J. T. de Grotthuss. Sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann. Chim. (Paris)*, LVII : 54–74, 1806.

[73] Joseph A. Morrone and Mark E. Tuckerman. *Ab initio* molecular dynamics study of proton mobility in liquid ethanol. *J. Chem. Phys.*, 117(9) : 4403–4413, 2002.

[74] Michiel Sprik. Hydrogen bonding and the static dielectric constant in liquid water. *J. Chem. Phys.*, 95(9) : 6762–6769, 1991.

Chapitre 6

Polarizable force field for the hydration of alkali and halide ions: Thermodynamics, solvent structure, and water exchange reactions

Tiré de :
Guillaume Lamoureux et Benoît Roux. *Manuscrit en préparation.*

Abstract

A simple polarizable force field for the hydration of alkali and halide ions is parametrized based one the previously developed SWM4-DP water model [G. Lamoureux, A. D. MacKerell, Jr., and B. Roux, *J. Chem. Phys.*, 119, p. 5185, 2003]. The ions are represented as polarizable Lennard-Jones centers and the parameters of the models are systematically studied. It is shown that, for the alkali series, the model can satisfyingly reproduce the gas-phase properties of the hydrates, provided the absolute hydration free energies of the series are adjusted relative to an absolute hydration free energy for the proton of about -247 kcal/mol.

The model is extensively simulated for structural and dynamical properties. The extraction of the residence numbers predicts a 4-coordinated lithium ion ($N_r = 4.0$) for which the water molecules in the first solvation shell exchange with the

bulk liquid through an associative mechanism. For sodium, the model predicts an alternative 5,6-coordination ($N_r = 5.5$) and a dissociative/associative water exchange mechanism. For potassium, it predicts a fluid coordination of 6 to 8 water molecules ($N_r = 6.3$). The poorer reproduction of the halide hydrates points to some limitations of the model, but these limitations are not affecting the structure of the fully hydrated anions.

6.1 Introduction

It has been recognized for many years that an accurate polarizable force field is a invaluable methodological tool for understanding the hydration structure and dynamics of small ions like lithium, sodium, potassium, fluoride, and chloride [1, 2, 3, 4]. Many polarizable (or non-additive) force fields have been designed for this specific purpose, and used to gather considerable insight. With the steady progresses in the application of "non-empirical" methods, such as quantum dynamics and hybrid quantum/classical dynamics, these empirical force fields can be tested against more reliable data: better understood thermodynamical properties and more accurate coordination structures. The hope is that a properly calibrated model will perform as well as an *ab initio* method, without being confined to the microscopic realm of fundamental solvation processes.

The present effort is aimed at designing a simple yet accurate polarizable force field for the alkali and halide ion series. The immediate purpose of the force field is to study the properties of hydrated ions. Its ultimate purpose, however, is the simulation of large biomolecular systems in which ions such as sodium, potassium, and chloride play a role. Although these three ionic species are the most relevant to natural biological processes, other ions of the series are often used for *in vitro* physiological studies. It is therefore important to get a sense of the strengths and limitations of a force field of a given functional form across the whole alkali and halide series. For the model to be used to compare the relative affinities of the ions, it has to consistently reproduce the hydration thermodynamics of the series.

The modelling approach used in the study is similar to that employed for the design of the SWM4-DP water model [5]: we use as few new "ingredients" as possible, so that the model remains compatible with existing biomolecular force fields [6, 7, 8, 9], but we examine them systematically. Section 6.2 presents the functional form of the polarizable model for the ions and the procedure by which the empirical parameters are adjusted. Some important methodological issues regarding the calculation of absolute hydration free energies are also discussed. Section 6.3 presents an assessment of the model for small alkali and halide hydrates. Section 6.4 analyzes and discusses the hydration structure produced by the model. Finally, Section 6.5 describes the water exchange reaction pathways predicted by the model for lithium, sodium, and potassium.

6.2 Model and methods

6.2.1 Polarizable model

The ions are represented as point charges of $\pm|e|$ with a "12-6" Lennard-Jones interaction with the oxygen atoms of the water molecules. They are made explicitly polarizable with a Drude oscillator [10]. As a solvent, we use the recently developed SWM4-DP polarizable water model [5]. The SWM4-DP model has a correct static dielectric constant, which makes it appropriate to study water-mediated interactions. For a single ion solvated in N water molecules, the total potential energy is

$$\begin{aligned} U(\{\mathbf{r}\}) &= \sum_{i=0}^{N} \tfrac{1}{2} k_{\mathrm{D}} |\mathbf{r}_i - \mathbf{r}_{i,\mathrm{D}}|^2 + \sum_{i=0}^{N} \sum_{j=i+1}^{N} \sum_{s,s'} \frac{q_s q_{s'}}{|\mathbf{r}_{i,s} - \mathbf{r}_{j,s'}|} \\ &\quad + \sum_{i=0}^{N} \sum_{j=i+1}^{N} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{|\mathbf{r}_i - \mathbf{r}_j|} \right)^{12} - \left(\frac{\sigma_{ij}}{|\mathbf{r}_i - \mathbf{r}_j|} \right)^{6} \right]. \end{aligned} \tag{6.1}$$

The indices i and j refer to the molecule, the ion being $i = 0$ and the N water molecules being $i = 1$ to N. A position vector $\mathbf{r}$ with a single index ($\mathbf{r}_i$) refers to the position of the heavy atom in the molecule: the ion itself for $i = 0$ and the oxygen

atom for $i = 1$ to N. These are the only sites participating in Lennard-Jones interactions. When necessary, a second index (s) refers to a specific site on the molecule. The ion has two sites: the atomic core "I" and the Drude particle "D". Each SWM4-DP water molecule has four sites: the oxygen atom "O", the hydrogen atoms "H_1" and "H_2", a non-atomic site "M", and the Drude particle "D" (attached to the oxygen atom). See Ref. [5] for the details on the water model. The spring constant k_D is set to 1000 kcal/mol/Å^2. It dictates the magnitude of the charges on the Drude particles: $q_D = \sqrt{\alpha k_D}$. For the SWM4-DP model, $\alpha = 1.04252$ Å^3, hence $q_D = 1.77185\ |e|$. We use the Lorentz-Berthelot mixing rule [11] for the Lennard-Jones parameters:

$$\epsilon_{IO} = \sqrt{\epsilon_I \epsilon_O} \tag{6.2}$$

and

$$\sigma_{IO} = \tfrac{1}{2}(\sigma_I + \sigma_O). \tag{6.3}$$

For each ion, there are only three parameters to adjust: the ionic polarizability α_I and the Lennard-Jones parameters ϵ_I and σ_I. As mentioned previously, the total ionic charge $\pm|e|$ is placed on the I site. Rashin *et al.* [12] have analyzed the electron density from DFT calculations of small hydrated ions (lithium, sodium, fluoride, and chloride) and shown that "ion–water clusters spherical volumes containing full ionic charge of $\pm e$ have radii similar to the classical ionic radii." [12]. This is an indication that charge transfer is negligible even for the smallest ions.

6.2.2 Ionic polarizabilities

The ideal picture of a monovalent ion as a particle with a uniquely defined polarizability does not hold very well in the condensed phase. From the analysis of the dispersion properties of alkali halide crystals [13, 14, 15, 16], it was found that the polarizabilities of the atoms are strongly affected by the dense packing of the crystalline environment. And although the studies do not agree on the magnitude of this polarizability renormalization, they show that the effect is important for

halide ions, and that it depends significantly on the size and coordination of the counterions. A truly transferable model should take this variation into account.

For the present model of the halide series, we assign the polarizability values with a simple scheme. We used the "best" estimates of the vacuum ionic polarizabilities and renormalize them by a factor 0.724. This factor corresponds to the renormalization for the molecular polarizability of the SWM4-DP water model [5], due to Pauli's exclusion principle opposing the full gas-phase induction in the condensed phase. The SWM4-DP model has $\alpha = 1.04252$ Å^3, which is 72.4% of 1.44 Å^3, the experimental value for a single water molecule in the gas phase. The "best" polarizabilities for the halogen ions are taken from the time-dependent MP2 *ab initio* calculations of Hättig and Heß [17], that include the relativistic effects for the larger Br^- and I^- ions. These authors report polarizabilities of 2.467 Å^3 for F^-, 5.482 Å^3 for Cl^-, 7.268 Å^3 for Br^-, and 10.275 Å^3 for I^-. Similar *ab initio* results are obtained by Jemmer *et al.* [16]: 2.495 Å^3 for F^-, 5.545 Å^3 for Cl^-, and 6.182 Å^3 for Br^-. These *ab initio* values are higher than the empirical values of Coker [13] (1.38 Å^3 for F^-, 3.94 Å^3 for Cl^-, 5.22 Å^3 for Br^-, and 7.81 Å^3 for I^-).

For chloride, 72.4% of 5.482 Å^3 is 3.969 Å^3, which corresponds remarkably to the solvent-renormalized value of 4 Å^3 inferred by Jungwirth and Tobias [18] (who, incidentally, obtain 5.49 Å^3 for the free chloride ion [18]). For fluoride, bromide, and iodide, the renormalized values are 1.786, 5.262, and 7.439 Å^3, respectively. These crudely renormalized polarizabilities are higher than what many authors previously used for molecular dynamics simulations. More common values are around 1 Å^3 for fluoride [19, 20, 21], between 4.5 and 4.8 Å^3 for bromide [22, 21, 23] (but sometimes [24] as high as 5.5–5.8 Å^3), and between 5.3 and 6.9 Å^3 for iodide [22, 21, 23, 25]. For fluoride, it could be argued that the polarizability should be smaller than 1.786 Å. Indeed, a smaller ion is more closely coordinated by the solvent, and its effective volume is smaller. However, a recent *ab initio* study [26] predicts a solvent-renormalized value somewhat closer to 1.786 Å^3 than to 1 Å^3: 1.588–1.875 Å^3 for "surface" hydration states and 1.346–1.636 Å^3 for "interior" hydration states (compared to 2.405 Å^3 in the gas phase). Because the 0.724

Table 6.I: Parameters of the polarizable ions.

	q ($\|e\|$)	α_{gas} (Å^3)	α (Å^3)	ϵ_{I} (kcal/mol)	σ_{I} (Å)
Li^+	+1	0.032	0.032	0.00300	2.441062
Na^+	+1	0.157	0.157	0.03151	2.583606
K^+	+1	0.830	0.830	0.18290	2.931056
Rb^+	+1	1.370	1.370	0.35190	3.118145
Cs^+	+1	2.360	2.360	0.35760	3.527958
F^-	−1	2.467	1.786	0.01000	4.026861
Cl^-	−1	5.482	3.969	0.07658	4.383221
Br^-	−1	7.268	5.262	0.10820	4.579218
I^-	−1	10.275	7.439	0.15910	4.917760

renormalization factor appears very appropriate for chloride, and because it is thought to decrease as the size of the ion increases [27], it is difficult to raise any *a priori* objection about the renormalized polarizabilities of bromide and iodide being too large. Although a constant renormalization factor is chosen for the sake of simplicity, our proposed values are similar to those derived by Pyper *et al.* [28], which are: 1.30 Å^3 for F^-, 3.76 Å^3 for Cl^-, 5.07 Å^3 for Br^-, and 7.41 Å^3 for I^-.

Unlike those of the halide ions, the polarizabilities of the alkali ions are not significantly changed by their coordination environment [13], and we use the gas-phase values reported by Mahan [29]: 0.032 Å^3 for Li^+, 0.157 Å^3 for Na^+, 0.830 Å^3 for K^+, 1.370 Å^3 for Rb^+, and 2.360 Å^3 for Cs^+. The exact α values are somewhat irrelevant for an empirical force field, because the alkali ions are not polarizable enough to affect the spherical symmetry of the solvent. Classical MD simulations of polarizable Na^+ in polarizable water [30] have shown that the electrostatic on the cation is not affected by any reasonable choice for the value of α.

Table 6.I displays the charges and polarizabilities for the model ions, compared with the gas-phase values of Mahan [29] and Hättig and Heß [17]. The values of the Lennard-Jones parameters are obtained from an empirical adjustment of the hydration free energy, and will be justified later on.

6.2.3 Absolute free energies of hydration

The persistent problem of determining the absolute free energies of ion hydration has been revisited in recent years [31, 32, 33, 34, 35, 36]. The intrinsinc difficulty of the problem is that thermodynamical measurements have only access to the free energies of processes involving neutral macroscopic systems, and are irreductibly left with *conventional* hydration free energies that are either the sum of the absolute free energies of an ion and a counter-ion, $\Delta G^{\text{abs}}_{\text{hydr}}(M^+) + \Delta G^{\text{abs}}_{\text{hydr}}(X^-)$, or the difference of the free energies of two ionic species of the same sign, $\Delta G^{\text{abs}}_{\text{hydr}}(M_1^+) - \Delta G^{\text{abs}}_{\text{hydr}}(M_2^+)$. Any contribution to the hydration free energy that is proportional to the charge of the ion cannot be resolved from calorimetric measurements alone. Traditionally, the absolute free energies were estimated from an extrathermodynamic assumption, based on an *a priori* notion of two specific ions having comparable absolute free energies. The most commonly used is the TATB hypothesis [37], that states that tetraphenylarsenium and tetraphenylborate ions have the same hydration free energies. But despite two ions of opposite charges having comparable (and large) sizes, the free energies associated with the solvent reorganization have no *a priori* reason of being comparable. In addition, the free energy contributions from the charge crossing the electrostatic potential ϕ at the air/water interface, $+F\phi$ for a monovalent cation and $-F\phi$ for a monovalent anion, are largely unknown. Because they cancel out for neutral macroscopic systems, they cannot be extracted from a standard calorimetry experiment and require some extrathermodynamic input.

In this paper, we denote the *absolute* hydration free energy estimates, independent of the phase potential the ion had to cross, by the superscript "abs" ($\Delta G^{\text{abs}}_{\text{hydr}}$) and the *real* hydration free energy estimates, including the contribution from the air/water interface potential, by the superscript "real" ($\Delta G^{\text{real}}_{\text{hydr}}$). For an ion of valence z, they are related through the phase potential ϕ:

$$\Delta G^{\text{real}}_{\text{hydr}} = \Delta G^{\text{abs}}_{\text{hydr}} + zF\phi. \tag{6.4}$$

From molecular simulations, ϕ is thought to be negative. This means that a positive test-charge going from air to bulk water across the air/water interface will experience a potential drop of $|\phi|$. The SWM4-DP model produces a air/water interface potential $\phi = -540$ mV [5], while the polarizable model of Dang and Chang [38] yields -500 mV [39]. The standard TIP3P model [40] produces -500 mV as well [41].

Pursuing along the original idea of Klots [32], Tissandier *et al.* [33] have extrapolated free energy measurements on small ionic hydrates to obtain hydration free energies of ions in bulk phase. Tissandier *et al.* argue that the gas-phase analyses do not include the phase potential (that is, that they provide with the absolute energies $\Delta G^{\text{abs}}_{\text{hydr}}$) because the measured clusters are too small for a genuine air/water interface to develop. We do not consider this argument sufficient, and instead we agree with Grossfield *et al.* [36] who emphasize that the quantity $\Delta G^{\circ}_{n,n+1}(M^+) - \Delta G^{\circ}_{n,n+1}(X^-)$ does imply an extrathermodynamic hypothesis: that the interface potential created by the reorganization of the outermost layers of water, however different it is from the bulk limit, is independent of the charge z of the ion. We know this hypothesis is invalid in the bulk limit. In the bulk limit, the reorganization would recreate the phase potential ϕ and contribute to an additional free energy term $zF\phi$. It is not yet clear how the free energy contribution from the surface reorganization of the small clusters can be accurately extrapolated [42, 43].

That said, Zhan and Dixon [34] computed a value for the "absolute" hydration free energy of the proton comparable to that of Tissandier *et al.* [33] (1.6 kcal/mol higher), using *ab initio* electronic structure calculations of a proton solvated by a small number of water molecules embedded in a reactive continuum. Embedding a quantum-mechanical description of the ion and its first solvation shell into a classical molecular dynamics simulation, Asthagiri *et al.* [35] (who are aware of the difficulties in interpreting the gas-phase measurements of Tissandier *et al.*) obtained absolute hydration free energies for a selection of small ions (H^+, Li^+, Na^+, and OH^-). For Li^+ and Na^+, their results are about 12 kcal/mol less negative than those of Tissandier *et al.* Using a high-level polarizable force field parametrized

from *ab initio* gas-phase properties, Grossfield *et al.* [36] predicted a somewhat comparable absolute hydration free energy for the proton, 11.5 kcal/mol less negative than that of Tissandier *et al.*

Figure 6.1 summarizes some representative results from the literature [31, 32, 44, 37, 33, 34, 35, 36]. All the energies have been converted in reference to the following standard states: an ideal gas of ions at a pressure of 1 atm (where 1 mol of ions occupies a volume of 24.465 ℓ), and an idealized 1 M ionic solution (with no ion–ion interactions). The entropic contribution associated with confining 1 mol of ions from a volume of 24.465 ℓ to a volume of 1 ℓ is $-k_{\mathrm{B}}T\ln(1/24.465) \simeq$ 1.9 kcal/mol, and this energy has to be subtracted from a calculation where no density change is implicitly taken into account. Most of the results fall within a band of 12 kcal/mol, which corresponds approximately to the contribution from the air/water potential for a monovalent ion ($\pm F\phi$). The arrows are illustrating the correction to the free energy for a air/water interface potential of -540 mV, that is, $+12.45$ kcal/mol for the monovalent cations and -12.45 kcal/mol for the monovalent anions. This simple graphical observation shows how acknowledging the existence of a phase potential can possibly reconcile the spread in various estimates of the "absolute" free energies. It also suggests that the cluster measurements of Klots [32] and Tissandier *et al.* [33] contain a contribution from the phase potential, as well as the calculations of Zhan and Dixon [34]. This was previously noted by Asthagiri *et al.* [35].

The classic results of Noyes [45] are presented in Fig. 6.1 as well, although they rely on somewhat inaccurate thermodynamic data and crude extrathermodynamic assumptions. The values represented are $\Delta F^0_{\mathrm{el}}+\Delta F^0_{\mathrm{neut}}$ (following the nomenclature of Noyes), where ΔF^0_{el} are taken from Table I of Ref. [45] and $\Delta F^0_{\mathrm{neut}}$ is estimated from the so-called "zero energy assumption" ($\Delta F^0_{\mathrm{neut}} = 1.325$ kcal/mol). The early electrochemical measurements of Randles [47] (based on the proof of concept of Latimer [49]) are presented as well, even though they rely on inaccurate estimations of the mercury work function.

Grossfield *et al.* [36] argued that the discrepancy between their prediction of

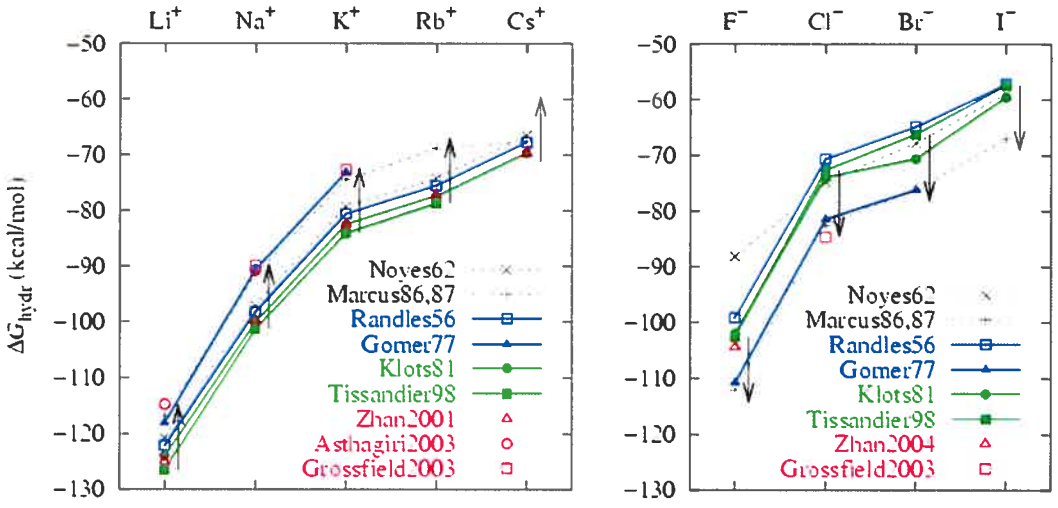

Figure 6.1: Survey of the absolute solvation free energies for the alkali metal and halide ions. Noyes62 (Ref. [45]) and Marcus86,87 (Ref. [46, 44, 37]) are based on "classical" extrathermodynamic hypothesis, Randles56 (Ref. [47]) and Gomer77 (Ref. [31]) are based on eletrochemical measurements, Klots81 (Ref. [32]) and Tissandier98 (Ref. [33]) are based on extrapolation of gas-phase cluster measurements, and Zhan2001 (Ref. [34]), Asthagiri2003 (Ref. [35]), Grossfield2003 (Ref. [36]), and Zhan2004 (Ref. [48]) are theoretical predictions.

$\Delta G_{\text{hydr}}(\text{H}^+)$ and that of Zhan and Dixon [34] could arise from an irreductible contribution from the potential at the ion/water interface. Indeed, Ashbaugh has shown [50] that, for fictitious ions having arbitrarily large sizes, the water molecules of the first solvation shell will form an inverted interface very similar to an air/water interface. This argument is useful for understanding why the TATB extrathermodynamic hypothesis is not valid, but is irrelevant for the solvation of small ions, because the rearrangement of the solvent at the surface of a small ion (like H^+) is an intrinsinc contribution to the hydration free energy. From Fig. 6.1, it appears that the difference might have an different origin: the calculation of Zhan and Dixon might include the artificial effect of a continuum/water interface. Because the continuum does not describe the geometry of hydrogen bonding, it might not prevent the explicit water molecules from forming an interface.

Figure 6.1 suggests an interpretation of the measured and calculated free energies that is sometimes in contradiction with the interpretations given by the

authors. For the alkali series, assuming that the contribution from the phase potential is indeed close to +12 kcal/mol, the low energies may be considered estimates of $\Delta G^{\text{real}}_{\text{hydr}}$ and the high energies may be considered estimates of $\Delta G^{\text{abs}}_{\text{hydr}}$ This consensus would require a revision of the interpretation of the measurements of Tissandier *et al.* [33] and the calculations of Zhan and Dixon [34] ($\Delta G^{\text{real}}_{\text{hydr}}$ instead of $\Delta G^{\text{abs}}_{\text{hydr}}$), but would confirm the interpretations of Asthagiri *et al.* [35] and Grossfield *et al.* [36].

6.2.4 Free energy calculations

The $\Delta G^{\text{real}}_{\text{hydr}}$ values calculated refer to the process of an ion going from the ideal gas phase at 1 atm to an idealized bulk solution at 1 M concentration. The free energies are obtained by the thermodynamic integration (TI) of the mutation of an argon-like atom (neutral and nonpolarizable) into a charged and polarizable ion. The mutation Ar $\rightarrow$ M^+ (or X^-) is controlled by a parameter λ going from 0 to 1. The free energy integral

$$\Delta G = \int_0^1 d\lambda \left\langle \frac{\partial E(\lambda)}{\partial \lambda} \right\rangle_{(\lambda)} \tag{6.5}$$

is computed as a discrete sum over 11 windows $\lambda = 0, 0.1, 0.2, \ldots, 1$. To correct for possible hysteresis errors, the mutation is conducted "forward" and "backward" and the sampling is averaged. Each window is simulated with 10 ps of molecular dynamics, of which only the last 9 ps are used for the averaging. The 22 simulations are performed sequentially, going from $\lambda = 0$ to 1 and back to 0. The final configuration of each window serves as the initial configuration for the next. The first window to run uses a configuration equilibrated for 20 ps. The λ parameter turns on the charges on both the nuclear site I and the Drude particle D, so that both the ionic charge and the ionic polarizability are gradually introduced as λ goes from 0 to 1. The λ parameter also controls the switching from the Lennard-Jones parameters of the argon model to the Lennard-Jones parameters of the ion.

To the mutation free energy ΔG of Eq. (6.5) is added the hydration free energy

or the argon-like atom, $\Delta G_{\text{hydr}}(\text{Ar})$. This quantity is extracted from a separate calculation of the potential of mean force (PMF) of the argon atom going across a slab of 125 SWM4-DP water molecules. The PMF is constructed using umbrella sampling and the WHAM procedure [51]. From three replicas of the PMF calculation, we get $\Delta G_{\text{hydr}}(\text{Ar}) = +2.8$ kcal/mol. The argon-like model itself uses the Lennard-Jones parameters $\epsilon = 0.2339$ kcal/mol and $\sigma = 1.9087 \times 2^{5/6} = 3.4009$ Å.

For the TI calculation, the ion is placed in the middle of a droplet of 250 explicit SWM4-DP water molecules, contained by a reactive spherical boundary potential representing implicit solvent (SSBP) [52]. Preliminary calculations in cavities with less explicit water molecules have shown that 250 molecules is enough to avoid any significant finite-system effect on the estimation of ΔG. Room temperature (298.15 K) is maintained using a Nosé–Hoover thermostat [53] on the atoms. The radius of the SSBP implicit cavity is adjusted dynamically to reproduce a correct internal pressure [52]. The Drude oscillators are maintained in the self-consistent field induction regime by being coupled to a second Nosé–Hoover thermostat at low temperature (1 K). See Ref. [10] for details on the simulation procedure. The equation of motions are integrated with a time step of 1 fs. All simulations are performed with the program CHARMM [54], using the VV2 molecular dynamics module [10].

The SSBP method produces an artificial phase potential at the interface between the explicit water molecules and the implicit solvent. As a net charge $\pm|e|$ is grown in the middle of the explicit ball of water, this interface potential contributes to the free energy by a term $\mp F|\phi_{\text{SSBP}}|$, where $|\phi_{\text{SSBP}}|$ is the potential drop created by the reorganization of the first few layers of water molecules inside the cavity in the SSBP continuum. This potential can be calculated by integrating the radial charge density obtained for a separate SSBP simulation with no solute. Remarkably, it is almost identical to the phase potential obtained from a simulation of a slab of explicit water in vacuum (-540 mV for the SWM4-DP model) [5]. In other words, although the structure of the first layers of explicit water in a SSBP cavity is different from that of the slab simulation, the presence of the

SSBP continuum does not significantly affect the potential building up in the bulk. It is comparable to the air/water interface potential within less than 1 kcal/mol. This observation holds for both the SWM4-DP model used in this work and the nonpolarizable TIP3P model, that is, $|\phi_{\mathrm{SSBP}}| \simeq -540$ for a SSBP simulation of the SWM4-DP model and $|\phi_{\mathrm{SSBP}}| \simeq -500$ for a simulation of the TIP3P model. Hence, the continuum/explicit interface potential can play the role of the physical air/water interface that the ion would have crossed going from the gas phase to the bulk water. The free energy computed from the charging of the solute in the SSBP represents the so-called "real" hydration free energy.

The contribution from this interface potential explains the large differences between the original free energy calculations [52] that led to the parametrization of the CHARMM force field for sodium, potassium, and chloride (using SSBP) and those reproduced by Grossfield *et al.* [36] using periodic boundary conditions. The difference corresponds almost exactly to the contribution from crossing the air/water interface of TIP3P water ($\phi = -500$ mV and $F\phi = \mp 11.53$ kcal/mol). Thus, the SSBP calculations are consistent with periodic-boundary-conditions calculations, provided the implicit/explicit interface potential it generates is taken into account.

6.2.5 Parametrization strategy

The functional form of the ion solvation model we are proposing in this work is not as flexible as the one of Grossfield *et al.* [36] and we will not attempt to make a prediction about the absolute hydration free energies. Instead, we will explore the range of solvation properties accessible from reasonable Lennard-Jones parameters. For each ionic species, we find the Lennard-Jones parameters that reproduce a set of (U_{min},d_{min}) pairs, where U_{min} is the interaction energy with a single water molecule and d_{min} is the ion–oxygen distance. These pairs lie on a rectangular grid in the d_{min}-U_{min} plane, spanning it in increments of 0.25 kcal/mol for the energy and of 0.025 Å for the distance. For each grid point (U_{min},d_{min}), we produce a ΔG value for the mutation of argon to the particular ion model.

Because of the "hard" repulsion of the Lennard-Jones 12–6 potential, it is not possible to reproduce any ($U_{\min}$,$d_{\min}$) pair on the grid with reasonable values of ϵ. We did not consider models requiring ϵ smaller than 0.001 kcal/mol or larger than 0.5 kcal/mol.

The statistical errors on the individual estimations of ΔG are averaged out by the least-squares fit of a quadratic response-surface model (using six coefficients, a_0 to a_5)

$$\Delta\widehat{G}(U_{\min}, d_{\min}) = a_0 + a_1 U_{\min} + a_2 d_{\min} + a_3 U_{\min}^2 + a_4 U_{\min} d_{\min} + a_5 d_{\min}^2. \quad (6.6)$$

This polynomial is used to predict the free energy of interpolating ionic models. As long as the model has properties ($U_{\min}, d_{\min}$) within the region spanned by the original grid, the prediction is more accurate than a single TI calculation.

6.2.6 Lennard-Jones parameters

Figure 6.2 presents the mapping of the response surfaces $\Delta\widehat{G}$ in the $d_{\min}$-ΔG_{hydr} plane. To compare with the reference energies of Fig. 6.1, the hydration free energies $\Delta G_{\text{hydr}}(\text{Ar}) + \Delta\widehat{G}$ are added $k_{\text{B}}T\ln(24.465) \simeq 1.9$ kcal/mol. The regions of the parameter plane where the free energy can be estimated from the response-surface model are hatched: a free energy calculation was done on each intersection of the grid. Their diagonal band shape is due to the upper and lower limits imposed on the ϵ parameter. For each ion, four ϵ levels are shown in dotted lines: $\epsilon = 0.3$, 0.1, 0.03, and 0.01 kcal/mol.

On the graph, the range of reasonable hydration free energies is defined by a band of 12.45 kcal/mol positioned at the "$\Delta G_{\text{hydr}}^{\text{real}}$" levels set to the free energies of Tissandier *et al.* [33] (the thick horizontal lines). The thinner lines are 12.45 kcal/mol higher for the cations and 12.45 kcal/mol lower for the anions, and correspond to consensus values for $\Delta G_{\text{hydr}}^{\text{abs}}$. The vertical lines connecting the two energy levels correspond to the *ab initio* ion–oxygen distances computed by Glendening and Feller [55]: 1.87 Å for Li^+, 2.26 Å for Na^+, 2.64 Å for K^+, 2.87 Å for

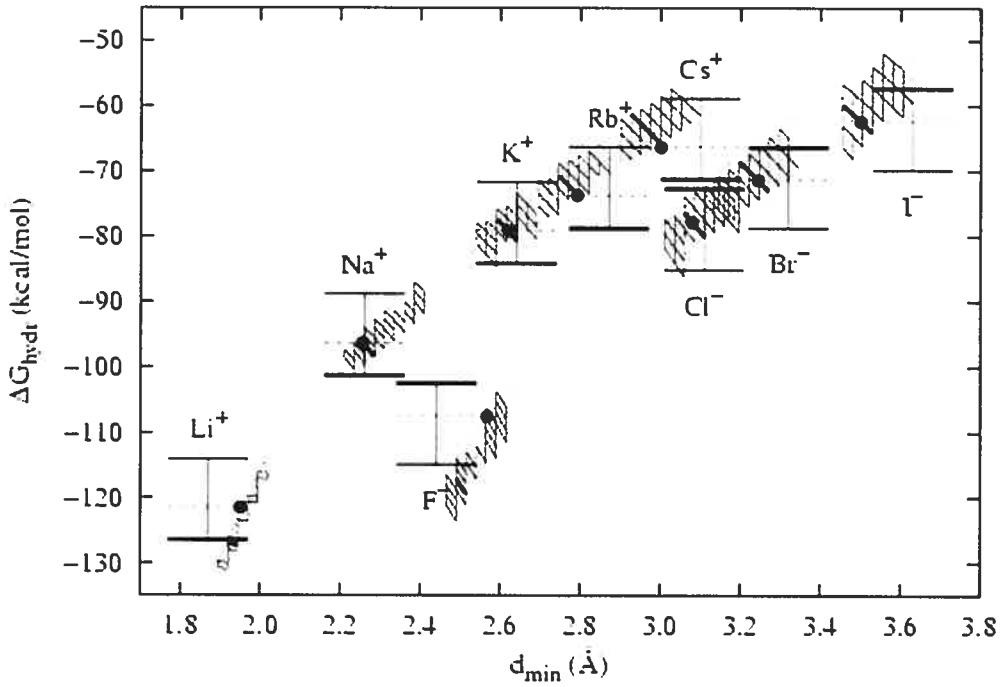

Figure 6.2: Mappings of the $(U_{\min}, d_{\min})$ properties on the SSBP free energies. The full circles correspond to the AH/SWM4-DP model, adjusted to some consensus "real" free energies corrected for 40% of the SWM4-DP air/water phase potential, and to the reference $U_{\min}$ values (except for lithium and fluoride, see text). The thick horizontal bars are Tissandier *et al.* results [33] and the thinner horizontal bars are corrected for 100% of the SWM4-DP phase potential. For each species, four oblique dotted curves show loci of constant ϵ (0.3, 0.1, 0.03, and 0.01 kcal/mol). The grids, rectangular in the $d_{\min}$-$U_{\min}$ plane, are positioned in reference to the target $U_{\min}$ values, shown using a thicker grid line.

Rb^+, and 3.10 Å for Cs^+. For the halide ions, they correspond to the *ab initio* distances of Kim *et al.* [56]: 2.44 Å for F^-, 3.11 Å for Cl^-, 3.32 Å for Br^-, and 3.63 Å for I^-.

For the alkali ions, the target $U_{\min}$ values are the binding enthalpies of Ďzidić and Kebarle [57]: −34.0 kcal/mol for Li^+, −24.0 kcal/mol for Na^+, −17.9 kcal/mol for K^+, −15.9 kcal/mol for Rb^+, and −13.7 kcal/mol for Cs^+. These values are in agreement with other experimental values [58, 59] and with *ab initio* calculations [60, 55, 61, 62, 63, 64, 65, 66]. For the halide ions, the target $U_{\min}$ values are the *ab initio* binding energies of Kim *et al.* [56]: −25.9 kcal/mol for F^-, −14.4 kcal/mol for Cl^-, −12.7 kcal/mol for Br^-, and −10.6 kcal/mol for I^-. These energies agree with the very accurate results of Xantheas and Dang for fluoride [20] and chloride [67]. Within each grid, the target $U_{\min}$ energies are highlighted with thicker gridlines.

According to the consensus free energy levels of Fig. 6.1, the $\Delta G^{\text{real}}_{\text{hydr}}$ values calculated with SSBP should lie on the thick lines of Fig. 6.2. But it is clearly not possible to reproduce both the binding energies to a single molecule and the consensus $\Delta G^{\text{real}}_{\text{hydr}}$'s. For the larger ions of the alkali series, it would require using values of the well-depth parameter ϵ well above anything reasonable. For the larger ions of the halide series, very small values of ϵ would have to be used, but would still prove inadequate for fluoride. Relaxing the accuracy of the match for the binding energies $U_{\min}$ would help for the anions, but would require very short ion–oxygen distances for the large cations. In regard of those constraints, we choose the target values for the absolute free energies to be the experimental energies of Tissandier *et al.* [33], corrected for 40% of the air/water potential according to the SWM4-DP water model. This arbitrary reference level is identified on Fig. 6.2 by dashed horizontal lines. With the exception of lithium and fluoride, the smallest ions of each series, it allows the target binding energies to be reproduced while keeping the free energies consistent with the thermodynamical data.

The chosen Lennard-Jones parameters are presented in Table 6.I. The parameters for lithium at fluoride are adjusted *ad hoc*, by fixing ϵ at a reasonable value and by adjusting σ to reproduce the hydration free energy. The properties of the

Table 6.II: Binding energies and ion–oxygen distances for the minimum energy monohydrates, and "real" hydration free energies. The "absolute" free energies are obtained by correcting the real energies by ±12.45 kcal/mol, the contribution from a SWM4-DP air/water interface.

	$U_{\min}$ (kcal/mol)	$d_{\min}$ (Å)	$\Delta G^{\text{real}}_{\text{hydr}}$ (kcal/mol)	$\Delta G^{\text{abs}}_{\text{hydr}}$ (kcal/mol)
Li^+	−32.538	1.951	−121.5	−109.1
Na^+	−24.000	2.255	−96.3	−83.9
K^+	−17.900	2.620	−79.1	−66.7
Rb^+	−15.900	2.792	−73.7	−61.3
Cs^+	−13.700	3.001	−66.3	−53.9
F^-	−23.317	2.567	−107.5	−120.0
Cl^-	−14.400	3.080	−77.8	−90.3
Br^-	−12.700	3.244	−71.4	−83.9
I^-	−10.600	3.501	−62.4	−74.9

ions in association with the SWM4-DP model are presented in Table 6.II. This set of parameters will be referred to as the AH/SWM4-DP model (AH standing for "alkali and halide").

The optimal values of the σ_I's can be related to the ionic radii. For the alkali series, these radii (R_I) are [68]: 0.73 Å for a 4-coordinated Li^+, 1.16 Å for a 6-coordinated Na^+, 1.60 Å for a 7-coordinated K^+, 1.75 Å for an 8-coordinated Rb^+, and 1.95 Å for a 10-coordinated Cs^+. It will be shown later that these coordination numbers are the integers closest to the actual coordination numbers obtained from the simulations. The σ_I/R_I ratios are smoothly converging as the ion size increases: 334% for Li^+, 223% for Na^+, 183% for K^+, 178% for Rb^+, and 181% for Cs^+. For the halide series, the ionic radii are [68]: 1.19 Å for F^-, 1.67 Å for Cl^-, 1.82 Å for Br^-, and 2.06 Å for I^- (for hexa-coordinated ions). The σ_I/R_I ratios are: 338% for F^-, 262% for Cl^-, 252% for Br^-, and 239% for I^-.

6.2.7 Calculation of the solvent structure and dynamics

For the purpose of calculating the solvent structure and the dynamic properties, a single ion of each species is simulated in a cubic box of 500 SWM4-DP water

molecules, with periodic boundary conditions. The electrostatic interaction energy is summed up using the particle-mesh Ewald (PME) method [69]. Constant temperature (298.15 K) and pressure (1 atm) are maintained by coupling the atoms to a Nosé–Hoover thermostat [53] and a modified Andersen–Hoover barostat [70] implemented in the VV2 module [10] of CHARMM. The dynamics of the Drude oscillators is integrated as described previously.

Five replicas of each simulation are equilibrated for 100 ps and average values are compiled for an additional 1000 ps, for a total of 5 ns of data for each ion type. The trajectories are saved every 0.1 ps, and analyzed in segments of 100 ps.

6.3 Gas-phase properties

The model was parametrized for the correct hydration energetics at infinite dilution. Because the present effort is concerned by the solvent structure and water exchange reactions, it is also important to look at how accurately the AH/SWM4-DP model reproduces the structures and the energies of small ion hydrates. The zero-temperature hydrates are very revealing of how the competition between ion–water and water–water interactions is resolved for the different ions [71].

Minimum energy clusters are generated by optimizing the geometry of many thousands of conformations of n molecules randomly placed and oriented around the ion. Tables 6.III and 6.IV present the binding energies for the AH/SWM4-DP model. Only the lowest-energy clusters are reported. They are identified following the "n_1+$n_2(S)$" notation, where n_1 is the number of water molecules in the first solvation shell, n_2 is the number of water molecules in the second solvation shell, and S is the point-group symmetry. The structures are shown in Figs. 6.3 and 6.4.

6.3.1 Small alkali hydrates: $M^+\cdot(H_2O)_{n=1-6}$

With the exception of lithium, the binding energies of the alkali monohydrates were adjusted to the experimental binding enthalpies of Ďzidić and Kebarle [57]. It is remarkable to observe that, having good binding energies for the monohydrates,

the polarizable force-field is able to reproduce the consensus *ab initio* energies of the larger clusters within a few percents.

Some important structural effects are equally well reproduced. The most stable dihydrate is the "linear" 2+0(D_{2d}) structure for lithium and sodium, but is the "bent" 2+0(C_2/C_s) structure for rubidium and cesium, as predicted by *ab initio* calculations [72, 55, 64, 66]. For potassium, the most stable structure is 2+0(C_2/C_s), but the 2+0(D_{2d}) structure is only 0.024 kcal/mol higher. Scanning the O–K–O angle shows that the energy surface is extremely flat (within 0.03 kcal/mol) from 180° to 120°. Trihydrates optimal structures are 3+0(D_3) for lithium, sodium, and potassium, and 2+1(C_{2v}) for rubidium and cesium, in agreement with *ab initio* calculations [55, 64, 66]. For rubidium, 3+0(D_3) is the second most stable structure. For cesium, the second most stable structure has a broken symmetry 3+0(C_3), because the three water molecules can optimize their mutual interaction by having the ion out of the plane.

The lowest-energy lithium tetrahydrate is the 4+0(S_4) structure, for which the four water molecules are in the first solvation shell. It is the most stable structure for sodium as well. For both lithium and sodium, the 3+1(C_{2v}) structure (with a water molecule in the second shell) has a higher energy. The *ab initio* calculations of Glendening and Feller [55] give the same ordering for lithium, but the reverse for sodium—although not by a large energy difference. This reversal, along with the fact that the energy of the 3+0(D_3) sodium trihydrate structure is low compared to the *ab initio* estimates, could be an indication that the sodium–first-shell-water interaction is slightly too strong for the water–water interaction. For potassium, the 3+1(C_{2v}) structure is more stable than the 4+0(S_4) structure, in agreement with *ab initio* results [55, 64, 66]. For rubidium, the lowest-energy tetrahydrate is a 3+1(C_2) structure "broken" into a 3+1(C_1) structure. The water molecule not involved in any water–water hydrogen bond is bent away from the mirror plane (see Fig. 6.3). Energetically, this structure is comparable to the *ab initio* 3+1(C_2) structure of Glendening and Feller [55]. For cesium, the lowest-energy structure is 4+0(C_4), where the four water molecules form a cyclic tetramer on one side of

Table 6.III: Lowest-energy alkali hydrates for $n = 1$–6 water molecules. See Fig. 6.3 for an illustration of the structures. U is the binding energy for the AH/SWM4-DP model and $U_{\text{a.i.}}$ corresponds to *ab initio* binding energies from the literature. The nondegenerated ion–oxygen distances are listed under the columns d and $d_{\text{a.i.}}$. For comparable distances, only the average is given. The energies are in kcal/mol and the distances in Å.

	n	Structure	$-U$	d	$-U_{\text{a.i.}}$	$d_{\text{a.i.}}$	$-U_{\text{a.i.}}$	$d_{\text{a.i.}}$	$-U_{\text{a.i.}}$	$d_{\text{a.i.}}$
					Ref. [55][a]		Ref. [60][b]		Ref. [63][c]	
Li^+	1	1+0(C_{2v})	**32.538**	1.951	**34.5**	1.850	**35.2**	1.847	**33.8**	1.849
	2	2+0(D_{2d})	**61.436**	1.968	**64.4**	1.878	**64.0**	1.884	**63.5**	1.875
		1+1(C_s)	**47.699**	1.927, 4.170	**51.1**	1.810, 3.977	**52.4**	1.835		
	3	3+0(D_3)	**84.346**	1.997	**87.5**	1.915	**86.8**	1.927	**86.5**	1.915
	4	4+0(S_4)	**101.780**	2.036	**104.1**	1.969	**104.3**	1.971	**103.2**	1.968
		3+1(C_{2v})	**98.092**	1.996, 3.694	**103.1**	1.911, 3.689	**103.3**	1.916, 1.928		
	5	4+1(C_2)	**114.943**	2.035, 3.768	**118.5**	1.966, 3.816	**119.3**	1.959, 1.970	**117.1**	1.966
	6	4+2(D_{2d})	**127.377**	2.037, 3.792					**130.3**	1.965
					Ref. [55][a]		Ref. [61][d]			
Na^+	1	1+0(C_{2v})	**24.000**	2.255	**24.3**	2.230	**23.7**	2.228		
	2	2+0(D_{2d})	**45.675**	2.273	**45.5**	2.249	**44.9**	2.253		
		1+1(C_s)	**37.651**	2.223, 4.395	**38.9**	2.185, 4.233				
	3	3+0(D_3)	**64.138**	2.295	**63.5**	2.277	**62.9**	2.282		
	4	4+0(S_4)	**79.522**	2.320	**77.5**	2.306	**77.5**	2.317		
		3+1(C_{2v})	**77.782**	2.293, 3.997	**78.0**	2.268, 4.027				
	5	4+1(C_2)	**92.597**	2.318, 4.044	**91.0**	2.300, 4.118	**90.7**	2.312, 4.149		
	6	4+2(D_{2d})	**105.298**	2.319, 4.064			**103.7**	2.307, 4.180		
					Ref. [55][a]		Ref. [64][e]		Ref. [66][f]	
K^+	1	1+0(C_{2v})	**17.900**	2.620	**18.9**	2.669	**17.4**	2.639	**17.8**	
	2	2+0(C_2)	**33.988**	2.642						
		2+0(C_s)	**33.988**	2.643						
		2+0(D_{2d})	*33.964	2.644	**35.8**	2.707	**32.9**	2.673	**33.5**	
		1+1(C_s)	**30.276**	2.575, 4.667	**31.7**	2.603, 4.611	**30.1**	2.567, 4.424		
	3	3+0(D_3)	**48.440**	2.661	**50.9**	2.730	**46.9**	2.690	**47.1**	
		2+1(C_{2v})	**47.890**	2.624, 4.348	**49.8**	2.662, 4.433	**47.1**	2.625, 4.354	**48.8**	
	4	3+1(C_{2v})	**61.818**	2.658, 4.366	**64.1**	2.719, 4.474	**60.2**	2.681, 4.384	**62.2**	
		4+0(S_4)	**61.083**	2.681	**63.7**	2.759	**58.9**	2.713	**59.1**	
	5	4+1(C_2)	**73.976**	2.677, 4.398	**76.4**	2.747, 4.542	**71.6**	2.703, 4.438		
	6	4+2(D_{2d})	**86.616**	2.673, 4.411			**84.0**	2.693, 4.451	**85.6**	
					Ref. [55][a]					
Rb^+	1	1+0(C_{2v})	**15.900**	2.792	**16.1**	2.902				
	2	2+0(C_2)	**30.238**	2.814	**30.7**	2.933				
		2+0(C_s)	**30.237**	2.813, 2.815	**30.7**	2.933, 2.934				
		1+1(C_s)	**27.785**	2.741, 4.795	**28.2**	2.832, 4.823				
	3	2+1(C_{2v})	**44.073**	2.791, 4.525	**44.2**	2.889, 4.672				
		3+0(D_3)	**43.212**	2.835	**43.7**	2.960				
	4	3+1(C_1)	**56.437**	2.829, 4.540	**56.6**	2.948, 4.699[g]				
		2+1+1(C_s)	**55.110**	2.780, 4.447, 6.613						
		4+0(S_4)	**54.802**	2.853	**55.1**	2.987				
	5	3+2(C_s)	**67.948**	2.791, 2.820, 4.576						
		4+1(C_2)	**67.569**	2.834, 2.859, 4.565						
	6	4+2(C_2)	**80.434**	2.812, 2.971, 3.874						
		4+2(D_{2d})	**80.078**	2.841, 4.575						
					Ref. [55][a]					
Cs^+	1	1+0(C_{2v})	**13.700**	3.008	**14.1**	3.136				
	2	2+0(C_s)	**26.122**	3.027, 3.031	**26.8**	3.168, 3.171				
		2+0(C_2)	**26.121**	3.029	**26.8**	3.169				
		1+1(C_s)	**25.037**	2.948, 4.962	**25.5**	3.055, 5.033				
	3	2+1(C_{2v})	**39.771**	2.999, 4.745	**39.8**	3.117, 4.909				
		3+0(C_3)	**37.398**	3.050	**38.2**	3.201[g]				
	4	4+0(C_4)	**50.817**	3.163	**53.0**	3.323				
		2+1+1(C_s)	**50.607**	2.986, 4.663, 6.832						
		3+1(C_1)	**50.452**	3.040, 4.765	**50.6**	3.152, 3.223, 4.934[h]				
	5	5+0(C_5)	**63.174**	3.198						
	6	5+1(C_1)	**74.561**	3.140, 3.167, 3.167, 3.190, 3.210, 3.981						
		4+2(C_2)	**74.407**							

* Unstable conformation.

[a] Ref. [55]. MP2/6-31+G*//RHF/6-31+G* level.

[b] Ref. [60]. MP2/CBS level for $n = 1$, MP2/aug-cc-pVQZ//MP2/aug-cc-pVTZ level from $n = 1$ to 2 (but MP2/aug-cc-pVDZ level for the 1+1(C_s) structure), and MP2/aug-cc-pVDZ for $n = 2$ to 3, $n = 3$ to 4, and $n = 4$ to 5.

[c] Ref. [63]. MP2/6-31++G(d,p)//HF/6-31++G(d,p) level with counterpoise correction.

[d] Ref. [61]. MP2/TZ2P level with BSSE correction.

[e] Ref. [64]. MP2/TZ2P level with BSSE correction.

[g] 3+0(D_3) structure.

[h] 3+1(C_2) structure.

the ion. 4+0(C_4) is the lowest-energy *ab initio* structure predicted by Glendening and Feller [55]. Somewhat surprizingly, the second most stable structure for both rubidium and cesium is the 2+1+1(C_s) structure, with a molecule in the third shell (see Fig. 6.3). This preference could be explained by the binding cooperativity in the SWM4-DP water model, enhanced by a significant polarization of the water molecules in the first and second shells. In the 2+1+1(C_s) hydrates, the water molecule involved in three H-bonds is actually more polarized than the two in proximity of the ion. To our knowledge, this 2+1+1 structure has never been reported in *ab initio* studies.

The most stable pentahydrate structures are 4+1(C_2) for lithium, sodium, and potassium, 3+2(C_s) for rubidium, and 5+0(C_5) for cesium (see Fig. 6.3). For rubidium, the 4+1(C_2) structure is only 0.379 kcal/mol above the absolute energy minimum. For cesium, it has an energy 3.015 kcal/mol above that of 5+0(C_5).

The symmetry of the lowest-energy hexahydrate structure is very sensitive to the size of the ion. The approximately tetrahedral coordination of the 4+2(D_{2d}) structure accomodates better a small ion, while the looser, asymmetric coordination of the 4+2(C_2) structure is compensated by six water–water hydrogen bonds (instead of four for the D_{2d} structure). For lithium, the AH/SWM4-DP model predicts that the D_{2d} structure has the lowest energy. The C_2 conformer is not even stable. For sodium, the energy of the 4+2(C_2) structure is 3.653 kcal/mol above the global minimum, in accordance with the *ab initio* calculations of Feller [73]. Within this energy interval, the AH/SWM4-DP model predicts many other conformers, including one with a water molecule in the third solvation shell. For potassium, the AH/SWM4-DP model predicts that D_{2d} has the lowest energy, but the C_2 structure is second-lowest, 1.000 kcal/mol above. This is contrasting with the calculations of Lee *et al.* [64], that predicted C_2 to be more stable than D_{2d} by a few kcal/mol. For rubidium, the two structures have comparable energies, C_2 being more stable by 0.355 kcal/mol. Cesium, which is big enough to allow water–water hydrogen bonding within its first solvation shell, is an exception: its energy minimum is a 5+1(C_1, pseudo-C_s) structure formed by a "book" water hex-

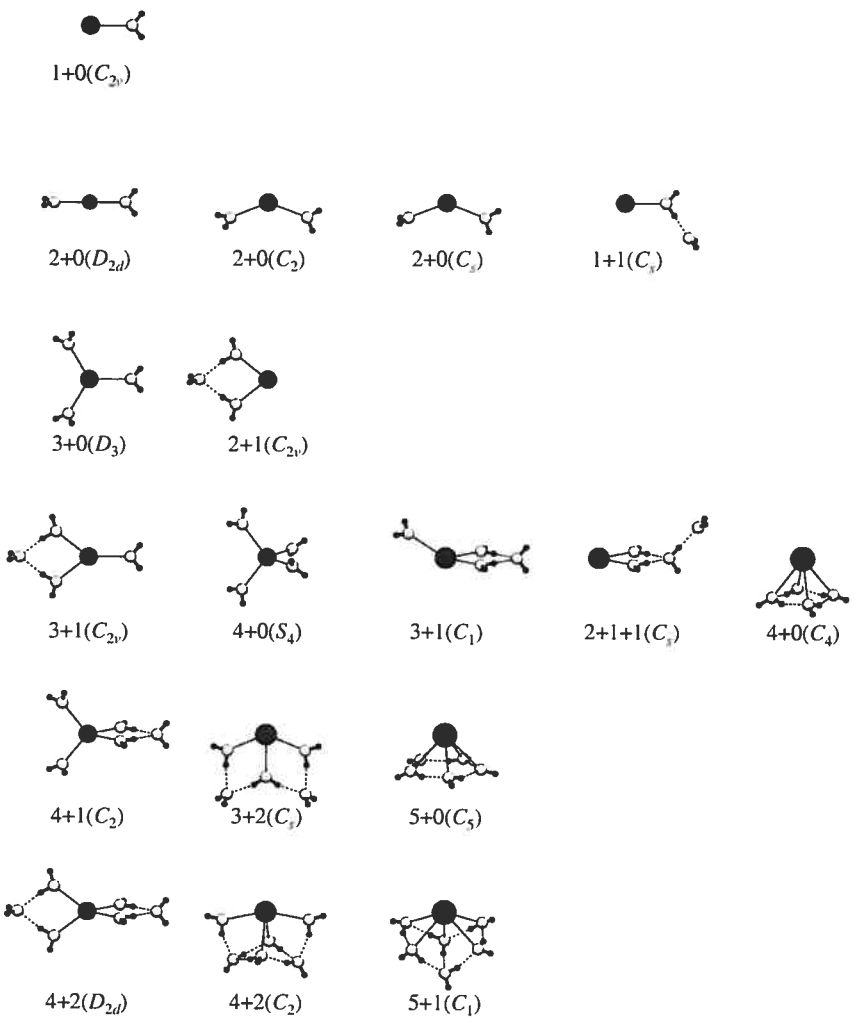

Figure 6.3: Illustration of the minimum-energy alkali-aqua clusters, with their structure names. All structures are shown for potassium, except the 2+0(D_{2d}) structure (for sodium), the 3+1(C_1), 2+1+1(C_1), 3+2(C_s), and 4+2(C_2) structures (for rubidium), and the 4+0(C_4), 5+0(C_5), and 5+1(C_1) structures (for cesium).

amer distorted to accomodate its contact with the ion. The second-lowest-energy structure is 4+2(C_2), 0.154 kcal/mol above. The 4+2(D_{2d}) structure of cesium is 2.139 kcal/mol above the global minimum, higher than many other conformers with water–water hydrogen bonds within the first solvation shell.

6.3.2 Small halide hydrates: $X^-\cdot(H_2O)_{n=1-4}$

Because the halide ions are known to severely disrupt the tetrahedral arrangement of water, the minimum-energy halide hydrates represent a stringent test of the AH/SWM4-DP model. The monohydrates have a C_s symmetry, with a hydrogen atom pointing toward the ion. For the dihydrates, it is known that the two water molecules form a distorted dimer on the same side of the ion—except for fluoride, that is strong enough to break the water dimer [74]. It turns out that the SWM4-DP water model, whose hydrogen sites were parametrized for normal hydrogen-bonding, is not very realistic for close hydrogen–hydrogen contacts forced by a strong anion.

Table 6.IV presents the lowest-energy halide hydrates structures. The monohydrates are as in Table 6.II. Except for the monohydrate energies (that were adjusted for) the binding energies are significantly overestimated. The structures are illustrated in Fig. 6.4. For the fluoride dihydrate, the AH/SWM4-DP model produces a O–F–O angle of 73.2°, much narrower than the *ab initio* predictions of about 100° [20, 74]. The *ab initio* calculations predict an open structure where the two water molecules are unbound, but the AH/SWM4-DP model produces a structure similar to those of chloride, bromide, and iodide dihydrates. The AH/SWM4-DP model is unable to stabilize a dihydrate structure with no water–water hydrogen bond. Figure 6.5a shows the energy profile of the AH/SWM4-DP model as the O–F–O angle is opened, compared to the same profile for an *ab initio* calculation at the MP2/6-311++G** level of theory (the same level used by Kim *et al.* [56] in their study of small halide hydrates), performed using Gaussian 98 [75]. The *ab initio* scan shows that the minimum energy angle is about 130° at the MP2/6-311++G** level (with rigid water molecules kept in the SWM4-DP geom-

Table 6.IV: Lowest-energy halide hydrates for $n = 1$–4 water molecules. See Fig. 6.4 for an illustration of the structures, and the caption of Table 6.III for the meaning of each column.

	n	Structure	$-U$	d	$-U_{\text{a.i.}}$	$d_{\text{a.i.}}$	$-U_{\text{a.i.}}$	$d_{\text{a.i.}}$
					Ref. [56][a]			
F^-	1	1+0(C_s)	**23.317**	2.567	**25.9 ± 1.5**	2.436		
	2	2+0(C_1)	**45.596**	2.587, 2.579				
	3	3+0(C_3)	**67.264**	2.600	**63.7 ± 3.7**	2.582		
	4	4+0(C_4)	**83.907**	2.647	**78.3 ± 4.9**	2.670		
		3+1(C_s)	**83.688**	2.588, 2.619, 3.669				
					Ref. [56][a]		Ref. [67][b]	
Cl^-	1	1+0(C_s)	**14.400**	3.080	**14.4 ± 1.2**	3.108	**13.8**	3.136
	2	2+0(C_1)	**29.965**	3.111, 3.082	**28.6 ± 2.7**	3.177	**27.5**	3.105, 3.276
	3	3+0(C_3)	**46.950**	3.119	**42.9 ± 4.2**	3.211	**41.9**	3.223
	4	3+1(C_s)	**61.714**	3.056, 3.126, 4.096				
		4+0(C_4)	**61.457**	3.149	**56.2 ± 5.7**	3.272	**54.1**	3.270
					Ref. [56][a]			
Br^-	1	1+0(C_s)	**12.700**	3.244	**12.7 ± 0.9**	3.316		
	2	2+0(C_1)	**27.076**	3.240, 3.286	**25.8 ± 2.1**	3.389		
	3	3+0(C_3)	**43.353**	3.288	**39.7 ± 3.6**	3.403		
	4	4+0(C_4)	**57.421**	3.316	**52.6 ± 5.0**	3.458		
		3+1(C_s)	**57.327**	3.203, 3.287, 4.248				
					Ref. [56][a]		Ref. [77][c]	
I^-	1	1+0(C_s)	**10.600**	3.501	**10.6 ± 0.6**	3.632	**10.43 ± 0.63**	3.59
	2	2+0(C_1)	**23.490**	3.483, 3.570	**22.3 ± 1.5**	3.718	**22.38 ± 1.49**	3.64
	3	3+0(C_3)	**38.863**	3.552	**35.4 ± 2.9**	3.700	**36.48 ± 2.37**	3.69
	4	4+0(C_4)	**52.309**	3.580	**48.0 ± 4.3**	3.743	**49.26 ± 3.66**	3.69
		3+1(C_s)	**51.466**	3.437, 3.548, 4.490			**47.09 ± 3.45**	3.62, 4.49

[a] Ref. [56]. MP2/6-311++G** level.

[b] Ref. [67]. MP2/aug-cc-pVDZ level with BSSE correction.

[c] Ref. [77]. MP2/aug-cc-pVDZ+diff level. For the 2+0(C_1) and 3+1(C_1) structures, the ion–oxygen distances are averages.

etry). But the MP2/6-31G* level cannot reproduce the breaking of the intra-shell hydrogen bond, and it has a minimum at about the same angle as the AH/SWM4-DP model. This unbalance is noticeable for chloride as well (see Fig. 6.5b), and although the model picks the correct minimum-energy conformation, its stability is overestimated compared to the MP2/6-311++G** level calculation. In contrast with *ab initio* predictions [76], the "bound" state is probably too stable for the "open" state to become thermodynamicall stable at room temperature.

The most stable halide trihydrates all have the 3+0(C_3) (see Fig. 6.4). The second lowest energy structure (not shown in Fig. 6.4) has one of the three water–

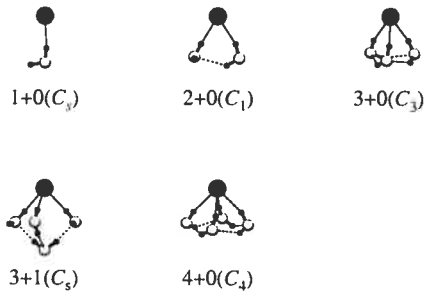

Figure 6.4: Illustration of the minimum-energy halide-aqua clusters, with their structure names. All structures are shown for chloride.

water bonds broken, but is 2.1 kcal/mol above the 3+0(C_3) minimum energy for fluoride and even less stable for the larger halide ions.

With the exception of chloride, the optimal halide tetrahydrates are 4+0(C_4) structures, as predicted by *ab initio* calculations [56]. Along with this well-documented minimum-energy structure is a pyramidal 3+1(C_s) structure (see Fig. 6.4). A pyramidal isomer was predicted for fluoride by Bryce *et al.* [78] (who used both *ab initio* calculations and a polarizable model) and shown consistent with vibrational spectroscopy measurements by Robertson *et al.* [79]. Using an additive molecular mechanical model, Duan and Zhang [80] found the lowest-energy chloride tetrahydrate to be a pyramidal structure as well. What is unusual about the structure predicted by the AH/SWM4-DP model is that the dangling hydrogen atoms of the two first-shell water molecules are pointing *toward* the oxygen of the third first-shell molecule. This, again, points to possible limitations in the hydrogen repulsive interactions in the SWM4-DP model.

The pentahydrates and hexahydrates have the interesting property of allowing thermodynamically stable *internal* hydration states [81, 82]. Because of the underestimation of the hydrogen–hydrogen short-range repulsion, the AH/SWM4-DP model does not indicate this transition to clusters with internal ions. Instead, the few most stable structures are "4+1" and "4+2" surface clusters with water

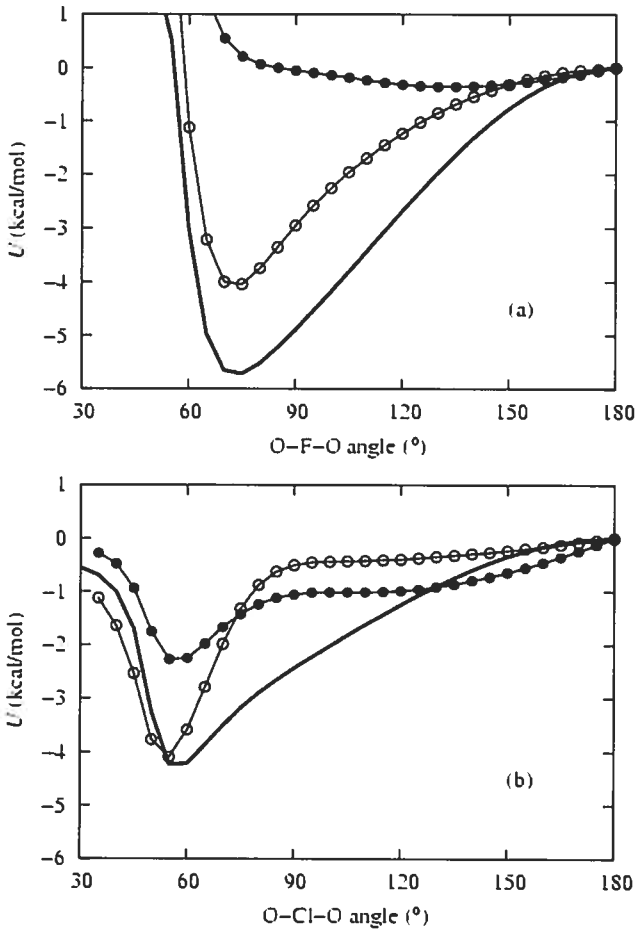

Figure 6.5: Binding energies for (a) fluoride dihydrate and (b) chloride dihydrate as a function of the O–Ion–O angle. The filled symbols are the MP2/6-311++G** calculations and the lines are the AH/SWM4-DP model. The open symbols are MP2/6-31G* calculations. The *ab initio* calculations are done with the water molecules maintained in the SWM4-DP geometry ($\ell_{OH} = 0.9572$ Å and $\theta_{HOH} = 104.52°$). The conformation with an O–Ion–O angle of 180° is the energy reference point.

molecules very densely packed on one side of the ion, all along the series from fluoride to iodide.

For bromide, the proposed AH/SWM4-DP model yields hydrates energies comparable to those of the POT_1 polarizable model of Ayala *et al.* [24], which was deemed inferior to their POT_2 and POT_3 models against the reproduction of most hydration properties. The POT_1 model overestimates the magnitude of the hydration enthalpy, probably for the same reason the present bromide model is overshooting the hydration free energy by about 5 kcal/mol. As will be seen in the following section, a fundamental difference is that the AH/SWM4-DP bromide model predicts an oxygen coordination number of 6.6 in aqeous phase, whereas the POT_1 model predicts a 10-coordinated bromide ion [24].

6.4 Solvent structure

Ion–oxygen and ion–hydrogen radial distribution functions (RDFs) $g_{\mathrm{IO}}(r)$ and $g_{\mathrm{IH}}(r)$, as well as the running coordination number $N(r) = \int dr' \, 4\pi r'^2 g_{\mathrm{IO}}(r')$, are extracted from the solvent box simulations (see Figs. 6.6 and 6.7). The positions and amplitudes of the extrema are presented in Table 6.V, as well as the oxygen coordination numbers for the first shell, $N_{\mathrm{c}} = N(r_{\mathrm{min1}})$, and second shell, $N_{\mathrm{c2}} = N(r_{\mathrm{min2}}) - N(r_{\mathrm{min1}})$.

6.4.1 Alkali ions

In accordance with the most recent computational studies [83, 84, 85, 86, 87], the AH/SWM4-DP model predicts that the lithium ion is essentially 4-coordinated (see Fig. 6.6 and Table 6.V). (Table I of Ref. [86] provides a survey of the literature.) The first peak of $g_{\mathrm{LiO}}(r)$ is 2.03 Å away from the ion, a distance comparable to recent results from quantum molecular dynamics (1.96 Å [84] and 1.93 Å [85]) and within the narrow experimental range obtained from neutron and X-ray diffraction (1.94–2.07, see Table I of Ref. [86] and references therein). The $g_{\mathrm{LiO}}(r)$ function has a deep first minimum (see Fig. 6.6 and Table 6.V), in agreement with the

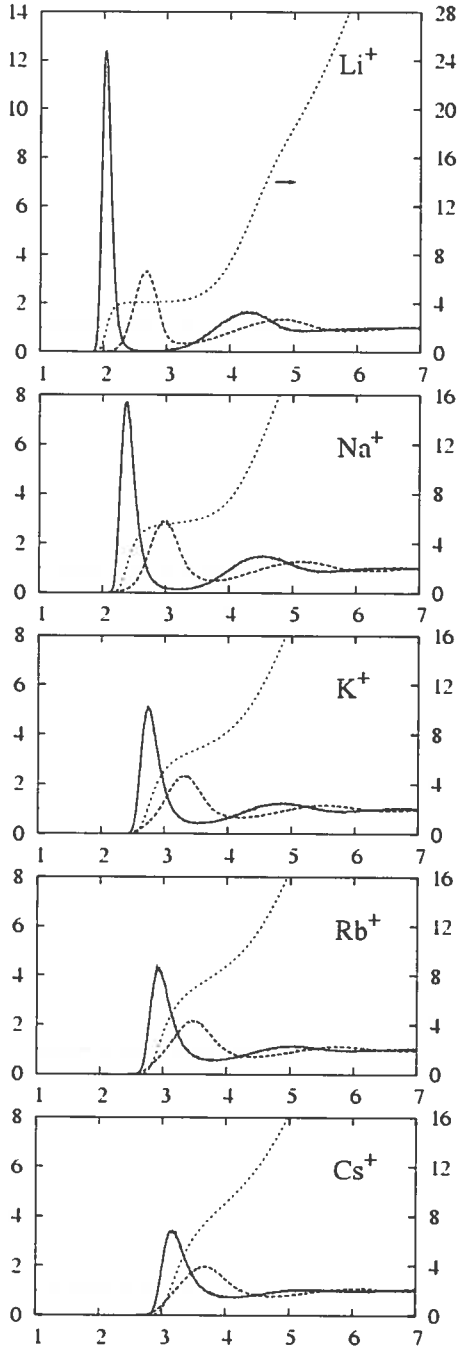

Figure 6.6: Radial hydration structure of the alkali ions for the AH/SWM4-DP model. The $g_{\mathrm{IO}}(r)$ functions are in solid lines, the $g_{\mathrm{IH}}(r)$ functions in dashed lines, and the $N(r)$ functions in dotted lines.

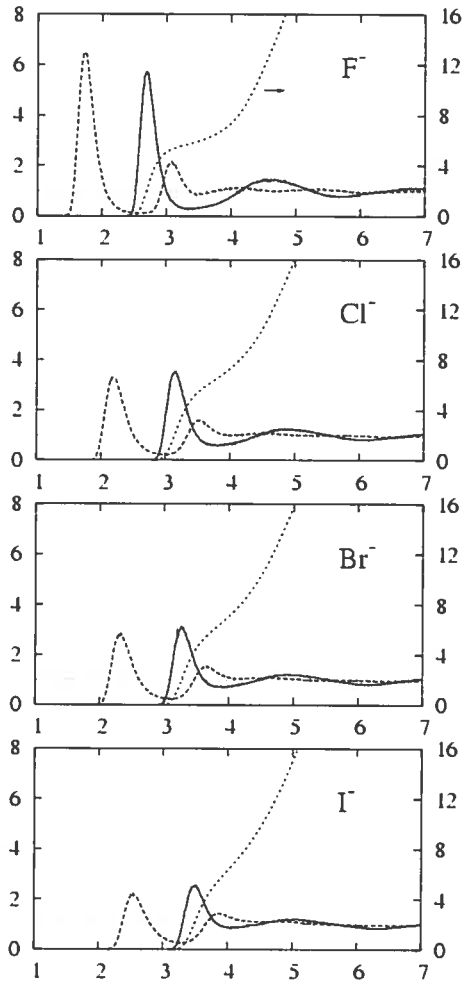

Figure 6.7: Radial hydration structure of the halide ions for the AH/SWM4-DP model. The $g_{\mathrm{IO}}(r)$ functions are in solid lines, the $g_{\mathrm{IH}}(r)$ functions in dashed lines, and the $N(r)$ functions in dotted lines.

Table 6.V: Summary of the radial hydration structures for the AH/SWM4-DP model.

	r_{max1} (Å)	g_{max1}	r_{min1} (Å)	g_{min1}	N_c	r_{max2} (Å)	g_{max2}	r_{min2} (Å)	g_{min2}	N_{c2}
Li^+	2.03	12.4	2.74	0.039	4.08	4.28	1.62	5.17	0.865	16.0
Na^+	2.37	7.65	3.21	0.141	5.67	4.53	1.47	5.50	0.868	18.3
K^+	2.74	5.05	3.55	0.431	6.90	4.84	1.23	5.80	0.915	20.2
Rb^+	2.92	4.23	3.80	0.576	8.00	5.04	1.13	5.94	0.955	20.7
Cs^+	3.15	3.40	4.10	0.728	9.55	5.25	1.03	6.05	0.982	20.2
F^-	2.69	5.69	3.36	0.287	5.74	4.61	1.43	5.69	0.784	20.0
Cl^-	3.14	3.51	3.79	0.592	6.50	4.88	1.25	6.07	0.831	23.3
Br^-	3.27	3.25	3.90	0.720	6.59	4.91	1.22	6.16	0.832	24.1
I^-	3.49	2.53	4.07	0.855	6.70	5.00	1.20	6.35	0.840	26.1

Car-Parrinello (CP) simulations of Lyubartsev *et al.* (who get $g_{min1} = 0.015$) [84], but somewhat in disagreement with the QM/MM simulations of Loeffler and Rode [85], who obtain a more populated inter-shell region. A strongly depopulated inter-shell region makes the exchange of water molecules between the first and second hydration shells less frequent. The AH/SWM4-DP model for lithium has 16.0 molecules in the second hydration shell, in accord with recent classical MD simulations of a nonadditive force field (16 and 18 molecules, depending on the model) [87], but in contrast with QM/MM simulations (13 molecules) [85]—which may suffer from a finite-size effect.

For sodium, the AH/SWM4-DP model predicts an average of 5.67 water molecules in the first solvation shell, in good agreement with the QM/MM simulations of Tongraar *et al.* [88] (5.6 ± 0.3), the *ab initio* MD simulations of White *et al.* [89] (5.2), and other classical polarizable MD simulations (5.7 [2] and 5.56 [90]), but in disagreement with the QM/MM simulations of Rempe and Pratt [91] (4.6). The AMOEBA polarizable force field of Grossfield *et al.* [36] gives $N_c = 6.0$. The position of the maximum is 2.37 Å, within the span of *ab initio* and classical nonadditive MD simulations (2.33 [88], 2.49 [89], and 2.37 Å [90]). The AH/SWM4-DP model has 18.3 molecules in the second solvation shell of sodium, compared to 16 for the *ab initio* MD of White *et al.* [89].

For both lithium and sodium, the first hydration shell for the AH/SWM4-DP model appears too rigidly structured. For lithium, the first peak of the $g_{\mathrm{LiO}}(r)$ distribution is high (12.4) compared to the *ab initio* simulations of Lyubartsev *et al.* (~ 9.5) [84] and of Loeffler and Rode (~ 8.5) [85]. MD simulations of nonadditive force-fields give comparable low maxima (~ 9.8 [85] and ~ 8.2 [86]). For sodium, the first peak of the $g_{\mathrm{NaO}}(r)$ distribution is too high and the first minimum is too low compared to *ab initio* simulations [88, 89] and X-ray results [92]. The model for sodium gives $g_{\mathrm{max1}} = 7.65$ and $g_{\mathrm{min1}} = 0.141$, while the experimental values are around 3.5 for g_{max1} and 0.5 for g_{min1} [92, 89]. Part of this enhanced structure can explained by the "hardness" of the of the Lennard-Jones potential, which does not allow large radial fluctuations for a molecule in the first shell of the ion. The rigidity of the molecular geometry of the SWM4-DP model could play a role in the finer coordination structure, but it is unclear weither allowing flexible bonds for water would enhance or weaken the structure [93, 94]. It should be stressed that no direct adjustment of the shape of the RDFs was made—no direct adjustment of any structural data, for that matter. Consequently, any defect in the shape of the repulsive interaction is repercuted on the structural properties of the model.

For potassium, the model predicts 6.90 water molecules in the first shell, in agreement with the *ab initio* Car-Parrinello results of Ramaniah *et al.* (6.75) [95]. The AMOEBA model [36] predicts a comparable 7.0. In contrast, the recent polarizable model of Carrillo-Tripp *et al.* [90] produces a significantly higher coordination (7.85 ± 0.48). A similar coordination is obtained in a QM/MM simulation by Tongraar *et al.* [88] (8.3 ± 0.3). In the same work, Tongraar *et al.* obtain a coordination of 7.8 ± 0.2 from a pair-potential model parametrized against an *ab initio* energy surface [88]. X-ray [96] and neutron [97] scattering experiments suggest that the potassium ion is hexa-coordinated, but a more sophisticated analysis [98] of the neutron data of Neilson and Skipper [97] proposes 7.5. The position of the first peak of the $g_{\mathrm{KO}}(r)$ function is 2.74 Å, closer than the consensus experimental value 2.80 Å proposed by Marcus [99], the CP simulations result 2.81 Å of Ramaniah *et al.* [95], and the MD results 2.79 Å of Carrillo-Tripp *et al.* [90]. The

first peak is higher than the CP result of Ramaniah *et al.*: 5.05 compared to 3.44 [95].

For rubidium and cesium, the model yields $N_c = 8.00$ and 9.55, respectively. X-ray absorbtion [100, 101, 102] and neutron scattering [103] studies on rubidium suggest a lower coordination of 6 at distances of about 2.9 Å.

The AH/SWM4-DP ions of the alkali series have hydration structures and hydration free energies overall similar to the standard nonpolarizable model of Åqvist [104], but have arguably better gas-phase properties for the small alkali ions (at least for the monohydrates reported by Åqvist). Compared to the nonpolarizable model for alkali ions of Lee and Rasaiah [105], the AH/SWM4-DP model yields comparable but slightly smaller coordination numbers.

6.4.2 Halide ions

Figure 6.7 presents the RDFs of oxygen and hydrogen with respect to the halide ions. It shows clearly that the anions are coordinated by hydrogen atoms. From the relative positions of the first peak of the g_{IO} distribution and first and second peaks of the g_{IH} distribution, it can be seen that each water molecule provides a single coordinating hydrogen: the second hydrogen populates a second maximum in g_{IH} located further apart.

The coordination numbers of the halide ions are remarkably similar across the series (see Table 6.V). Although the positions of the first oxygen-coordination peak vary from 2.69 Å for fluoride to 3.49 Å for iodide (corresponding to an increase of 0.80 Å in the hydration radius), the increase in the surface of the coordination shell allows only one additional water molecule to insert ($N_c = 5.74$ for fluoride and 6.70 for iodide). In comparison, sodium has a coordination number of 5.67 for $r_{min1} = 2.37$ Å while cesium has $N_c = 9.55$ for r_{min1} only 0.78 Å longer. This reflects the difficulty of creating a compact hydration shell around anions.

Fluoride has $N_c = 5.74$ and chloride has $N_c = 6.50$, in agreement with most of the theoretical estimates and within the broad experimental range (see Table 2 of Ref. [106] and Table 8 of Ref. [107], and references cited therein). The QM/MM

simulations of Tongraar and Rode [106] give 4.6 for fluoride and 5.6 for chloride, but their pure MM simulations (using a pairwise additive force field) give both ions comparable coordination numbers (5.8 and 5.9, respectively). The position of the first oxygen-coordination peak of fluoride (2.69 Å) is comparable to most results from the literature (see Ref. [106] and Ref. [107]) but that of chloride (3.14 Å) is relatively short: distances between 3.2 and 3.3 Å are more commonly obtained [106, 107].

For bromide, the first oxygen-coordination peak is at 3.27 Å and produces a coordination number $N_c = 6.59$. It is shorter than the data from Table 8 of Ref. [107] and than the results of Merkling *et al.* [108] for the POT_2 model [24] (3.44 ± 0.07 Å and $N_c = 6 \pm 0.5$), but comparable to the CP results of Raugei and Klein [109] (3.26 Å and $N_c = 6.3$).

The AH/SWM4-DP model predicts an oxygen coordination of 6.70 for iodide. Photoelectron spectroscopy suggests an optimal coordination of 6 [110], but this result could be tainted by the fact that iodide strongly prefers the surface of small water clusters [21]. In comparison, Ignaczak *et al.* [107] obtain $N_c = 9.1$.

6.5 Water exchange reactions

6.5.1 Residence times

It is clear from the RDF of Fig. 6.6 that Li^+ creates a strong barrier for a water molecule to exchange from the bulk liquid to the first solvation shell. The region between the first and second solvation shells is heavily depopulated, and water exchanges are rather infrequent. The kinetics of the exchange processes can be defined in terms of the residence correlation function [111]

$$N_r(t) = \left\langle \sum_{i=1}^{N} P_i(t', t; r_{min1}, t_*) \right\rangle, \tag{6.7}$$

where P_i equals 1 if the oxygen atom of water molecule i is in the first shell at time t' and at time $t' + t$, without leaving the shell for a period longer than t_*, and equals 0 otherwise. The first shell is defined as the sphere of radius r_{min1} (2.74 Å for lithium, see Table 6.V) centered around the ion. Following Impey *et al.* [111], t_* is set to 2 ps, approximately the time it takes in the bulk region for two water molecules initially in close contact to diffuse away. Apart from a fast sub-picosecond component, this correlation function decreases as an almost perfect exponential (see Fig. 6.8). A least-square fit of the expression $N_{\mathrm{r}}e^{-t/\tau_{\mathrm{r}}}$ onto the $N_{\mathrm{r}}(t)$ curve for t between 1 and 20 ps gives a residence time $\tau_{\mathrm{r}} = 78.6$ ps for lithium, in relatively good accord with the CP simulations of Lyubartsev *et al.* [84] (who estimated the residence time at 20–50 ps, and who found $N_{\mathrm{c}} = 4$), with the nonadditive MD simulations of Spångberg *et al.* [86] (112 ± 16 ps, $N_{\mathrm{c}} = 4.1$), and with the additive MD simulations of Duan and Zhang [80] (15–57 ps, with a coordination number of 4.6).

For sodium, the residence time τ_{r} is 27.3 ps, similar to the molecular dynamics results of Guàrdia and Padró [93] (23.8 ps for their rigid water model, with $N_{\mathrm{c}} = 5.8$), but longer than the original results of Impey *et al.* (9.9 ps, with $N_{\mathrm{c}} = 7.5$) [111].

The $g_{\mathrm{OO}}(r)$ distribution of a pure SWM4-DP liquid has a first minimum at 3.35 Å and has a coordination number of 4.63 [5]. Performing the same analysis on trajectories of a liquid of 250 SWM4-DP molecules (the same trajectories that were used to compute the self-diffusion constant in Ref. [5]), one gets a residence time τ_{r} of 6.19 ps for a water molecule in the first shell of another water molecule.

Because the least-squares fit ignores the molecules occupying the first shell for less than 1 ps, the values of N_{r} are smaller than the coordination numbers of Table 6.V: 4.0 for lithium (2% smaller than N_{c}), 5.5 for sodium (3% smaller), 6.3 for potassium (9% smaller), 7.1 for rubidium (11% smaller), and 7.9 for cesium (17% smaller). [In comparison, the residence number for pure SWM4-DP water is 3.6 (22% smaller than the coordination number).] This reflects the fact that, for the larger ions, a significant contribution to the coordination number N_{c} comes from

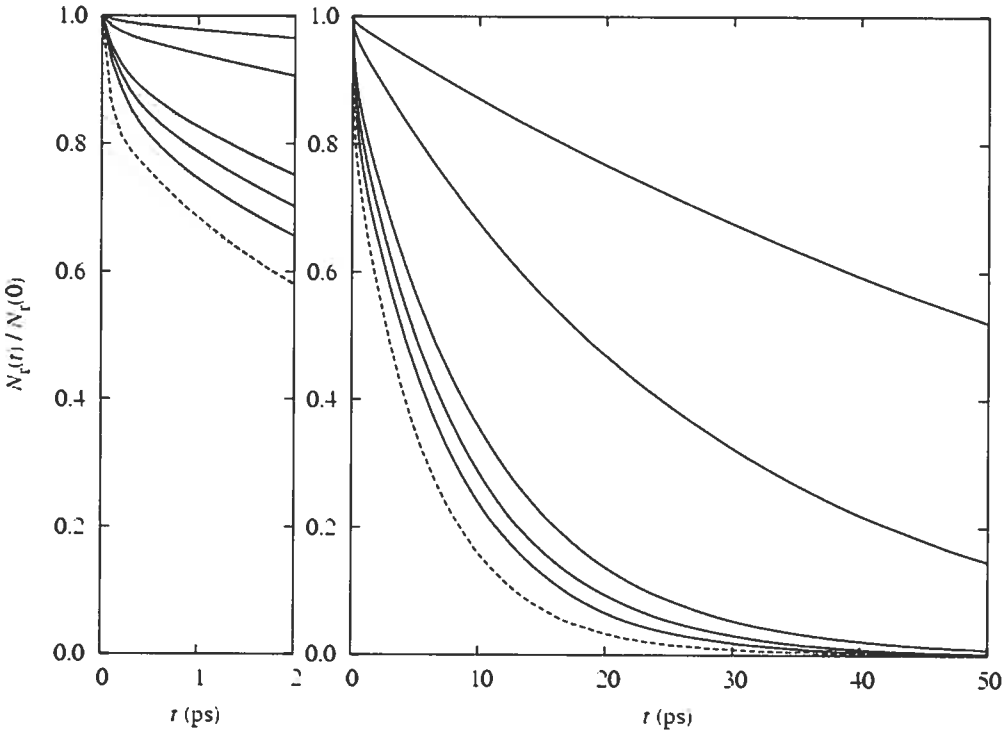

Figure 6.8: Normalized residence correlation functions $N_r(t)/N_r(0)$ for the alkali metal ions (solid lines) and for SWM4-DP water (dashed line). The first graph shows the first 2 ps of the functions, and how each exponential curve $N_r e^{-t/\tau}$ adjusted for $t = 1$ to 20 ps produces a *residence number* N_r smaller than the coordination number $N_c = N_r(0)$.

Table 6.VI: Dynamical properties of the alkali ions from the AH/SWM4-DP model: residence number N_r (to be compared to the coordination numbers reported from Table 6.V), residence time τ_r, and exchange time $\tau_x \equiv \tau_r/N_r$. Results for the SWM4-DP water model are shown for comparison.

	N_c	N_r	τ_r (ps)	τ_x (ps)
Li^+	4.1	4.0	78.6	19.5
Na^+	5.7	5.5	27.3	4.93
K^+	6.9	6.3	10.7	1.69
Rb^+	8.0	7.1	8.91	1.25
Cs^+	9.6	7.9	7.91	1.00
F^-	5.7	5.5	23.2	4.22
Cl^-	6.5	5.8	11.6	2.00
Br^-	6.6	5.7	10.9	1.92
I^-	6.7	5.6	9.28	1.66
H_2O	4.6	3.6	6.19	1.73

molecules staying less than 1 ps in the first shell, and obeying a different kinetics. The *residence number* N_r describes the number of water molecules forming a stable coordination around the ion—by opposition to an *instantaneous* coordination. Impey *et al.* [111] used the expression "persisting coordination", but did not define it exactly in these terms. In contrast to the residence time, the residence number is insensitive to the exact value of the tolerance time t_*. It effectively ignores the molecules that are not forming a stable hydration shell, decaying according to the kinetic process associated to τ_r.

The residence time divided by the residence number gives the *exchange time* τ_x, which is the characteristic time between two exchange reactions (see Table 6.VI). Interestingly, the exchange time of potassium (1.69 ps) is comparable to that of water (1.73 ps).

Even though the halide ions have almost identical residence numbers, they have well-separated residence times: 23.2 ps for fluoride, 11.6 ps for chloride, 10.9 ps for bromide, and 9.28 ps for iodide (see Table 6.VI). In comparison, the original MD simulations of Impey *et al.* [111], using the nonpolarizable MCY water model, give

a τ_r of 20.3 ps for fluoride and of 4.5 for chloride. From CP molecular dynamics of HBr in 31 water molecules, Raugei and Klein [109] found the residence time of bromide to be about 19 ps.

6.5.2 Detailed solvation structure

For the alkali ions, we compute the "n-th closest atom" radial distributions, $g_{IO}^{(n)}(r)$, corresponding to the contribution of the n-th closest oxygen atom to the $g_{IO}(r)$ distribution. At every instant of the MD simulation, the oxygen atoms are ranked according to their distance to the ion, and the position of n-th closest atom is accounted for in the $g_{IO}^{(n)}(r)$ distribution. The n-th closest water molecule (according to the position of its oxygen atom) will be denoted W_n and the oxygen-ion distance will be denoted r_n. The "n-th closest atom" radial distributions have the following obvious property:

$$\sum_n g_{IO}^{(n)}(r) = g_{IO}(r). \tag{6.8}$$

These distributions are presented in Fig. 6.9, for $n = 1$ to 10. For lithium, they show that the first solvation shell is almost totally populated with the four closest water molecules (W_1 to W_4). However, the $g_{LiO}^{(5)}(r)$ function has a tail into the first solvation shell, indicating that a fifth molecule, W_5, exceptionally enters the first solvation shell. For sodium, W_5 is most likely within the first solvation shell, and W_7 is most likely out, but W_6 populates both shells and can be as far as 4 Å from the ion. This explains the coordination number of sodium, between 5 and 6 (see Table 6.V). For potassium, W_6 is contained within the first shell, and W_9 is in the bulk liquid, but W_7 and W_8 exchange easily between the first and second shells. For rubidium and cesium, the n-th closest atom distributions are so much overlapping in the region between the first and second shells that the very definition of an exchange pathway is uncertain.

To show more clearly the nature of the exchange reactions, Fig. 6.10 presents the joint distributions of the distances of the n-th and n+1-th closest molecules (n chosen so that the n-th closest molecule is mostly in the first shell and the n+1-th

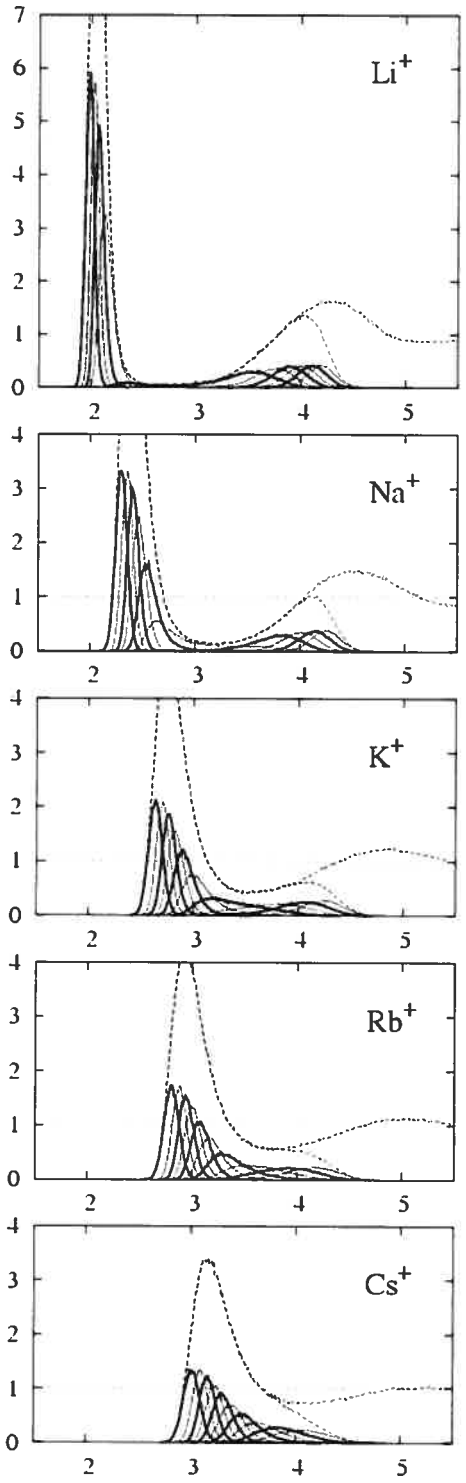

Figure 6.9: "n-th closest oxygen" radial distribution functions (for $n = 1$ to 10) for the alkali series. The odd-n distributions are drawn with thicker lines. Both the full RDF and the sum of the ten presented partial distributions are shown. For lithium, the four closest water molecules are always in the first shell. The fifth closest molecule is out in the second shell most of the time, but can be in the first shell occasionally. For sodium, the sixth closest molecule is mostly populating the first shell, but has a significant second shell "tail".

closest molecule is mostly out). The distributions are shown as contour plots with the levels separated by a factor e, that is, each level represents an energy k_BT.

For lithium, the free energy minimum is clearly when r_4 is about 2.1 Å and r_5 is about 3.6 Å, that is, when the ion is coordinated by four water molecules. At the transition state of the exchange reaction, $r_5 = r_4 \simeq 2.2$ Å, and a barrier of 3–4 k_BT has been crossed. The nature of the exchange reaction can be clearly read from the distribution: W_5 is coming into the first shell, while W_4 is hardly getting out (going from 2.1 to 2.2 Å). It is an *associative* exchange. Using a nonadditive force field specialized for lithium hydration [112], Spångberg *et al.* found [86] a similar associative mechanism, but with a significant interchange pathway as well. From their relatively short CP simulations, Lyubartsev *et al.* [84] observed two water exchanges (from which the 20–50 ps water residence time was estimated) and both were associative. Because the 5-coordinated Li^+ state is a local free energy minimum, the excedent molecule is not immediately dissociating.

The joint distribution for the sodium ion shows that the majority of the exchange reactions are dissociative/associative: W_6 exchanges with W_7 during the time it is out of the first shell, and a new W_6 reintegrates the shell. The most probable exchange happens when $r_6 = r_7 \simeq 3.8$ Å. Along this path, the free energy barrier is approximately 1 k_BT. The alternate associative mechanism for which W_7 comes in is very unlikely, but an *interchange* reaction for which W_6 and W_7 would exchange at mid-distance is possible.

For potassium (and for the larger alkali ions; data not shown), the joint distribution is almost featureless: very little free energy is required for the exchange reaction, and the reaction is of no preferred nature. These results are similar to the classical MD simulations of Cascella *et al.* [113] (using the Åqvist parameters [104] for K^+ and the TIP3P water model [40]). These authors find an identical coordination number (6.9, compared to 6.90 for this work) and a very flexible hydration structure able to exchange water molecules with the bulk using a variety of mechanisms.

For the lithium ion, it is of interest to look at how W_5 is affecting the struc-

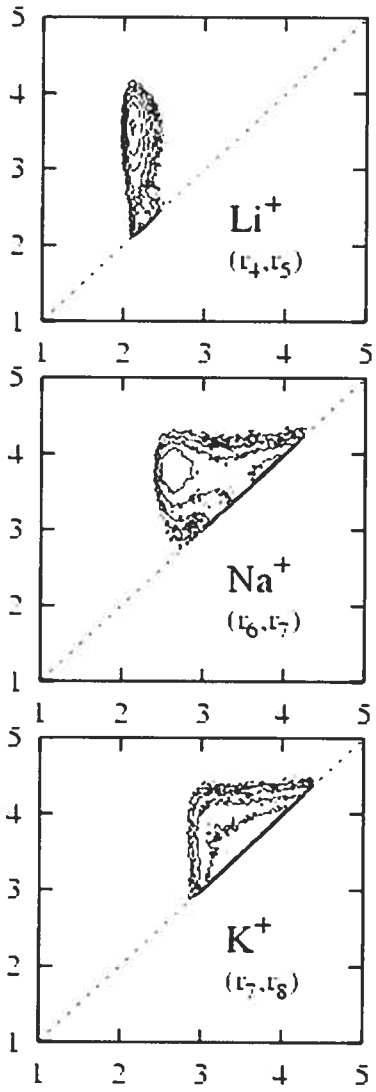

Figure 6.10: Joint probability distributions of r_n and r_{n+1} for lithium ($n = 4$), sodium ($n = 6$), and potassium ($n = 7$). It shows an associative exchange mechanism for lithium and a dissociative/associative mechanism of sodium. The contour scale is logarithmic and each level corresponds to $k_B T$. The distances are in Angstroms.

ture of the four closest molecules within the first solvation shell. We compute the distribution of the angles θ_{ij} between the i-th closest oxygen, the ion, and the j-th closest oxygen (for $i, j = 1$–4), as a function of r_5 (see Fig. 6.11). For $r_5 > 3$ Å, the four closest water molecules are forming a tetrahedral structure, with a signature 109° angle. The Car-Parrinello simulation of Lyubartsev *et al.* [84] and the QM/MM simulation of Loeffler and Rode [85] produce very similar distributions. As r_5 goes down to 2.2 Å, the distribution of angles presents a maximum at 90°, signalling that the tetrahedral structure is breaking into a bipyramidal structure. To complete the picture, we plot the distribution of θ_{i5} as a function of r_5 (see Fig. 6.11). It shows that as W_5 penetrates the first solvation shell, one molecule is on the opposite side of the ion ($\theta = 180°$) and the three others are at right angle.

A similar analysis for Na^+ (see Fig. 6.12) shows that, when W_6 is in the first shell (i.e., when $r_6 < 3.21$ Å), the molecules form an octahedral cage, whose signature is a major peak at $\theta_{i6} = 90°$ and a minor peak at $\theta_{i6} = 180°$. These characteristic angles get smaller for $r_6 > 3.21$ Å, indicating that the volume left by W_6 is being partially occupied. Figure 6.12b shows that the dissociated state is too short-lived for the remaining five coordinating water molecules to form a consistent bipyramidal structure. For any value of r_6, the θ_{ij} angle distribution displays peaks at 90 and 180°, and for $r_6 > 3.21$ Å, the 60–120° signature of the trigonal bipyramid is very weak. Figure 6.13 shows the distribution of angles between W_6 and W_7. For a dissociated W_6, W_7 has its most probable orientation at 45°, in position to exchange with W_6 from the same side. However, the whole range of θ_{67} angles from 45 to 135° is populated, and the new W_6 could enter the first solvation shell from any face of the octahedron. The 6-coordination octahedral structure would have to be reformed with a Berry pseudorotation [68]. These exchange mechanisms were described specifically for Na^+ by Rey and Hynes [114, 115].

The water exchange reactions produced by the AH/SWM4-DP for sodium are similar to the results from *ab initio* molecular dynamics by White *et al.* [89] and from MD simulations by Cascella *et al.* [113] Both works find comparably low sodium coordination numbers (5.2 and 5.7, respectively, compared to 5.67 for this

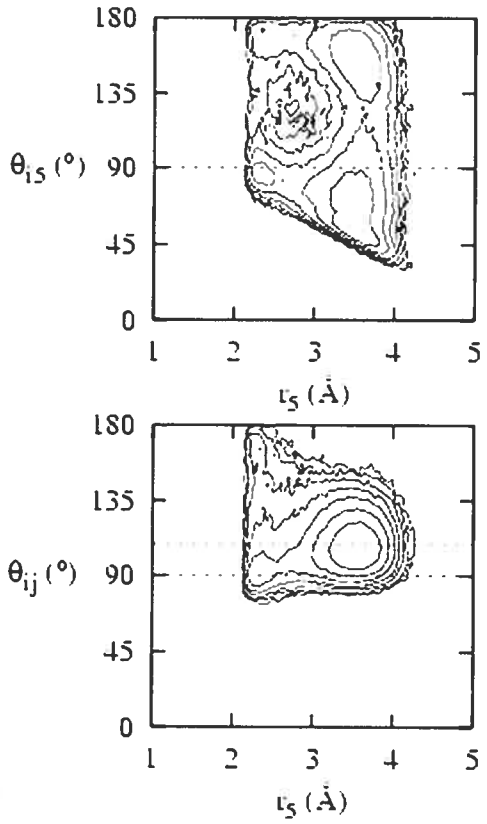

Figure 6.11: Effect of an approaching W_5 on the first shell structure of the tetracoordinated lithium ion. Combined joint distributions of (a) r_5 and θ_{i5}, and (b) r_5 and θ_{ij} (for $i, j = 1$–4). Each consecutive inner density contour is a factor e higher than the outer one. The distributions show that the molecules in the inner shell rearrange from a tetrahedron to a triangular bipyramid as W_5 comes in. The horizontal dashed lines are at 90° and at 109°.

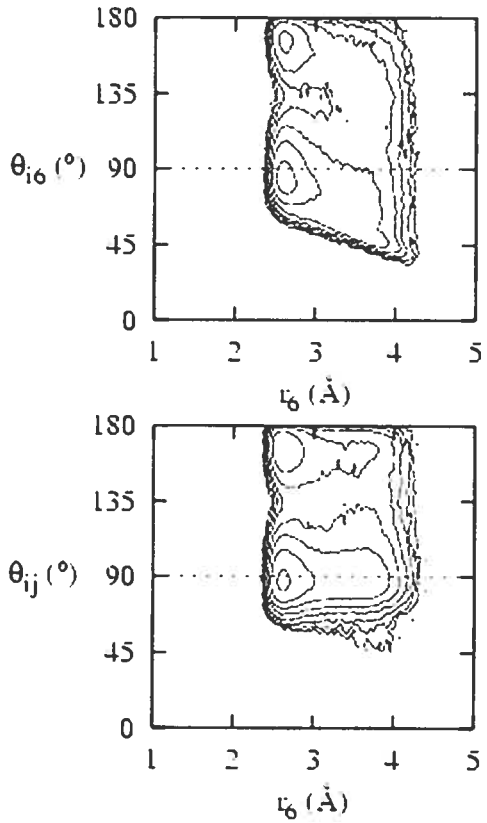

Figure 6.12: Effect of the dissociation of W_6 on the first shell structure of the sodium ion. Combined joint distributions of (a) r_6 and θ_{i6}, and (b) r_6 and θ_{ij} (for $i, j = 1$–5). The distributions show that, although the "equatorial" water molecules are moving toward W_6, the octahedral coordination structure is not significantly destroyed by the pre-exchange dissociation.

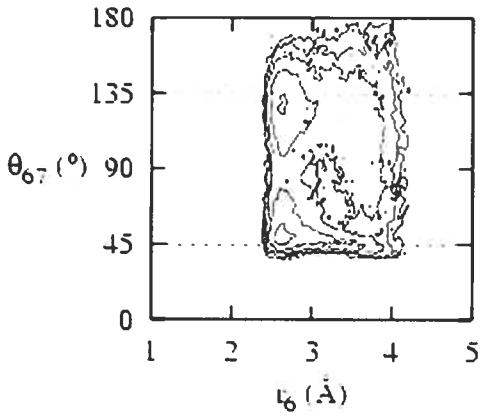

Figure 6.13: "Attack angles" of W_7 on the first solvation shell of sodium as W_6 is dissociating. The combined joint distribution of r_6 and θ_{67} shows that, as W_6 dissociates, W_7 is most likely to be inserted from the same side. The $\theta_{67} > 90°$ region is significantly populated, and a insertion from the opposite side (with the five resident molecules rearranging with a Berry pseudorotation [68]) is possible as well.

work), and find octahedral 6-coordinations alternating with trigonal-bipyramidal 5-coordinations as W_6 moves from first to second hydration shell. In contrast, the MD simulations by Spångberg, Wojcik, and Hermansson [116, 117] show a completely different water exchange mechanism for sodium, based on the association of a seventh water molecule. Accordingly, these simulations produce a coordination number of 6.5.

6.6 Summary and concluding remarks

The parameters of the simple functional Lennard-Jones form that was chosen were systematically varied. This procedure has shown that, with the exception of lithium, all alkali ions are bound to produce absolute hydration free energies about 5 kcal/mol higher than the consensus values emerging from Fig. 6.1. According to this consensus, the absolute free energy of the proton would be around −251.5 kcal/mol, but the AH/SWM4-DP model is consistent with $\Delta G^{\text{abs}}_{\text{hydr}}(\text{H}^+) \simeq -247$ kcal/mol. The halide ions, on the other side, are adjusted 5 kcal/mol lower. This ensures that the counter-ions are thermodynamically consistent, i.e., that

the solvation free energies of a neutral pair of ions correspond to the well-defined experimental values.

The model was then tested against the reproduction of zero-temperature hydrates. The model is successful for the microscopic hydration of alkali ions because the small cations are not significantly modifying the hydrogen-bonding structure of water. It is less successful for halide hydration probably because the first solvation shell of the halide ions undergoes severe distortions that are not representative of the bulk liquid water. This problem could arguably be tackled using a more robust representation of the hydrogen-bonding energetics, with stronger hydrogen-hydrogen repulsion and with explicit lone pairs instead of the effective leverage created by the "M" site of the SWM4-DP model. The precise shape of the oxygen–oxygen repulsive interaction could also be revisited. For that matter, a more complex water model could be tested against water hydrogen-bonding within the first shell of a halide ion. That being said, these limitations do not affect the properties of the anions in bulk water. Moreover, the somewhat high values used for the atomic polarizabilities of Br^- and I^- appear to produce reasonable hydration properties: the water coordination of bromide is similar that from recent *ab initio* MD studies [109]. Overall, the model of the halide series is reasonable for the purpose of providing fully hydrated counter-ions.

For the alkali ions, the water exchange reaction pathways are characterized in terms of the water coordinates relative to an n-th closest atom. The mechanisms are consistent with the residence numbers: a $4 \rightarrow 5 \rightarrow 4$ associative mechanism for Li^+ ($N_r = 4.0$) and a dissociative/associative mechanism for Na^+ ($N_r = 5.5$). The analysis suggests that the sodium ion is rattling inside an octahedral cage of six water molecules, of which one is more loosely bound. Potassium has a coordination number of 6.9 but a residence number of 6.3. Figure 6.9 suggests the picture of a first hydration shell of potassium containing "ideally" 6 water molecules, but being flexible enough to accomodate 7 or even 8 molecules at a small thermodynamical cost. Given the gained transferability for the sodium and potassium models, it is encouraging to see that the bulk properties, most importantly the coordination

numbers and the exchange mechanisms, are comparable to both the original Åqvist model [104] for the alkali series [113] and the recent CP studies of the hydrated ions [89, 95].

Finally, Fig. 6.1 serves as a warning about the interpretation of calculations and measurements of absolute hydration energies. Calculations combining both an implicit and explicit representation of the solvent may create an artificial interface whose interaction with the ionic charge should be accounted for. For the spherical solvent boundary potential (SSBP) [52], this interface potential was found to be almost identical to the air/water interface, and the charging free energies were interpreted accordingly, that is, as $\Delta G^{\text{real}}_{\text{hydr}}$. Given that the air/water potential is a difficult quantity to measure [118], and that the simulations do not agree on its value, it becomes important to distinguish between the $\Delta G^{\text{real}}_{\text{hydr}}$ obtained from a simulation that reproduces or mimicks the interaction with the air/water interface, and the $\Delta G^{\text{abs}}_{\text{hydr}}$ obtained from a simulation that does not include any interface (e.g., that uses periodic boundary conditions). For those two values to be rigorously compared, a correction for the phase potential should be agreed upon.

References

[1] Terry P. Lybrand and Peter A. Kollman. Water-water and water-ion potential functions including terms for many body effects. *J. Chem. Phys.*, 83(6) : 2923–2933, 1985.

[2] Liem X. Dang, Julia E. Rice, James Caldwell, and Peter A. Kollman. Ion Solvation in Polarizable Water : Molecular Dynamics Simulations. *J. Am. Chem. Soc.*, 113(7) : 2481–2486, 1991.

[3] Benoît Roux. Non-additivity in cation-peptide interactions. A molecular dynamics and ab initio study of Na+ in the gramicidin channel. *Chem. Phys. Lett.*, 212(3,4) : 231–240, 1993.

[4] Benoît Roux and Martin Karplus. Potential Energy Function for Cation-Peptide Interactions : An *Ab initio* study. *J. Comput. Chem.*, 16(6) : 690–704, 1995.

[5] Guillaume Lamoureux, Alexander D. MacKerell, Jr., and Benoît Roux. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.*, 119(10) : 5185–5197, 2003.

[6] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., 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, M. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, 102(18) : 3586–3616, 1998.

[7] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, 117(19) : 5179–5197, 1995.

[8] W. F. van Gunsteren, X. Daura, and A. E. Mark. The GROMOS force field. In P. von Ragué Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, III, and P. R. Schreiner (editors) : Encyclopedia of Computational Chemistry, volume 2, pages 1211–1216. Wiley & Sons, New York, 1998.

[9] William L. Jorgensen and Julian Tirado-Rives. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.*, 110(6) : 1657–1666, 1988.

[10] Guillaume Lamoureux and Benoît Roux. Modeling induced polarization with classical Drude oscillators : Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.*, 119(6) : 3025–3039, 2003.

[11] M. P. Allen and D. J. Tildesley. Computer Simulation of Liquids. Clarendon Press, Oxford, 1987.

[12] A. A. Rashin, I. A. Topol, G. J. Tawa, and S. K. Burt. Charge distributions in water and ion-water clusters. *Chem. Phys. Lett.*, 335 : 327–333, 2001.

[13] Howard Coker. Empirical Free-Ion Polarizabilities of the Alkali Metal, Alkaline Earth Metal and Halide Ions. *J. Phys. Chem.*, 80(19) : 2078–2084, 1976.

[14] G. D. Mahan. van der Waals coefficient between closed shell ions. *J. Chem. Phys.*, 76(1) : 493–497, 1982.

[15] P. W. Fowler and P. A. Madden. In-crystal polarizabilities of alkali and halide ions. *Phys. Rev. B*, 29(2) : 1035–1042, 1984.

[16] Patrick Jemmer, Patrick W. Fowler, Mark Wilson, and Paul A. Madden. Environmental Effects on Anion Polarizability : Variation with Lattice Parameter and Coordination Number. *J. Phys. Chem. A*, 102(43) : 8377–8385, 1998.

[17] Christof Hättig and Bernd Artur Heß. TDMP2 calculation of dynamic multipole polarizabilities and dispersion coefficients for the halogen anions F^-, Cl^-, Br^- and I^-. *J. Chem. Phys.*, 108(10) : 3863–3870, 1998.

[18] Pavel Jungwirth and Douglas J. Tobias. Chloride Anion on Aqueous Clusters, at the Air-Water Interface, and in Liquid Water : Solvent Effects on Cl^- Polarizability. *J. Phys. Chem. A*, 106(2) : 379–383, 2002.

[19] Lalith Perera and Max L. Berkowitz. Structures of $Cl^-(H_2O)_n$ and $F^-(H_2O)_n$ ($n = 2, 3, \ldots, 15$) clusters. Molecular dynamics computer simulations. *J. Chem. Phys.*, 100(4) : 3085–3093, 1994.

[20] Sotiris S. Xantheas and Liem X. Dang. Critical Study of Fluoride-Water Interactions. *J. Phys. Chem.*, 100(10) : 3989–3995, 1996.

[21] Pavel Jungwirth and Douglas J. Tobias. Ions at the Air/Water Interface. *J. Phys. Chem. B*, 106(25) : 6361–6373, 2002.

[22] Gil Markovich, Lalith Perera, Max L. Berkowitz, and Ori Cheshnovsky. The solvation of Cl^-, Br^-, and I^- in acetonitrile clusters : Photoelectron spectroscopy and molecular dynamics simulations. *J. Chem. Phys.*, 105(7) : 2675–2685, 1996.

[23] Liem X. Dang. Computational Study of Ion Binding to the Liquid Interface of Water. *J. Phys. Chem. B*, 106(40) : 10388–10394, 2002.

[24] Regla Ayala, José M. Martínez, Rafael R. Pappalardo, Humberto Saint-Martin, Ivan Ortega-Blake, and Enrique Sánchez Marcos. Development of first-principles interaction model potentials. An application to the study of the bromide hydration. *J. Chem. Phys.*, 117(23) : 10512–10524, 2002.

[25] T. N. V. Nguyen and G. H. Peslherbe. Microsolvation of alkali and halide ions in acetonitrile clusters. *J. Phys. Chem. A*, 107(10) : 1540–1550, 2003.

[26] E. Valderrama and R. J. Wheatley. An Environmental Pseudopotential Approach to Molecular Interactions : Implementation in MOLPRO. *J. Comput. Chem.*, 24(16) : 2075–2082, 2003.

[27] Howard Coker. Polarizability Changes on Ion Hydration. *J. Phys. Chem.*, 80(19) : 2084–2091, 1976.

[28] N. C. Pyper, C. G. Pike, and P. P. Edwards. The polarizabilities of species present in ionic solutions. *Mol. Phys.*, 76(2) : 353–372, 1992.

[29] G. D. Mahan. Modified Sternheimer equation for polarizability. *Phys. Rev. A*, 22(5) : 1780–1785, 1980.

[30] M. A. Carignano, G. Karlström, and P. Linse. Polarizable Ions in Polarizable Water : A Molecular Dynamics Study. *J. Phys. Chem. B*, 101(7) : 1142–1147, 1997.

[31] R. Gomer and G. Tryson. An experimental determination of absolute half-cell emf's and single ion free energies of solvation. *J. Chem. Phys.*, 66(10) : 4413–4424, 1977.

[32] Cornelius E. Klots. Solubility of Protons in Water. *J. Phys. Chem.*, 85(24) : 3585–3588, 1981.

[33] Michael D. Tissandier, Kenneth A. Cowen, Wan Yong Feng, Ellen Gundlach, Michael H. Cohen, Alan D. Earhart, James V. Coe, and Thomas R. Tuttle, Jr.. The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster-Ion Solvation Data. *J. Phys. Chem. A*, 102(40) : 7787–7794, 1998.

[34] Chang-Guo Zhan and David A. Dixon. Absolute Hydration Free Energy of the Proton from First-Principles Electronic Structure Calculations. *J. Phys. Chem. A*, 105(51) : 11534–11540, 2001.

[35] D. Asthagiri, Lawrence R. Pratt, and H. S. Ashbaugh. Absolute hydration free energies of ions, ion–water clusters, and quasichemical theory. *J. Chem. Phys.*, 119(5) : 2702–2708, 2003.

[36] Alan Grossfield, Pengyu Ren, and Jay W. Ponder. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *J. Am. Chem. Soc.*, 125(50) : 15671–15682, 2003.

[37] Yizhak Marcus. The Thermodynamics of Solvation of Ions. Part 4.—Application of the Tetraphenylarsonium Tetraphenylborate (TATB) Extrathermodynamic Assumption to the Hydration of Ions and to Properties of Hydrated Ions. *J. Chem. Soc., Faraday Trans. 1*, 83(9), 1987.

[38] Liem X. Dang and Tsun-Mei Chang. Molecular dynamics study of water clusters, liquid, and liquid-vapor interface of water with many-body potentials. *J. Chem. Phys.*, 106(19) : 8149–8159, 1997.

[39] Liem X. Dang and Tsun-Mei Chang. Molecular Mechanism of Ion Binding to the Liquid/Vapor Interface of Water. *J. Phys. Chem. B*, 106(2) : 235–238, 2002.

[40] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2) : 926–935, 1983.

[41] Scott E. Feller, Richard W. Pastor, Atipat Rojnuckarin, Stephen Bogusz, and Bernard R. Brooks. Effects of Electrostatic Force Truncation on Interfacial and Transport Properties of Water. *J. Phys. Chem.*, 100(42) : 17011–17020, 1996.

[42] James V. Coe. Connecting Cluster Anion Properties to Bulk : Ion Solvation Free Energy Trends with Cluster Size and the Surface vs Internal Nature of Iodide in Water Clusters. *J. Phys. Chem. A*, 101(11) : 2055–2063, 1997.

[43] Thomas R. Tuttle, Syd Malaxos, and James V. Coe. A New Cluster Pair Method of Determining Absolute Single Ion Solvation Energies Demonstrated in Water and Applied to Ammonia. *J. Phys. Chem. A*, 106(6) : 925–932, 2002.

[44] Yizhak Marcus. The Thermodynamics of Solvation of Ions. Part 2.—The Enthalpy of Hydration at 298.15 K. *J. Chem. Soc., Faraday Trans. 1*, 83(2) : 339–349, 1987.

[45] Richard M. Noyes. Thermodynamics of Ion Hydration as a Measure of Effective Dielectric Properties of Water. *J. Am. Chem. Soc.*, 84(4) : 513–522, 1962.

[46] Yizhak Marcus. The Hydration Entropies of Ions and their Effects on the Structure of Water. *J. Chem. Soc., Faraday Trans. 1*, 82(1) : 233–242, 1986.

[47] J. E. B. Randles. The real hydration energies of ions. *Trans. Faraday Soc.*, 52 : 1573–1581, 1956.

[48] Chang-Guo Zhan and David A. Dixon. Hydration of the Fluoride Anion : Structure and Absolute Hydration Free Energy from First-Principles Electronic Structure Calculations. *J. Phys. Chem. A*, 108(11) : 2020–2029, 2004.

[49] Wendell M. Latimer. The Energy of Solution of Gaseous Ions in Relation to the Effect of a Charge Upon the Dielectric. *J. Am. Chem. Soc.*, 48(5) : 1234–1239, 1926.

[50] Henry S. Ashbaugh. Convergence of Molecular and Macroscopic Continuum Descriptions of Ion Hydration. *J. Phys. Chem. B*, 104(31) : 7235–7238, 2000.

[51] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13 : 1011–1021, 1992.

[52] Dmitrii Beglov and Benoît Roux. Finite representation of an infinite bulk system : Solvent boundary potential for computer simulations. *J. Chem. Phys.*, 100(12) : 9050–9063, 1994.

[53] William G. Hoover. Canonical dynamics : Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3) : 1695–1697, 1985.

[54] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM : A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4 : 187–217, 1983.

[55] Eric D. Glendening and David Feller. Cation-Water Interactions : The $M^+(H_2O)_n$, Clusters for Alkali Metals, M = Li, Na, K, Rb, and Cs. *J. Phys. Chem.*, 99(10) : 3060–3067, 1995.

[56] Jongseob Kim, Han Myoung Lee, Seung Bum Suh, D. Majumdat, and Kwang S. Kim. Comparative *ab initio* study of the structures, energetics and spectra of $X^-\cdot(H_2O)_{n=1-4}$ [X = F,Cl,Br,I] clusters. *J. Chem. Phys.*, 113(13) : 5259–5272, 2000.

[57] I. Džidić and P. Kebarle. Hydration of the Alkali Ions in the Gas Phase. Enthalpies and Entropies of Reactions $M^+(H_2O)_{n-1} + H_2O = M^+(H_2O)_n$. *J. Phys. Chem.*, 74(7) : 1466–1474, 1970.

[58] Kenzo Hiraoka, Susumu Mizuse, and Shinichi Yamabe. Solvation of Halide Ions with H_2O and CH_3CN in the Gas Phase. *J. Phys. Chem.*, 92(13) : 3943–3952, 1988.

[59] M. T. Rodgers and P. B. Armentrout. Collision-Induced Dissociation Measurements on $Li^+(H_2O)_n$, $n = 1 - 6$: The First Direct Measurement of the Li^+–OH_2 Bond Energy. *J. Phys. Chem. A*, 101(7) : 1238–1249, 1997.

[60] David Feller, Eric D. Glendening, Rick A Kendall, and Kirk A. Peterson. An extended basis set *ab Initio* study of $Li^+(H_2O)_n$, n = 1–6. *J. Chem. Phys.*, 100(7) : 4981–4997, 1994.

[61] Jongseob Kim, Sik Lee, Seung Joo Cho, Byung Jin Mhin, and Kwang S. Kim. Structures, energetics, and spectra of aqua-sodium(I) : Thermodynamic effects and nonadditive interactions. *J. Chem. Phys.*, 102(2) : 839–849, 1995.

[62] Janet E. Del Bene. Basis Set and Correlation Effects on Computed Lithium Ion Affinities. *J. Phys. Chem.*, 100(15) : 6284–6287, 1996.

[63] Kenro Hashimoto and Tetsuya Kamimoto. Theoretical Study of Microscopic Solvation of Lithium in Water Clusters : Neutral and Cationic $Li(H_2O)_n$ (n = 1–6 and 8). *J. Am. Chem. Soc.*, 120(15) : 3560–3570, 1998.

[64] Han Myoung Lee, Jongseob Kim, Sik Lee, Byung Jin Mhin, and Kwang S. Kim. Aqua-potassium(I) complexes : *Ab initio* study. *J. Chem. Phys.*, 111(9) : 3995–4004, 1999.

[65] Ahmed M. El-Nahas and Kimihiko Hirao. Complexation of Li^+ and Cu^+ with HX (X = F, Cl, OH, SH, NH_2, and PH_2) Molecules by B3LYP and CCSD(T) Methods. *J. Phys. Chem. A*, 104(1) : 138–144, 2000.

[66] Oleg Borodin, Richard L. Bell, Yi Li, Dmitry Bedrov, and Grant D. Smith. Polarizable and nonpolarizable potentials for K^+ cation in water. *Chem. Phys. Lett.*, 336 : 292–302, 2001.

[67] Sotiris S. Xantheas. Quantitative Description of Hydrogen Bonding in Chloride-Water Clusters. *J. Phys. Chem.*, 100(23) : 9703–9713, 1996.

[68] James E. Huheey, Ellen A. Keiter, and Richard L. Keiter. Inorganic Chemistry : Principles of Structure and Reactivity. Harper Collins, Fourth edition, 1993.

[69] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald : An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12) : 10089–10092, 1993.

[70] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, 101(5) : 4177–4189, 1994.

[71] Alex J. Honeycutt and Richard J. Saykally. Building Solutions—One Molecule at a Time. *Science*, 299 : 1329–1330, 2003.

[72] Martin Kaupp and Paul v. R. Schleyer. Do Low-Coordinated Group 1-3 Cations $M^{n+}L^m$ ($M^{n+} = K^+$, Rb^+, Cs^+, Ca^{2+}, Sr^{2+}, Ba^{2+}, Sc^{3+}, Y^{3+}, La^{3+}; $L = NH_3$, H_2O, HF ; $m = 1$–3) with a Formal Noble-Gas Electron Configuration Favor Regular or "Abnormal" Shapes? *J. Phys. Chem.*, 96(18) : 7316–7323, 1992.

[73] David Feller. *Ab initio* Study of M^+ :18-Crown-6 Microsolvation. *J. Phys. Chem. A*, 101(14) : 2723–2731, 1997.

[74] Patrick Ayotte, Steen B. Nielsen, Gary H. Weddle, Mark A. Johnson, and Sotiris S. Xantheas. Spectroscopic Observation of Ion-Induced Water Dimer Dissociation in the $X^-\cdot(H_2O)_2$ (X = F, Cl, Br, I) Clusters. *J. Phys. Chem. A*, 103(50) : 10665–10669, 1999.

[75] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, 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, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople. Gaussian 98, Revision A.9. Gaussian, Inc., Pittsburgh PA, 1998.

[76] M. Masamura. Structures, Energetics, and Spectra of $Cl^-(H_2O)_n$ Clusters, $n = 1$–6 : Ab Initio Study. *J. Phys. Chem. A*, 106(38) : 8925–8932, 2002.

[77] Han Myoung Lee and Kwang S. Kim. Structures and spectra of iodide-water clusters $I^-(H_2O)_{n=1-6}$: An *ab initio* study. *J. Chem. Phys.*, 114(10) : 4461–4471, 2001.

[78] Richard A. Bryce, Mark A. Vincent, Nathaniel O. J. Malcolm, Ian H. Hillier, and Neil A. Burton. Cooperative effects in the structuring of fluoride water clusters : Ab initio hybrid quantum mechanical/molecular mechanical model incorporating polarizable fluctuating charge solvent. *J. Chem. Phys.*, 109(8) : 3077–3085, 1998.

[79] William H. Robertson, Eric G. Diken, Erica A. Price, Joong-Won Shin, and Mark A. Johnson. Spectroscopic Determination of the OH^- Solvation Shell in the $OH^- \cdot (H_2O)_n$ Clusters. *Science*, 299 : 1367–1372, 2003.

[80] Zhenhao Duan and Zhigang Zhang. Solvation properties of Li^+ and Cl^- in water : molecular dynamics simulation with a non-rigid model. *Mol. Phys.*, 101(10) : 1501–1510, 2003.

[81] Jaime E. Combariza and Neil R. Kestner. Microscopic Study of Fluoride-Water Clusters. *J. Phys. Chem.*, 98(13) : 3513–3517, 1994.

[82] Jaime E. Combariza, Neil R. Kestner, and Joshua Jortner. Energy-structure relationships for microscopic solvation of anions in water clusters. *J. Chem. Phys.*, 100(4) : 2851–2864, 1994.

[83] Susan B. Rempe, Lawrence R. Pratt, Gerhard Hummer, Joel D. Kress, Richard L. Martin, and Antonio Recondo. The Hydration Number of Li^+ in Liquid Water. *J. Am. Chem. Soc.*, 122(5) : 966–967, 2000.

[84] A. P. Lyubartsev, K. Laasonen, and A. Laaksonen. Hydration of Li^+ ion. An *ab initio* molecular dynamics simulation. *J. Chem. Phys.*, 114(7) : 3120–3126, 2001.

[85] Hannes H. Loeffler and Bernd M. Rode. The hydration structure of the lithium ion. *J. Chem. Phys.*, 117(1) : 110–117, 2002.

[86] Daniel Spångberg, Rossend Rey, James T. Hynes, and Kersti Hermansson. Rate and Mechanisms for Water Exchange around Li^+(aq) from MD Simulations. *J. Phys. Chem. B*, 107(18) : 4470–4477, 2003.

[87] Hannes H. Loeffler. Many-body effects on structure and dynamics of aqueous ionic solutions. *J. Comput. Chem.*, 24(10) : 1232–1239, 2003.

[88] Anan Tongraar, Klaus R. Liedl, and Bernd M. Rode. Born-Oppenheimer ab Initio QM/MM Dynamics Simulations of Na^+ and K^+ in Water : From Structure Making to Structure Breaking Effects. *J. Phys. Chem. A*, 102(50) : 10340–10347, 1998.

[89] Jody A. White, Eric Schewegler, Guilia Galli, and François Gygi. The solvation of Na^+ in water : First-principles simulations. *J. Chem. Phys.*, 113(11) : 4668–4673, 2000.

[90] Mauricio Carrillo-Tripp, Humberto Saint-Martin, and Iván Ortega-Blake. A comparative study of the hydration of Na^+ and K^+ with refined polarizable model potentials. *J. Chem. Phys.*, 118(15) : 7062–7073, 2003.

[91] Susan B. Rempe and Lawrence R. Pratt. The hydration number of Na^+ in liquid water. *Fluid Phase Equilibria*, 183 : 121–132, 2000.

[92] N. T. Skipper and G. W. Neilson. X-ray and neutron-diffraction studies on concentrated aqueous-solutions of sodium-nitrate and silver-nitrate. *Journal of Physics—Condensed Matter*, 1(26) : 4141–4154, 1989.

[93] E. Guàrdia and J. A. Padró. Molecular Dynamics Simulation of Single Ions in Aqueous Solutions : Effects of the Flexibility of the Water Molecules. *J. Phys. Chem.*, 94(15) : 6049–6055, 1990.

[94] Clint G. Guymon, Matthew L. Hunsaker, John N. Harb, Douglas Henderson, and Richard L. Rowley. Effects of solvent model flexibility on aqueous electrolyte behavior between electrodes. *J. Chem. Phys.*, 118(22) : 10195–10202, 2003.

[95] Lavanya M. Ramaniah, Marco Bernasconi, and Michele Parrinello. *Ab Initio* molecular-dynamics simulation of K^+ in water. *J. Chem. Phys.*, 111(4) : 1587–1591, 1999.

[96] G. Palinkas, T. Radnai, and F. Hajdu. Ion-Solvent and Solvent-Solvent Interactions—X-Ray Study of Aqueous Alkali Chloride Solutions. *Z. Naturforsch. A*, 35(1) : 107–114, 1980.

[97] G. W. Neilson and N. Skipper. K^+ coordination in aqueous solution. *Chem. Phys. Lett.*, 114(1) : 35–38, 1985.

[98] G. N. Chuev and M. V. Fedorov. Wavelet treatment of radial distribution functions of solutes. *Phys. Rev. E*, 68(027702) : 1–4, 2003.

[99] Yizhak Marcus. Ionic Radii in Aqueous Solutions. *Chem. Rev.*, 88 : 1475–1498, 1988.

[100] Y. Kubozono, A. Hirano, H. Maeda, S. Kashino, S. Emura, and H. Ishida. An EXAFS investigation of local-structure around Rb^+ in aqueous-solution. *Zeitschrift fur Naturforschung Section A—A Journal of Physical Sciences*, 49(6) : 727–729, 1994.

[101] J. L. Fulton, D. M. Pfund, S. L. Wallen, M. Newville, E. A. Stern, and Y. J. Ma. Rubidium ion hydration in ambient and supercritical water. *J. Chem. Phys.*, 105(6) : 2161–2166, 1996.

[102] Adriano Filipponi, Simone De Panfilis, Cecilia Oliva, Maria Antonietta Ricci, Paola D'Angelo, and Daniel T. Bowron. Ion Hydration under Pressure. *Phys. Rev. Lett.*, 91(16) : 165505, 2003.

[103] Johan E. Enderby. Ion solvation *via* neutron scattering. *Chem. Soc. Rev.*, 24(3) : 159–168, 1995.

[104] Johan Åqvist. Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *J. Phys. Chem.*, 94 : 8021–8024, 1990.

[105] Song Hi Lee and Jayendran C. Rasaiah. Molecular Dynamics Simulation of Ion Mobility. 2. Alkali Metal and Halide Ions Using the SPC/E Model for Water at 25°C. *J. Phys. Chem.*, 100(4) : 1420–1425, 1996.

[106] Anan Tongraar and Bernd Michael Rode. The hydration structures of F^- and Cl^- investigated by *ab initio* QM/MM molecular dynamics simulations. *Phys. Chem. Chem. Phys.*, 5(2) : 357–362, 2003.

[107] Anna Ignaczak, J. A. N. F. Gomes, and M. N. D. S. Cordeiro. Quantum and simulation studies of $X^-(H_2O)_n$ systems. *Electrochimica Acta*, 45(4–5) : 659–673, 1999.

[108] Patrick J. Merkling, Regla Ayala, José M. Martínez, Rafael R. Pappalardo, and Enrique Sánchez Marcos. Interplay of computer simulations and x-ray absorption spectra in the study of the bromide hydration structure. *J. Chem. Phys.*, 119(13) : 6647–6654, 2003.

[109] Simone Raugei and Michael L. Klein. Dynamics of Water Molecules in the Br^- Solvation Shell : An ab Initio Molecular Dynamics Study. *J. Am. Chem. Soc.*, 123(38) : 9484–9485, 2001.

[110] Gil Markovich, Rina Giniger, Menachem Levin, and Ori Cheshnovsky. Photoelectron spectroscopy of iodine anion solvated in water clusters. *J. Chem. Phys.*, 95(12) : 9416–9419, 1991.

[111] R. W. Impey, P. A. Madden, and I. R. McDonald. Hydration and Mobility of Ions in Solution. *J. Phys. Chem.*, 87(25) : 5071–5083, 1983.

[112] Daniel Spångberg and Kersti Hermansson. Effective three-body potentials for $Li^+(aq)$ and $Mg^{2+}(aq)$. *J. Chem. Phys.*, 119(14) : 7263–7281, 2003.

[113] Michele Cascella, Leonardo Guidoni, Amos Maritan, Ursula Rothlisberger, and Paolo Carloni. Multiple Steering Molecular Dynamics Applied to Water Exchange at Alkali Ions. *J. Phys. Chem. B*, 106(50) : 13027–13032, 2002.

[114] Rossend Rey and James T. Hynes. Hydration Shell Exchange Kinetics : An MD Study for Na^+(aq). *J. Phys. Chem.*, 100(14) : 5611–5614, 1996.

[115] Rossend Rey and James T. Hynes. Hydration shell exchange dynamics for Na^+ in water. *J. Phys. : Condens. Matter*, 8 : 9411–9416, 1996.

[116] Daniel Spångberg, Mark Wojcik, and Kersti Hermansson. Pressure dependence and activation volume for the water exchange mechanism in NaCl(aq) from MD simulations. *Chem. Phys. Lett.*, 276 : 114–121, 1997.

[117] Kersti Hermansson and Mark Wojcik. Water Exchange around Li^+ and Na^+ in LiCl(aq) and NaCl(aq) from MD Simulations. *J. Phys. Chem. B*, 102(31) : 6089–6097, 1998.

[118] Maria Paluch. Electrical properties of free surface of water and aqueous solutions. *Adv. Colloid Interface Sci.*, 84 : 27–45, 2000.

Chapitre 7

Calcul de l'énergie libre

L'énergie *interne*, la somme de l'énergie potentielle et de l'énergie cinétique, représente le contenu énergétique total d'un système. Cependant, seule une fraction de cette énergie, l'énergie *libre*, est disponible pour effectuer un travail. Ce court chapitre développe le concept d'énergie libre et explique certaines façons de la calculer.

7.1 Rappels de thermodynamique

Si on effectue un travail qui amène un système de conditions thermodynamiques X à des conditions thermodynamiques Y, on augmente son énergie interne. Si on relâche les contraintes Y et on laisse le système retourner à des conditions X, une partie de l'énergie interne accumulée est reconvertie en travail, mais une partie est dissipée sous forme de chaleur. On récupère le maximum de travail lorsque la reconversion se déroule suffisamment lentement que le système est à tout moment dans un état de quasi-équilibre thermodynamique avec les contraintes. Même dans ce cas idéal, un travail exercé en variant une certaine contrainte ne nous sera rendu au complet que si le système n'interagit avec son environnement qu'à travers cette contrainte. Si le système est couplé à un *réservoir de chaleur* (qui garde sa température à une valeur fixe), un travail mécanique tel une compression ne sera pas restitué au complet : une fraction sera dissipée en travail sur le bain de chaleur.

7.1.1 Première loi de la thermodynamique

D'un point de vue strictement thermodynamique, l'état d'un système est régi par le fait que l'énergie interne E est une fonction *extensive* des autres variables extensives du système : l'entropie S_{B}, le volume V et le nombre de particules N. Considérant un système ayant une énergie $E(S_{\mathrm{B}}, V, N)$, un système semblable λ fois plus grand aura une énergie

$$E(\lambda S_{\mathrm{B}}, \lambda V, \lambda N) = \lambda E(S_{\mathrm{B}}, V, N). \tag{7.1}$$

À cause du théorème d'Euler, on a pour une telle fonction extensive :

$$E(S_{\mathrm{B}}, V, N) = \left(\frac{\partial E}{\partial S_{\mathrm{B}}}\right)_{V,N} S_{\mathrm{B}} + \left(\frac{\partial S_{\mathrm{B}}}{\partial V}\right)_{N,S_{\mathrm{B}}} V + \left(\frac{\partial S_{\mathrm{B}}}{\partial N}\right)_{S_{\mathrm{B}},V} N. \tag{7.2}$$

Les dérivées partielles, coefficients de S_{B}, V et N, sont nécessairement des variables *intensives*. On définit les variables intensives conjuguées suivantes :

$$T \equiv \left(\frac{\partial E}{\partial S_{\mathrm{B}}}\right)_{V,N}, \qquad -p \equiv \left(\frac{\partial E}{\partial V}\right)_{S_{\mathrm{B}},N} \qquad \text{et} \qquad \mu \equiv \left(\frac{\partial E}{\partial N}\right)_{S_{\mathrm{B}},V}, \tag{7.3}$$

qui sont la température, la pression (ou plutôt, *moins* la pression) et le potentiel chimique. Ainsi, on a l'équation fondamentale :

$$E(S_{\mathrm{B}}, V, N) = TS_{\mathrm{B}} - pV + \mu N. \tag{7.4}$$

Comme cette fonction ne dépend que de S_{B}, V et N, sa différentielle exacte s'écrit

$$dE = T\,dS_{\mathrm{B}} - p\,dV + \mu\,dN. \tag{7.5}$$

En d'autres termes, l'énergie interne d'un système peut être augmentée en y ajoutant de la chaleur ($dS_{\mathrm{B}} > 0$), en le comprimant ($dV < 0$) ou en y insérant des particules ($dN > 0$).

7.1.2 Potentiels thermodynamiques

On définit le *potentiel de Helmholtz* F comme la transformation de Legendre de E suivante :

$$\begin{aligned} F(T,V,N) &\equiv E(S_{\mathrm{B}},V,N) - TS_{\mathrm{B}} \\ &= -pV + \mu N. \end{aligned} \tag{7.6}$$

De manière générale, une transformation de Legendre d'une fonction extensive E de trois variables extensives X_1, X_2 et X_3 peut former une nouvelle fonction extensive $\widetilde{E}$ indépendante de la variable extensive X_i mais dépendante de la variable intensive conjuguée $p_i = \partial E/\partial X_i$:

$$\widetilde{E}(p_i) = E - p_i X_i. \tag{7.7}$$

Le potentiel de Helmholtz $F \equiv \widetilde{E}(T)$, qui dépend de la température plutôt que de l'entropie, est approprié pour décrire l'état d'un système dans l'ensemble canonique. Pour un processus isotherme (et *non-diffusif*), sa variation correspond au travail mécanique fourni par compression :

$$dF = -\underbrace{S_{\mathrm{B}}\,dT}_{0} - p\,dV + \underbrace{\mu\,dN}_{0} = -p\,dV. \tag{7.8}$$

C'est aussi le travail idéalement restitué lorsque le système se dilate spontanément. On définit l'*enthalpie* H de la façon suivante :

$$\begin{aligned} H(S_{\mathrm{B}},-p,N) &\equiv E(S_{\mathrm{B}},V,N) + pV \\ &= TS_{\mathrm{B}} + \mu N. \end{aligned} \tag{7.9}$$

Pour un processus isobare (et non-diffusif), la variation de l'enthalpie correspond à la chaleur ajoutée ou enlevée au système :

$$dH = T\,dS_{\mathrm{B}} + \underbrace{V\,dp}_{0} + \underbrace{\mu\,dN}_{0} = T\,dS_{\mathrm{B}}. \tag{7.10}$$

On définit le *potentiel de Gibbs* G à partir du potentiel de Helmholtz ou de l'enthalpie :

$$\begin{aligned} G(T, -p, N) &\equiv F(T, V, N) + pV \\ &\equiv H(S_{\mathrm{B}}, -p, N) - TS_{\mathrm{B}} \\ &= \mu N. \end{aligned} \tag{7.11}$$

Le potentiel de Gibbs, qui dépend de la pression plutôt que du volume, est approprié pour décrire l'état d'un système dans l'ensemble isotherme–isobare. Pour un processus isotherme–isobare (et non-diffusif), sa variation est strictement nulle :

$$dG = -\underbrace{S_{\mathrm{B}}\,dT}_{0} + \underbrace{V\,dp}_{0} + \underbrace{\mu\,dN}_{0} = 0. \tag{7.12}$$

En d'autres termes, c'est la quantité conservée—et *échangée*—lors d'une réaction à température et pression constantes. Pour un processus diffusif ($dN \neq 0$), elle correspond à l'énergie requise pour insérer ou enlever des particules du système.

Les fonctions F et G sont appelées des *énergies libres* parce qu'elles représentent le maximum théorique de l'énergie disponible pour effectuer un travail selon une contrainte thermodynamique donnée. À température constante, une réaction spontanée $A \longrightarrow B$ dégage une énergie libre (de Helmholtz) ΔF qui, idéalement, peut être entièrement convertie en travail. À température et pression constantes, comme c'est habituellement le cas pour les processus biochimiques *in vivo*, la même réaction dégage une énergie libre (de Gibbs) ΔG. Si la réaction n'est pas spontanée, l'énergie libre correspond au travail minimum qu'il faut fournir pour la provoquer. Dans un organisme vivant, l'énergie accumulée dans un lien chimique ou dans une affinité moléculaire se mesure donc en « unités » de G. L'énergie libre F [équation (7.6)] correspond à l'énergie interne E moins TS_{B}, la quantité minimum de chaleur contenue dans un système à température T. L'énergie libre G [équation (7.11)] contient aussi pV, le travail nécessaire pour maintenir le système à pression p.

Pour établir une connection avec la mécanique statistique, il est intéressant de reprendre le développement précédent avec l'entropie (plutôt que l'énergie) comme fonction extensive fondamentale. En écrivant l'entropie comme une quantité sans dimensions $S \equiv S_{\mathrm{B}}/k_{\mathrm{B}}$ et en définissant $\beta \equiv 1/k_{\mathrm{B}}T$, on a, suivant l'équation (7.4) :

$$S(E, V, N) = \beta E + \beta p V - \beta \mu N. \tag{7.13}$$

Les variables intensives conjuguées sont β, βp et $-\beta\mu$, équivalentes aux variables T, $-p$ et μ. L'équivalent du potentiel de Helmholtz est

$$\begin{aligned} -\beta F(\beta, V, N) &\equiv S(E, V, N) - \beta E \\ &= \beta p V - \beta \mu N, \end{aligned} \tag{7.14}$$

l'équivalent de l'enthalpie est

$$\begin{aligned} -\beta H(E, \beta p, N) &\equiv S(E, V, N) - \beta p V \\ &= \beta E - \beta \mu N \end{aligned} \tag{7.15}$$

et l'équivalent du potentiel de Gibbs est

$$\begin{aligned} -\beta G(\beta, \beta p, N) &\equiv -\beta F(\beta, V, N) - \beta p V \\ &\equiv -\beta H(E, \beta p, N) - \beta E \\ &= -\beta \mu N. \end{aligned} \tag{7.16}$$

Suivant les contraintes thermodynamiques en présence, l'état d'équilibre est représenté par les extréma des fonctions S, $-\beta F$, $-\beta H$ ou $-\beta G$. Dans l'ensemble microcanonique, l'équilibre thermodynamique est atteint quand l'entropie S est maximum par rapport à toute variation virtuelle des conditions E, V ou N. Dans l'ensemble canonique, il est atteint quand $-\beta F$ est maximum par rapport à V et N (tout comme S) mais *minimum* par rapport à β. Dans l'ensemble isotherme–isobare, l'équilibre est atteint quand $-\beta G$ est maximum par rapport à N mais

minimum par rapport à β et βp. Dans les termes plus courants des transformées de l'énergie, l'équilibre thermodynamique est atteint quand, à température constante, l'energie libre de Helmholtz F diminue pour toute variation de T et augmente pour toute variation de V ou de N. À température et pression constantes, l'équilibre est atteint quand l'energie libre de Gibbs G diminue pour toute variation de T ou de p et augmente pour toute variation de N.

7.1.3 Mécanique statistique

Si on conçoit le système thermodynamique discuté jusqu'à présent comme un fluide de N atomes identiques décrit par $3N$ coordonnées $\{q\}$ et $3N$ impulsions conjuguées $\{p\}$, l'énergie interne instantanée du système est $E(\{q\},\{p\})$. L'entropie de Boltzmann est liée à la fonction de partition microcanonique Ω :

$$\begin{aligned} e^S &= \Omega(E,V,N) \\ &= \frac{1}{N!}\int d\{\hat{q}\}d\{\hat{p}\}\,\delta[E(\{q\},\{p\})-E]. \end{aligned} \tag{7.17}$$

L'accent circonflexe au-dessus de chaque élément de longueur dx indique symboliquement que l'intégrale sur le domaine de la variable x est normalisée. Cette relation, aussi écrite $S_{\mathrm{B}} = k_{\mathrm{B}} \ln \Omega$, est le lien fondamental entre la thermodynamique et la mécanique statistique. L'énergie libre de Helmholtz est liée à la fonction de partition canonique Q, qui est une transformée de Laplace de la fonction Ω :

$$\begin{aligned} e^{-\beta F} &= Q(\beta,V,N) = \int dE\,\Omega(E,V,N)e^{-\beta E} \\ &= \frac{1}{N!}\int d\{\hat{q}\}d\{\hat{p}\}\,e^{-\beta E(\{q\},\{p\})}. \end{aligned} \tag{7.18}$$

L'énergie libre de Gibbs est liée à la fonction de partition isotherme–isobare Δ, qui est une transformée de Laplace de la fonction Q :

$$e^{-\beta G} = \Delta(\beta,\beta p,N) = \int d\hat{V}\,Q(\beta,V,N)e^{-\beta pV}$$

$$= \frac{1}{N!}\int d\hat{V}\int d\{\hat{q}\}d\{\hat{p}\}\, e^{-\beta E(\{q\},\{p\})}e^{-\beta pV}. \tag{7.19}$$

Grace à cette connexion avec les intégrales de la mécanique statistique, un potentiel thermodynamique $\widetilde{E}(p_i)$ peut servir de *fonction génératrice* des *cumulants* d'une variable extensive X_i :

$$\langle X_i^m \rangle_{\mathrm{c}} = \left(\frac{\partial^m \widetilde{E}(p_i)}{\partial(-p_i)^m}\right)_{\{p',X'\}}. \tag{7.20}$$

La notation $\{p', X'\}$ signifie que toutes les autres variables dont le potentiel dépend sont gardées constantes. En pratique, cela signifie que plusieurs propriétés du système peuvent être mesurées ou calculées à partir de la variation de l'énergie libre par rapport à la température ou la pression. Par exemple, pour un système à température constante, on peut extraire l'énergie interne à partir de l'énergie libre de Helmholtz :

$$E = \left(\frac{\partial(-\beta F)}{\partial(-\beta)}\right)_{V,N}. \tag{7.21}$$

On peut aussi extraire la capacité calorifique à volume constant, C_V, définie comme la dérivée de l'énergie interne en fonction de la température :

$$C_V \equiv \left(\frac{\partial E}{\partial T}\right)_{V,N} = \frac{1}{k_{\mathrm{B}}T^2}\left(\frac{\partial E}{\partial(-\beta)}\right)_{V,N} = \frac{1}{k_{\mathrm{B}}T^2}\left(\frac{\partial^2(-\beta F)}{\partial(-\beta)^2}\right)_{V,N} = \frac{\langle E^2\rangle_{\mathrm{c}}}{k_{\mathrm{B}}T^2}. \tag{7.22}$$

Pour un système à température et pression constantes, on peut obtenir l'enthalpie à partir de l'énergie libre de Gibbs :

$$H = \left(\frac{\partial(-\beta G)}{\partial(-\beta)}\right)_{\beta p,N}. \tag{7.23}$$

La capacité calorifique à pression constante, C_p, est définie comme la dérivée de l'enthalpie en fonction de la température :

$$C_p \equiv \left(\frac{\partial H}{\partial T}\right)_{\beta p,N} = \frac{1}{k_{\mathrm{B}}T^2}\left(\frac{\partial H}{\partial(-\beta)}\right)_{\beta p,N} = \frac{1}{k_{\mathrm{B}}T^2}\left(\frac{\partial^2(-\beta G)}{\partial(-\beta)^2}\right)_{\beta p,N} = \frac{\langle H^2\rangle_{\mathrm{c}}}{k_{\mathrm{B}}T^2}. \tag{7.24}$$

À pression constante, la capacité calorifique est liée aux fluctuations de l'enthalpie (sachant que $\langle H^2 \rangle_c = \langle H^2 \rangle - \langle H \rangle^2$). À volume constant, elle est liée aux fluctuations de l'énergie.

7.2 Énergie libre d'une simulation de dynamique moléculaire

À température et pression constantes, un système moléculaire dont l'énergie interne instantanée s'écrit $E(\{q\},\{p\}) = U(\{q\}) + K(\{p\})$ a une énergie libre

$$G = -k_B T \ln \left[\frac{1}{N!} \int d\hat{V} \int d\{\hat{q}\} d\{\hat{p}\}\, e^{-\beta(E+pV)} \right]. \tag{7.25}$$

La quantité entre crochets est la fonction de partition isotherme–isobare Δ, et correspond à $1/\langle e^{+\beta(E+pV)} \rangle$:

$$\begin{aligned} \frac{1}{\Delta} &= \frac{1}{\frac{1}{N!} \int d\hat{V} \int d\{\hat{q}\} d\{\hat{p}\}\, e^{-\beta(E+pV)}} = \frac{\frac{1}{N!} \int d\hat{V} \int d\{\hat{q}\} d\{\hat{p}\}\, e^{+\beta(E+pV)} e^{-\beta(E+pV)}}{\frac{1}{N!} \int d\hat{V} \int d\{\hat{q}\} d\{\hat{p}\}\, e^{-\beta(E+pV)}} \\ &= \langle e^{+\beta(E+pV)} \rangle. \end{aligned} \tag{7.26}$$

On obtient donc

$$G = k_B T \ln \left\langle e^{+\beta(E+pV)} \right\rangle. \tag{7.27}$$

Calculer l'énergie libre revient finalement à calculer, au cours d'une simulation de dynamique moléculaire, la valeur moyenne de $e^{+\beta(E+pV)}$, l'exponentielle de l'enthalpie instantanée. Il apparaît un problème fondamental : plus une région de l'espace des phases a une enthalpie élevée, plus elle contribue à la valeur moyenne, mais moins la dynamique moléculaire peut efficacement l'explorer. De fait, la moyenne de la quantité $e^{+\beta(E+pV)}$ au cours d'une simulation de dynamique moléculaire converge extrêmement lentement. Comment, donc, calculer efficacement l'énergie libre ?

7.2.1 Intégration thermodynamique

Comme l'énergie libre est une *variable d'état*, elle ne dépend pas de la façon dont un système a été construit, ou de la façon que son équilibre thermodynamique a été atteint. On pourrait en principe calculer la différence d'énergie libre d'une réaction d'association $A+B \rightleftharpoons AB$ en préparant deux systèmes thermodynamiques : un où les deux espèces A et B sont isolées (le côté « $A+B$ » de la réaction), et un où les deux espèces se mélangent et sont libres de s'associer (le côté « AB »). L'énergie libre dégagée par l'association de A et B serait simplement $\Delta G = G_{AB} - G_{A+B}$. La quantité ΔG est indépendante du mécanisme de la réaction de mélange et d'association des espèces A et B. S'il est difficile de calculer chaque terme de cette différence, on peut néanmoins imaginer un paramètre λ contrôlant l'état thermodynamique du système de sorte que $\lambda = 0$ correspond à l'état $A+B$ et que $\lambda = 1$ correspond à l'état AB. Dans un tel cas, la différence d'énergie libre est

$$\Delta G = \int_0^1 d\lambda \frac{\partial G}{\partial \lambda}; \tag{7.28}$$

peu importe la signification physique des états pour des valeurs intermédiaires $0 < \lambda < 1$. L'intérêt de cette réécriture est que $\partial G/\partial\lambda$ est une quantité potentiellement plus facile à calculer que G. Reprenant la définition de l'énergie libre (7.19), on a :

$$\frac{\partial G}{\partial \lambda} = -\frac{1}{\beta\Delta}\frac{\partial \Delta}{\partial \lambda} = \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda . \tag{7.29}$$

H est l'enthalpie instantanée $E + pV$ et la notation $\langle \cdots \rangle_\lambda$ précise que la moyenne est prise dans l'ensemble isotherme–isobare à la valeur λ. Le paramètre λ contrôle l'interaction des espèces A et B à l'aide d'une continuation analytique de l'énergie potentielle $U(\{q\}, \lambda)$, et on a simplement

$$\Delta G = \int_0^1 d\lambda \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda . \tag{7.30}$$

Le difficile calcul de $\langle e^{\beta H} \rangle$ peut donc être remplacé par une intégration nu-

mérique de la fonction $\langle \partial U/\partial\lambda \rangle_\lambda$. L'intégration numérique pourrait utiliser $n+1$ « fenêtres » de largeur uniforme $\Delta\lambda = 1/(n+1)$, dont les contributions seraient additionnées suivant la règle trapézoïdale [1, page 131] :

$$\Delta G \simeq \tfrac{1}{2}\Delta\lambda \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\Delta\lambda/2} + \sum_{i=1}^{n-1} \Delta\lambda \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{(i+1/2)\Delta\lambda} + \tfrac{1}{2}\Delta\lambda \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{(n+1/2)\Delta\lambda} . \quad (7.31)$$

Les $n+1$ simulations sont faites pour les valeurs λ centrées sur l'intervalle de chaque fenêtre. Imaginons que les atomes de la simulations sont divisés en trois blocs : les atomes formant l'espèce A, ceux formant l'espèce B et ceux formant le solvant (s). La continuation analytique de l'énergie potentielle pourrait alors s'écrire

$$U(\lambda) = U_s + U_A + U_B + U_{sA} + U_{sB} + \lambda U_{AB}. \quad (7.32)$$

Dans ce cas, la quantité $\langle \partial U/\partial\lambda \rangle = \langle U_{AB} \rangle$ peut être calculée sans problème, au fur et à mesure de la dynamique.

Idéalement, la forme de l'extension $U(\lambda)$ devrait être telle que $\langle \partial U/\partial\lambda \rangle$ soit constant de $\lambda = 0$ à $\lambda = 1$. Dans ce cas idéal, l'équation (7.31) se réduit à $\Delta G \simeq \langle \partial U/\partial\lambda \rangle$, ce qui signifie que ΔG pourrait être calculé à l'aide d'une seule fenêtre prise pour n'importe quelle valeur de λ. On peut en principe régulariser l'intégration thermodynamique *a posteriori* en formulant un paramètre $\mu = \mu(\lambda)$ qui soit tel que

$$\left\langle \frac{\partial U}{\partial \mu} \right\rangle_\mu = \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda(\mu)} \frac{\partial \lambda}{\partial \mu} = \text{constante}. \quad (7.33)$$

Sans aller jusqu'à cette extrémité d'aucun intérêt pratique—Elle exige de connaître la forme exacte de $\langle \partial U/\partial\lambda \rangle_\lambda$.—, on peut néanmoins préférer de deux formes alternatives $U(\lambda)$ et $U(\mu)$ celle qui a la variation la plus régulière pour une variété de systèmes de simulation—et donc qui s'intègre avec un plus petit nombre de fenêtres.

7.2.2 Méthode perturbative

L'idée de base de cette méthode est la suivante : si l'on dispose d'une trajectoire de dynamique moléculaire générée avec la fonction potentielle $U(\{\mathbf{r}\})$, on peut calculer la contribution à l'énergie libre d'une perturbation $\delta U(\{\mathbf{r}\})$ de la façon suivante :

$$\begin{aligned}
\Delta G_{U\to U+\delta U} &= -k_{\mathrm{B}}T\ln\left(\frac{\Delta_{U+\delta U}}{\Delta_U}\right) = -k_{\mathrm{B}}T\ln\left[\frac{\int d\{\mathbf{r}\}\, e^{-\beta(U+\delta U)}}{\int d\{\mathbf{r}\}\, e^{-\beta U}}\right] \\
&= -k_{\mathrm{B}}T\ln\left(\frac{\int d\{\mathbf{r}\}\, e^{-\beta\delta U}e^{-\beta U}}{\int d\{\mathbf{r}\}\, e^{-\beta U}}\right) \\
&= -k_{\mathrm{B}}T\ln\langle e^{-\beta\delta U}\rangle_U. \qquad (7.34)
\end{aligned}$$

Cette relation est précise dans la mesure où δU demeure une petite perturbation de l'état U. Si δU est significativement plus grand que $k_{\mathrm{B}}T$, des configurations statistiquement importantes dans l'ensemble $U+\delta U$ ne sont pas visitées dans l'ensemble U, et le résultat est biaisé. Si les états U et $U+\delta U$ ne se recoupent pas suffisamment dans l'espace des configurations, on obtient une énergie libre différente pour le calcul perturbatif inverse :

$$\Delta G_{U+\delta U\to U} \neq -\Delta G_{U\to U+\delta U}. \qquad (7.35)$$

Dans ce cas, on devra utiliser l'équation (7.34) de façon séquentielle :

$$\Delta G_{U_0\to U_n} = \sum_{i=0}^{n-1}\Delta G_{U_i\to U_{i+1}}, \qquad (7.36)$$

et vérifier que $\Delta G_{U_0\to U_n} \simeq -\Delta G_{U_n\to U_0}$.

La méthode perturbative présente un biais supplémentaire lorsque le potentiel n'est pas additif. Pour un potentiel U non-additif, une perturbation δU induit une réponse linéaire dans le système, qui se traduit par une augmentation de l'énergie

de l'ordre de $(\partial U/\partial \delta U)\delta U$. Ce terme de couplage n'est pas représenté dans l'équation (7.34) et manque au calcul de ΔG. Notons que la méthode de l'intégration thermodynamique ne souffre pas de ce problème, puisque la fonction de réponse $\partial U/\partial \lambda$ est exacte, que U soit additif ou non. Cela dit, dans les cas où il est difficile d'exprimer la forme analytique de $\partial U/\partial \lambda$, il est toujours possible d'utiliser la méthode perturbative avec un plus grand nombre de fenêtres. Pour une perturbation δU suffisamment petite, la réponse $(\partial U/\partial \delta U)\delta U$ devient négligeable par rapport à $k_{\mathrm{B}}T$.

7.2.3 Utilisation d'un potentiel de biaisage

On peut calculer l'énergie libre *partielle* suivant une coordonnée de réaction $\xi(\{\mathbf{r}\})$:

$$\begin{aligned} W(\xi) &= -k_{\mathrm{B}}T \ln \rho(\xi) \\ &= -k_{\mathrm{B}}T \ln \left[\frac{\int d\{\mathbf{r}\}\, \delta[\xi'(\{\mathbf{r}\}) - \xi] e^{-\beta U(\{\mathbf{r}\})}}{\int d\{\mathbf{r}\}\, e^{-\beta U(\{\mathbf{r}\})}} \right]. \end{aligned} \tag{7.37}$$

Cette fonction W s'appelle le *potentiel de force moyenne* (PMF) [2]. On obtient l'énergie libre totale de la réaction en l'intégrant de $\xi = 0$ à $\xi = 1$. Si aucune importante barrière d'énergie ne doit être franchie lors de la réaction, il est envisageable de calculer la densité de probabilité $\rho(\xi)$ sur tout l'intervalle $[0, 1]$ à l'aide d'une seule longue simulation. Autrement, il est préférable d'utiliser une technique de biaisage comme celle du *umbrella sampling* de Torrie et Valleau [3]. Pour échantillonner la dynamique moléculaire autour d'un point ξ_i dont la probabilité est faible, on peut effectuer une simulation séparée avec un potentiel de biaisage $w_i(\xi)$ ajouté à l'énergie $U(\{\mathbf{r}\})$. Ce potentiel devra forcer le système à occuper la région ξ_i (la *fenêtre*) en imposant une pénalité pour des conformations $\xi(\{\mathbf{r}\})$ trop éloignées de ξ_i. Sa forme la plus courante est

$$w_i(\xi) = \tfrac{1}{2}K(\xi - \xi_i)^2. \tag{7.38}$$

La densité biaisée $\widetilde{\rho}_i(\xi)$ obtenue pour cette simulation ne peut pas directement être ajoutée à un échantillonnage $\rho(\xi)$ obtenu précédemment sans potentiel de biaisage, mais doit être *débiaisée*, c'est-à-dire reconvertie pour le potentiel original U. La relation de principe entre la densité biaisée $\widetilde{\rho}$ et la densité débiaisée ρ est la suivante [3] :

$$\rho_i(\xi) = e^{\beta[w_i(\xi)-f_i]}\widetilde{\rho}_i(\xi). \tag{7.39}$$

La quantité f_i est une inconnue qui représente l'énergie libre ajoutée au système par l'introduction du potentiel de biaisage.

$$e^{-\beta f_i} = \left\langle e^{-\beta w_i} \right\rangle. \tag{7.40}$$

Cette moyenne dans l'ensemble non-biaisé n'est évidemment jamais calculée.

Si, par contre, on dispose de n simulations $i = 1$ à $i = n$, biaisées par des potentiels de la forme (7.38) répartis sur tout l'intervalle d'intérêt de ξ, on peut obtenir les n inconnus f_i de façon auto-cohérente à l'aide de la méthode WHAM (pour *weighted histogram analysis method*) [4, 5]. On écrit la distribution $\rho(\xi)$ comme une combinaison pondérée des densitées débiaisées $\rho_i(\xi)$:

$$\rho(\xi) = \sum_{i=1}^{n} p_i(\xi)\rho_i(\xi). \tag{7.41}$$

Les poids p_i sont des inconnus que l'on choisit de telle sorte que l'erreur statistique sur ρ est minimale en tout point de la réaction, c'est-à-dire que

$$\frac{\partial}{\partial p_i}\sigma^2[\rho(\xi)] = 0. \tag{7.42}$$

Cette condition d'optimalité met les poids p_i en relation avec les inconnus f_i :

$$p_i(\xi) = \frac{n_i e^{-\beta[w_i(\xi)-f_i]}}{\sum_{j=1}^{n} n_j e^{-\beta[w_j(\xi)-f_j]}}. \tag{7.43}$$

La démonstration en est faite notamment par Souaille et Roux [5]. Le facteur n_i

correspond au poids statistique de chaque simulation—le nombre de configurations statistiquement indépendantes utilisées pour le moyennage. En terme des densités biaisées $\widetilde{\rho}_i$, on a donc

$$\rho(\xi) = \sum_{i=1}^{n} \frac{n_i}{\sum_{j=1}^{n} n_j e^{-\beta[w_j(\xi)-f_j]}} \widetilde{\rho}_i(\xi). \tag{7.44}$$

On transforme cette équation en un système d'équations auto-cohérentes en se rappelant que $\rho(\xi)$ permet de calculer les quantités f_k :

$$e^{-\beta f_k} = \int d\xi \, \rho(\xi) e^{-\beta w_k(\xi)} \tag{7.45}$$

et donc que

$$e^{-\beta f_k} = \int d\xi \sum_{i=1}^{n} \frac{n_i e^{-\beta w_k(\xi)}}{\sum_{j=1}^{n} n_j e^{-\beta[w_j(\xi)-f_j]}} \widetilde{\rho}_i(\xi). \tag{7.46}$$

Quand on dispose des histogrammes biaisés $\{\widetilde{\rho}_i(\xi)\}$ et des formes fonctionnelles des potentiels de biaisage $\{w_i(\xi)\}$, ces équations permettent de résoudre itérativement les valeurs $\{f_i\}$. Les valeurs $\{f_i\}$ optimales permettent ensuite de calculer une densité débiaisée $\rho(\xi)$, directement reliée au potentiel de force moyenne $W(\xi)$.

La méthode WHAM est une extension de la méthode des histogrammes de Ferrenberg et Swendsen [6]. Très versatile, elle peut être généralisée à plusieurs coordonnées de réaction [7, 2]. Elle a aussi l'avantage de pouvoir traiter un nombre arbitraire de fenêtres i, choisies arbitrairement le long de la coordonnée de réaction. Pour que le système d'équations auto-cohérentes (7.46) soit fortement couplé et qu'il converge bien d'un point de vue numérique, on devra s'assurer que les densités biaisées des fenêtres se recouvrent suffisamment, c'est-à-dire que le dénominateur de l'équation (7.46) ait une valeur appréciable en tout point de la réaction. Une bonne pratique consiste à vérifier que la somme des densités biaisées, $\sum_i \widetilde{\rho}_i(\xi)$, assure une « couverture » minimale en tout point ξ, et à combler les lacunes dans l'échantillonnage en ajoutant des fenêtres.

Des propositions plus récentes ont étendu les possibilités des calculs d'énergie libre à des simulations biaisées de manière *dynamique*. La première consiste à périodiquement adapter le potentiel de biaisage à la réponse du système obtenue en cours de simulation [8, 9, 10]. Les diverses formes de cette proposition permettent d'explorer des réactions possédant des coordonnées de réaction multidimensionnelles. La seconde proposition est une généralisation des principes de calcul à un potentiel de biaisage (ou une perturbation) dynamique qui entraîne le système hors d'équilibre [11, 12, 13]. Elle repose sur une relation exacte entre l'énergie libre d'une réaction à l'équilibre thermodynamique et la moyenne du facteur de Boltzmann du travail dissipé par une force biaisage dynamique qui contraint la réaction à se produire [14]. Cette méthode permet d'obtenir un PMF à partir d'un ensemble de trajectoires hors-équilibre simulées indépendamment plutôt qu'à partir d'une séquence de simulations à l'équilibre le long d'une trajectoire comparable.

Bibliographie

[1] William H. Press, Saul A. Teukolsky, William T. Vetterling et Brian P. Flannery. Numerical Recipes in C : The Art of Scientific Computing. Cambridge University Press, New York, Second édition, 1992.

[2] Benoît Roux. The calculation of the potential of mean force using computer simulations. *Comp. Phys. Comm.*, 91(1–3) : 275–282, 1995.

[3] G. M. Torrie et J. P. Valleau. Non-physical sampling distributions in Monte-Carlo free-energy estimation – Umbrella sampling. *J. Comp. Phys.*, 23(2) : 187–199, 1977.

[4] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman et J. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13 : 1011–1021, 1992.

[5] Marc Souaille et Benoît Roux. Extension to the weighted histogram analysis method : combining umbrella sampling with free energy calculations. *Comp. Phys. Comm.*, 135(1) : 40–57, 2001.

[6] Alan M. Ferrenberg et Robert H. Swendsen. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, 63 : 1195–1198, 1989.

[7] Erik M. Boczko et Charles L. Brooks, III. Constant-Temperature Free Energy Surfaces for Physical and Chemical Processes. *J. Phys. Chem.*, 97(17) : 4509–4513, 1993.

[8] Christian Bartels et Martin Karplus. Multidimensional Adaptive Umbrella Sampling : Applications to Main Chain and Side Chain Peptide Conformations. *J. Comp. Chem.*, 18(12) : 1450–1462, 1997.

[9] Alessandro Laio et Michele Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci.*, 99(20) : 12562–12566, 2002.

[10] Marcella Iannuzzi, Alessandro Laio et Michele Parrinello. Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics. *Phys. Rev. Lett.*, 90(23) : 238302, 2003.

[11] Gerhard Hummer et Atilla Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci.*, 98(7) : 3658–3661, 2001.

[12] D. A. Hendrix et C. Jarzynski. A "fast growth" method of computing free energy differences. *J. Chem. Phys.*, 114(14) : 5974–5981, 2001.

[13] Sanghyun Park et Klaus Schulten. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.*, 120(13) : 5946–5961, 2004.

[14] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.*, 78(14) : 2690–2693, 1997.

Chapitre 8

Conclusion

8.1 Rappel des résultats principaux

Parce que la fonction potentielle polarisable est destinée à être utilisée pour simuler des systèmes de plusieurs dizaines de milliers d'atomes, il était nécessaire dans un premier temps de développer un algorithme de simulation plus rapide que celui couramment utilisé jusque-là. Il était en effet impensable de recalculer l'induction mutuelle des dipôles à chaque instant de la dynamique à l'aide d'une méthode itérative. La méthode dynamique que nous avons développée au chapitre 2, qui constitue une application nouvelle des simulations de Lagrangien étendu, permet de reproduire précisément le régime dynamique de la méthode itérative avec un ordre de grandeur moins d'opérations arithmétiques.

Cet algorithme a été indispensable pour réaliser le grand nombre de simulations qui ont mené à la calibration d'un nouveau modèle pour l'eau liquide (chapitre 4). Par l'étude systématique des paramètres du modèle, nous avons pu prouver que ce modèle était optimal vis-à-vis des propriétés physiques que nous avions choisi de reproduire. À notre connaissance, cette approche de paramétrisation empirique n'avait jamais été tentée pour un modèle moléculaire. Elle a notamment mis en évidence le fait que, pour que le modèle reproduise correctement les propriétés diélectriques du liquide, la polarisabilité moléculaire de l'eau devait être inférieure

à sa valeur en phase gazeuse. Cette idée a tout récemment été confirmée sous un angle différent par Giese et York [1].

L'étude des ions alcalins et halogènes a montré que le modèle polarisable proposé pouvait modéliser à la fois des propriétés microscopiques telles l'énergie d'interaction avec une unique molécule d'eau et des propriétés collectives telles la structure de solvatation. Malgré sa simplicité, le modèle ionique s'est avéré compatible avec de nombreuses études de dynamique *ab initio*. Et *grâce* à sa simplicité, nous avons pu caractériser avec précision des processus d'échange relativement lents.

C'est l'étude de l'hydratation hydrophobe de l'éthanol (chapitre 5) qui a le plus directement révélé l'utilité d'un modèle polarisable. Bien que les modèles d'eau et d'éthanol aient été calibrés séparément (pour reproduire les propriétés du liquide pur), le modèle polarisable crée une réponse variant avec la composition du milieu (voir la figure 5.10), et reproduit fidèlement les propriétés diélectrique des mélanges eau–éthanol. Les simulations ont permis de dégager de nouvelles propositions quant l'origine des propriétés de mélange anormales des solutions d'alcool et d'eau, et quand à la structure de l'eau autour de groupes hydrophobes.

8.2 Remise en contexte

Les quelques systèmes physico-chimiques étudiés dans cette thèse se voulaient représentatifs d'une plus vaste classe de molécules. L'eau a servi de prototype de molécule formant des liaisons hydrogène et a permis d'étudier des questions fondamentales concernant la modélisation électrostatique des molécules en phase condensée. L'éthanol a servi de prototype de composé à la fois polaire (pouvant former des liaisons hydrogène) et non-polaire, et son interaction avec l'eau a montré comment pouvait se restructurer le réseau naturel de liaisons hydrogène de l'eau autour d'un groupe hydrophobe. Les ions, quant à eux, ont servi à tester la force relative des interactions eau–eau et ion–eau. Nous n'avons pas étudié de composés non-polaires purs (comme des alcanes), qui auraient pu servir de modèles pour les membranes lipidiques. Il est attendu qu'un modèle polarisable de

lipides aura un effet déterminant sur les propriétés diélectriques des membranes, puisque, les chaînes carbonées ne possèdant pas de dipôles permanents, la seule réponse diélectrique des molécules vient de leur polarisabilité électronique. Malgré que cette réponse ne soit absolument pas décrite par un modèle non-polarisable, des modèles empiriques de lipides non-polarisables existent depuis plusieurs années et sont utilisés avec succès—quoique des calculs d'énergie libre récents montrent que des lipides non-polarisables ont un effet déstabilisant sur la perméation du potassium au travers de la gramicidine A [2]. Il serait donc intéressant de voir l'influence d'une membrane explicitement polarisable sur le fonctionnement et la structure des protéines transmembranaires.

Les résultats concernant la coordinence de l'ion potassium dans l'eau présentent une signification particulière pour les canaux potassiques. En effet, MacKinnon *et al.* ont récemment émis l'idée que la symétrie à quatre branches du canal KcsA [3] était adaptée à la coordinence naturelle de l'ion dans l'eau, et que le filtre de sélectivité du KcsA constituait un passage naturel pour l'ion. Si l'on en croit les simulations réalisées avec le champ de force polarisable, il semblerait que ce soit plutôt l'inverse : que l'ion potassium possède une structure de solvatation suffisamment flexible pour accommoder la variété de types de coordinence qu'il rencontre le long de son passage au travers du KcsA. Le fait qu'un ion potassium puisse traverser le filtre de sélectivité du KcsA sans rencontrer de véritable barrière d'énergie provient d'un équilibre des forces plus subtil, impliquant l'interaction avec des ions voisins [4], de même que la répulsion mutuelle des groupes carbonyle [5].

Plusieurs avenues sont envisageables pour améliorer le réalisme du modèle d'eau. Des simulations à des températures voisines de la température de la pièce nous ont montré que le modèle SWM4-DP avait un coefficient d'expansion thermique trop élevé. Un modèle disposant de plus de paramètres libres pourrait viser à reproduire cette quantité exactement, et ainsi améliorer le domaine de validité du modèle. Compte tenu de la variété de phases de l'eau, cette entreprise est un programme de recherche en soi. À notre avis, l'amélioration la plus pressante est

celle mentionnée au chapitre 6, concernant la robustesse de la représentation des interactions moléculaires ne constituant pas de véritables liaisons hydrogène.

Un immense travail reste à faire pour pouvoir disposer d'un champ de force polarisable « complet » (pour les protéines, les acides nucléiques et les membranes lipidiques). L'annexe A présente un aperçu du travail entrepris pour les acides nucléiques. Le même effort doit être fourni pour les acides aminés et les phospholipides. Une attention particulière devra être portée à la calibration des groupes chimiques susceptibles de coordonner des ions. Le travail présenté dans cette thèse laisse présager qu'il est possible de formuler une représentation unifiée de la mécanique moléculaire entre des ions, des groupes polaires et des groupes non-polaires.

Bibliographie

[1] T. J. Giese et D. M. York. Many-body force field models based solely on pairwise Coulomb screening do not simultaneously reproduce correct gas-phase and condensed-phase polarizability limits. *J. Chem. Phys.*, 120(21) : 9903–9906, 2004.

[2] Toby W. Allen, Olaf S. Andersen et Benoît Roux. Energetics of ion conduction trough the gramicidin channel. *Proc. Natl. Acad. Sci.*, 101(1) : 117–122, 2004.

[3] Declan A. Doyle, João Morais Cabral, Richard A. Pfuetzner, Anling Kuo, Jacqueline M. Gulbis, Steven L. Cohen, Brian T. Chait et Roderick MacKinnon. The Structure of the Potassium Channel : Molecular Basis of K^+ Conduction and Selectivity. *Science*, 280(5360) : 69–77, 1998.

[4] Simon Bernèche et Benoît Roux. Energetics of ion conduction trough the K^+ channel. *Nature*, 414(6859) : 73–77, 2001.

[5] Sergey Yu. Noskov, Simon Bernèche et Benoît Roux. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature*, 431(7010) : 830–834, 2004.

Annexe A

Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator

Tiré de :
Victor M. Anisimov, Guillaume Lamoureux, Igor V. Vorobyov, Niu Huang, Benoît Roux, Alexander D. MacKerell, Jr. *J. Chem. Theory Comput.*, 1(1) : 153–168, 2005.

Abstract

A procedure to determine the electrostatic parameters has been developed for a polarizable empirical force field based on the classical Drude oscillator model. Atomic charges and polarizabilities for a given molecule of interest were derived from restrained fitting to quantum-mechanical electrostatic potentials (ESP) calculated at the B3LYP/cc-pVDZ or B3LYP/aug-cc-pVDZ levels on grid points located on concentric Connolly surfaces. The determination of the atomic polarizabilities requires a series of perturbed ESP maps, each one representing the electronic response of the molecule in the presence of a background charge placed on Connolly surfaces primarily along chemical bonds and lone pairs. Reference values for the partial atomic charges were taken from the CHARMM27 additive all-atom force field and those for the polarizabilities were based on adjusted Miller's ahp atomic polariz-

ability values. The fitted values of atomic polarizabilities were scaled to reflect the reduced polarization expected for the condensed media and/or to correct for the systematic underestimation of experimental molecular polarizabilities by B3LYP calculations. Following correction of the polarizabilities, the atomic charges were adjusted to reproduce gas-phase dipole moments. The developed scheme has been tested on a set of small molecules representing functional moieties of nucleic acids. The derived electrostatic parameters have been successfully applied in a preliminary polarizable molecular dynamics simulation of a DNA octamer in a box of water with sodium counterions. Thus, this study confirms the feasibility of the use of a polarizable force field based on a classical Drude model for simulations of biomolecules in the condensed phase.

A.1 Introduction

Computer simulations based on empirical force fields are now a standard procedure to investigate biological phenomena [1]. Empirical force field calculations, due to the simplicity of the potential energy function, allow for atomic detail studies of biomolecules with explicit representation of the condensed phase environment to be performed. However, it is essential that the force field accurately reproduces the experimental regimen to insure the quality of results of such calculations.

The majority of force fields consist of electrostatic, van der Waals and bonding energy terms calculated in a pairwise additive fashion [2]. The induced polarization, which arises from a perturbation of the electronic structure of the molecular species in response to the external electric field, is typically incorporated implicitly by using enhanced fixed partial atomic charges reflecting the average polarization taking place in the condensed phase. Despite the apparent simplification, additive empirical force fields have been remarkably successful in modeling complex molecular systems for the last two decades [3]. However, there are shortcomings in the additive model [4], emphasizing the need to account for many-body induced

polarization effects in an explicit way, motivating the development of polarizable force fields.

Current polarizable models can be classified into three major categories: point dipole, charge transfer, and classical Drude oscillator (or shell models); information about each of these models can be found elsewhere [4, 5]. While a variety of efforts are ongoing to apply the point dipole and charge transfer methods to biological systems [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27], the classical Drude oscillator approach has only seen minimal attention [28, 29, 30]. The Drude oscillator polarizability model was first introduced by Paul Drude in 1900 as a simple approach to describe the dispersive properties of materials [31]. The classical version of this model has been successfully used in statistical mechanical studies of condensed systems to treat electronic polarizability [5] and has been recently implemented [32] into the CHARMM program [33, 34].

Creation of a force field for biomolecular systems traditionally starts with the development of a water model. The polarizable SWM4-DP water model based on the classical Drude oscillator formalism has recently been presented [30]. It was parameterized to reproduce properties of liquid water under ambient conditions as well as some gas-phase properties such as the dipole moment and the interaction energy of the water dimer. In the SWM4-DP water model the polarizability of the oxygen atom, which in this case is equivalent to the molecular polarizability (since no Drude particles for hydrogen atoms were considered) was found to be 1.04 $\AA^3$, which is 0.724 of the experimental gas phase molecular polarizability of water, 1.44 $\AA^3$ [35]. Such reduced polarizability, which is essential to reproduce liquid-phase properties of water, including the dielectric constant, has been attributed to the energy cost of overlapping electron clouds in the condensed phase opposing induction [30, 36, 37, 38, 39, 40, 41].

In the present work, steps towards systematic development of a polarizable force field based on the classical Drude oscillator model are presented. The determination of the electrostatic parameters is considered as the first step towards this goal. This effort, which includes fitting of partial atomic charges and atomic

polarizabilities to a series of electrostatic potentials (ESP) around a molecule, each in the presence of an individual background charge, is the main focus of the present study. The theoretical background for the classical Drude oscillator model and the methodological details of the electrostatic parameter fitting are discussed in the next section. Validation of the theoretical level used in the derivation procedure is presented followed by a discussion of the selection of the reference values for charges and polarizabilities. Then, an example is given for the application of that procedure to the cytosine base, a model compound for nucleic acids. Finally, the validity of the proposed approach is illustrated via a condensed phase simulation of a DNA octamer using a preliminary classical Drude polarizable force field.

A.2 Theory and methods

A.2.1 Classical Drude oscillator model

According to the classical Drude oscillator model, the polarizability is introduced by adding massless charged particles attached to each polarizable atom by a harmonic spring (Figure A.1). Thus, a finite induced dipole is created and the partial atomic charge of atom A, $q(A)$, is redistributed between the Drude particle, $q_{\mathrm{D}}(A)$, and the atomic core, $q_{\mathrm{c}}(A)$. The positions of the Drude particles relative to the corresponding atomic centers are determined self-consistently by seeking the minimum energy consistent with the Born-Oppenheimer approximation. For the equilibrium position of the Drude particles, the atomic polarizability of atom A, $\alpha(A)$, is related to the charge $q_{\mathrm{D}}(A)$ via the equation

$$\alpha(A) = \frac{q_{\mathrm{D}}^2(A)}{k_{\mathrm{D}}}, \tag{A.1}$$

where k_{D} is the force constant of the harmonic spring connecting a Drude particle to its corresponding atomic core. The magnitude of k_{D} is chosen to achieve small displacements of Drude particles from their corresponding atomic positions, $\mathbf{r}_{\mathrm{D}}$, as required to remain close to the point-dipole approximation for the induced dipole

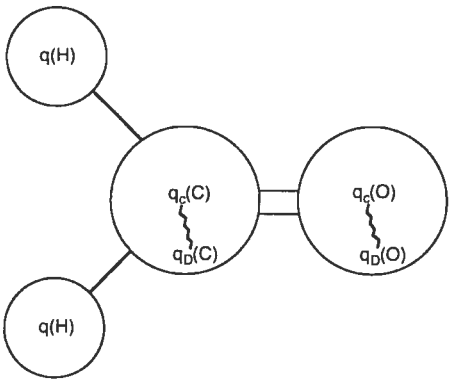

Figure A.1: Classical Drude oscillator model on the example of a formaldehyde molecule. The displacements of Drude particles attached to the atomic centers are exaggerated and correspondd to the positive values of Drude charges for the visualization clarity.

associated with the atom–Drude pair [32]. Consequently, the atomic polarizability is determined by the amount of charge assigned to the Drude particle. Thus, the only adjustable parameters to be determined during parameterization of a polarizable atom A in the Drude model are partial atomic charges $q(A) = q_c(A) + q_D(A)$. The development of a consistent protocol for the determination of $q_c(A)$ and $q_D(A)$ (or $\alpha(A)$) for a series of model compounds is, in part, the subject of the present study.

One of the major advantages of the classical Drude oscillator model is that it preserves the simple functional form of the pairwise additive force field and yet explicitly accounts for the electronic polarizability [32]. The electrostatic energy from the additive force field is substituted in the polarizable model by the Coulombic energy terms describing interactions between atomic cores and Drude particles, and the self-energy of a polarizable atom treated via a harmonic term:

$$E_{\text{Drude}} = \sum_{A<B}^{N} \frac{q_D(A)q_c(B)}{|\mathbf{r}_D(A) - \mathbf{r}(B)|} + \sum_{A<B}^{N} \frac{q_D(A)q_D(B)}{|\mathbf{r}_D(A) - \mathbf{r}_D(B)|} + \frac{1}{2}\sum_{A}^{N} k_D|\mathbf{r}_D(A) - \mathbf{r}(A)|^2, \tag{A.2}$$

where N and N_D are the number of real atoms and Drude particles, respectively, q_c and q_D are the atomic core and Drude particle charges, respectively, and $\mathbf{r}$ and $\mathbf{r}_D$ are the positions of the real atoms and Drude particles, respectively. Modifying

the classical force field energy function using equation (A.2) facilitates performing dynamical simulations with only minor modifications to existing programs, because the original energy functional form remains essentially unchanged. Computationally this means that the Drude particle positions for a given atomic configuration have to be self-consistently adjusted for each step of the dynamics simulation; however such calculations are inefficient and rather expensive [32]. Therefore a molecular dynamics simulation algorithm based on the extended Lagrangian formalism [42, 43] has been implemented, in which a small mass is attributed to the Drude particles, and the amplitude of the oscillations is controlled with a low temperature thermostat [32]. This technique allows the computationally expensive self-consistent field (SCF) regimen of molecular dynamics simulations to be avoided. Tests of the extended Lagrangian algorithm have shown that stable and accurate molecular dynamics trajectories can be generated, yielding liquid properties equivalent to the SCF regimen of molecular dynamics at a fraction of the computational cost [32]. Therefore, the classical Drude oscillator model for simulating atomic polarizability can be applied for molecular modeling in condensed media including macromolecular systems such as fully solvated nucleic acids, proteins and lipid aggregates. However, the existing parameters from the non-polarizable force fields need to be adjusted to take into account the presence of Drude particles.

A.2.2 Charge fitting scheme

The electrostatic properties of a molecular mechanics model with Drude polarizabilities are represented by atomic core charges $q_{\mathrm{c}}(A)$ and Drude charges $q_{\mathrm{D}}(A)$ producing effective atomic charge $q(A)$ as their sum. Indeed, in the classical Drude oscillator polarizable model, the determination of atomic polarizabilities $\alpha(A)$ can be reduced to the determination of the partial charge of Drude particles, $q_{\mathrm{D}}(A)$. Both $q(A)$ and $q_{\mathrm{D}}(A)$ can be determined simultaneously in a single fitting step.

Partial atomic charges are often obtained by optimizing the fit of an electrostatic potential ϕ^{MM} derived for the molecular mechanics (MM) model to a potential map ϕ^{QM} generated by quantum-mechanical (QM) calculations on a set of

grid points $\{\mathbf{r}_g\}$ placed around the molecule. Although partial atomic charges of a nonpolarizable model can be extracted from a single potential map, adjusting the polarizabilities requires a series of response potential maps ϕ_p^{QM}, each one representing the altered charge distribution for the molecule in the presence of a small perturbing point charge z_p at a given position $\mathbf{r}_p$. A similar approach was used by Friesner *et al.* to derive parameters for the fluctuating charge and polarizable dipole models [17, 37, 44]. In our calculations the value of the perturbing charge was arbitrarily chosen to be +0.5 e. The molecular mechanics potential at the g-th grid point (at the coordinate $\mathbf{r}_g$) for the molecule under the influence of a point-charge perturbation at position $\mathbf{r}_p$ is

$$\phi_{pg}^{\mathrm{MM}} = \sum_{A}^{N} \left(\frac{q_{\mathrm{c}}(A)}{|\mathbf{r}(A) - \mathbf{r}_g|} + \frac{q_{\mathrm{D}}(A)}{|\mathbf{r}(A) + \mathbf{d}_p(A) - \mathbf{r}_g|} \right) + \frac{z_p}{|\mathbf{r}_p - \mathbf{r}_g|}, \tag{A.3}$$

where $\mathbf{d}_p(A)$ is the Drude particle displacement from the corresponding atomic center position $\mathbf{r}(A)$ in response to the perturbation p. The last term is the contribution from the perturbation charge itself.

During the fitting procedure, all core atomic and Drude charges have to be adjusted to minimize the discrepancy between the QM and MM potential maps, i.e., to minimize the following function:

$$\chi_\phi^2[q_{\mathrm{c}}, q_{\mathrm{D}}] = \sum_{p,g} \left(\phi_{pg}^{\mathrm{QM}} - \phi_{pg}^{\mathrm{MM}} \right)^2. \tag{A.4}$$

The function ϕ_{pg}^{MM} has three unknown parameters: q_{c}, q_{D}, and $\mathbf{d}_p$. The first two are the subject of the standard least-squares fitting procedure, but the Drude particle displacement, $\mathbf{d}_p$, requires special consideration. Because of the implicit charge-dependence of the displacements $\mathbf{d}_p(A)$, the system of equations

$$\frac{\partial \chi_\phi^2}{\partial q(A)} = 0, \tag{A.5}$$

where $q(A)$ designates either q_{c} or q_{D} assigned to an atom A, has to be solved

iteratively. We use the Levenberg–Marquardt algorithm [45], specially designed to minimize χ^2 functions (see below). First, Drude displacements, $\mathbf{d}_p$, are optimized to minimize total energy of the molecular system using atomic charges from the initial guess. This is followed by an ESP fitting step using the current positions of the atoms and Drude particles. The new set of fitted charges is again used to optimize the coordinates of the Drude particles. The iterative procedure is continued until equation (A.5) is satisfied.

Because the charge fitting problem is underdetermined, directly solving equation (A.5) usually leads to partial charges having poor chemical significance [46]. This is mainly due to the low contribution of some charges to the overall electrostatic potential associated with the screening of the charge on buried atoms by atoms located on the periphery of the molecule. In order to optimize individual charge contribution to the minimization function, it is necessary to penalize charge deviations from "chemically intuitive" reference values, as long as the penalty does not significantly deteriorate the quality of the fit. This requirement motivated the inclusion of restraints during charge fitting, referred to as restrained electrostatic potential (RESP) fitting. The original RESP scheme of Bayly *et al.* [46] minimizes

$$\chi^2 = \chi^2_\phi + \chi^2_\mathrm{r} \tag{A.6}$$

through adding a penalty term in one of the two following forms

$$\chi^2_\mathrm{r} = w \sum_A^N \left[q(A) - \bar{q}(A)\right]^2, \tag{A.7}$$

$$\chi^2_\mathrm{r} = w \sum_A^N \left[\sqrt{q^2(A) + b^2} - b\right]^2, \tag{A.8}$$

where w is a weighting constant. The first restraint (A.7) is forcing the charges $q(A)$ to their "reference" values $\bar{q}(A)$, and the second restraint (A.8) favors smaller magnitude charges, where b is a hyperbolic stiffness parameter.

The RESP scheme can be generalized for the presence of Drude particles, yield-

ing the following parabolic (A.9) and hyperbolic (A.10) equations, respectively:

$$\chi^2_{\mathrm{r}} = \sum_{A}^{N} \left\{ w \left[q(A) - \bar{q}(A) \right]^2 + w_{\mathrm{D}} \left[q_{\mathrm{D}}(A) - \bar{q}_{\mathrm{D}}(A) \right]^2 \right\}, \tag{A.9}$$

$$\chi^2_{\mathrm{r}} = \sum_{A}^{N} \left\{ w \left[\sqrt{q^2(A) + b^2} - b \right]^2 + w_{\mathrm{D}} \left[\sqrt{q^2_{\mathrm{D}}(A) + b^2_{\mathrm{D}}} - b_{\mathrm{D}} \right]^2 \right\}, \tag{A.10}$$

where w and w_{D} are weighting factors for real atoms and Drude particles, respectively, b and b_{D} are the respective stiffness constants, $q(A)$ is the atomic charge representing the sum of atomic core and Drude particle charges, and $\bar{q}(A)$ is the reference charge. Due to the Drude-charge–atomic-polarizability formal equality (A.1) postulated by the classical Drude oscillator model, equations (A.9) and (A.10) effectively lead to restraining the atomic polarizabilities.

To allow for additional flexibility of the fitted charges and polarizabilities, flat well potentials can be introduced into the fitting procedure. The parabolic restraint can be used only on the amount of charge deviating from $\bar{q}$ by more than a fixed charge tolerance q_{flat}. This allows the charge q to vary at no cost within the interval $[\bar{q} - q_{\mathrm{flat}}, \bar{q} + q_{\mathrm{flat}}]$, and creates a restraint only when the deviation is larger than q_{flat}. The charge fitting algorithm outlined above is implemented in the module FITCHARGE in the latest release of the CHARMM program [33, 34].

A.2.3 Grid generation and placement of perturbation charges

Electrostatic charge fitting procedures traditionally use QM electrostatic potentials determined on a cube based grid, with the grid points placed at an equidistant separation from each other. A limitation of this approach is the non-unique definition for the selection of the axes for the cube and lack of control over significance of each grid point. One approach to eliminate the orientation dependence is to use random point generation within the defined cube. However, this solution is not ideal due to reproducibility issues. In addition, for computational efficiency it

seems reasonable to avoid placing grid points in regions having minimal chemical relevance. This is especially important when the molecular shape is significantly different from a spherical form. The well-known surface reflecting molecular shape, the Connolly surface [47], is ideal for hosting the grid points placement. Such a structurally-aware grid reduces the number of required grid points, facilitating the least-squares fitting by reducing the number of linear equations to solve, as well as maximizing the information content in selected regions around a molecule with the most chemical significance. The methodology is expected to be particularly advantageous for large molecules.

In practice, a Connolly surface is generated by overlapping atomic spheres and preserving only those regions of the spheres located on a periphery of a molecule. Surface building is based on using the van der Waals radii for the atomic spheres. Multiplication of all the atomic radii of the molecule by a constant creates a Connolly surface at the desired distance from the atomic centers. Choosing several different multiplication constants allows for creation of a set of non-intersecting Connolly surfaces. Similar considerations, as discussed above regarding the grid point placement, also apply to the placement of perturbation charges around the molecule. Multiple Connolly surfaces carrying the perturbation charges are required to probe molecular polarizability at different distances from the atomic centers. The chosen solution provides a simple mechanism to control the locations and number of point charges to place.

The simplest technique for point placement on a particular Connolly surface is to generate equidistant points on the atomic spheres at a selected density and to delete all the points which are within the overlapping spheres. The algorithm for placing grid points on the atomic sphere moves an atom to the center of the coordinate system, assigns the atom a unit radius and generates points according the sphere equation $x^2 + y^2 + z^2 = 1$. Next, the atomic sphere is scaled to the desired radius and coordinates of the grid points are translated accordingly. Then a back translation of the atomic center to its original position is performed with simultaneous translation of coordinates of the generated grid points. These steps

are applied to each atom of the molecule and grid points in overlapping atomic spheres are deleted. This creates the final set of points situated on the Connolly surface.

By changing the number of grid points on a particular Connolly surface one can increase or decrease the contribution of that surface to the fitted molecular properties. Typically, large contributions from the nearest and most distant Connolly surfaces should be avoided; the nearest surface is approaching distances where the deviation from the atomic point-charge approximation of the electronic distribution is still non-negligible, while the contribution of the most distant surface may shift the accuracy of charge fitting to larger distances than those at which hydrogen bonding occurs. The following rationale may be used to identify the minimal number of Connolly surfaces for the placement of perturbation charges and grid points. The first layer of perturbation charges may be placed at distances typical for hydrogen bonds followed by a layer of grid points to detect an immediate ESP change caused by a perturbation charge. Several layers of perturbation charges are needed to adequately capture the orientational dependence of molecular polarizability. Therefore at least one more layer of perturbation charges is necessary followed by one more layer of grid points. Finally one additional distant layer of grid points is necessary to resolve the molecular dipole moment, which is a far-field molecular property. These operations yield five non-intersecting Connolly surfaces in total: two for perturbation charges and three for grid points, illustrated as 1) charges, 2) grid, 3) charges, 4) grid, and 5) grid, according to increasing distance from the atomic centers. This number can be expanded as required based on computational tests.

The charge placement method employs a set of additional rules to the Connolly surface construction that further reduces the number of the charges. These additional rules are designed to assign perturbation charges to places of chemical significance, i.e., along covalently connected atoms and lone pairs where polarizability is expected to be largest. The perturbation points are prioritized into three groups according to the order of their generation. Charges placed along chemical

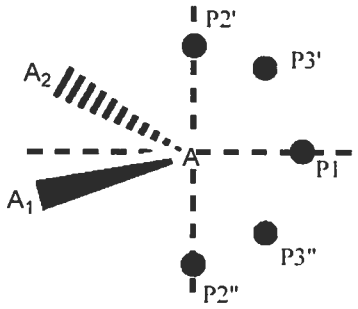

Figure A.2: Orientation of the placement of perturbation charges (P) for polarizability probing the region around lone pairs of sp^3 hybridized oxygen or sp^2 hybridized nitrogen atoms (A).

bonds are first generated. If two atoms, A and B, are linked by a covalent bond, two perturbation points along the line A–B are generated that intersect with the corresponding Connolly surfaces on opposite sides of the bond. A newly generated charge will not be saved if it is too close to a previously generated charge. The distance criteria imposed in this study is 1.5 Å. The second group of perturbation points is created to sample regions around lone pairs on sp^3 hybridized oxygen or sp^2 hybridized nitrogen atoms, A, when these atoms have just two covalently bound neighbors (A1 and A2). In this case a bisector line is drawn in the plane of the covalent bonds of atom A, dividing the valence angle (A1–A–A2) in half (Figure A.2). On the side of the lone pairs a perturbation point, P1, is generated along this line where it intersects with the Connolly surface. Two more points, P2′ and P2″, are generated on the line through point A, perpendicular to the plane defined by atoms A1, A, and A2 on the two opposite sides of the plane. Two other points, P3′ and P3″, are generated on the bisector line of angles P1–A–P2′ and P1–A–P2″. All these points are placed on the intersection of the Connolly surface with the corresponding line.

The last group of charges is created to close gaps between the previously placed perturbation points providing nearly equivalent coverage of the molecular shape by perturbation charges. The generation of the grids and the placement of pertur-

bation charges are performed using the stand-alone program CGRID developed in our laboratory.

A.2.4 Molecular dipole and polarizability

The dipole moment and polarizability are important molecular properties which may be used to validate the quality of derived electrostatic parameters through comparison with experimental and/or QM values. To calculate molecular dipole moments, $\boldsymbol{\mu}$, positions of Drude particles must be optimized self-consistently. Then $\boldsymbol{\mu}$ can be calculated from the sum over all charges using the following equation:

$$\boldsymbol{\mu} = \sum_{A}^{N} q_{\mathrm{c}}(A)\mathbf{r}(A) + \sum_{B}^{N_{\mathrm{D}}} q_{\mathrm{D}}(B)\mathbf{r}_{\mathrm{D}}(B), \tag{A.11}$$

where N and N_{D} are the number of real atoms and Drude particles, respectively; $q_{\mathrm{c},i}$ and $q_{\mathrm{D},j}$ are the atomic core and Drude particle charges, respectively; and $\mathbf{r}_i$ and $\mathbf{r}_{\mathrm{D},j}$ are the positions of the real atoms and Drude particles, respectively.

Polarizability is a measure of the response of a molecular system to an external electric field. Experimental determination of isotropic molecular polarizability is conventionally conducted by measuring the refractive index η [48]

$$\alpha = \frac{3}{4\pi n}\frac{\eta^2 - 1}{\eta^2 + 2}, \tag{A.12}$$

where n is the number density of the gas or liquid. In the present MM picture, the fast linear response of the electronic density to the excitation of a beam of light is modelled by the quasi-instantaneous readjustement of the Drude particles. The total molecular polarizability $\boldsymbol{\alpha}$ of the MM model can be calculated in analogy with its standard QM definition, as the sum over all excitation modes of the square of the response dipole divided by the excitation energy. Being a linear-response coefficient, it can be summed over the vibration modes ν of the molecule obtained from a normal mode analysis of the polarizable MM model, complete with atomic cores and Drude particles. In terms of the cartesian normal mode vectors $\mathbf{A}_\nu$, the

components of the tensor $\boldsymbol{\alpha}$ are

$$\alpha_{ij} = \sum_{\nu > \nu_0} \frac{\partial \mu_i}{\partial \mathbf{A}_\nu} \cdot \frac{\partial \mu_j}{\partial \mathbf{A}_\nu} \left(\frac{\partial^2 U}{\partial A_\nu^2} \right)^{-1}, \tag{A.13}$$

where μ_i are components of the molecular dipole moment and U is the potential energy of the model. The isotropic polarizability values α_{iso} can be calculated as a trace of the molecular polarizability tensor, i.e., $\alpha_{\text{iso}} = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$. This formula was implemented in the VIBRAN module of the CHARMM program. Only the very-high frequency normal modes, attributed to Drude particle excitations, are summed, since the polarizability is modeled through the movement of these auxiliary particles. Given the much lighter mass of the Drude particles, a frequency cutoff at $\nu_0 = 5000$ cm^{-1} ensures that the normal modes associated with the vibrations of nuclei are excluded. The zero-frequency rotation and translation modes are ignored as well. It should be noted that the calculation of molecular polarizabilities requires full optimization of the molecular structure, since the harmonic approximation used in the vibrational analysis is valid only for equilibrium geometries.

A.2.5 Computational details

The determination of electrostatic parameters for the polarizable force field is a multi-step iterative procedure and employs both QM and empirical force field calculations. QM calculations were performed using the Gaussian 98 suite of programs [49]. Geometry optimizations were performed at the MP2(fc)/6-31G(d) level of theory for neutral species and at the MP2(fc)/6-31+G(d,p) level for ions [50, 51, 52, 53]. This level of theory provides molecular geometries consistent with available gas phase experimental data and it has been previously utilized during optimization of the CHARMM27 all-atom empirical force field for nucleic acids [54]. Geometry optimization of adenine, cytosine, and guanine bases were conducted with the amino group planarity enforced, since it is believed that amino group will acquire an approximately planar geometry due to hydrogen bonding

with complementary bases and solvent molecules. No constraints were imposed during optimizations of other test molecules.

Selection of the appropriate level of theory for determination of the ESPs required for calculation of the charges and polarizabilities is important. Ideally, in order to assure high accuracy in the ESP, a highly correlated method in conjunction with a very large basis set should be used, since computed dipole moments and molecular polarizabilities strongly depend on the level of theory and size of the basis set employed. However, when developing a force field for biomolecular systems it is necessary to have a large number of model compounds in the training set, with many of those compounds being relatively large, i.e., greater than 20 non-hydrogen atoms. All these requirements are unlikely to be satisfied in full and therefore a compromise between accuracy of the theoretical model and its computational performance is necessary.

QM calculations of the molecular electrostatic potentials were performed on MP2 optimized geometries using the B3LYP hybrid functional [55, 56, 57] and the correlation-consistent double-ζ Dunning cc-pVDZ and aug-cc-pVDZ basis sets [58]. Single-point energy B3LYP calculations were performed with a tight convergence criteria producing the target QM ESP maps. Cartesian coordinates for grid points and perturbation charges for the ESP calculation were generated by the program CGRID described above and read by Gaussian from external files. The generated QM ESP maps were extracted from the Gaussian output to use as input to the electrostatic parameter fitting by the FITCHARGE module of the CHARMM program. Restrained fitting using the RESP algorithm employed a penalty function with 10^{-5} $\mathring{A}^{-2}$ for the weighting factor. The flat well potential described above was applied allowing penalty-free deviation of charges by 0.1 e in both directions from the corresponding reference values. Coordinates of the Drude particles were self-consistently adjusted after each change of the optimized charges during the charge fitting procedure to minimize potential energy of the system whereas coordinates of real atoms were fixed to the corresponding MP2 geometry.

Following the fitting procedure the atomic charges and polarizabilities were

scaled. Polarizability scaling is necessary to reflect the reduced polarization expected for the condensed media and/or to correct for systematic underestimation of the experimental values by B3LYP/cc-pVDZ calculations (see below). Atomic charge scaling for neutral compounds can be performed to reproduce experimental or high-level QM gas-phase target molecular dipole moments. In addition, rounding of fitted and scaled charges and polarizabities to three decimal places was performed to facilitate their transferability while preserving the value of the net molecular charge. Technical details of the scaling and rounding procedure are described in the Supporting Information. In this study, similarly to the SWM4-DP water model, the polarization of only "heavy" (i.e., non-hydrogen) atoms is considered, although the method can be easily extended to all atoms at the increase of the computational expense.

In all Drude polarizable CHARMM calculations, the Drude particles were attached to the real atoms via a harmonic spring with a force constant of a 500 kcal/mol/Å^2. This force constant is of sufficient magnitude to prevent large displacement of the Drude particle from its atom and thus, ensure the validity of the point dipole approximation [32]. In this scenario, the atomic polarizabilities unequivocally determine the magnitude of Drude charges from equation (A.1). In addition, the sign of the charges on Drude particles is irrelevant due to the point dipole approximation. We chose q_D to be negative by analogy with the electron charge.

Molecular dynamics (MD) simulations were performed at 300 K and 1 atm pressure using the new velocity Verlet integrator [32] implemented into CHARMM [33, 34]. The extended Lagrangian double-thermostat formalism [32] was used in all polarizable MD simulations where a mass of 0.1 amu was transferred from real atoms to the corresponding Drude particles. The amplitude of their oscillation was controlled with a separate low-temperature thermostat (at a $T = 1$ K) to ensure that their time course stays close to the SCF regimen [32]. A Nosé–Hoover thermostat with a relaxation time of 0.1 ps was applied to all real atoms to control the global temperature of the system. A modified Andersen–Hoover barostat with

a relaxation time of 0.1 ps was used to maintain the system at constant pressure. Condensed-phase MD simulations were performed using periodic boundary conditions and SHAKE to constrain covalent bonds involving hydrogens [59]. A 1 fs time step was used for extended Lagrangian polarizable MD simulation. The electrostatic interactions were treated using particle-mesh Ewald (PME) summation [60] with a coupling parameter 0.34 and sixth order spline for mesh interpolation. Nonbond pair lists were maintained out to 14 Å, and a real space cutoff of 12 Å was also used for the Lennard-Jones parameters within the atom-based force switch algorithm [61]. The long-range van der Waals correction recently implemented into CHARMM program was also applied [62, 63].

Validation of the electrostatic parameters optimization method, along with proof of concept that the classical Drude oscillator model is applicable to biomolecular condensed phase simulations was obtained via a MD simulation of DNA in solution. The pre-equilibrated GAGTACTC duplex DNA structure solvated in a box of water with sodium ions was taken from our previous study [64]. 1746 water molecules and 14 sodium ions were used. The solvated molecular system contains 9586 atoms including Drude particles. Starting from the CHARMM27 additive force field equilibrated system [64], the Drude particles and then the solvent molecules were minimized for 200 steps using the steepest descent (SD) algorithm with all DNA real atoms fixed. Then the minimized structure was subjected to a 20 ps NPT MD simulation with all DNA atoms, excluding the Drude particles, harmonically constrained with a mass-weighted force constant of 2 kcal/mol/Å to equilibrate the solvent around the DNA. The final structure from that simulation was then subjected to two 200 step SD minimizations, first for Drude particles and then for all atoms prior to initialization of the production trajectory. The simulation was run for 1000 ps with coordinates saved every 2 ps for analysis.

A.3 Results and Discussion

A.3.1 QM calculations of molecular dipoles and polarizabilities

The performance of B3LYP/cc-pVDZ and B3LYP/aug-cc-pVDZ calculations on MP2 optimized geometries in reproducing experimental gas-phase dipole moments and polarizabilities was first verified. Model compounds for which experimental values of molecular polarizabilities and/or dipole moments are available were chosen to represent different chemical classes as well as building blocks of biomolecules. The experimental and B3LYP calculated dipole moments are summarized in Table A.I. The data indicate that both B3LYP/cc-pVDZ and B3LYP/aug-cc-pVDZ single-point energy calculations generally provide reasonable estimates of the dipole moment magnitudes. In many cases some underestimation of the experimental values can be noted for the B3LYP/cc-pVDZ calculations, whereas augmenting this basis set with diffuse functions in general results in a slight overestimation of the gas-phase molecular dipole moments. The average ratio of calculated to experimental dipole moments for compounds listed in Table A.I, excluding water, was 0.94 and 1.05 for the cc-pVDZ and aug-cc-pVDZ basis sets, respectively.

Experimental and calculated molecular polarizabilities for selected compounds are summarized in Table A.II. In all cases the experimental gas-phase molecular polarizabilities are underestimated by the B3LYP/cc-pVDZ calculations. The degree of the discrepancy is the largest for water (almost 50%) and substantially smaller for the remaining compounds. Importantly, for the majority of compounds the ratio of B3LYP to experimental polarizabilities is quite uniform, with an average value of 0.83 ± 0.06. As data in Table A.II demonstrate, the substantial deviation of calculated from experimental values of polarizabilities can be corrected by augmenting the cc-pVDZ basis set by diffuse functions, which provides better quantitative agreement with the experimental data. Thus, B3LYP/aug-cc-pVDZ QM calculations may be considered the method of choice for calculation of the ESP data for electrostatic parameter fitting. However, QM calculations

Table A.I: Summary of experimental and QM calculated dipole moments (in Debyes).

Molecule		Gas phase experimental	B3LYP cc-pVDZ	ratio	B3LYP aug-cc-pVDZ	ratio
water	H_2O	1.855 ± 0.004	1.910	1.03	1.854	1.00
propane	C_3H_8	0.084 ± 0.001	0.070	0.83	0.096	1.15
isobutane	C_4H_10	0.132 ± 0.002	0.110	0.83	0.147	1.12
pentene-1	C_5H_10	0.500*	0.400	0.80	0.447	0.89
toluene	C_7H_8	0.375 ± 0.01	0.380	1.01	0.405	1.08
fluoromethane	CH_3F	1.858 ± 0.002	1.757	0.95	1.873	1.01
fluorobenzene	C_6H_5F	1.600 ± 0.08	1.394	0.87	1.603	1.00
chlorobenzene	C_6H_5Cl	1.690 ± 0.03	1.667	0.99	1.749	1.03
methanol	CH_4O	1.700 ± 0.02	1.576	0.93	1.680	0.99
ethanol (trans)	C_2H_6O	1.440 ± 0.03	1.483	1.03	1.598	1.11
ethanol (gauche)	C_2H_6O	1.680 ± 0.03	1.538	0.92	1.727	1.03
dimethylether	C_2H_6O	1.300 ± 0.01	1.200	0.92	1.306	1.00
tetrahydrofuran	C_4H_8O	1.750 ± 0.04	1.695	0.97	1.865	1.07
trimethylamine	C_3H_9N	0.612 ± 0.003	0.450	0.74	0.591	0.97
dimethyl sulfide	C_2H_6S	1.554 ± 0.004	1.471	0.95	1.620	1.04
ethanethiol (trans)	C_2H_6S	1.580 ± 0.08	1.557	0.99	1.669	1.06
acetaldehyde	C_2H_4O	2.750 ± 0.006	2.600	0.95	2.965	1.08
acetone	C_3H_6O	2.880 ± 0.03	2.789	0.97	3.167	1.10
acetic acid	$C_2H_4O_2$	1.700 ± 0.03	1.613	0.95	1.826	1.07
methylformate	$C_2H_4O_2$	1.770 ± 0.04	1.778	1.00	1.931	1.09
dimethylamine	C_2H_7N	1.010 ± 0.02	0.890	0.88	1.043	1.03
imidazole	$C_3H_4N_2$	3.800 ± 0.4	3.668	0.97	3.773	0.99
pyrazole	$C_3H_4N_2$	2.200 ± 0.01	2.217	1.01	2.316	1.05
pyridine	C_5H_5N	2.215 ± 0.01	2.050	0.93	2.299	1.04
trimethylphosphate	$C_3H_9O_4P$	3.18**	3.475	1.09	3.723	1.17
Average				0.94		1.05
Standard deviation				0.08		0.06

- Ratios are calculated with respect to experimental values. Data on water was not included in the calculation of the average ratio and standard deviation. Calculations were performed using MP2(fc)/6-31G(d) geometries except for water, for which experimental gas-phase geometry was utilized.
- Experimental dipole moments are from Ref. [35]. The experimental uncertainties are indicated where available.

* Questionable results because of undetermined error sources.

** Liquid phase measurements, which may have large errors because of association effects.

with the aug-cc-pVDZ basis set are much more computationally expensive and become impractical for large and flexible molecules such as nucleosides. For example, for the guanine nucleoside cc-pVDZ and aug-cc-pVDZ basis sets consist of 331 and 554 basis functions, respectively, and single-point energy B3LYP calculations using 1.5 GHz CPU take around 2 and 12 hours, respectively. Approximately 100 such calculations are required for each conformation of the model compound to generate the perturbed maps of the ESP. Thus, for large systems, B3LYP/cc-pVDZ calculationss combined with the appropriate scaling, can be used for determination of ESP target data.

Based on the data in Tables A.I and A.II, the following QM approach is suggested for determination of the electrostatic parameters. B3LYP/aug-cc-pVDZ calculations will be used for the calculation of QM response electrostatic potential maps. The scale factor 0.724 is applied to the fitted values of atomic polarizabilities to reflect the reduced polarizability required for the condensed phase simulations [30]. For larger molecules, B3LYP/cc-pVDZ calculations are recommended. The underestimation of experimental polarizabilities by this level of theory must be corrected by applying the inverse of the average ratio of calculated to experimental gas phase molecular polarizabilities 1/0.83 to the empirical values of $\alpha(A)$ as additional scale factor. Combining this scale factor with the 0.724 factor yields an overall scaling factor of 0.87 that should be applied to the atomic polarizabilities after the fitting procedure. In addition, care must be taken when applying this level of theory due to the presence of outliers with respect to the calculated molecular polarizabilities (e.g., fluoromethane, Table A.II).

A.3.2 Reference values for atomic charges and polarizabilities

Ideally, charges and polarizabilities can be determined via free fitting, as the resultant charges and polarizabilities are theoretically invariant to the initial guess. However, free fitting often results in physically unrealistic values of the charges [46]

Table A.II: Summary of experimental and calculated molecular polarizabilities (in $Å^3$).

Molecule		Expt.	ahp Miller	B3LYP cc-pVDZ	ratio	B3LYP aug-cc-pVDZ	ratio
water	H_2O	1.45	1.41	0.78	0.54	1.39	0.96
ethane	C_2H_6	4.47	4.44	3.61	0.81	4.31	0.96
propane	C_3H_8	6.29	6.28	5.27	0.84	6.14	0.98
isobutane	C_4H_{10}	8.14	8.11	6.94	0.85	7.98	0.98
butane	C_4H_{10}	8.20	8.11	6.94	0.85	7.99	0.97
pentene-1	C_5H_{10}	9.65	9.76	8.42	0.87	9.88	1.02
pentene-2	C_5H_{10}	9.84	9.76	8.61	0.87	10.00	1.02
cyclohexane	C_6H_{12}	11.00	11.01	9.60	0.87	10.68	0.97
benzene	C_6H_6	10.00	10.43	8.57	0.86	10.35	1.04
toluene	C_7H_8	11.80	12.27	10.48	0.89	12.39	1.05
fluoromethane	CH_3F	2.97	2.52	1.92	0.64	2.53	0.85
fluorobenzene	C_6H_5F	10.30	10.34	8.57	0.83	10.35	1.00
chlorobenzene	C_6H_5Cl	12.25	12.36	10.20	0.83	12.51	1.02
methanol	CH_4O	3.32	3.25	2.41	0.73	3.19	0.96
ethanol	C_2H_6O	5.11	5.08	4.11	0.80	5.04	0.99
dimethylether	C_2H_6O	5.29	5.08	4.14	0.78	5.09	0.96
ethanethiol	C_2H_6S	7.41	7.44	5.70	0.77	7.30	0.99
acetaldehyde	C_2H_4O	4.59	4.53	3.69	0.80	4.58	1.00
acetone	C_3H_6O	6.39	6.37	5.28	0.83	6.37	1.00
acetic acid	$C_2H_4O_2$	5.10	5.17	4.05	0.79	5.14	1.01
methylformate	$C_2H_4O_2$	5.05	5.17	4.06	0.80	5.12	1.01
pyrazole	$C_3H_4N_2$	7.23	7.72	5.76	0.80	7.32	1.01
pyridine	C_5H_5N	9.18	9.73	7.87	0.86	9.55	1.04
adenine	$C_5H_5N_5$	13.10	15.05	11.82	0.90	14.44	1.10
cytosine	$C_4H_5N_3O$	10.30	11.12	9.32	0.90	11.60	1.13
guanine	$C_5H_5N_5O$	13.60	15.68	12.56	0.92	15.40	1.13
thymine	$C_5H_6N_2O_2$	11.23	12.11	10.34	0.92	12.44	1.11
trimethylphosphate	$C_3H_9O_4P$	10.86	10.86	9.27	0.85	11.23	1.03
Average					0.83		1.01
Standard deviation					0.06		0.06

- Ratios are calculated with respect to experimental values. Data on water were not included in the calculation of the average ratio and standard deviation. Calculations were performed using MP2(fc)/6-31G(d) geometries except for water, for which experimental gas-phase geometry was utilized.
- Experimental polarizabilities are from Ref. [35]. When several experimental estimates were available, the most recent data was used.
- Calculated empirical polarizabilities are those obtained from atomic component values using the additive ahp scheme from Ref. [65].

and polarizabilities. For instance, the polarizability tends to be attracted to one or a few atoms typically in the central region of a molecule and negative charges are often obtained for hydrogen atoms attached to aliphatic carbons from such fitting (see Table A.VIII of the Supporting Information). Therefore, the use of restraints is necessary to obtain chemically meaningful charges and polarizabilities. The task of careful selection of initial values for charges and polarizabilities becomes an important step since in restrained fitting these charges and polarizabilities are used as reference values in equation (A.9).

The classical Drude oscillator formalism employs the concept of atomic polarizability and assigns a corresponding unique parameter to individual atoms. Unlike molecular dipoles and polarizabilities, the atomic charges and polarizabilities of atoms in molecules are not well-defined quantities that can be unambiguously determined. A variety of different schemes have been proposed to obtain values of atomic polarizabilities, using both QM and experimental molecular polarizabilities [65, 66, 67, 68, 69, 70, 71]. They can be categorized into element, group, bond and hybrid polarizability schemes [65]. These methods can be also classified as additive, where the molecular polarizability is considered as a sum of atomic contributions, and non-additive, which usually rely on the iterative solution of non-linear equations to obtain molecular polarizability values. Non-additive atomic polarizability schemes of Applequist [66] and Thole [67] have been commonly used to obtain polarizabilities in polarizable force fields employing point dipole induction models [21, 25, 72]. However, the atomic polarizability values from these as well as the majority of other non-additive schemes can be considered as parameters of that model, and, therefore, may be inappropriate for use in other polarization models. To the best of our knowledge, no atomic polarizability values for molecules other than water have been suggested for the classical Drude oscillator model.

Based on the classical Drude oscillator formalism, the atomic polarizability, $\alpha(A)$, of atom A is theoretically independent on the polarizability of the other atoms. However, in practice, atomic polarizabilities will be interdependent. Such interdependence is systematically minimized in force fields through assignment of

atom types, which are expected to show transferability within a given class of chemical compounds. In order to facilitate transferability, the initial guess for atomic polarizability should also be taken from an additive polarizability scheme. The atomic hybrid polarizability (ahp) scheme of Miller [65] is a good example of such a scheme and, after some modifications, it has been used in this work to provide an initial guess of atomic polarizabilities.

In the original Miller scheme the atomic polarizability contributions were obtained from least-squares fitting to experimental gas-phase molecular polarizabilities for approximately 400 organic compounds [65]. In addition, the Miller atomic hybrid polarizabilities $\alpha_A(\text{ahp})$ depend not only on the identity but also on the hybridization state of a particular atom, which is similar to an empirical force field atom type concept. According to the Miller scheme, molecular polarizabilities are obtained by summing up atomic hybrid polarizabilities $\alpha_A(\text{ahp})$ [65]. In most cases the sum of ahp polarizabilities is very close to experimental values of molecular polarizabilities (see Table A.II). Values of $\alpha_A(\text{ahp})$ are available for most atomic hybrids which are encountered in biological compounds. Besides, the additivity of the model allows summing up atomic polarizabilities for functional groups and easy adjustment of $\alpha_A(\text{ahp})$ values by applying scale factors (set below). Thus, the $\alpha_A(\text{ahp})$ values were used as the initial estimates for atomic polarizabilities in the electrostatic parameter fitting procedure. It should be noted that molecular polarizability in the classical Drude oscillator model is not the additive sum of the atomic polarizabilities. However, $\alpha_A(\text{ahp})$ values can still be used as an initial guess and then adjusted during the fitting procedure.

Since no polarizability is assigned to hydrogens in the current model, the initial guess for atomic polarizabilities was constructed by adding ahp polarizabilities of the heavy atom and their covalently bound hydrogen atoms, thereby constituting the "united atom" approximation for atomic polarizability. If B3LYP/cc-pVDZ calculations are used for the calculation of the response ESP maps, these "united-atom" atomic polarizabilities have to be scaled by the factor of 0.83 introduced above. Thus, their sum will reproduce the molecular polarizability at the

B3LYP/cc-pVDZ level of theory, which will avoid biasing during the restrained fitting. Following the fitting procedure the scale factors described above are applied. No initial atomic polarizability scaling is required if the B3LYP/aug-cc-pVDZ level of theory is employed. Both unscaled and scaled atomic polarizabilities for biologically important atom types are given in Table A.III.

A few small adjustments to Miller's polarizabilities were introduced to improve agreement with the QM results. A new hybrid atom type for anionic oxygen (OTA) was introduced to reflect the substantially higher electronegativity of anions and was derived based on the polarizability of the acetate anion. The atomic polarizability of the P atom was substantially increased to obtain a better estimate for different phosphate species, especially those with methyl groups attached to ester oxygen atoms. The comparison between B3LYP/aug-cc-pVDZ calculated and the sum of corrected ahp values from Table A.III for a few relevant compounds are given in Table A.VI of the Supporting Information. Overall, the satisfactory agreement between QM and empirical values of molecular polarizabilities was obtained for most compounds. It should be emphasized that these values represent initial estimates of atomic polarizability values, which are then adjusted in the fitting procedure. Since in the classical Drude oscillator model the atomic polarizability is directly related to a partial charge assigned to the Drude particle, the atomic and Drude charges are determined in one step through charge fitting to the series of perturbed ESP maps obtained from QM calculations.

It is also necessary to have a good choice of reference atomic charges used in equation (A.9). Three different initial guesses were considered, namely CHARMM27 [54], NBO [73], and Mulliken [74] charges. Initial guesses for the polarizabilities were kept the same in all three cases. In general, it was found that the reference charge selection affects the final atomic charges but not the polarizability values. An example for such calculations for the cytosine base is given in Table A.VII of the Supporting Information. It was also found that some molecular properties, such as residue–water interaction energies, were not very sensitive to the reference atomic charges scheme chosen (see also Table A.VII). In addition, NBO and Mul-

Table A.III: Initial values of atomic polarizabilities.

Atom (group)		polarizabilities (Å^3)	
Symbol	Miller[1]	unscaled[2]	scaled[3]
$CH_3(sp^3)$	CTE+3H	2.222	1.844
$CH_2(sp^3)$	CTE+2H	1.835	1.523
$CH(sp^3)$	CTE+H	1.448	1.202
$C(sp^2)$	CTR	1.352	1.122
$CH(sp^2)$	CTR+H	1.739	1.443
$CH_2(sp^2)$	CTR+2H	2.126	1.765
$C(sp^2,br)$	CBR	1.896	1.574
–OH	OTE+H	1.024	0.850
–O–	OTE	0.637	0.529
=O	OTR4	0.569	0.472
$-O^{(-)}$	OTA[4]	0.858	0.712
$NH_2(sp^3)$	NTE+2H	1.738	1.443
$NH(sp^3)$	NTE+H	1.351	1.121
$N(sp^3)$	NTE	0.964	0.800
$NH_2(sp^2)$	NPI2+2H	1.864	1.547
$NH(sp^2)$	NPI2+H	1.477	1.226
$N:(sp^2)$	NTR2	1.030	0.855
P	PTE	2.063[5]	1.712

[1] Hybrid names used by Miller [65] are given in the second column along with the number of attached hydrogen atoms.

[2] Unscaled united atom values of ahp polarizabilities can be used as reference values in the charge fitting to the B3LYP/aug-cc-pVDZ potentials.

[3] United atom values of ahp polarizabilities were scaled by the factor 0.83 reflecting the underestimation of experimental gas phase polarizabilities by B3LYP/cc-pVDZ calculations.

[4] The anionic oxygen atom type (OTA) was not present in the original Miller's ahp scheme and was added to reflect its higher polarizability compared to the ester oxygen (OTE).

[5] The atomic polarizability of phosphorous atom was substantially changed from its original value in Miller's ahp scheme (1.538) to obtain a better estimate of the B3LYP molecular polarizabilities for phosphates. See details in the text.

liken atomic charges as initial guesses have some drawbacks. For instance, NBO charges on aliphatic hydrogens are often very large (0.20–0.30) even in non-polar compounds such as alkanes. This results in the undesirable overpolarization of the C–H bonds. Mulliken charge assignment has a variety of shortcomings including a strong dependence on the basis set, and unrealistic population assignment in the presence of diffuse and polarization functions, among others [75]. Moreover, charges obtained from QM models are conformation dependent, thereby complicating selection of the initial guesses. Therefore, the CHARMM27 charges were selected as the initial guess. This choice facilitates transfer of already existing internal and Lennard-Jones CHARMM27 force field parameters to the polarizable model, although optimization of these parameters will be necessary due to the new charge distribution and inclusion of explicit polarization. A disadvantage of using CHARMM27 charges as the initial guess is that they are limited to molecules for which the force field parameterization has been performed. Currently there is no simple scheme to derive these charges for new molecular residues although fitting procedures based on electronegativity equalization schemes [5] may be useful.

A.3.3 Atomic charge and polarizability derivation for model compounds

The methodology described above has been applied for the derivation of the electrostatic parameters for the model compounds shown on Figure A.3, which comprise the most important functional moieties of nucleic acids. Details of the fitting procedure will be discussed for the cytosine base. The geometry was optimized at the MP2/6-31G(d) level of theory with the amino group geometry constrained to be planar. Three Connolly surfaces for grid points and two surfaces for perturbation charges were generated (see Figure A.4). The total number of grid points and perturbation charges was 1327 and 57, respectively. Surface parameters and the number of points placed on each surface are shown in Table A.IV. In total, 57 QM ESP maps were calculated, one for each placement of the perturbation charge, at the B3LYP/aug-cc-pVDZ level of theory. For other model

compounds a similar strategy for the grid generation has been used. Reference partial atomic charges and atomic polarizabilities for the fitting procedure are shown in Table A.IX of the Supporting Information. Fitting was performed under the RESP parabolic restraint with a 10^{-5} Å^{-2} weighting factor. Charges and polarizabilities were restrained separately to their corresponding initial values. The restraint was combined with a flat well potential with a half-width of 0.1 e centered at the initial charge value. The restrained ESP fitting produced a final RMS error of $6.8 \cdot 10^{-4}$ e/Å with respect to the B3LYP potential, which is close to the RMS error of $3.9 \cdot 10^{-4}$ e/Å for the unrestrained fitting. The resultant atomic polarizabilities were then scaled by factor 0.724 to reflect reduced polarizability expected for the condensed phase (see above). This was followed by scaling the atomic charges to reproduce the B3LYP/aug-cc-pVDZ gas-phase dipole moment. Final optimized partial atomic charges and atomic polarizabilities are shown in Table A.IX of the Supporting Information for cytosine as well as other model compounds. In general, the fitted values of atomic charges decreased in magnitude with respect to initial values. This is expected as polarization was taken into account implicitly in the CHARMM27 charges, whereas explicit polarization in the Drude polarizable CHARMM force field naturally results in decreased values of the atomic charges in most cases (see Table A.IX). In the fitting procedure redistribution of the polarizability also takes place. As a rule, the polarizability tends to accumulate on atoms that are located near the center of the molecules, which is mainly an artifact of the fitting procedure resulting from the inadequate sampling of the electrostatic potential for the "buried" central atom(s). The use of restraints allows a substantial reduction of such effect (see Table A.VIII).

In situations where model compounds have several stable conformers (e.g., dimethylphosphate or ethanol), the grid generation and QM ESP calculations have been performed for the most relevant (i.e., the lowest-energy) non-equivalent conformers. For instance, gg and tg conformers have been used for the dimethylphosphate electrostatic parameter derivation. The tt conformer of the DMP anion has not been taken into consideration since it lies substantially higher in energy

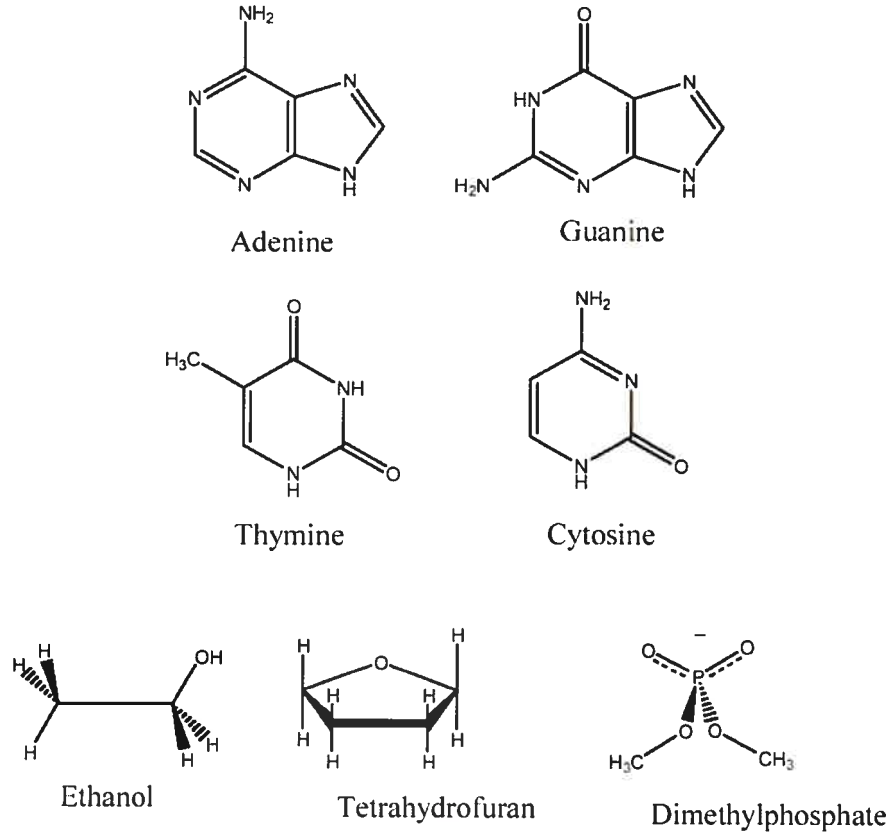

Figure A.3: Model compounds used for the preliminary parameter development of the Drude polarizable CHARMM force field for nucleic acids.

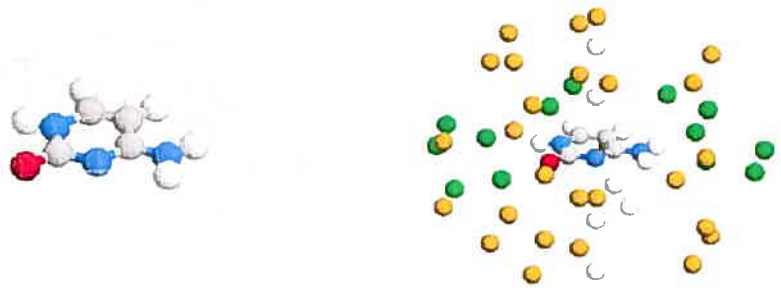

Figure A.4: Electrostatic potential grid created based on Connolly surfaces around the cytosine base and placement of perturbation charges. For visualization purposes only grid 1 and grid 2 Connolly surfaces are displayed (see Table A.IV for surface parameters). Perturbation charges are placed along chemical bonds (green spheres), around the N lone pair (white spheres) and in gaps between other charges (brown spheres) on the corresponding Connolly surfaces.

Table A.IV: Parameters used for the grid point and perturbation charge generation for the cytosine base.

Connolly surface	Size factor[1] (f)	Density factor (d)	Number of points
Perturbation charges 1	2.2	1.1	46
Grid 1	3.0	1.3	505
Perturbation charges 2	4.0	0.1	11
Grid 2	5.0	0.6	564
Grid 3	6.0	0.2	258

[1] Size factors (multiplication constants) multiplied by the van der Waals radii of corresponding atoms (1.2 Å for H, 1.5 Å for C, 1.4 Å for N, 1.35 Å for O) define the distance from that atom to the generated Connolly surface. Density factors determine the relative density of points on a particular Connolly surface.

($\sim$ 3 kcal/mol compared to gg) and the corresponding conformation is not populated in nucleic acids. Then the fitting procedure was performed by simultaneously including the target ESP data for both conformers resulting in one set of charges and polarizabilities. The procedure tends to decrease the quality of the fit (i.e., increases the RMSE) but minimizes bias in electrostatic parameters towards one conformer.

The quality of the derived set of atomic charges and polarizabilities can be evaluated by the ability of the Drude polarizable model to reproduce the gas-phase dipole and polarizability. The corresponding data are presented in Table A.V. The atomic charges and polarizabilities derived from the fitting procedure reproduce the reference experimental (if available) or QM gas-phase dipole moments within 3% for nucleic bases and THF and within 10% for the stable conformers of the dimethylphosphate anio (DMP) and ethanol, for which multiple conformers were used for fitting. The scaling of atomic polarizabilities results in the increase in the molecular dipole moments for the model compounds, which is corrected by the adjustment of partial atomic charges for all compounds except DMP. Since DMP is an ion the direct uniform charge scaling is not applicable. However the

reproduction of the gas-phase dipole moment for DMP was considered to not be essential as the dipole moment of ionic species is undetermined, i.e., depends on the molecular orientation.

Ideally, the experimental and/or high-level QM gas-phase dipole moment values should be reproduced in Drude calculations for all neutral model compounds through the application of the atomic charge fitting. However, since the molecular dipole moment is also a function of the molecular geometry it may be necessary to perform fitting in an iterative fashion following adjustment of Lennard-Jones and internal parameters. In general it should be noticed that the polarizable Drude model better reproduces gas-phase molecular dipole moments than additive CHARMM, which often overestimate them (see Table A.V) in order to reproduce condensed-phase properties of model compounds.

Concerning the molecular polarizabilities, those from the Drude model are underestimated by 1–4% for nucleic bases, $\sim$ 5% for ethanol and THF, and $\sim$ 14% for DMP compared to reference QM B3LYP/aug-cc-pVDZ data (see Table A.V). These differences may be a consequence of the different approaches used for determination of the QM and classical molecular polarizabilities as well as an artifact of the fitting procedure, which does not allow polarizability proper sampling for "buried" atoms. Future work will address the causes of these differences. However, the relative magnitudes of experimental and QM molecular polarizabilities are well reproduced by Drude polarizable CHARMM calculations.

A.3.4 Polarizable condensed-phase molecular dynamics simulation of DNA

The ability to perform condensed-phase simulations of biologically relevant molecular systems is the ultimate goal of any macromolecular empirical force field. As a proof of concept of the Drude model and the presented electrostatic parameter derivation procedure, a MD of the DNA duplex GAGTACTC in a SWM4-DP water box, including sodium counterions, was performed. Atomic charges and polarizabilities (Table A.IX of the Supporting Information) of the model compounds

Table A.V: Summary of calculated molecular dipoles and polarizabilities for a set of model compounds.

Model[1]	Expt.	B3LYP/ aug-cc-pVDZ	B3LYP/ cc-pVDZ	CH27[2]	Drude unscaled[3]	Drude scaled[4]
Molecular dipoles (in Debyes)						
ADE		2.43	2.35	2.94	2.43	2.25
CYT		6.72	6.24	7.88	6.70	6.92
GUA		6.97	6.84	7.61	6.96	6.81
THY		4.54	4.11	4.51	4.51	4.65
THF	1.75	1.87	1.70	2.34	1.91	1.72
DMPA gg		5.43	4.88	–[5]	5.42	–[5]
DMPA gt		4.82	4.31	–[5]	4.75	–[5]
DMPA tt		2.70	2.35	–[5]	2.43	–[5]
ETOH t	1.44	1.60	1.48	2.36	1.70	1.68
ETOH g	1.68	1.73	1.54	2.40	1.81	1.73
Isotropic molecular polarizability (in Å^3)						
ADE	13.10	14.44	11.82		13.98	10.36
CYT	10.30	11.60	9.32		11.19	8.16
GUA	13.60	15.40	12.56		15.29	11.09
THY	11.23	12.44	10.34		12.09	8.72
THF		7.81	6.76		7.44	5.70
DMPA gg		10.83	8.03		9.26	6.63
DMPA gt		10.86	8.07		9.27	6.64
DMPA tt		10.91	8.13		9.38	6.68
ETOH t	5.11	5.04	4.11		4.79	3.47
ETOH g		5.04	4.10		4.76	3.49

[1] ADE = adenine, CYT = cytosine, GUA = guanine, THY = thymine, THF = tetrahydrofuran, DMPA = dimethylphosphate anion, ETOH = ethanol.

[2] CHARMM27 (CH27) calculations were performed using the standard additive CHARMM force field for nucleic acids [54] and fully optimized geometry of the model compounds.

[3] The fitted partial atomic charges and atomic polarizabilities were not scaled. MP2 optimized geometry was used, and only positions of Drude particles were optimized to make a direct assessment of the quality of the fit to the QM ESP.

[4] The fitted atomic polarizabilities were scaled by factor of 0.724. The full geometry optimization of model compounds was performed for molecular dipole and polarizability calculations.

[5] Molecular dipole of ionic species depends on the molecular orientation and therefore cannot be compared directly for DMP using the QM and CHARMM optimized geometries.

shown on Figure A.3 have been used. The values of atomic charges for terminal atoms had to be adjusted upon creating the covalent bonds in an oligonucleotide. For example, in pyrimidine bases the charge of the hydrogen H1 was summed into the N1 charge. The experimental polarizability of the Na^+ ion (0.157 $Å^3$) [76] scaled by 0.724 (see above) has been used.

Lennard-Jones (LJ) and bonded parameters for model compounds were also required for the simulation. Preliminary optimization of LJ parameters was performed to reproduce minimum interaction energies and geometries of model compounds with water and for base pairs. High-level *ab initio* LMP2/cc-pVQZ//MP2/6-31G(d) calculations for nucleic bases as well as THF and LMP2/aug-cc-pVTZ//MP2/6-31+G(d) calculations for the DMP were used as a reference data [77]. In both the QM and empirical calculations the respective monomer geometries were fixed. The QM interaction distances were determined from the constrained MP2 optimization and then interaction energies were calculated at the LMP2 level with a larger and more flexible basis set. In CHARMM the corresponding hydrogen bonding distances were sampled at a resolution of 0.01 Å, and the minimum interaction energies were obtained. The positions of Drude particles were self-consistently adjusted at every step of the potential energy scan. The hydrogen bond angles have been also sampled for some orientations of nucleic acid bases with water molecule, where it acts as a hydrogen bond donor to the carbonyl oxygen of the base [54]. The LJ parameters for model compounds were adjusted to minimize the difference between QM and CHARMM interaction energies and to reduce RMS deviation for both distances and energies across different complex orientations. The comparison of results with available CHARMM27 additive force field data [54] was also performed. The final values of the minimized base pair interaction energies are summarized in Table A.X of the Supporting Information and those for the nucleic acid bases, DMP and THF with water are given in Tables A.XI and A.XII of the Supporting Information. For the nucleic acid bases the LJ optimization was performed mainly based on the *ab initio* LMP2/cc-pVQZ//MP2/6-31G(d) base pairing interaction energies. The optimization of Lennard-Jones parameters for

the DMP and THF was performed based on the LMP2//MP2 data on interactions with water.

Selected internal parameters were also optimized. Equilibrium bond lengths and angle values were adjusted to reproduce average values of the bases targeting a survey of nucleic acid crystal structures from the NDB [78]. Force constants were adjusted to reproduce MP2/6-31G(d) frequencies scaled by 0.9434 [79]. Torsional parameters for the C–O–P–O dihedral angles were adjusted to reproduce MP2/6-31+G(d) conformational energies for the DMP^- ion [80].

Using this zero-generation polarizable model a MD simulation of the GAG-TACTC DNA duplex in a box of SWM4-DP water with sodium counterions was run for 1 ns. The results of this simulation were compared with a previously published MD simulation of this DNA sequence using the additive CHARMM27 force field and the TIP3P water model [64]. RMSD values for non-hydrogen atoms of the six central residues of the DNA duplex were calculated with respect to canonical A and B DNA structure for this duplex over the course of the simulation. These data are shown in Figure A.5 and demonstrate that DNA structures from both Drude polarizable and CHARMM27 additive force field simulations remain closer to the B form versus the A form of DNA. More specific information may be obtained from the analysis of the base pairing interactions, e.g., through monitoring of the N1–N3 distance between Watson–Crick bases. The average values over the course of the simulation are presented in Table A.XIII of the Supporting Information. These data indicate that the current Drude polarizable model gives reasonable agreement with the additive CHARMM27 force field. However, larger fluctuations of the N1–N3 distances for most base pairs from the Drude polarizable DNA simulation occur. Another test of the validity of the performed simulation is the analysis of the backbone dihedral parameters. Probability distributions of the backbone dihedral angles are presented in Figure A.7 of the Supporting Information. The plots indicate that reasonable backbone dihedral angle probability distributions are obtained, being similar to those from the CHARMM27 MD simulation as well as data from a survey of the nucleic acid databank (NDB) [78] in most

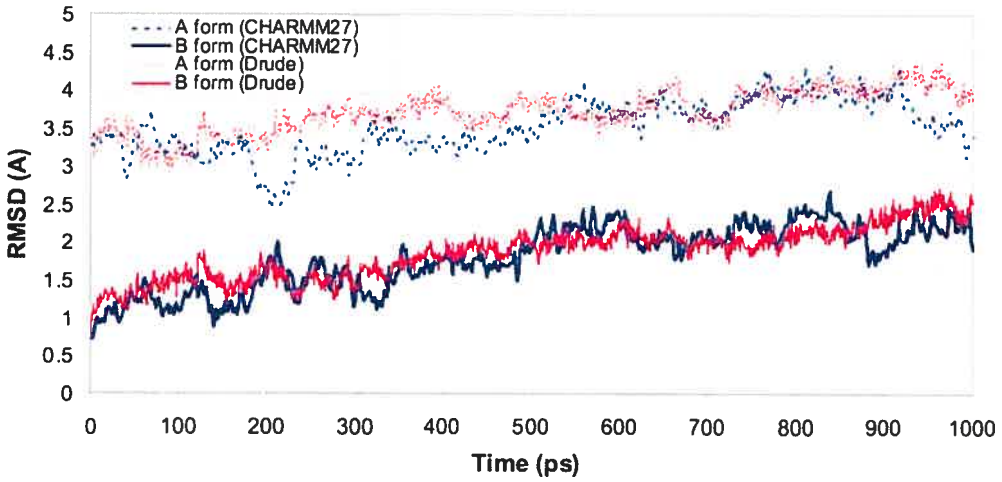

Figure A.5: Root mean square deviation (RMSD) of heavy atoms for six central residues of the GAGTACTC DNA duplex with respect to crystal structures of A and B forms of DNA (dashed and solid lines) during the course of the MD simulation. RMSD from simulations using CHARMM27 pairwise additive force field and TIP3P water model (blue lines) and Drude polarizable CHARMM force field with SWM4-DP water model (red lines) are presented.

cases. However, for some dihedrals there are noticeable differences in the relative conformer populations. The observed differences in the N1–N3 distances and in the dihedral distributions emphasize potential differences in molecular properties associated with the polarizable versus the additive CHARMM27 model. However, such differences also indicate the need for careful force field parameterization to be performed in order to properly implement a classical Drude based polarizable force field; such efforts are ongoing in our laboratory.

Importantly, the MD simulation was run in a reasonable amount of time. For instance, 50 ps of the extended Lagrangian DNA simulation took 23 CPU hours to run on the 3 GHz Pentium IV computer. For comparison, the additive CHARMM27 force field simulation for the same system consumed approximately 5 hours of CPU time. It should be noted that a larger time step of 2 fs was used in the CHARMM27 simulation, whereas only 1 fs time step was used for the ex-

tended Lagrangian simulation, and that that the TIP3P water model contains only 3 particles versus 4 particles in the SWM4-DP model.

Thus, applying the presented electrostatic parameter optimization methodology along with preliminary optimization of selected internal and LJ parameters produces a polarizable model of DNA that yields results comparable to those from the additive CHARMM27 nucleic acid force field. The final solvated DNA structure after 1000 ps is shown in Figure A.6. In addition, the simulation system remained stable over the course of the trajectory as evidenced by a lack of significant drift in the molecular volume and potential energy of the system (not shown). These results indicate the applicability of the developed method for derivation of the electrostatic parameters and the resulting classical Drude-based model of electronic polarizability for MD simulation studies of biological macromolecules in the condensed phase.

A.4 Conclusions

A general procedure for the determination of the electrostatic parameters for the classical Drude oscillator polarizable model, the partial atomic charges and atomic polarizabilities, is presented. This task is performed through fitting to a series QM electrostatic potentials for a test molecule obtained in the presence of perturbation charges. Consequently, the partial atomic and Drude particle charges, where the latter are responsible for the atomic polarizabilities, are determined simultaneously in a single step. The QM level of theory for the electrostatic parameter fitting was determined via comparison of dipole moments and molecular polarizabilities for a variety of small organic molecules. The B3LYP/aug-cc-pVDZ level was selected based on this comparison, although the computationally less expensive B3LYP/cc-pVDZ level may be used for larger molecules with the appropriate scaling factors. Special emphasis was placed on the creation of the grid required for the electrostatic potential and placement of the perturbation charges. Versus the commonly used cubic grid, an approach based on the placement of grid points on a pre-determined

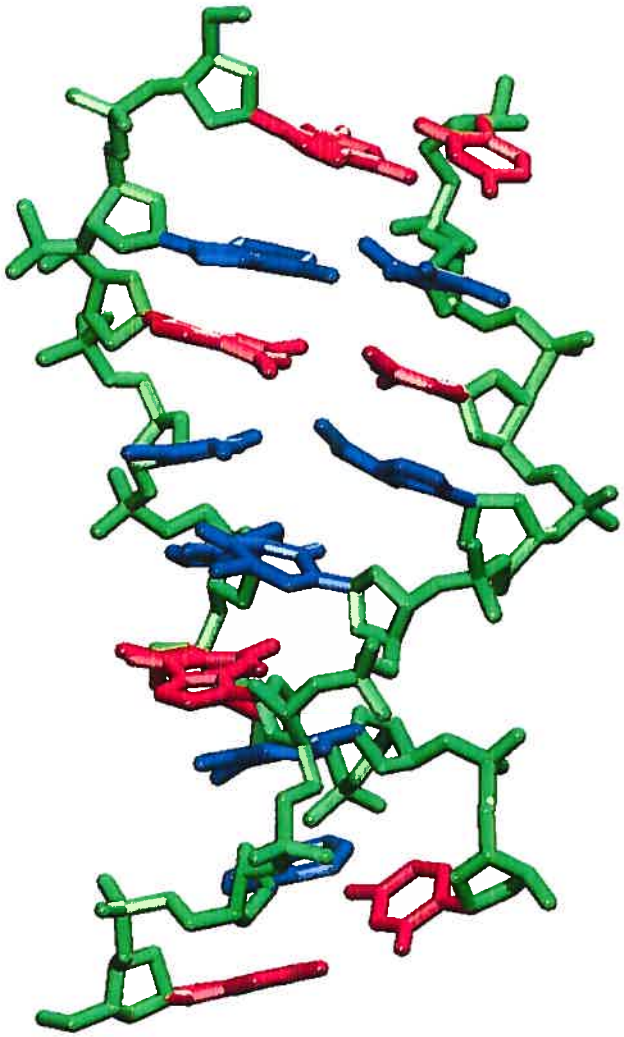

Figure A.6: GAGTACTC duplex DNA molecule. Final structure from the 1000 ps Drude polarizable extended Lagrangian molecular dynamics simulation in a box of SWM4-DP water with sodium counterions.

series of non-intersecting Connolly surfaces was developed. This approach reduces the number of grid points by optimizing their placement around the test molecules. Perturbation charges were placed along chemical bonds, lone pairs and in gaps between previously placed charges to provide equal coverage of the corresponding Connolly surfaces.

Consistent with previous work [46], it was found that restraints were needed during the fitting procedure to avoid unphysical atomic charges and polarizabilities. Thus, generic reference values for the atomic charges and polarizabilities become important. Atomic charges from the CHARMM27 additive force field and atomic polarizabilities obtained using the atomic hybrid polarizability scheme of Miller [65] were identified as suitable reference values. However, it was necessary to adjust Miller's ahp atomic polarizability values to take into account the united-atom polarizability model used in this study, i.e., polarizabilities of H atoms are added to that of the corresponding heavy atom to which they are bonded. Furthermore, fitted values of atomic polarizabilities were scaled to reflect the reduced polarization which appears to take place in the condensed phase [30, 36, 37, 38, 39, 40, 41]. Studies on the SWM4-DP [30] and other polarizable models [36, 37] have shown such scaling to be necessary to reproduce condensed-phase properties. In this study we used the same scaling factor as was used for SWM4-DP water model. Studies are underway in our laboratory to determine if the scale factor based on the water molecule (0.724) is appropriate for other small organic molecules. These tests will involve the ability of the scaling approach to reproduce pure solvent properties, including dielectric constants and free energies of solvation. Use of scaling of the polarizabilities required readjustment of the partial atomic charges to insure that gas-phase dipole moments were reproduced. The quality of the fits was evaluated via comparison of calculated molecular dipole moments and polarizabilities with the available gas phase experimental and QM values.

The developed scheme for the determination of atomic charges and polarizabilities has been tested on a set of small molecules representing functional moieties of nucleic acids. All other parameters have been taken from the all-atom additive

CHARMM27 force field for nucleic acids. Selected Lennard-Jones parameters were adjusted to reproduce QM data on interactions of model compounds with water as well as nucleic acid base pairing interactions. Selected internal parameters were optimized to reproduce experimental and QM data on molecular geometries, vibrational frequencies and rotational barriers. The resulting zero-generation force field has been successfully applied in a 1 ns polarizable MD simulation of a DNA octamer in aqueous solution. This simulation validates the feasibility of the developed methodology for the determination of partial atomic charges and polarizabilities as well as the use of the Drude oscillator model to include electronic polarizability in biomolecular systems. Future efforts will apply the methodology developed in this work, along with an iterative parameter optimization scheme that includes the internal and LJ parameters, to develop a non-additive empirical force field for molecules of biological and pharmacological interest.

Acknowledgements

We gratefully acknowledge the financial support from the NIH (GM51501). Computer time allocations were received from the Pittsburg Supercomputing Center and DOD ACS Major Shared Resource Computing and High Performance Computing.

Supporting information

[See Tables A.VI to A.XIII and Figure A.7 on following pages.]

Scaling and rounding strategy

The average ratios of calculated to experimental dipole moments at the B3LYP/aug-cc-pVDZ and B3LYP/cc-pVDZ levels for compounds listed in Table A.I, excluding water, were 1.05 and 0.94, respectively. In principle, these factors could be used for the scaling of partial atomic charges after the electrostatic parameter fitting

procedure. Moreover, the atomic charges may need to be scaled since polarizability scaling, as discussed in the main text, changes the molecular dipole thereby requiring adjustment of the final charges to reproduce the gas-phase dipole. It is also desirable to round off the obtained charges and molecular polarizabilities to, for example, three digits after the decimal points, as additional figures will not significantly impact the molecular properties. However, rounding off and charge scaling introduce an error in the formal molecular charge, where the sum of atomic charges is not equal to the total molecular charge. This error should be minimized and possibly equally distributed over all atoms. Because of rounding constraints it is not feasible to distribute the charge error in equal increments to all atoms. Instead by taking into account an equivalence constraint applied to selected atoms during fitting, the total charge error will be minimized, with the charge difference assigned to the first non-equivalent atom in the list.

The procedure for scaling and rounding off corrections is performed as follows. Atomic $q(A)$ and Drude $q_{\mathrm{D}}(A)$ charges are obtained from the FITCHARGE procedure. Then the atomic polarizabilities are calculated from the Drude charges according to equation (A.1). Next, the polarizabilities are scaled by the factor $p = 0.724$ or 0.87 if the B3LYP/aug-cc-pVDZ or B3LYP/cc-pVDZ calculations, respectively, are used for obtaining QM ESP maps (see main text for more information). Then the scaled polarizabilities are rounded to three digits after the decimal point producing the final scaled atomic polarizabilities, which are the Drude model parameters:

$$\alpha'(A) = \mathrm{round}[\alpha(A) \cdot p]. \tag{A.14}$$

The second part of the scaling and rounding procedure derives the final atomic charges. Here the Drude charges are calculated from the Drude polarizabilities $\alpha'(A)$:

$$q'_{\mathrm{D}}(A) = \sqrt{k_{\mathrm{D}} \cdot \alpha'(A)}. \tag{A.15}$$

Since the polarizability scaling changes the Drude charges initially produced by the FITCHARGE routine, we compensate for the introduced error by recalculation of

atomic core charges $q_c(A)$ in order to retain the effective atomic charges $q(A)$:

$$q'_c(A) = q_c(A) + q_D(A) - q'_D(A). \tag{A.16}$$

The obtained atomic core and Drude charges create a dipole moment $\boldsymbol{\mu}_{emp}$ that slightly differs from the dipole obtained directly from the charge fitting to QM data. Since the final charges should reproduce the target gas-phase dipole moment $\boldsymbol{\mu}_{target}$, the effective atomic charges $q(A)$ need to be scaled by the following factor:

$$c = \frac{\mu_{target}}{\mu_{emp}}. \tag{A.17}$$

The final effective atomic charges $q'(A)$ are obtained through applying the charge scale factor, c, and rounded to three digits after the decimal point:

$$q'(A) = \text{round}[q(A) \cdot c]. \tag{A.18}$$

Since we do not want the polarizability to be changed during this operation, we preserve the Drude charges, which are obtained from the polarizability scaling, and recalculate only the atomic core charges:

$$q'_c(A) = q'(A) - q'_D(A). \tag{A.19}$$

The charges $q'_c(A)$ and $q'_D(A)$ are final results of the charge fitting procedure. As mentioned above, the total molecular charge is preserved by adding the small error accumulated after charge rounding to the first non-equivalent atom. The rounding error was 0.001 e in the case of the cytosine base.

A number of points should be considered when performing charge scaling. Scaling cannot been applied to ions, such as dimethylphosphate anion (DMP^-), since the total molecular charge will not be conserved, along with the dipole moment for ionic species being dependent on the molecular orientation [79]. The scaling factors presented in Table A.I are not fully systematic, which might argue against using a

uniform scaling factor. Also, the scaling factor cannot be applied if experimental values of dipole moment are not available or of low accuracy. In the latter case very high-level QM calculations (e.g., CISD(T) with large basis sets) can be used for the derivation of the scaling factors for atomic charges. Charge scaling is also not recommended for molecular species with dipole moments less than 0.5 Debye (e.g., alkanes, see Table A.I), since capabilities of charge fitting to identify unique atomic charges is essentially diminished for molecules with small dipole moments.

Table A.VI: Comparison between QM polarizabilities and polarizabilities obtained using scaled and corrected ahp values.

Molecule		Polarizabilities (in Å^3)		
		QM	Empirical	Diff. (%)
ethane	C_2H_6	4.31	4.44	3.16
propane	C_3H_8	6.14	6.28	2.21
isobutane	C_4H_{10}	7.98	8.11	1.64
ethanol	C_2H_6O	5.04	5.08	0.79
methanol	CH_4O	3.19	3.25	1.76
dimethylether	C_2H_6O	5.09	5.08	−0.20
tetrahydrofuran	C_4H_8O	7.81	7.98	2.16
cyclohexane	C_6H_{12}	10.68	11.01	3.06
diethylether	$C_4H_{10}O$	8.87	8.75	−1.30
propanol	C_3H_8O	6.85	6.92	0.96
ribose	$C_5H_{10}O_5$	12.20	12.36	1.29
acetone	C_3H_6O	6.37	6.37	0.00
acetic acid	$C_2H_4O_2$	5.14	5.17	0.56
benzene	C_6H_6	10.35	10.43	0.81
toluene	C_7H_8	12.39	12.27	−0.96
dimethylamine	C_2H_7N	5.79	5.80	0.10
pyridine	C_5H_5N	9.55	9.73	1.88
adenine	$C_5H_5N_5$	14.44	15.05	4.22
guanine	$C_5H_5N_5O$	15.40	15.68	1.81
cytosine	$C_4H_5N_3O$	11.60	11.12	−4.12
thymine	$C_5H_6N_2O_2$	12.44	12.11	−2.67
trimethylphosphate	Me_3PO_4	11.23	11.21	−0.15

QM molecular polarizabilities were calculated at the B3LYP/aug-cc-pVDZ level using MP2/6-31G(d) optimized geometries. Empirical molecular polarizabilities were calculated using united-atom corrected Miller ahp values (see text).

Table A.VII: Effect of the initial charge selection on the example of the cytosine base.

a) Fitted atomic charges (in e)

Atom type	CHARMM27		NBO		Mulliken	
	initial	final	initial	final	initial	final
N1	−0.500	−0.350	−0.650	−0.500	−0.077	0.016
H1	0.370	0.202	0.423	0.233	0.095	0.132
C6	0.050	−0.044	0.097	0.030	0.164	0.040
H6	0.170	0.266	0.226	0.286	0.014	0.136
C2	0.520	0.600	0.851	0.874	0.174	0.244
O2	−0.490	−0.545	−0.641	−0.630	−0.275	−0.458
N3	−0.660	−0.602	−0.642	−0.687	−0.248	−0.344
C4	0.650	0.610	0.491	0.565	0.101	0.056
N4	−0.750	−0.708	−0.801	−0.625	−0.050	−0.158
H41	0.350	0.369	0.405	0.352	0.095	0.185
H42	0.350	0.369	0.405	0.352	0.095	0.185
C5	−0.130	−0.212	−0.399	−0.320	−0.058	−0.158
H5	0.070	0.045	0.235	0.070	−0.032	0.124

b) Fitted atomic polarizabilities (in Å^3)

Atom type	Initial	CHARMM27	NBO	Mulliken
N1	1.470	1.105	0.965	1.175
C6	1.731	1.227	1.281	1.328
C2	1.346	1.078	1.080	1.081
O2	0.566	0.444	0.478	0.349
N3	1.025	0.774	0.658	0.831
C4	1.346	0.870	0.866	0.873
N4	1.855	1.223	1.223	1.222
C5	1.731	1.143	1.128	1.145

c) Interaction energies with the water molecule (in kcal/mol)

	QM	Drude calculations. Initial charges		
Orientation	LMP2	CHARMM27	NBO	Mulliken
H1	−5.83	−5.32	−5.50	−5.06
O2_1	−7.01	−7.32	−7.69	−7.04
O2_2	−9.73	−9.17	−9.61	−8.46
N3	−8.20	−6.60	−6.73	−5.74
H41	−5.51	−4.44	−4.31	−3.14
H42	−5.43	−5.20	−5.15	−3.67
H5	−2.37	−2.14	−2.25	−2.62
H6	−2.83	−3.68	−3.86	−3.02

NBO and Mulliken charges were calculated at the B3LYP/aug-cc-pVDZ//MP2(fc)/6-31G(d) level. More details on the parameters if charge and polarizability fitting can be found in the footnote for Table A.VIII and in the main text. More details on the calculation of the cytosine–water interactions can be found in the footnote of Table A.XI.

Table A.VIII: Effect of constraints on the fitted values of atomic charges and polarizabilities for selected model compounds. Charges are in e, polarizabilities in $Å^3$, and RMSE in e/Å.

Tetrahydrofuran (THF)				Dimethylphosphate (DMPA)				Ethanol (ETOH)			
	initial	free fit	restr.		initial	free fit	restr.		initial	free fit	restr.
Charges				Charges				Charges			
O4′	−0.500	−0.568	−0.394	P1	1.500	0.978	1.326	C1	0.050	0.494	0.100
C4′	0.070	0.455	0.117	O3	−0.780	−0.740	−0.938	O1	−0.660	−0.573	−0.490
C3′	−0.180	0.021	−0.144	O4	−0.780	−0.740	−0.938	HO1	0.430	0.304	0.325
C2′	−0.180	0.021	−0.144	O1	−0.570	−0.422	−0.500	H11	0.090	−0.065	0.054
C1′	0.070	0.455	0.117	O2	−0.570	−0.422	−0.500	H12	0.090	−0.065	0.054
H11′	0.090	−0.068	0.037	C1	−0.170	−0.100	−0.136	C2	−0.270	−0.278	−0.250
H12′	0.090	−0.068	0.037	H11	0.090	0.091	0.137	H21	0.090	0.061	0.069
H21′	0.090	−0.028	0.075	H12	0.090	0.091	0.137	H22	0.090	0.061	0.069
H22′	0.090	−0.028	0.075	H13	0.090	0.091	0.137	H23	0.090	0.061	0.069
H31′	0.090	−0.028	0.075	C2	−0.170	−0.100	−0.136				
H32′	0.090	−0.028	0.075	H21	0.090	0.091	0.137				
H41′	0.090	−0.068	0.037	H22	0.090	0.091	0.137				
H42′	0.090	−0.068	0.037	H23	0.090	0.091	0.137				
Polarizabilities				Polarizabilities				Polarizabilities			
O4′	0.637	0.018	0.575	P1	2.063	0.000	2.277	C1	1.835	3.874	1.895
C4′	1.835	2.537	1.904	O3	0.858	2.113	0.973	O1	1.024	0.261	0.904
C3′	1.835	1.695	1.995	O4	0.858	2.113	0.973	C2	2.222	0.315	2.000
C2′	1.835	1.695	1.995	O1	0.637	2.480	0.742				
C1′	1.835	2.537	1.904	O2	0.637	2.480	0.742				
				C1	2.222	0.800	1.893				
				C2	2.222	0.800	1.893				
RMSE		7.9E−4	9.2E−4	RMSE		8.0E−4	1.6E−3	RMSE		1.0E−3	1.4E−3

All atom names correspond to those used in the CHARMM27 topology files (downloadable from http://www.pharmacy.umaryland.edu/faculty/amackere/). The initial charges are CHARMM27 charges, the initial polarizabilities are Miller's ahp values [65] summed to heavy atoms (see text). The fitting was performed to the B3LYP/aug-cc-pVDZ response electrostatic potential maps using MP2/6-31+G(d) for the DMP^- or MP2/6-31G(d) geometries for the THF and ETOH. The scheme for the placement of the ESP grid points and perturbation ions is described in the text. The gg and tg conformers of the DMP^- and trans and gauche(+) conformers of the ETOH were used for fitting. Fitting to the same charge and polarizability values was imposed for chemically equivalent atoms. No other restraints were used for the free fitting. The parabolic RESP restraints with the weighting factor of 10−5 $Å^{-2}$ were imposed to all charges and polarizabilities in the restrained fitting. Additionally, the flat well potential with the half-width 0.1 e was used for atomic charges and polarizabilities. The fitted atomic charges and polarizabilities values were not scaled. The root mean square errors (RMSE) are in the potential units (e/Å).

Table A.IX: Initial and final (fitted and scaled) values of atomic charges and polarizabilities for the model compounds.

a) Nucleic bases

Cytosine (CYT)			Thymine (THY)			Adenine (ADE)			Guanine (GUA)		
	initial	final		initial	final		initial	final		initial	final
charges			charges			charges			charges		
N1	−0.500	−0.350	N1	−0.800	−0.660	C5	0.280	0.145	N9	−0.380	−0.088
H1	0.370	0.202	H1	0.460	0.347	N7	−0.710	−0.531	C4	0.260	0.361
C6	0.050	−0.044	C6	0.170	0.259	C8	0.340	0.259	H9	0.360	0.198
H6	0.170	0.266	H6	0.170	0.191	H8	0.120	0.120	N3	−0.740	−0.639
C2	0.520	0.600	C2	0.510	0.597	N9	−0.370	−0.381	C2	0.750	0.643
O2	−0.490	−0.545	O2	−0.410	−0.505	H9	0.320	0.273	N1	−0.340	−0.407
N3	−0.660	−0.602	N3	−0.460	−0.342	N1	−0.740	−0.546	H1	0.260	0.283
C4	0.650	0.610	H3	0.360	0.211	C2	0.500	0.320	N2	−0.680	−0.789
N4	−0.750	−0.708	C4	0.500	0.449	H2	0.130	0.060	H21	0.320	0.384
H41	0.370	0.369	O4	−0.450	−0.404	N3	−0.750	−0.615	H22	0.350	0.384
H42	0.330	0.369	C5	−0.150	−0.152	C4	0.430	0.435	C6	0.540	0.491
C5	−0.130	−0.212	C5M	−0.110	−0.204	C6	0.460	0.357	O6	−0.510	−0.493
H5	0.070	0.045	H51	0.070	0.071	N6	−0.770	−0.542	C5	0.000	−0.072
			H52	0.070	0.071	H61	0.380	0.323	N7	−0.600	−0.527
			H53	0.070	0.071	H62	0.380	0.323	C8	0.250	0.138
									H8	0.160	0.133
polarizabilities			polarizabilities			polarizabilities			polarizabilities		
N1	1.470	1.105	N1	1.470	0.963	C5	1.566	1.025	N9	1.470	1.174
C6	1.731	1.227	C6	1.731	1.240	N7	1.025	0.833	C4	1.566	1.027
C2	1.346	1.078	C2	1.346	0.921	C8	1.731	1.143	N3	1.025	0.660
O2	0.566	0.444	O2	0.566	0.347	N9	1.345	1.077	C2	1.346	0.872
N3	1.025	0.774	N3	1.470	0.966	N1	1.025	0.657	N1	1.470	0.965
C4	1.346	0.870	C4	1.346	1.080	C2	1.731	1.373	N2	1.730	1.134
N4	1.855	1.223	O4	0.566	0.477	N3	1.025	0.756	C6	1.346	1.018
C5	1.731	1.143	C5	1.346	0.878	C4	1.566	1.176	O6	0.566	0.475
			C5M	2.212	1.459	C6	1.346	0.872	C5	1.566	1.242
						N6	1.855	1.214	N7	1.025	0.832
									C8	1.731	1.143

Table A.IX. (Continued)

b) Phosphodiester backbone model compounds

Tetrahydrofuran (THF)			Dimethylphosphate (DMP)			Ethanol (ETOH)		
	initial	final		initial	final		initial	final
charges			charges			charges		
O4′	−0.500	−0.350	P1	1.500	1.326	C1	0.050	0.094
C4′	0.070	0.105	O3	−0.780	−0.938	O1	−0.660	−0.471
C3′	−0.180	−0.130	O4	−0.780	−0.938	HO1	0.430	0.312
C2′	−0.180	−0.130	O1	−0.570	−0.500	H11	0.090	0.052
C1′	0.070	0.105	O2	−0.570	−0.500	H12	0.090	0.052
H11′	0.090	0.033	C1	−0.170	−0.136	C2	−0.270	−0.240
H12′	0.090	0.033	H11	0.090	0.137	H21	0.090	0.067
H21′	0.090	0.067	H12	0.090	0.137	H22	0.090	0.067
H22′	0.090	0.067	H13	0.090	0.137	H23	0.090	0.067
H31′	0.090	0.067	C2	−0.170	−0.136			
H32′	0.090	0.067	H21	0.090	0.137			
H41′	0.090	0.033	H22	0.090	0.137			
H42′	0.090	0.033	H23	0.090	0.137			
polarizabilities			polarizabilities			polarizabilities		
O4′	0.637	0.417	P1	2.063	1.648	C1	1.835	1.372
C4′	1.835	1.379	O3	0.858	0.705	O1	1.024	0.654
C3′	1.835	1.444	O4	0.858	0.705	C2	2.222	1.448
C2′	1.835	1.444	O1	0.637	0.537			
C1′	1.835	1.379	O2	0.637	0.537			
			C1	2.222	1.370			
			C2	2.222	1.370			

All atom names correspond to those used in the CHARMM27 topology files, which can be obtained from http://www.pharmacy.umaryland.edu/faculty/amackere/force_fields.htm. The initial charges are CHARMM27 charges, the initial polarizabilities are "united atom" corrected Miller's ahp values [65] (see text). The fitting was performed to the B3LYP/aug-cc-pVDZ perturbed electrostatic potentials using MP2/6-31+G(d) for the DMP^- or MP2/6-31G(d) geometries for the THF and ETOH. The scheme for the placement of the ESP grid points and perturbation ions is described in the text. The gg and tg conformers of the DMP^- and trans and gauche(+) conformers of the ETOH were used for fitting. Fitting to the same charge and polarizability values was imposed for chemically equivalent atoms. No other restraints were used for the free fitting. The parabolic RESP restraints with the weighting factor of 10−5 $Å^{-2}$ were imposed to all charges and polarizabilities in the restrained fitting. Additionally, the flat well potential with the half-width 0.1 e was used for atomic charges and polarizabilities. Final values of polarizabilities were scaled by 0.724 (see text) following which the charges were scaled for all compounds but DMP^- to reproduce the experimental or B3LYP/aug-cc-pVDZ gas-phase dipole moment.

Table A.X: Base pairing interaction energies (in kcal/mol).

	LMP2[1]	Drude[2]	Drude−LMP2[3]	CH27[4]	CH27−LMP2[5]
ATRH	−11.75	−11.35	0.40	−12.66	−0.91
ATRWC	−11.77	−10.33	1.44	−12.07	−0.30
ATH	−8.86	−11.45	−2.60	−12.93	−4.08
ATWC	−8.22	−10.48	−2.27	−12.35	−4.13
AC1	−11.18	−11.20	−0.02	−12.87	−1.69
AC2	−10.49	−11.47	−0.99	−12.63	−2.14
GA1	−12.50	−12.83	−0.33	−13.93	−1.43
GA2	−9.42	−9.15	0.27	−11.11	−1.69
GA3	−10.60	−12.25	−1.65	−12.59	−1.99
GA4	−10.27	−10.45	−0.18	−11.79	−1.52
GT1	−8.95	−12.67	−3.72	−12.96	−4.01
GT2	−9.56	−12.30	−2.75	−12.48	−2.92
GCWC	−18.47	−22.59	−4.12	−24.36	−5.89
GC1	−10.18	−12.28	−2.10	−15.01	−4.82
TC1	−9.26	−8.04	1.22	−9.86	−0.60
TC2	−9.17	−8.25	0.93	−10.31	−1.13
GG1	−16.74	−21.54	−4.79	−21.33	−4.59
GG3	−10.16	−15.93	−5.77	−18.22	−8.07
GG4	−9.09	−10.32	−1.23	−11.81	−2.72
AA1	−11.72	−9.37	2.35	−11.31	0.41
AA2	−10.35	−8.86	1.48	−10.64	−0.29
AA3	−9.01	−7.96	1.04	−9.88	−0.87
TT1	−6.13	−9.32	−3.19	−9.62	−3.50
TT2	−5.93	−9.57	−3.64	−10.03	−4.10
TT3	−6.68	−9.06	−2.38	−9.22	−2.54
CC	−12.35	−15.38	−3.03	−17.31	−4.96
Average absolute difference			2.07		2.74
RMS difference			2.14		1.97

[1] LMP2/cc-pVQZ energy calculation based on MP2/6-31G(d) all-atom optimized geometries of the base pairs [77].

[2] Interaction energy calculation based on fully optimized geometry of base pairs by classical Drude oscillator polarizable force field.

[3] DRUDE−LMP2 difference in interaction energy of the DNA bases.

[4] CHARMM27 interaction energy based on CHARMM27 fully optimized geometry of the base pairs.

[5] CHARMM27−LMP2 difference in interaction energy of the DNA base pairs.

Table A.XI: Base–water interactions.[1]

	R_{min} (Å)			Energy (kcal/mol)				R_{min} (Å)			Energy (kcal/mol)		
Residue	MP2[2]	Drude[3]	CH27[4]	LMP2[5]	Drude	CH27	Residue	MP2	Drude	CH27	LMP2	Drude	CH27
Cyt–H1	1.91	1.78	1.84	−5.83	−5.32	−6.43	Thy–H1	1.90	1.80	1.82	−6.60	−7.04	−7.51
Cyt–O2′	2.02	1.84	1.79	−7.01	−7.32	−8.45	Thy–O2′	1.90	1.92	1.77	−6.99	−7.11	−8.33
Cyt–O2″	1.84	1.76	1.72	−9.73	−9.17	−10.37	Thy–O2″	1.92	1.96	1.78	−5.62	−4.52	−6.35
Cyt–N3	1.98	2.02	1.88	−8.20	−6.60	−9.71	Thy–H3	1.88	1.72	1.83	−5.90	−5.79	−6.54
Cyt–H41	1.93	2.00	1.86	−5.51	−4.44	−5.81	Thy–O4′	1.91	1.92	1.76	−7.64	−4.86	−6.73
Cyt–H42	2.02	2.02	1.92	−5.43	−5.20	−6.00	Thy–O4″	1.98	2.00	1.79	−7.97	−5.00	−6.29
Cyt–H5	2.37	2.50	2.45	−2.37	−2.14	−2.84	Thy–H6	2.25	2.32	2.34	−3.30	−4.03	−4.30
Cyt–H6	2.27	2.34	2.35	−2.83	−3.68	−4.11							
Ave. diff.		0.01	0.07		−0.38	0.85	Ave. diff.		0.01	0.09		−0.81	−0.29
Ave. abs. diff.		0.09	0.11		0.67	0.85	Ave. abs. diff.		0.06	0.12		1.18	1.03
RMS diff.		0.10	0.10		0.71	0.45	RMS diff.		0.08	0.09		1.41	1.04
Gua–H1	1.94	1.94	1.89	−8.62	−5.73	−7.21	Ade–N1	2.02	1.70	1.89	−7.23	−7.29	−6.97
Gua–O6′	1.88	1.88	1.74	−7.43	−7.97	−9.94	Ade–H2	2.38	2.50	2.45	−1.47	−0.44	−1.55
Gua–O6″	2.05	2.00	1.80	−3.58	−5.07	−5.47	Ade–N3	2.05	1.70	1.90	−7.15	−7.75	−7.07
Gua–N7	2.04	2.04	1.95	−9.79	−6.50	−5.05	Ade–H9	1.90	1.74	1.89	−7.21	−5.83	−7.16
Gua–H8	2.27	2.20	2.36	−4.43	−1.97	−2.81	Ade–H8	2.25	2.38	2.37	−2.79	−2.29	−3.27
Gua–H9	1.91	1.84	1.84	−8.40	−5.37	−6.85	Ade–N7	2.00	2.00	1.84	−6.94	−5.51	−7.25
Gua–N3	2.06	1.88	1.92	−7.31	−5.11	−4.72	Ade–H61	1.96	2.00	1.89	−5.21	−3.07	−4.54
Gua–H21	2.01	2.02	1.89	−8.38	−6.19	−8.25	Ade–H62	1.90	2.00	1.85	−5.50	−4.58	−5.35
Gua–H22	1.95	2.00	1.87	−7.33	−4.57	−6.24							
Ave. diff.		0.14	0.09		−1.87	−0.97	Ave. diff.		0.05	0.05		−0.84	−0.04
Ave. abs. diff.		0.14	0.11		2.32	1.95	Ave. abs. diff.		0.15	0.10		1.01	0.26
RMS diff.		0.06	0.09		1.59	2.08	RMS diff.		0.18	0.10		0.82	0.33

[1] Base–water interaction calculations were performed on fixed MP2/6-31G* optimized geometry of the DNA bases and fixed gas-phase geometry of water molecule. Only base–water intermolecular distance and water orientation angle with base oxygen atom were optimized.

[2] MP2/6-31G(d) partial optimization of water–base interaction.

[3] Classical Drude oscillator polarizable force field partial optimization of water–base interaction. SWM4-DP polarizable water model was used in these calculations.

[4] CHARMM27 partial optimization of water-base interaction on CHARMM27 optimized geometry of DNA bases and TIP3P water molecule.

[5] LMP2/cc-pVQZ base-water interaction energy calculation on MP2/6-31G* partially optimized geometry of the complex.

Table A.XII: Interactions of the dimethylphosphate anion (DMP) and tetrahydrofuran (THF) with water.

	Orientation	R_{min}			E_{min}		
		Ab initio[1]	CH27[2]	Drude[2]	*Ab initio*	CH27	Drude
DMP gg	P1IN	3.32	3.29	3.53	−15.1	−16.7	−15.3
DMP gg	P1OUT	3.61	3.60	3.85	−8.5	−11.1	−9.6
DMP gg	O3ANI	1.83	1.66	1.90	−12.4	−13.3	−12.5
DMP gg	O4ANI	1.83	1.66	1.90	−12.4	−13.3	−12.5
DMP gg	O1EST	1.90	1.86	1.78	−9.3	−9.5	−9.2
DMP gg	O2EST	1.90	1.86	1.78	−9.3	−9.5	−9.2
DMP gg	AAD[3]		0.08	0.14		1.0	0.3
DMP gg	RMSD[3]		0.07	0.14		0.9	0.4
DMP gt	P1IN	3.33	3.29	3.54	−15.5	−16.8	−16.3
DMP gt	P1OUT	3.63	3.60	3.89	−8.7	−11.0	−9.3
DMP gt	O3ANI	1.83	1.66	1.88	−12.6	−13.3	−13.3
DMP gt	O4ANI	1.82	1.66	1.88	−12.4	−13.6	−13.3
DMP gt	O1EST	1.92	1.87	1.79	−8.9	−9.1	−8.4
DMP gt	O2EST	1.87	2.5	1.79	−9.9	−4.4	−8.4
DMP gt	AAD		0.18	0.13		1.8	0.8
DMP gt	RMSD		0.27	0.14		2.5	0.9
THF	O4′IN	1.97	1.85	1.84	−5.1	−5.9	−4.5
THF	O4′OUT	1.95	1.85	1.83	−5.4	−6.1	−4.6

[1] *Ab initio* calculations were performed at the LMP2/cc-pVQZ//MP2/6-31G(d) level for the THF, LMP2/aug-cc-pVTZ//MP2/6-31+G(d) level for the DMP.

[2] Fully optimized monomer geometries were used for the CHARMM27 (CH27) and classical Drude polarizable (Drude) empirical force field calculations. TIP3P water was used for CHARMM27, whereas SWM4-DP water model was used for Drude polarizable calculations.

[3] AAD is average absolute difference, RMSD is root mean square deviation.

Table A.XIII: Comparison of the average N1···N3 base pairing distance from the 1 ns CHARMM27 and Drude polarizable GAGTACTC duplex DNA molecular dynamics simulation.

Interacting bases		N1···N3 distance (Å)	
Strand1	Strand2	CHARMM27	Drude
G1	C8	3.68 ± 0.87	3.22 ± 0.26
A2	T7	4.97 ± 0.70	4.29 ± 0.58
G3	C6	3.85 ± 0.73	3.63 ± 0.63
T4	A5	2.94 ± 0.12	3.25 ± 0.71
A5	T4	2.96 ± 0.12	3.05 ± 0.43
C6	G3	3.01 ± 0.12	3.19 ± 0.23
T7	A2	2.98 ± 0.12	2.97 ± 0.30
C8	G1	3.02 ± 0.14	3.19 ± 0.21

The N1···N3 base pairing distances were calculated every 1000 steps of the MD simulations and averaged over the first 1000 ps of the simulation.

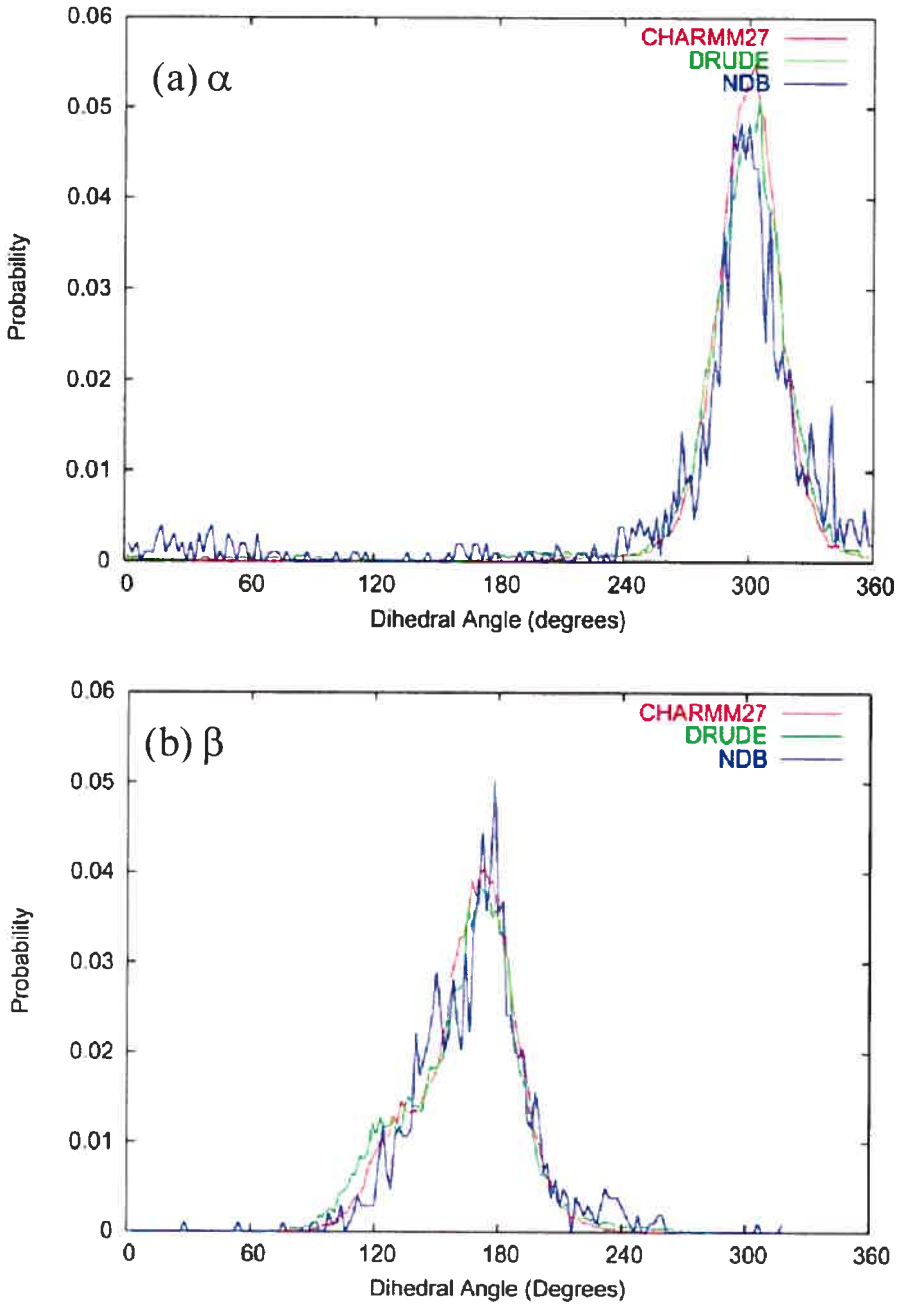

Figure A.7: Probability distributions from the solution molecular dynamics simulations of the GAGTACTC duplex DNA using additive CHARMM27 [54] and Drude polarizable CHARMM (DRUDE) force fields and from a survey of B DNA crystal structures in the nucleic acid databank (NDB) [78] as a function of dihedral angles α, β, γ, δ, ϵ, χ, and ξ.

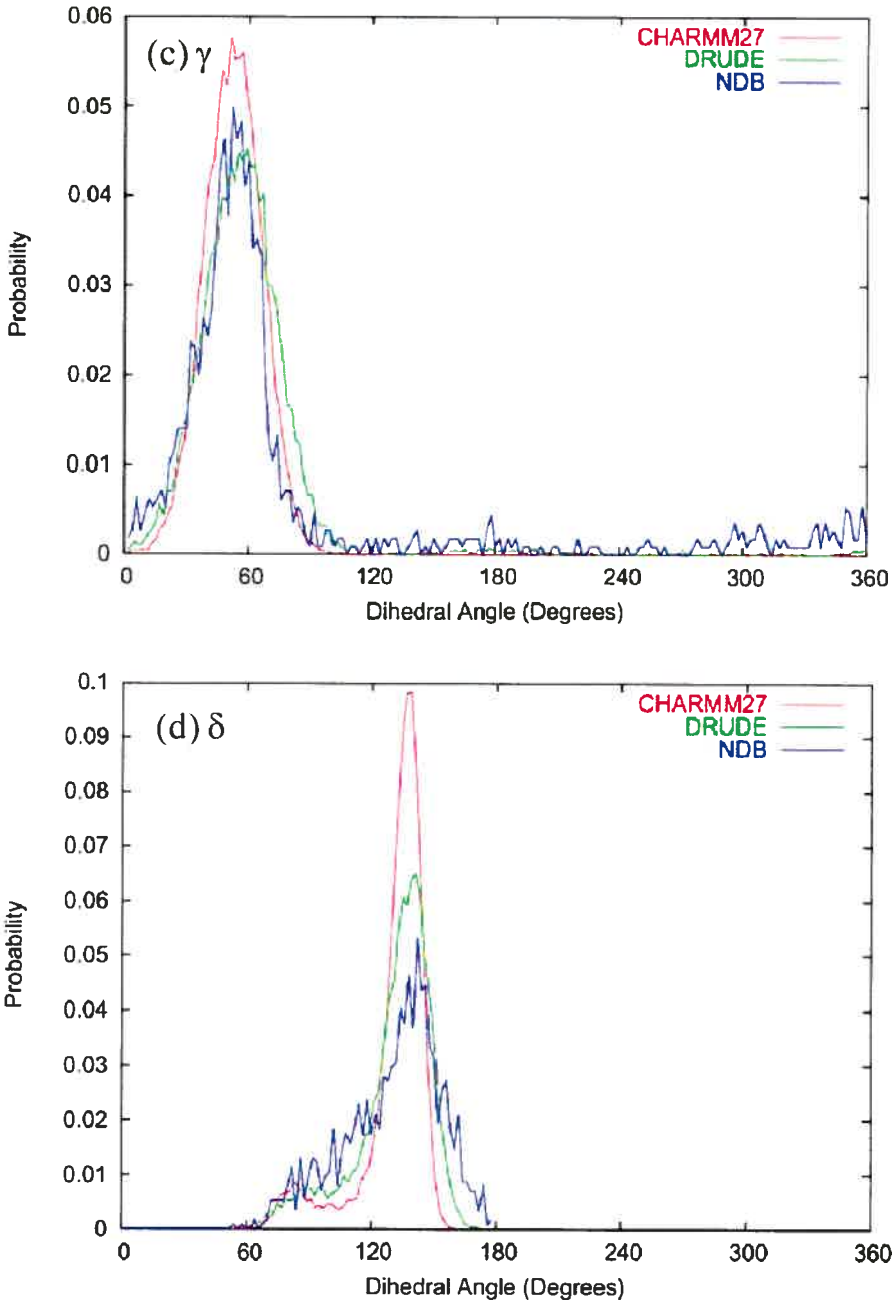

Figure A.7. (Continued)

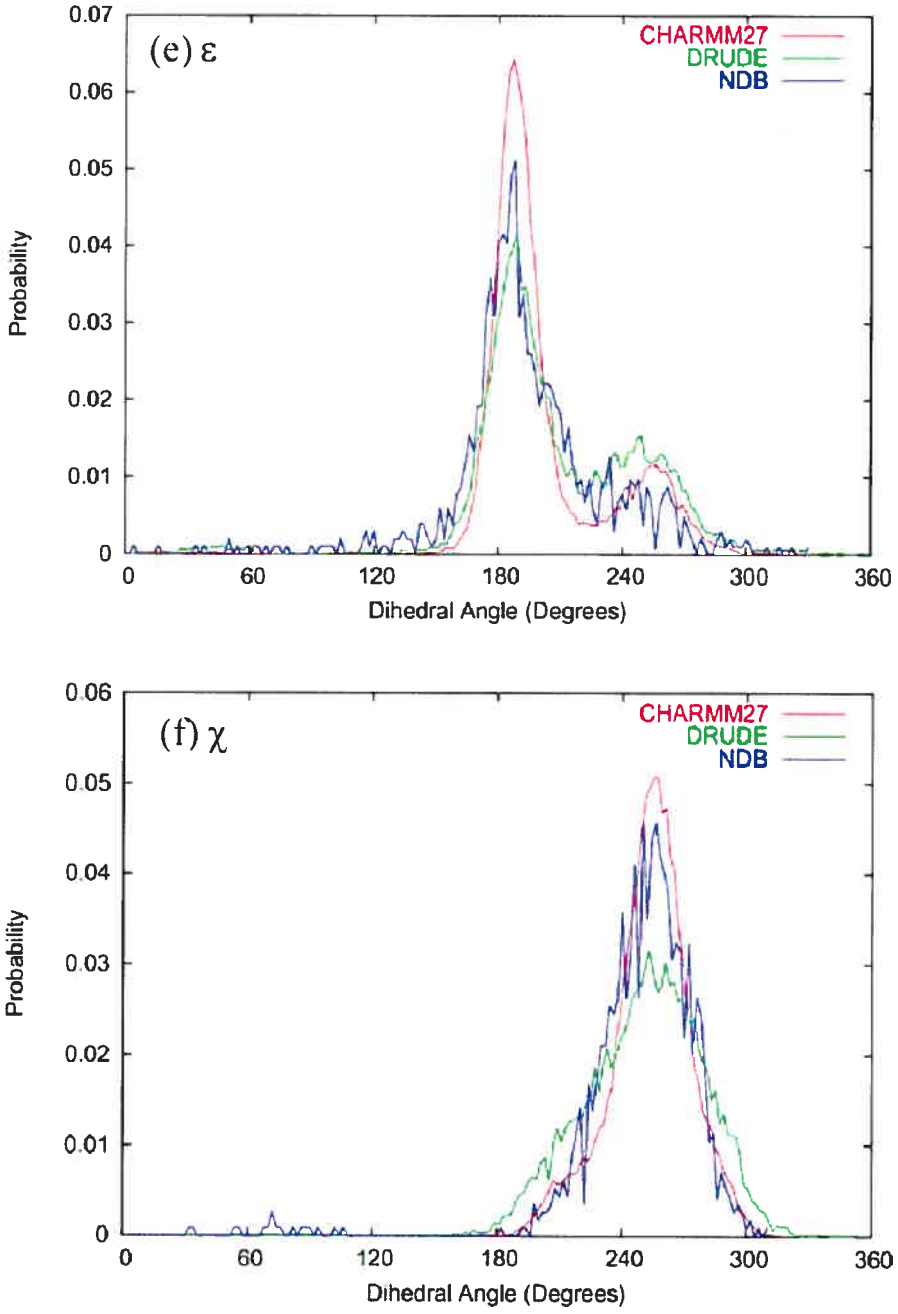

Figure A.7. (Continued)

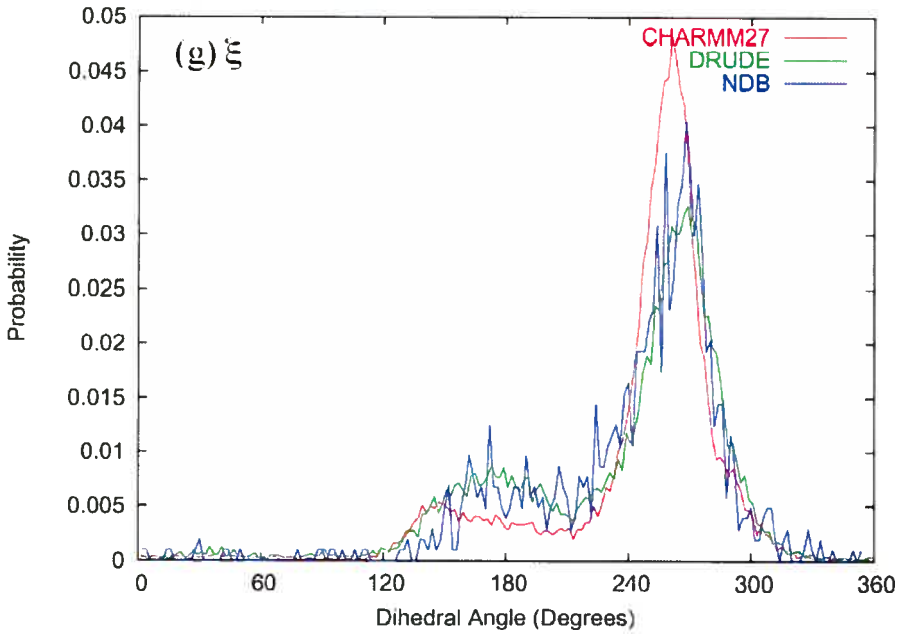

Figure A.7. (Continued)

References

[1] MacKerell, A. D., Jr. Atomistic Models and Force Fields. In Computational Biochemistry and Biophysics; Becker, O. M., MacKerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Marcel Dekker, Inc.: New York, 2001; pp 7.

[2] MacKerell, A. D., Jr. *J. Comput. Chem.* **2004**, 35, 1584.

[3] Saiz, L.; Klein, M. L. *Acc. Chem. Res.* **2002**, 35, 482.

[4] Halgren, T. A.; Damm, W. *Curr Opin Struct Biol.* **2001**, 11, 236.

[5] Rick, S. W.; Stuart, S. J. *Rev. Comp. Chem.* **2002**, 18, 89.

[6] Sprik, M.; Klein, M. L. *J. Chem. Phys.* **1988**, 89, 7556.

[7] Caldwell, J.; Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, 112, 9144.

[8] Wallqvist, A.; Berne, B. J. *J. Phys. Chem.* **1993**, 97, 13841.

[9] Bernardo, D. N.; Ding, Y.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1994**, 98, 4180.

[10] Dang, L. X. *J. Phys. Chem. B* **1998**, 102, 620.

[11] Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1996**, 118, 672. Bryce98

[12] Bryce, R. A.; Vincent, M. A.; Malcolm, N. O. J.; Hillier, I. H.; Burton, N. A. *J. Chem. Phys.* **1998**, 109, 3077.

[13] Patel, S.; Brooks, C. L., III *J. Comput. Chem.* **2003**, 25, 1.

[14] Patel, S.; MacKerell, A. D., Jr.; Brooks, C. L., III *J. Comput. Chem.* **2004**, 25, 1504.

[15] Chen, B.; Xing, J.; Siepmann, J. I. *J. Phys. Chem. B* **2000**, 104, 2391.

[16] Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **2001**, 115, 2237.

[17] Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, 103, 4730.

[18] Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, 107, 5933.

[19] Grossfield, A.; Ren, P.; Ponder, J. W. *J. Am. Chem. Soc.* **2003**, 125, 15671.

[20] Shelley, J. C.; Sprik, M.; Klein, M. L. *Langmuir* **1993**, 9, 916.

[21] Gao, J.; Habibollazadeh, D.; Shao, L. *J. Phys. Chem.* **1995**, 99, 16460.

[22] Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1995**, 99, 6208.

[23] Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, 117, 4177.

[24] Freindorf, M.; Gao, J. *J. Comput. Chem.* **1996**, 17, 386.

[25] Cieplak, P.; Caldwell, J. W.; Kollman, P. A. *J. Comput. Chem.* **2001**, 22, 1048.

[26] Dang, L. X. *J. Phys. Chem. B* **1999**, 103, 8195.

[27] Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, 23, 1515.

[28] Stuart, S. J.; Berne, B. J. *J. Phys. Chem.* **1996**, 100, 11934.

[29] van Maaren, P. J.; van der Spoel, D. *J. Phys. Chem. B* **2001**, 105, 2618.

[30] Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Phys.* **2003**, 119, 5185.

[31] Drude, P.; Mann, C. R.; Millikan, R. A. The theory of optics; Longmans, Green, and Co.: New York [etc.], **1902**.

[32] Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, 119, 3025.

[33] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, 4, 187.

[34] MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Paramerization with an Overview of the Program. In Encyclopedia of Computational Chemistry; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, **1998**; Vol. 1; p. 271.

[35] CRC Handbook Chemistry and Physics; 84th ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, **2003**.

[36] Giese, T. J.; York, D. M. *J. Chem. Phys.* **2004**, 120, 9903.

[37] Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, 108, 621.

[38] Morita, A. *J. Comput. Chem.* **2002**, 23, 1466.

[39] Morita, A.; Kato, S. *J. Chem. Phys.* **1999**, 110, 11987.

[40] in het Panhuis, M.; Popelier, P. L. A.; Munn, R. W.; Ángyán, J. G. *J. Chem. Phys.* **2001**, 114, 7951.

[41] Tu, Y.; Laaksonen, A. *Chem. Phys. Lett.* **2000**, 329, 283.

[42] Rick, S. W.; Stuart, S. J.; and Berne, B. J. *J. Chem. Phys.* **1994**, 101, 6141.

[43] van Belle, D.; Couplet, I.; Prevost, M.; Wodak, S. J. *J. Mol. Biol.* **1987**, 198, 721.

[44] Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, 110, 741.

[45] Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. Numerical Recipes in C; Cambridge University Press: Cambridge, **1988**.

[46] Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, 97, 10269.

[47] Connolly, M. L. *Science* **1983**, 221, 709.

[48] Bonin, K. D.; Kresin, V. V. Electric-dipole polarizabilities of atoms, molecules, and clusters; World Scientific: Singapore River Edge, NJ, **1997**.

[49] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98; Gaussian, Inc.: Pittsburgh, PA, **1998**.

[50] Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, 46, 618.

[51] Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, 153, 503.

[52] Hariharan, P. C.; Pople, J. A. Theoretica Chimica Acta (Berlin) **1973**, 28, 213.

[53] Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; von Ragué Schleyer, P. *J. Comput. Chem.* **1983**, 4, 294.

[54] Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, 21, 86.

[55] Becke, A. D. *Phys. Rev. A* **1988**, 38, 3098.

[56] Becke, A. D. *J. Chem. Phys.* **1993**, 98, 5648.

[57] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, 37, 785.

[58] Dunning, T. H. *J. Chem. Phys.* **1989**, 90, 1007.

[59] Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, 23, 327.

[60] Darden, T. A.; York, D.; Pedersen, L. G. *J. Chem. Phys.* **1993**, 98, 10089.

[61] Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, 15, 667.

[62] Lagüe, P.; Pastor, R. W.; Brooks, B. R. *J. Phys. Chem. B* **2004**, 108, 363.

[63] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, **1987**.

[64] Pan, Y.; MacKerell, A. D., Jr. *Nucl. Acid Res.* **2003**, 31, 7131.

[65] Miller, K. J. *J. Am. Chem. Soc.* **1990**, 112, 8533.

[66] Applequist, J.; Carl, J. R.; Fung, K.-K. *J. Am. Chem. Soc.* **1972**, 94, 2952.

[67] Thole, B. T. *Chem. Phys.* **1981**, 59, 341.

[68] van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, 102, 2399.

[69] Ewig, C. S.; Waldman, M.; Maple, J. R. *J. Phys. Chem. A* **2002**, 106, 326.

[70] Stout, J. M.; Dykstra, C. E. *J. Phys. Chem. A* **1998**, 102, 1576.

[71] Zhou, T.; Dykstra, C. E. *J. Phys. Chem. A* **2000**, 104, 2204.

[72] Ding, Y.; Bernardo, D. N.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1995**, 99, 11575.

[73] Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, 83, 735.

[74] Mulliken, R. S. *J. Chem. Phys.* **1955**, 23, 1833.

[75] Leach, A. R. Molecular Modelling: Principles and Applications; Longman: Harlow, **1996**.

[76] Mahan, G. D. *Phys. Rev. A* **1980**, 22, 1780.

[77] Huang, N.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2002**, 106, 7820.

[78] Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, 63, 751.

[79] Foresman, J. B.; Frisch, A. Exploring Chemistry with Electronic Structure Methods; Gaussian, Inc: Pittsburgh, PA, **1996**.

[80] Banavali, N. K.; MacKerell, A. D., Jr. *J. Am. Chem. Soc.* **2001**, 128, 6747.