Développement de champs de forces polarisables : vers la dynamique moléculaire SIBFA

Christophe Narth

▶ **To cite this version:**

Christophe Narth. Développement de champs de forces polarisables : vers la dynamique moléculaire SIBFA. Chimie théorique et/ou physique. Université Pierre et Marie Curie - Paris VI, 2015. Français. NNT : 2015PA066739 . tel-02406358

HAL Id: tel-02406358

https://theses.hal.science/tel-02406358

Submitted on 12 Dec 2019

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

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

UPMC SORBONNE UNIVERSITÉS

École Doctorale 388

Développement de champs de forces polarisables : vers la dynamique moléculaire SIBFA

THÈSE

présentée et soutenue publiquement le 29 septembre 2015

pour l'obtention du

Doctorat de l'Université Pierre et Marie Curie

par

Christophe Narth

Composition du jury

Rapporteurs :	Yves-Henri Sanejouand (Nantes) Jean-François Dufrêche (Montpellier)
Examinateurs :	Aurélien de la Lande (Orsay) Rodolphe Vuilleumier (Paris) Nohad Gresh (Paris)
Directeur de thèse :	Jean-Philip Piquemal

Laboratoire de Chimie Théorique – UMR 7616

Mis en page avec la classe thesul.

Je dédie cette thèse
à mes parents.
Oui, à Papa,
et Maman.

Sommaire

Introduction générale **1**

Partie I Introduction à la Dynamique Moléculaire **7**

Chapitre 1

Dynamique Moléculaire

1.1 Approches classiques 10

1.1.1 Interactions intramoléculaires 10

1.1.2 Interactions intermoléculaires 14

1.2 Vers des simulations réalistes : la solvatation 16

1.2.1 Solvatation explicite 17

1.2.2 Solvatation implicite 20

1.3 Principe de la Dynamique Moléculaire 22

1.3.1 Introduction 22

1.3.2 Intégration des vitesses 23

1.3.3 Contrôle thermodynamique 24

1.3.4 Exploitation des résultats 25

1.4 Conclusions 30

Chapitre 2

Introduction à SIBFA

2.1 Vers un meilleur traitement des interactions électroniques 34

2.1.1 Oscillateur de Drude 35

2.1.2 Charges fluctuantes 36

2.1.3 Pour aller plus loin 37

2.2 Le potentiel SIBFA 38

2.2.1 Electrostatique 40

2.2.2 Répulsion 41

2.2.3 Polarisation 42

2.2.4 Transfert de charge 43

2.2.5 Dispersion 43

2.3 Calibration 44

2.4 Conclusions & Perspectives 47

Partie II Vers la Dynamique Moléculaire SIBFA 51

Chapitre 1

Electrostatique

1.1 Introduction 54

1.2 Description charge atomique 58

1.3 Pénétration Electrostatique 61

1.3.1 Point de départ : quelle distribution multipolaire ? 75

1.3.2 Comparaison & discussion de cp-GDMA [1] et cp-MEP 80

1.3.3 Application des corrections de pénétration de charge 82

1.4 Conditions périodiques 88

1.5 Conclusions & Perspectives 132

Chapitre 2

Interactions de van der Waals et d'Induction :

Energies et Forces

2.1 Interactions de van der Waals 136

2.1.1 Introduction 136

2.1.2 Répulsion 138

2.1.3 Dispersion 149

2.2 Traitement de la mobilité électronique : Induction 154

2.2.1 Introduction 154

2.2.2 Polarisation 154

2.2.3 Transfert de charge 160

2.3 Résultats numériques & Discussion 166

2.4 Conclusions & Perspectives 171

Conclusions & Perspectives **175**

Table des figures **183**

Liste des tableaux **187**

Annexes

Annexe A

Résultats SAPT

Annexe B

Chapitre NCIplot :

Etude des différents types d'interactions non-covalentes

Annexe C

Paramètres et constantes du potentiel SIBFA

C.1 Electrostatique 238

C.2 Répulsion & Dispersion 239

Bibliographie **241**

Introduction générale

La chimie est la science de la matière. Elle permet de prédire et/ou comprendre ses transformations. Elle joue donc un rôle clef dans notre compréhension de la vie au sens large. De plus, elle se place au centre de plusieurs domaines, entre sciences physiques et biologie ou encore entre notions fondamentales et applications. On peut distinguer deux grands domaines : la chimie moléculaire et la chimie supramoléculaire [1]. Malgré ce rôle pivot de la chimie, ce fut d'abord la synthèse organique qui se distingua au cours des deux derniers siècles. Citons la synthèse de l'urée en 1828 [2] ou encore de la vitamine B12 [3]. Ce champ de recherche est la chimie moléculaire. Elle est fondée sur les liaisons chimique fortes, c'est-à-dire les interactions intramoléculaires : avec notamment la formation ou la cassure de liaison covalentes.

Au-delà de la liaison covalente, existe un autre monde : la chimie supramoléculaire [1]. Elle s'intéresse aux édifices moléculaires complexes et aux systèmes polymoléculaires maintenus par des interactions intermoléculaires dites faibles. Elle emprunte évidemment des notions de chimie moléculaire, plus particulièrement organique quant à la « construction » des molécules en jeux mais aussi à la chimie de coordination. En 1894, Emil Fisher a introduit la célèbre image « clef serrure » qui définit la relation entre un substrat et un récepteur [4] (voir figure 1).

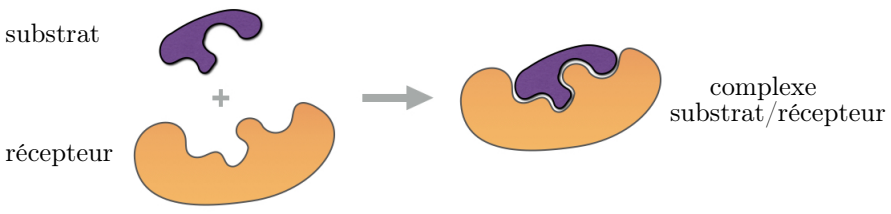


Figure 1 – schéma association ligand-récepteur (« clef - serrure »)

Celle-ci repose sur trois concepts :

(i) fixation : sélectivité des interactions entre les deux espèces

(ii) reconnaissance : complémentarité de forme et taille

(iii) coordination : l'affinité du substrat avec récepteur

On parle plus généralement de reconnaissance moléculaire. Ce concept est présent en biologie structurale dans le cadre des complexes protéine-ligand, eux-aussi majoritairement gouvernés par des interactions non-covalentes.

La biologie structurale a elle aussi connu un développement rapide ces dernières décennies. En effet, grâce aux méthodes de cristallographie et de résonance magnétique nucléaire (RMN), il est possible d'obtenir la structure en trois dimensions du système étudié. La visualisation permet des interprétations, certes, mais reste discutable quant à la résolution de la manipulation dans le cas de la cristallographie. De plus, toutes les protéines ne sont pas cristallisables. La RMN de macromolécules renseigne sur un ensemble de structures possibles.

La combinaison de ces méthodes a notamment permis la construction de bases de données, telle que la *Protein Data Bank* (PDB) fondée en 1971 qui regroupe aujourd'hui plus de 110.000 structures référencées. Ces données expérimentales sont accessibles gratuitement dans un format préétabli et se destinent principalement à la bioinformatique.

En disposant ainsi des coordonnées de ces systèmes, cela laisse entrevoir de nombreuses applications numériques, bien au-delà de l'analyse structurelle post-expérimentale. Parmi les approches computationnelles, seule la chimie quantique permet d'atteindre un très haut niveau de précision quant

à la stabilité d'une molécule . Seulement, ces systèmes comportent des milliers d'atomes, ce qui rend hors de portée de telles simulations, que ce soit en terme de temps ou de mémoire. Cependant, des avancées ont été réalisées ces dernières années, avec l'avénement de la théorie de la fonctionnelle de la densité (DFT) qui peut traiter un système d'une centaine d'atomes avec une bonne précision en un temps raisonnable [5, 6].

En s'intéressant à la composition des ces systèmes protéiques, on observe des répétitions de plusieurs motifs formant ainsi une séquence. Ces motifs sont les acides aminés reliés entre eux par des liaisons peptidiques. Ces acides aminés sont constitués d'atomes de carbone, d'oxygène, d'hydrogène et de souffre. Finalement, une approche d'origine empirique est envisageable vu le faible nombre d'atomes à paramétrer. L'objectif étant de mieux comprendre des systèmes d'intérêt biologique, d'autres atomes sont à prendre à compte. Le diagramme secteur ci-dessous (voir figure 2) décrit la composition atomique du corps humain, montrant ainsi les autres espèces à considérer, dont notamment des cations. Le même raisonnement peut-être appliqué à l'acide désoxyribonucléique (ADN) et l'acide ribonucléique (ARN), composés par des bases azotées, au nombre de quatre.

Ainsi, les méthodes de simulation alors envisagées sont les champs de forces, qui reposent sur un potentiel analytique en se basant sur la mécanique moléculaire. Plusieurs approches existent, qui sont classées par raffinement du modèle. On parle de champs de forces de première et de seconde génération. Cette différence se situe au niveau du traitement des interactions non-covalentes. SIBFA (Sum of Interactions Between Fragments *ab-initio* computed) [8–13] est l'un des potentiels polarisables les plus raffinés et complet permettant de décomposer l'énergie intermoléculaire. On pourrait même parler de champs de forces de troisième génération.

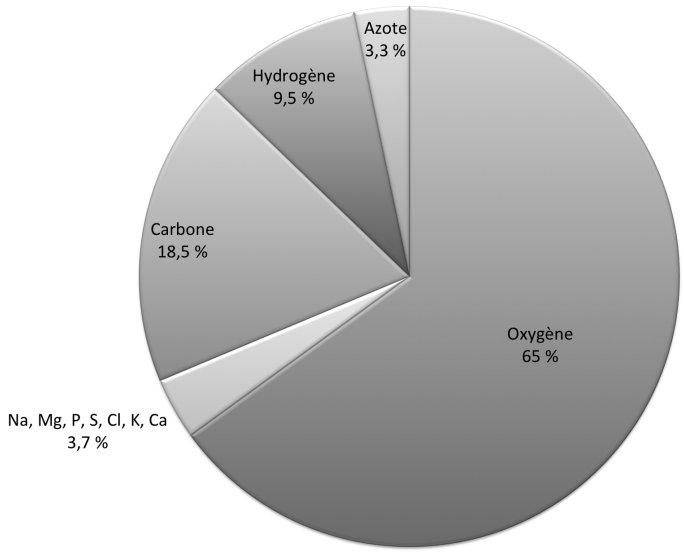


FIGURE 2 – diagramme secteur de la composition atomique du corps humain en pourcentage relatif [7].

L'objectif de cette thèse est la revisite méthodologique du modèle* SIBFA *et son implémentation dans un programme de dynamique moléculaire,* TINKER *[14], en exploitant dans un premier temps une partie du potentiel polarisable* AMOEBA *(Atomic Multipole Optimized Energetics for Biomolecular Simulation) [15]. Cette approche permet de décomposer l'énergie intermoléculaire en cinq termes : Electrostatique, Répulsion, Polarisation, Transfert de charge et Dispersion. Cette décomposition énergétique a l'ambition de reproduire les résultats *ab-initio* mais laisse aussi entrevoir* SIBFA *comme une potentielle approche interprétative permettant ainsi de mieux comprendre les origines physiques des interactions intermoléculaires.

Ce manuscrit de thèse s'organise autour de deux parties. La première est dédiée à la dynamique moléculaire. Tout d'abord, l'approche classique sera exposée, ainsi que les notions et principes importants autorisant à réaliser de telles simulations dans un contexte réaliste. L'utilité des résultats et analyses issus de la dynamique moléculaire sera ensuite discutée. Enfin, un second chapitre traitera des limites de cette approche en introduisant les champs de forces de seconde génération et notamment SIBFA, ainsi que les méthodes de calibration.

Par la suite, la seconde partie traitera plus concrètement de l'implémentation du potentiel SIBFA dans un programme de dynamique moléculaire. Dans un premier temps, nous nous attarderons sur la composante électrostatique, qui est souvent considérée comme étant la principale responsable de l'association de systèmes complexes. La reproductibilité des résultats référence via une approche de chimie quantique sera explorée, ainsi que son applicabilité sur de larges systèmes tout en préservant un faible impact computationnel. Enfin, un second chapitre traitera des interactions de VAN DER WAALS d'induction dans le cadre de la dynamique moléculaire, c'est-à-dire en détaillant forces et énergies. Ces composantes seront détaillées dans le contexte de SIBFA en considérant séparément les énergies de répulsion et de dispersion, ainsi que celles de polarisation et de transfert de charge.

Première partie

Introduction à la Dynamique Moléculaire

Chapitre 1

Dynamique Moléculaire

Sommaire

1.1 Approches classiques 10
1.1.1 Interactions intramoléculaires 10
1.1.2 Interactions intermoléculaires 14
1.2 Vers des simulations réalistes : la solvatation 16
1.2.1 Solvatation explicite 17
1.2.2 Solvatation implicite 20
1.3 Principe de la Dynamique Moléculaire 22
1.3.1 Introduction 22
1.3.2 Intégration des vitesses 23
1.3.3 Contrôle thermodynamique 24
1.3.4 Exploitation des résultats 25
1.4 Conclusions 30

1.1 Approches classiques

Le calcul de l'énergie des champs de forces classiques sont donc uniquement fonction des positions atomiques, qui sont finalement les seuls variables du système. Il existe plusieurs approches classiques dépendant de leur calibration. Certains potentiels sont ajustés à partir de propriétés expérimentales observables (enthalpie de vaporisation, densité, etc) [16, 17], tandis que d'autres le sont à partir de calculs de chimie quantique et de résultats expérimentals afin de reproduire l'énergie d'un système. En allant plus loin, on peut même distinguer des méthodes calibrées selon des catégories de systèmes (molécules organiques, macromolécules, etc) [18–20] . Ce traitement exprime l'énergie totale en deux types d'interactions : intra- et intermoléculaires. De plus, chacune d'elles contient plusieurs termes tendant à décrire les différentes contributions.

$$E_{\text{totale}} = E_{\text{intra}} + E_{\text{inter}} \tag{1.1}$$

1.1.1 Interactions intramoléculaires

Les interactions intramoléculaires se distinguent selon trois grandes « familles » de degrés de liberté :

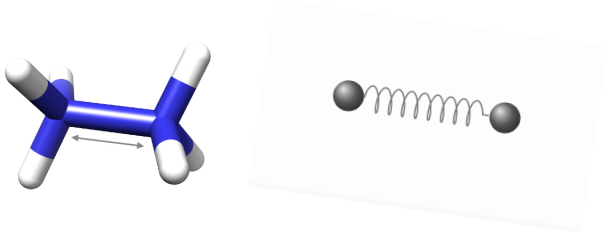
$$E_{\text{intra}} = E_{\text{liaisons}} + E_{\text{angles}} + E_{\text{dièdres}} \tag{1.2}$$

Les termes de liaisons et d'angles sont représentés par des potentiels harmoniques.

$$E_{\text{liaisons}} = \sum_{\text{liaisons}} K_l (l - l_0)^2 \tag{1.3}$$

Ce calcul se fait donc par paires d'atomes. A partir d'une longueur de liaison à l'équilibre l_0 estimée par optimisation géométrique ab-initio, il est alors possible d'évaluer l'énergie selon la position des atomes impliqués dans une liaison covalente. K_l est une constante de force calibrée à partir de calculs référence. Physiquement elle décrit la rigidité de cette liaison (voir figure

1.1). En pratique, l_0 et K_l sont spécifiques aux types d'atomes à chaque paire d'atomes.

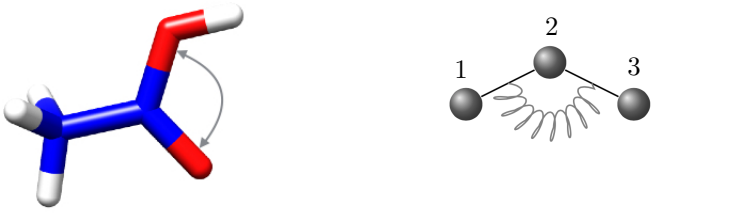


(a) (b)

FIGURE 1.1 – Schéma élongation liaison. Cas de l'éthane.

$$E_{angles} = \sum_{angles} K_\theta(\ \theta - \ \theta_0)^2 \tag{1.4}$$

Ce deuxième terme en 1-3 (voir figure 1.2) tend à décrire la capacité de deux liaisons à « s'ouvrir » et à se contracter selon la valeur de θ. De la même manière que dans l'équation précédente, K_θ est une constance de force associée à un couple de trois atomes spécifiques. Enfin, θ_0 est la valeur à l'équilibre de l'angle formé par ces deux liaisons.



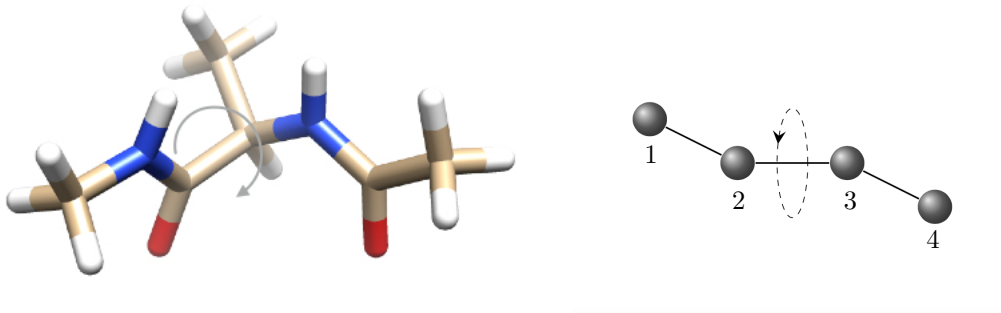
(a) (b)

FIGURE 1.2 – Schéma déformation angulaire. Cas d'un acide.

Ces deux premiers termes intramoléculaires correspondent aux degrés de liberté laissant peu de fluctuation énergétique possible. En effet, leurs constantes de force K_l et K_θ sont respectivement les plus importantes de la composante intramoléculaire. Ceci traduit bien le caractère covalent puisque la déformation de longueur de liaison ou d'angle requiert de grandes quantités énergétiques à fournir au système.

Parmi les termes intramoléculaires, les variations de dièdres sont les plus susceptibles de se produire au cours d'une dynamique moléculaire en franchissant des barrières de rotation, laissant ainsi apparaitre une structure « réorientée » qui peut changer de manière substantielle l'interaction d'un ligand dans un site actif par exemple.

$$E_{\text{dièdres}} = \sum_{\text{dièdres}} \sum_{n} V_n[1 + cos(n\phi - \phi_0)] \tag{1.5}$$



(a) (b)

FIGURE 1.3 – Schéma déformation diédrale. Cas du dipeptide de l'alanine.

Ce terme fait intervenir quatre atomes consécutifs séparés par trois liaisons covalentes. Il décrit donc la susceptibilité de rotation autour de la liaison formée par les atomes 2-3 (voir figure 1.3). Celle-ci est aussi d'origine intermoléculaire et notamment électrostatique en modélisant l'interaction des atomes 1 et 4. La périodicité de cette torsion est contrôlée par $n = 1, 2, 3$ et par une constante V_n associée, ce qui permet de mieux décrire la barrière

de rotation selon l'interaction des atomes 1 et 4. ϕ est l'angle de torsion instantané formé par les quatre atomes et ϕ_0 est la valeur à l'équilibre. Les paramètres V_n, n et ϕ_0 dépendent de la nature de la liaison et des atomes considérés dans le calcul.

Ces trois termes ne sont pas toujours suffisants pour prédire une géométrie au niveau quantique. Afin d'affiner l'évaluation des forces intramoléculaires, deux termes sont ajoutés dans les champs de forces classiques modernes. Le premier traite les dièdres impropres. Trois atomes sont définis dans un plan et le quatrième l'est hors (voir figure 1.4). Ceci est approprié dans la description d'un carbone hybridé sp^2 afin d'assurer la planéité de ce fragment.

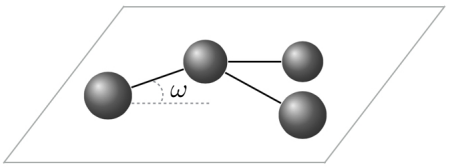


FIGURE 1.4 – schéma dièdre impropre

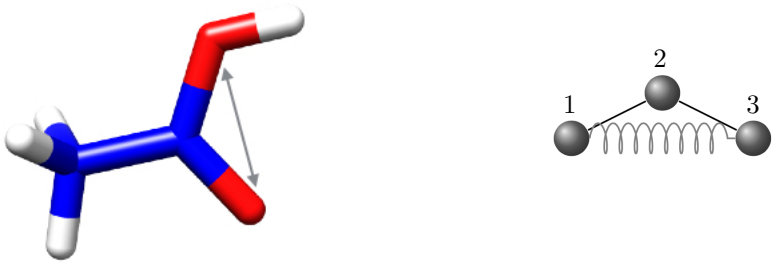
$$E_{\text{dièdres impropres}} = \sum_{\text{dièdres impropres}} K_\omega(\omega - \omega_0)^2 \tag{1.6}$$

Enfin le terme dit croisé d'UREY-BRADLEY (voir figure 1.5) est une correction des approximations de valences présentées dans les équations 1.3 et 1.4.

$$E_{\text{UREY-BRADLEY}} = \sum_{\text{UREY-BRADLEY}} K_{\text{UB}}(r - r_0)^2 \tag{1.7}$$

où r est la distance instantanée entre les atomes 1 et 3 et r_0 celle à l'équilibre,

c'est-à-dire à θ_0 (voir equation 1.4) et K_{UB} la constante de force associé.



(a) (b)

FIGURE 1.5 – Schéma correction d'UREY-BRADLEY. Cas de l'acide acétique.

Ces termes intramoléculaires ne seront pas abordés au cours de cette thèse mais ces modèles sont utilisés dans la majorité des champs de forces classiques.

1.1.2 Interactions intermoléculaires

Selon l'approche supramoléculaire [1], les interactions intermoléculaires sont les principales forces motrices des complexations. Physiquement, ce sont les électrons et plus particulièrement leur mobilité qui est responsable de ce type d'interaction. Selon l'approximation de BORN-OPPENHEIMER [21], leur mouvement est négligé. Parmi ces interactions non-covalentes, la plus importante est l'interaction électrostatique, avec les liaisons hydrogène typiquement [22, 23]. Pour modéliser ce type d'interaction, une simple loi coulombienne est appliquée. Des charges partielles sont affectées à chaque atome de manière spécifique selon leur type et leur environnement proche. Ces charges reflètent finalement l'électronégativité, qui décrit la tendance d'un atome à céder ou accepter un électron d'un voisin pour former une liaison. Plus la différence entre deux atomes est élevée, plus la liaison sera

forte. Ces charges classiques ont donc pour but de représenter la susceptibilité électronique d'attraction ou de répulsion entre atomes non consécutifs.

$$E_{\text{électrostatique}} = \sum_{i>j} \frac{q_i q_j}{r_{ij}} \tag{1.8}$$

Ce calcul s'effectue donc par paires d'atomes notées i et j et r_{ij} est la distance les séparant.

Un traitement à court et longue distance est aussi utilisé pour les interactions dites de VAN DER WAALS. Elles sont le plus souvent modélisées par un potentiel de Lennard Jones qui décrit selon deux termes : la répulsion à courte distance et à plus longue distance, la dispersion [24]. Leur dénomination VAN DER WAALS pourrait être perçue comme un abus de langage puisque ces interactions sont à longue distance.

$$E_{\text{Lennard Jones}} = \sum_{\text{atomes non liées}} K_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{1.9}$$

En pratique, selon les champs de forces, d'autres potentiels sont proposés mais la forme reste similaire, c'est-à-dire deux termes dont celui avec la puissance la plus élevée pour modéliser le caractère répulsif du couple d'atomes et celui avec la plus petite, son caractère dispersif qui lui est attractif (voir figure 1.6).

Comme dans le cas des interactions intramoléculaires, chaque potentiel décrit ci-dessus est uniquement fonction des positions des atomes qui sont finalement les seules variables des approches classiques.

Les différentes approximations menant aux champs de forces classiques ont donc pour but de traiter de grands systèmes, de plusieurs milliers

d'atomes, d'intérêt biologique. Pour se placer dans ce cadre, il est donc nécessaire de modéliser un environnement pertinent en allant au delà du modèle phase gazeuse.

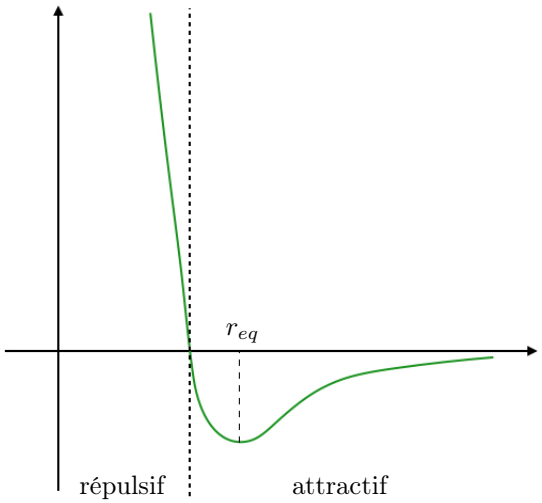


FIGURE 1.6 – tracé schéma d'un potentiel du type 12-6

1.2 Vers des simulations réalistes : la solvatation

Un environnement plus proche de la réalité biologique est donc nécessaire pour permettre de faire le lien avec les expérimentateurs notamment et ainsi exploiter les résultats de dynamique moléculaire. Pour cela, l'inclusion d'un solvant est indispensable. Cependant, une étude préliminaire en phase gaz reste intéressante pour comparer ensuite à des simulations du même système solvaté. Ceci permet d'évaluer l'effet du solvant sur le système étudié d'un point de vue qualitatif en interprétant notamment les variations de certains degrés de liberté. Dans le cadre de simulation de type ligand-récepteur, la représentation en solution aqueuse est un bon compromis au vue de sa présence dans notre système sanguin (voir tableau 1.1). Le plasma sanguin

est donc un fluide aqueux, composé de cations, d'anions et de protéines mais aussi de gaz dissous (O_2, CO_2, N_2) [25].

espèces		concentration (g/dl)
eau		90 - 92
protéines		7 - 8
cations	Na^+	0.310 - 0.340
	K^+	0.016 - 0.021
	Ca^{2+}	0.009 - 0.011
	Mg^{2+}	0.002 - 0.003
anions	Cl^-	0.360 - 0.390
	bicarbonates	0.020 - 0.024
	phosphates	0.003 - 0.004

TABLE 1.1 – Composition du plasma sanguin en g/dl [25]

Pour ce faire, plusieurs approches sont possibles [26]. Il n'est pas juste suffisant d'ajouter quelques molécules autour du soluté. En effet, on rencontrerait alors des effets de bords et ces molécules pourrait dériver par rapport au soluté. Physiquement, l'organisation de l'eau autour du soluté, voire même dans le site actif ne serait pas représentative. Il est donc nécessaire d'ajouter un très grand nombre de molécules de solvant. Une autre alternative serait de modéliser le solvant, qui n'est constitué que d'un même type de molécule, par un continu qui produit un champ électrostatique sur le soluté.

1.2.1 Solvatation explicite

L'ajout d'un grand nombre de molécules de solvant est donc nécessaire pour assurer une modélisation rigoureuse d'un point de vue des interactions intermoléculaires, avec notamment l'organisation du solvant (voir figure 1.7). Cependant, on se heurte tout de même aux effets de bords qui biaisent les résultats. L'autre problème est le coût de calcul. En effet, le système soluté-solvant contient un très grand nombre d'atomes et voit donc le nombre d'interactions à calculer augmenter.

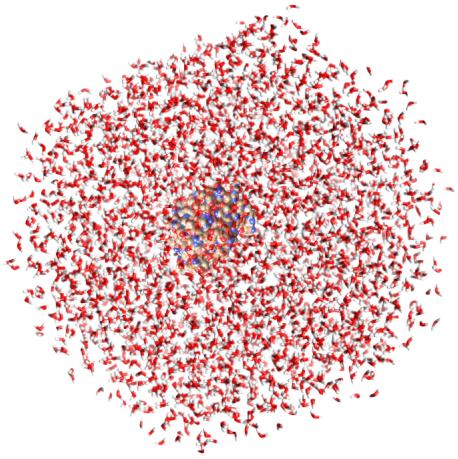


FIGURE 1.7 – Exemple de solvatation explicite d'un système.

Afin de pallier cette limite, les champs de forces se placent dans le cadre d'un système périodique en exploitant la sommation d'Ewald et plus particulièrement un algorithme issu de ces travaux : Smooth Particle Mesh Ewald (SPME) [27]. Cette méthode permet d'éviter les effets de bords puisque le système est traité en conditions périodiques mais aussi d'abaisser le coût de calcul en $N \log(N)$ (voir section 1.4 du chapitre 1 de la partie 2).
En pratique, cet algorithme est appliqué à l'énergie électrostatique qui est initialement de complexité N^2 pour N atomes et traite ce potentiel de manière schématique, tel que :

$$E_{\text{électrostatique}} = E_{\text{courte distance}}(r) + E_{\text{longue distance}}(r) \tag{1.10}$$

L'autre aspect de la solvatation explicite est la qualité du modèle de solvant utilisé. Le plus utilisé en simulation numérique est l'eau. Il existe donc différents modèles d'eau classiques, c'est-à-dire différente calibration et même formulation en introduisant des corrections de flexibilité par exemple [28–39].

Le premier modèle d'eau proposé est le modèle SPC, pour *single point charge* en anglais. Son extension, SPC/E ajoute un terme à la fonction d'énergie propre à cette molécule en considérant le dipole de l'eau (voir équation 1.11 ci-dessous).

$$E_{\text{pol}}^{spc/e} = \frac{1}{2}\sum_{i}\frac{(\mu-\mu^0)^2}{\alpha_i} \tag{1.11}$$

Le modèle TIP3P, pour *transferrable intermolecular potential three-sites*, reprend le même schéma. Enfin le modèle TIP4P, dérivé du précédent, ne dispose pas de charge sur l'oxygène mais sur un point fictif noté M sur la figure 1.8. Comme on peut l'observer sur le tableau 1.2, les paramètres ajustés pour les différents modèles présentés restent du même ordre de grandeur sans grande variation. Toutefois, l"utilisation d'un modèle d'eau dépend fortement de la fonction d'énergie employée, la transférabilité n'est donc pas assurée. Peut-être arrive-t-on aujourd'hui aux limites des approches classiques [40]. En effet, ces modèles d'eau sont essentiellement calibrés statistiquement sur des dynamiques moléculaires et pas forcément avec un regard plus fin sur les interactions intermoléculaire.

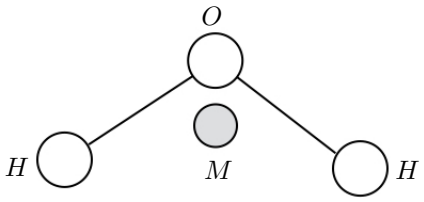


FIGURE 1.8 – schéma modèle d'eau en lien avec le tableau 1.2

Ces potentiels d'eau ont donc pour but de reproduire un environnement réaliste mais aussi des interactions intermoléculaires, c'est-à-dire des effets globaux et locaux respectivement.

	SPC	SPC/E	TIP3P	TIP4P
$r(OH)$	1,0	1,0	0,9572	0,9572
H-O-H	109,47	109,47	104,52	104,52
A	629,4	629,4	582,0	600,0
C	625,5	625,5	595,0	610,0
$q(O)$	-0,82	-0,8472	-0,834	0,0
$q(H)$	0,41	0,4238	0,417	0,52
$q(M)$	0,0	0,0	0,0	-1,04

TABLE 1.2 – Tableau de paramètres des modèles d'eau. Distance en Angströms, angle en degrés. A et C sont les constantes au numérateur de l'expression du potentiel de LENNARD-JONES. [41]

1.2.2 Solvatation implicite

L'alternative à une modélisation explicite du solvant est son traitement implicite [42–45]. Le solvant qui est composé de plusieurs molécules est modélisé sous forme d'un ensemble qui est décrit par une constante diélectrique (voir figure 1.9). Le champ de cet ensemble se substitue alors aux effets intermoléculaires du solvant explicite.

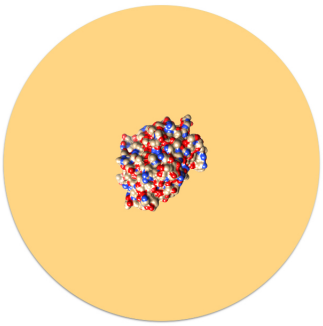


FIGURE 1.9 – Exemple de solvatation implicite d'un système.

Ce modèle tire son avantage du coût de calcul. Seulement, contrairement au modèle explicite, cette approche peut être vue comme une bonne représentation d'un effet global de solvatation. En effet, des interactions plus fines

comme l'organisation du solvant ou une coopérativité d'un réseau de molécules du solvant peuvent être négligées selon cette méthode. De même, une inclusion du solvant au sein d'un site actif participant à la complexation ligand-récepteur n'est pas prise en compte.

En conclusion de ces deux sections traitant des différentes approches de la solvatation, on peut dire que chacune a des avantages sur l'autre et inversement. Un bon compromis serait un traitement en deux couches (voir figure 1.10) :

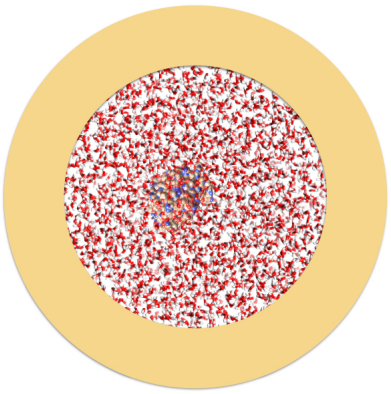


FIGURE 1.10 – exemple d'application d'un schéma deux couches de solvatation.

Toutefois, ce que l'on peut ajouter est que le choix de l'utilisation de l'une ou l'autre approche dépend de l'étude et des résultats attendus qui sont souvent, dans les approches classiques, plus qualitatifs que quantitatifs.

1.3 Principe de la Dynamique Moléculaire

1.3.1 Introduction

Un potentiel classique, ainsi défini, selon la mécanique moléculaire permet d'avoir une vision que l'on peut qualifier de « statique » du système étudié. Cette vision est loin de la réalité chimique en perpétuel mouvement, tout comme peut l'être une étude phase gaz dans un contexte biologique. Afin de se rapprocher des conditions réelles, c'est-à-dire en respect avec les conditions physiques d'un environnement vivant (comme la température et la pression), des simulations d'un système dynamique (en mouvement) sont donc essentielles. Pour cela, on utilise la dynamique moléculaire qui repose sur la seconde loi de NEWTON dont l'équation est la suivante :

$$\sum_{j \neq i}^{N} \mathbf{F_i} = m_i \mathbf{a_i} \tag{1.12}$$

où $\mathbf{F_i}$ correspond aux forces exercées par les atomes j (au nombre de N) sur i ayant une masse m_i et possédant une accélération $\mathbf{a_i}$ à un temps t. Le système est décrit par ses coordonnées cartésiennes, on peut donc facilement accéder à l'accélération $\mathbf{a_i}$ d'une particule i à l'instant t qui correspond à la dérivée seconde de sa position $\mathbf{r_i}$ (voir équation 1.13).

$$\mathbf{a_i} = \frac{d^2 \mathbf{r_i}}{dt^2} \tag{1.13}$$

Aussi, les forces $\mathbf{F_i}$ subites par i se déduisent du gradient (∇) de l'énergie potentielle, telle que :

$$\mathbf{F_i} = -\nabla_r E_{pot}(\mathbf{r_i}, t) \tag{1.14}$$

où E_{pot} est l'énergie potentielle associée à la particule i à l'instant t. Nous disposons dès lors d'une méthode pour résoudre l'équation de la deuxième loi de NEWTON. Cependant un « ingrédient » est manquant. En effet, comment évaluer la dynamique d'un système statique... ? En pratique, des vitesses sont générées aléatoirement par une distribution de MAXWELL-BOLTZMANN

afin de démarrer la trajectoire [46–48]. Toutefois, celles-ci ne peuvent rester constantes au cours de la simulation. Le second pilier de la dynamique moléculaire, après la loi de NEWTON, est l'intégration du temps, c'est-à-dire la propagation du mouvement. Enfin, reste alors à reproduire les conditions physiques du vivant à l'aide des ensembles statistiques : NVT, NVE et NPT. Ceux-ci seront abordés un peu plus loin dans la section 1.3.3. Il existe aussi d'autres approches, telles que la dynamique de Langevin ou Brownienne qui ne seront pas abordées ici. Rappelons toutefois que ces éléments de dynamique moléculaire sont applicables à toute fonction d'énergie, classique ou quantique. Ceci autorise donc notamment des dynamiques DFT, CAR-PARRINELLO, etc.

1.3.2 Intégration des vitesses

L'algorithme de VERLET fait partie des méthodes les plus utilisées d'intégration des équations du mouvement [49]. Le principe consiste à déterminer les vitesses et les positions des particules à l'instant $t + \Delta t$ à partir de celles à l'instant t, en appliquant un développement de Taylor au troisième ordre en $t + \Delta t$.

$$\mathbf{r_i}(t + \Delta t) = \mathbf{r_i}(t) + \mathbf{v_i}(t)\Delta t + \frac{1}{2}\frac{\mathbf{F_i}(t)}{m_i}\Delta t^2 + \frac{1}{6}\frac{d^3\mathbf{r_i}(t)}{dt^3}\Delta t^3 + O(\Delta t)^4 \quad (1.15)$$

et de la même manière en $t - \Delta t$.

$$\mathbf{r_i}(t - \Delta t) = \mathbf{r_i}(t) - \mathbf{v_i}(t)\Delta t + \frac{1}{2}\frac{\mathbf{F_i}(t)}{m_i}\Delta t^2 - \frac{1}{6}\frac{d^3\mathbf{r_i}(t)}{dt^3}\Delta t^3 + O(\Delta t)^4 \quad (1.16)$$

En combinant les équations 1.15, 1.16 et 1.13, on obtient alors, à l'ordre 4 près :

$$\mathbf{r_i}(t + \Delta t) = 2\mathbf{r_i}(t) - \mathbf{r_i}(t - \Delta t) + \mathbf{a_i}(t)(\Delta t)^2 \quad (1.17)$$

Ce schéma est réversible en temps, ce qui convient particulièrement aux systèmes conservatifs tels que les dynamiques moléculaires. De plus, il s'agit

d'un schéma itératif à deux pas, c'est-à-dire que l'intégration ne peut démarrer que si $\mathbf{r_i}(t)$ et $\mathbf{r_i}(t-\Delta t)$ sont connus. Or à $t = 0$, ne sont connus que $\mathbf{r_i}(t = 0)$ et $\mathbf{v_i}(t = 0)$. Il est alors admis, afin d'initier l'itération, que :

$$\mathbf{r_i}(t = -\Delta t) = \mathbf{r_i}(t = 0) - \mathbf{v_i}(t = 0)\Delta t + \mathbf{a_i}(t = 0)\frac{(\Delta t)^2}{2} \tag{1.18}$$

En pratique, ce Δt correspond au pas d'intégration, plus souvent appelé pas de temps. D'un point de vue physique, ce dernier doit être plus petit que la période de vibration la plus courte du système afin de décrire rigoureusement la trajectoire. Ces vibrations les plus rapides correspondent aux liaisons du type hétéroatome - hydrogène de l'ordre de la dizaine de femtosecondes. Un pas de temps Δt de 1 fs est un bon compromis. Cependant, cette approche n'est pas adaptée dans le cas d'étude de constante cinétique. En effet, la formulation de la vitesse n'est qu'approximée. Toutefois, son avantage est son faible coût en mémoire.

Notons qu'il existe d'autres algorithmes d'intégration du mouvement, tels que LEAP-FROG [50], VERLET-vitesses [51] ou encore BEEMAN [52] qui dérivent de l'algorithme de VERLET.

1.3.3 Contrôle thermodynamique

Comme énoncé dans la section 1.3.1 de ce chapitre, les ensembles statistiques permettent se rapprocher de l'environnement du vivant. En pratique, en dynamique moléculaire, trois ensembles thermodynamiques sont majoritairement utilisés. Dans les trois cas ci-dessous, N représente le nombre de particules considérées qui est toujours constant.

NVT

Cet ensemble dit canonique permet de contrôler essentiellement la température du système à volume constant. Il est utilisé pour amener un système à température ambiante. En pratique, elle se situe autour de 300 Kelvin.

Cette première phase de simulation correspond à l'équilibration. Cette dernière n'est pas exploitable mais sert à préparer le système.
Maintenir une température constante et forcer son système à y évoluer permet d'observer sa stabilité thermodynamique en explorant l'espace des phases au sein d'un tel environnement. Les thermostats les plus courants sont Nosé-Hoover [53, 54], Andersen [55] et Berendsen [56].

NVE

Cet ensemble microcanonique autorise le système à évoluer sans contrôle finalement [57]. Il conserve une énergie totale, somme des énergies cinétique et potentielle, constante tout au long de la dynamique. Typiquement, cet ensemble est utilisé pour s'assurer de la conservation de l'énergie après modification du potentiel ou implémentation de nouveaux termes. Cette conservation permet de s'assurer de la qualité des gradients d'énergie. Cette approche laisse aussi la possibilité d'étudier la réorganisation d'un système en démarrant la dynamique à partir d'une géométrie optimisée pour un autre état électronique par exemple [58–60].

NPT

Cet ensemble dit isotherme-isobare décrit un système en contact avec un thermostat et un barostat. Dans ce cas particulier, le volume n'est pas constant. Typiquement, cet ensemble est utilisé lors de la relaxation et réorganisation du solvant. Il est également utilisé dans l'étude des changements de phases.

1.3.4 Exploitation des résultats

Dans cette section, la volonté est de montrer l'intérêt de résultats de dynamique moléculaire, en tant qu'outil et en lien avec des techniques expérimentales (anisotropie de fluorescence et dichroïsme circulaire entre autres)

dans un contexte biologique.

Exemple d'analyse « brute »

Le système, alors étudié, est un cofacteur redox : le tryptophane tryptophylquinone (TTQ), présenté sur la figure 1.11. Présent au sein de l'enzyme methylamine deshydrogénase, le groupement TTQ constitue le premier maillon de la chaine respiratoire d'une bactérie : *paracoccus denitrificans*. Des simulations de dynamique moléculaire de ces deux états électroniques ont été réalisées selon deux méthodes : DFT et DFTB [5, 61, 62]. Au-delà du contexte de ces travaux [58–60], une analyse conformationnelle systématique du TTQ selon ces deux approches a été conduite.

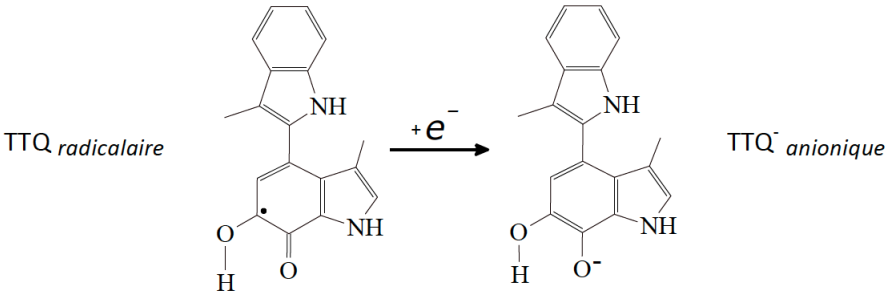


FIGURE 1.11 – schéma des deux états redox du cofacteur TTQ

Le tableau 1.3 reprend les longueurs de liaisons moyennes ainsi que les écart-type au cours des dynamiques moléculaires, ceci réalisé sur un fragment moléculaire du TTQ d'intérêt dans le cadre de cette étude.
Une simple analyse conformationnelle a permis de rationaliser les deux formes redox en identifiant les sites sur lesquels porter son attention, avec un écart-type plus important, de l'ordre de 10^{-1} alors que pour les régions de moindres intérêt finalement, il est de 10^{-2}. Ces sites d'intérêt sont signalés en couleur dans tableau 1.3. Certes, cette observation n'est pas pertinente dans le cas d'une rotation par exemple, qui éloignerait ou rapprocherait deux

atomes. Une rapide visualisation de la trajectoire lève le voile sur ce genre de situation. De plus, les résultats et observations étant similaires entre les deux méthodes, ces simulations ont aussi permis de valider l'utilisation de la DFTB par rapport à la méthode référence DFT.

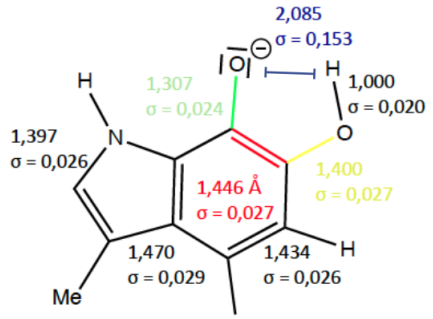
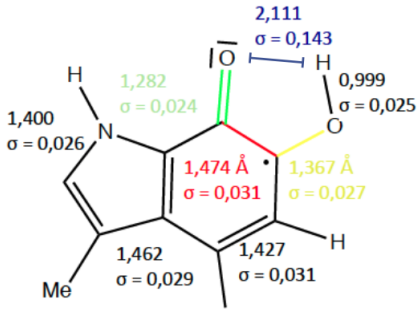
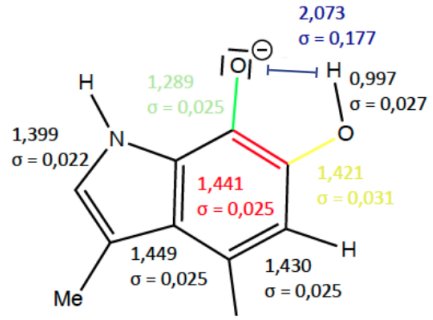
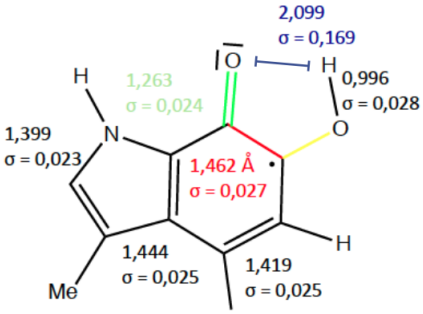
	TTQ anionique	TTQ radicalaire
DFT		
DFTB		

TABLE 1.3 – Longueurs de liaisons et écart-types en Angström du TTQ moyennés sur l'ensemble des trajectoires.

Anisotropie de fluorescence

La fluorescence permet d'évaluer les variations de l'environnement par des sondes fluorescentes intrinsèques (principalement le tryptophane pour les protéines) ou extrinsèques (fluorophores synthétiques greffés sur les molécules). Lorsqu'une molécule est excitée, l'un de ses électrons passe de son état fondamental à un niveau d'énergie d'excitation plus élevé. Ce changement d'orbitale est instable et l'électron retombe rapidement à son niveau d'énergie fondamental. Cette excitation se traduit par un phénomène de

fluorescence puis l'émission d'un photon lorsque l'électron revient à son état fondamental. L'excitation et l'émission sont le plus souvent illustrées par le diagramme de Jablonski (voir figure 1.12), représentant ainsi le transfert d'énergie entre molécules fluorescentes.

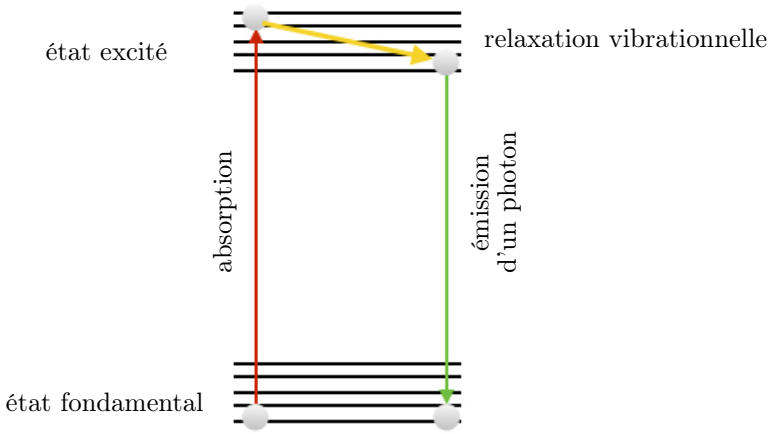


Figure 1.12 – schéma simplifié illustrant un diagramme de Jablonski

Lors de l'excitation par une source lumineuse polarisée, les fluorophores, dont les dipoles sont orientés parallèlement au vecteur de la lumière incidente, sont excités de manière préférentielle. On parle de photosélection des fluorophores. Si les molécules excitées étaient totalement immobiles, les photons libérés seraient émis suivant un vecteur parallèle au vecteur électrique de la lumière incidente. La fluorescence n'étant pas instantanée, les fluorophores vont subir un mouvement angulaire entre l'excitation et l'émission et l'orientation du vecteur d'émission des photons sera donc modifiée. Ce changement d'orientation conduit à une dépolarisation de la lumière émise. L'anisotropie est donc une mesure de la dépolarisation de fluorescence et est de ce fait directement reliée à la dynamique des macromolécules [63]. Elle fournit alors des informations sur la taille, la forme, la flexibilité et les

interactions des macromolécules. Elle permet aussi de déterminer les paramètres de liaison : constante de dissociation, coopérativité et stœchiométrie de liaison [64–66]. Ainsi, les analyses et observations issues de dynamiques moléculaires peuvent être comparées à celles données par l'anisotropie de fluorescence.

Dichroïsme circulaire

Le dichroïsme circulaire est une méthode de spectroscopie d'absorption dans l'ultraviolet (UV) qui fournit des informations sur la structure secondaire et les variations conformationnelles des biopolymères (ADN, protéines, peptides etc). Un échantillon possédant un chromophore chiral ou placé dans un environnement asymétrique, absorbe différemment la lumière polarisée circulairement à droite et à gauche. Le spectre dichroïque correspond à la différence d'absorbance entre ces deux types de lumière, pour chaque longueur d'onde. L'allure des spectres dichroïques varie donc suivant le type de repliement en hélices α, feuillets β et structures désordonnées (voir figure 1.13) [67, 68]. Le chromophore, dans le cas des protéines, est la liaison peptidique et plus précisément son groupement amide. Il se caractérise par une délocalisation des électrons sur ces atomes C, O et N mais aussi par deux doublets d'électrons libres sur l'oxygène du carbonyle. Le signal en dichroïsme circulaire des protéines résulte de la perturbation de ce chromophore par le carbone α asymétrique voisin à la liaison peptidique. Ajoutons que dans les structures secondaires, la somme des moments de transition des groupements peptidiques va engendrer un spectre caractéristique dit signature de la structure secondaire. Pour l'hélice α, la signature se caractérise par une bande positive à 190 nm et deux bandes négatives à 208 et 222 nm. L'intensité de la bande à 222 nm nous renseigne sur le pourcentage d'hélice dans le polypeptide alors que le spectre de la structure désordonnée (pelote aléatoire) est caractérisé par une bande négative intense à 195-200 nm et

une autre de plus faible intensité négative ou positive à 210-215nm (voir figure 1.13). Des méthodes de « déconvolution » des spectres de dichroïsme circulaire permettent de quantifier ces différents éléments de structure secondaire [69].
Des analyses conformationnelles de dynamiques moléculaires peuvent donc être comparées aux informations obtenues par dichroïsme circulaire.

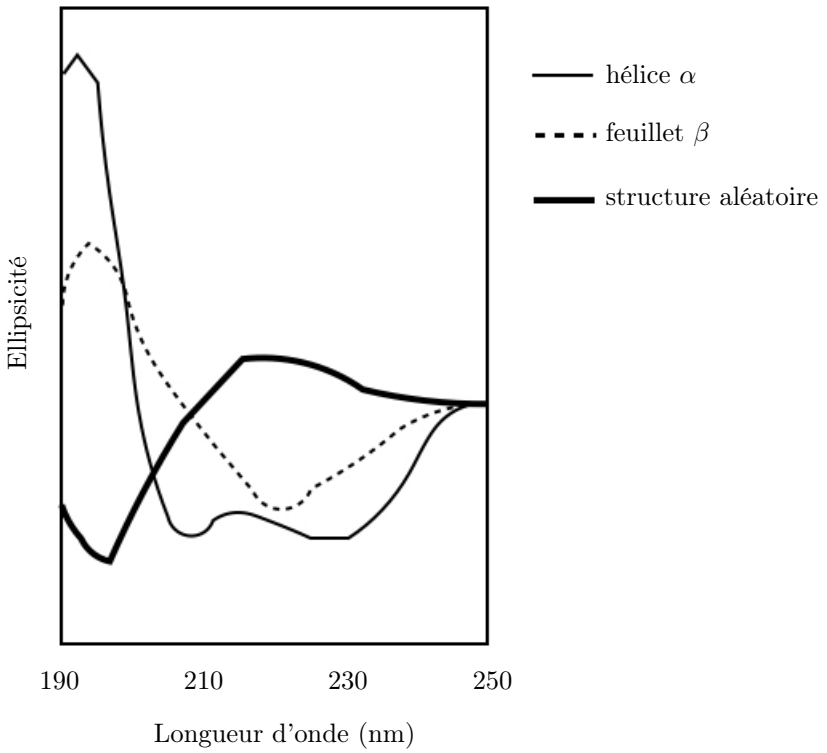


FIGURE 1.13 – Illustration de spectres de différentes structures secondaires obtenus par dichroïsme circulaire

1.4 Conclusions

La dynamique moléculaire est un outil puissant pour la biochimie, mêlant ainsi chimie et physique. (i) Chimie dans la compréhension des interac-

tions, non-covalentes particulièrement. (ii) Physique, grâce aux ensembles thermodynamiques permettant ainsi de se rapprocher des conditions biologiques et aux intégrateurs de temps de réaliser des dynamiques moléculaires. Toutefois, il est nécessaire de s'intéresser à ces deux domaines pour bien comprendre quels méthodes ou algorithmes utilisés afin de réaliser des dynamiques dans un cadre rigoureux.

De plus, la relation des phénomènes microscopiques (simulations atomiques) aux méso- et macroscopiques (méthodes expérimentales) existe. La section 1.3.4 de ce chapitre ne reprend que deux de ces approches expérimentales, mais sont orientées dans le contexte des biomolécules. De telles études préliminaires sont peu coûteuses par rapport aux méthodes expérimentales.

Cependant, le modèle classique s'appuie sur une méthodologie qui n'a plus connu de profonds changements depuis des décennies et arrive peut-être à ses limites [40]. D'autres méthodes existent, ajoutant par exemple des corrections de polarisation (voir sections 2.1.1 et 2.1.2 du chapitre 2) tout en gardant le même modèle, tendant ainsi vers un caractère de plus en plus empirique. Pour autant, des potentiels dits de deuxième génération voire même troisième génération vont plus loin dans la compréhension et la modélisation des interactions intermoléculaires, qui sont l'objet du prochain chapitre.

Chapitre 2

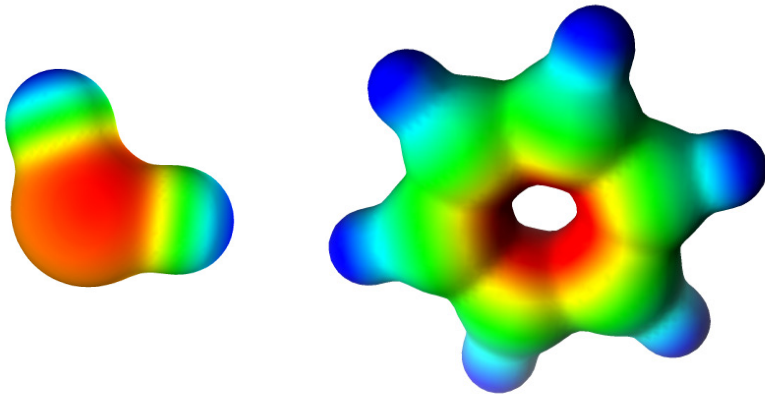
Introduction à SIBFA

Sommaire

2.1 Vers un meilleur traitement des interactions électroniques 34
2.1.1 Oscillateur de Drude 35
2.1.2 Charges fluctuantes 36
2.1.3 Pour aller plus loin 37
2.2 Le potentiel SIBFA 38
2.2.1 Electrostatique 40
2.2.2 Répulsion 41
2.2.3 Polarisation 42
2.2.4 Transfert de charge 43
2.2.5 Dispersion 43
2.3 Calibration 44
2.4 Conclusions & Perspectives 47

2.1 Vers un meilleur traitement des interactions électroniques

Les interactions responsables de la formation de complexes supramoléculaires sont les interactions intermoléculaires [1]. Leur origine est purement électronique. Or, elles ne sont traitées que sous la forme de deux potentiels (voir sous-section 1.1.2) représentant trois types d'interactions. Pour rappel, il s'agit de l'interaction électrostatique, de la répulsion et de la dispersion. De plus, au-delà des approximations dues aux approches classiques, il n'est pas rare de retrouver dans la littérature, en biochimie ou biophysique, des interprétations d'interactions non-covalentes en se basant simplement sur des mesures de distances entre hétéroatomes. Même si ce type d'approche a donné lieu à de bons résultats, des améliorations du traitement des interactions intermoléculaires ont leur place dans le développement de nouvelles méthodes. En effet, reproduire des effets quantiques électroniques à partir de potentiels classiques nécessite une analyse plus fine, tout comme le sont ces interactions. La mobilité électronique intra-fragment moléculaire de la densité électronique n'est pas prise en compte par les champs de force de première génération. Il s'agit de la polarisation. Comme on peut le voir sur la figure 2.1, la répartition des charges sur les fragments représentés n'est pas uniforme. D'un point de vue global, on voit clairement que la molécule d'eau forme un dipole bissecteur de l'angle H-O-H, alors que le benzène est composé de six dipoles orientés selon les axes des liaisons C-H mais dont la somme est nulle. Cette simple observation donne tout son sens quant au schéma adopté pour une meilleure description de la densité électronique (voir section 2.2.1).



(a) eau (b) benzene

FIGURE 2.1 – Potentiel électrostatique moléculaire.

Selon les approches classiques, cette distribution restera la même au cours d'une dynamique moléculaire. Physiquement, cela signifierait qu'elle ne se retrouverait pas affectée par ses voisins et réciproquement : elle-même n'affecterait pas ses voisins. En restant dans le cadre classique (avec une énergie électrostatique basé sur des charges ponctuelles), des modèles de polarisations existent : l'oscillateur de Drude [38, 70] et les charges fluctuantes [71–73].

2.1.1 Oscillateur de Drude

Dans le modèle de l'oscillateur de Drude, la polarisation est prise en compte par une charge partielle reliée à chaque atome du système, exceptés les hydrogènes, par un oscillateur harmonique comme dans le cas des potentiels des interactions intramoléculaires (voir sous-section 1.1.1). L'approximation faite sur les hydrogènes se justifie par la bonne description de sa structure électronique avec une simple charge. Ces points fictifs portant une charge peuvent donc se déplacer par rapport à l'atome auquel ils sont reliés

au cours d'une minimisation de l'énergie ou d'une dynamique moléculaire. Finalement, le mouvement de ces points fictifs représente la polarisation. Celui-ci est dépendant de deux paramètres : la constante associée à ce potentiel harmonique ainsi que la distance d'équilibre. Cette constante est déterminée à partir de la polarisabilité α (voir équation 2.1).

$$\alpha = \frac{q_D^2(A)}{k_D} \tag{2.1}$$

Ce modèle a été développé pour améliorer les modèles d'eau existant (voir sous-section 1.2.1). Trois charges sont centrées sur les atomes de la molécule, comme un modèle TIP3P [74, 75]. A l'image du modèle TIP4P [75], une charge maintenue par un oscillateur est reliée à l'oxygène, laissant ainsi plus de liberté qu'un point fixe fictif. La polarisation est prise en compte de manière globale dans le calcul de l'énergie électrostatique. De plus, ce modèle bénéficie d'une fonction de THOLE d'atténuation (en anglais : *Thole's damping*) à courte distance [76].

$$\begin{aligned} E_{\text{électrostatique}} = &\sum_{A<B}^{N} \frac{q_C(A)q_C(B)}{|r_C(A) - r_C(B)|} + \sum_{A<B}^{N,N_D} \frac{q_D(A)q_C(B)}{|r_D(A) - r_C(B)|} \\ &+ \sum_{A<B}^{N_D} \frac{q_D(A)q_D(B)}{|r_D(A) - r_D(B)|} + \frac{1}{2}\sum_{A}^{N_D} k_D(r_D(A) - r_C(A))^2 \end{aligned} \tag{2.2}$$

où N_D et N sont le nombre de particules de Drude et les atomes non hydrogènes, q_D et q_C sont les charges de Drude et celle de son porteur, r_D et r_C leurs positions et k_D la force associée au potentiel harmonique entre la particule de Drude et son porteur.

2.1.2 Charges fluctuantes

Enfin, le modèle des charges fluctuantes est basé sur la méthode d'équilibration des charges (CHEQ) [77]. De la même manière que précédemment avec l'oscillateur de Drude, le potentiel électrostatique coulombien se voit

modifier. Cette modification tend ainsi à représenter la réponse de l'environnement sur une charge partielle. Cette réponse est donc différente selon l'environnement.

Finalement, peu de champs de forces distribués contiennent un terme de polarisation de façon explicite et non-additive.

2.1.3 Pour aller plus loin

Au delà du traitement de la polarisation explicite ou d'une approximation de celle-ci, des méthodes de champs de forces plus raffinées existent. Citons NEMO [78–80], AMOEBA [39, 81, 82], et SIBFA [8–13] notamment.

Ces trois champs de forces de seconde génération sont basés sur l'utilisation de multipoles distribués afin de décrire la distribution de charge. Cependant, leur schéma de distribution est différent. En effet, alors que NEMO et AMOEBA emploient des centroides atomiques, SIBFA étend ses sites multipolaires jusqu'au milieu de liaisons covalentes. L'idée est de mieux reproduire les interactions dues aux liaisons π riches en électrons et pouvant attirer un hydrogène d'un autre fragment par exemple.

Encore une fois selon des schémas différents, la polarisation est traitée de manière explicite et se voulant non-additive. NEMO est le seul champ de forces polarisables à considérer les moments multipolaires jusqu'aux quadrupoles dans le calcul de la polarisation et ce de manière systématique. En effet, SIBFA étend son modèle jusqu'aux quadrupoles pour le traitement des cations métalliques. Ceci est rendu possible par les hyperpolarisabilités. Les polarisabilités, ainsi que les moments multipolaires sont obtenus grâce à la procédure LOPROP [83]. AMOEBA et SIBFA partagent les méthodes de calculs des multipoles par l'analyse des multipoles distribués [84]. Ceci étant, le niveau de calcul quantique préliminaire à cela n'est pas le même (voir

sections 1.2 et 1.3 dans le chapitre 1 de la partie 2).

De plus, la formulation de la polarisation l'est aussi. AMOEBA se place dans l'approche d'APPLEQUIST [85] et SIBFA dans celle de GARMER et STEVENS [86].

Enfin, seuls SIBFA et NEMO décomposent l'énergie intermoléculaire au-delà de l'électrostatique et de la polarisation. Rappelons que SIBFA rajoute une correction à courte distance à l'énergie électrostatique [11, 12].
NEMO possède un terme d'induction global alors que SIBFA sépare cette contribution en polarisation et transfert de charge.

2.2 Le potentiel SIBFA

Le potentiel SIBFA s'appuie sur quatre principes :

séparabilité

L'énergie intermoléculaire s'exprime sous la forme de potentiels explicitement séparés. Ceci implique que chaque composante de cette énergie soit ajustée sur des valeurs référence quantiques obtenues par analyse de décomposition de l'énergie intermoléculaire (voir section 2.3).

Anisotropie

Le potentiel doit être en mesure de reproduire les composantes de l'énergie intermoléculaire quelles que soient les orientations spatiales, telles que des variations angulaires ou des approches hors plan, entre fragments évalués.

Non-additivité

La modélisation et le calcul d'un effet physique au sein d'un système ne doivent être compris que par un terme énergétique. Sinon cela montre la

coopérativité ou à l'inverse l'anti-coopérativité entre contributions.

Transférabilité

Le potentiel, ayant été calibré sur des valeurs référence de dimères, doit être capable de passer à des édifices polymoléculaires sans changer ces paramètres. Cette propriété est primordiale pour les multipoles.

Le potentiel SIBFA a été développé en France dans les années 80 par Nohad GRESH, Pierre CLAVERIE et Alberte PULLMAN [9]. Depuis une quinzaine d'années, il est essentiellement co-développé par Nohad GRESH et Jean-Philip PIQUEMAL ainsi que leur collaborateurs [8,87]. Cette approche permet de formuler l'énergie intermoléculaire sous cinq composantes (voir équation 2.3 et figure 2.2).

$$\begin{aligned} E_{\text{intermoléculaire}} = E_{\text{électrostatique}} \; + \; E_{\text{répulsion}} \; + \; E_{\text{polarisation}} \\ + \; E_{\text{transfert de charge}} \; + E_{\text{dispersion}} \end{aligned} \tag{2.3}$$

$E_{\text{électrostatique}}$: contribution électrostatique multipolaire avec correction à courte portée.

$E_{\text{répulsion}}$: échange-répulsion, principalement à courte-distance.

$E_{\text{polarisation}}$: la polarisation.

$E_{\text{transfert de charge}}$: le transfert de charge et

$E_{\text{dispersion}}$: dispersion et échange-dispersion.

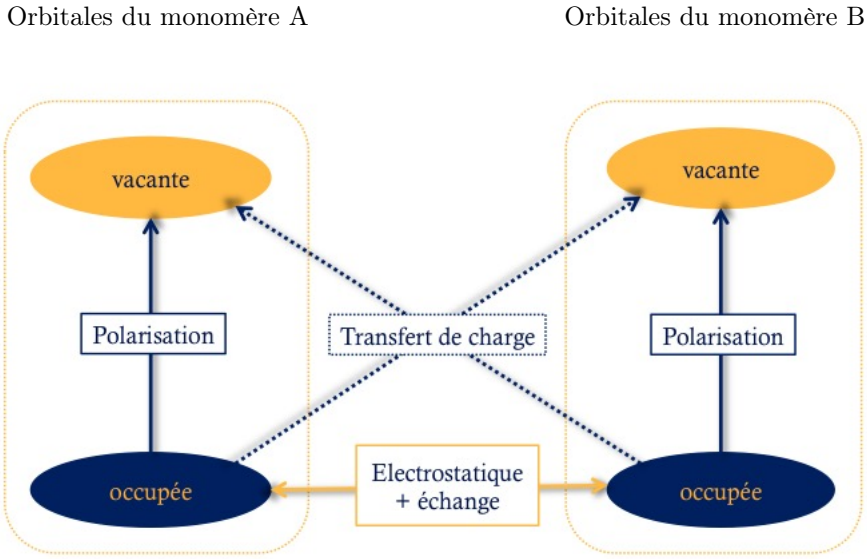


FIGURE 2.2 – schéma RVS (*Reduced Variational Space*) illustrant la décomposition des interactions intermoléculaires [88].

De plus, au-delà de l'utilisation de multipoles et de polarisabilité, l'une des principales caractéristiques de SIBFA est de considérer les paires électroniques libres, chacune porteuse de deux électrons, dans le calcul de la répulsion, la dispersion et le transfert de charge.

2.2.1 Electrostatique

$E_{\text{électrostatique}}$ se décompose sous la forme de six termes distincts :

$$\begin{aligned} E_{\text{électrostatique}} = E_{\text{mono-mono}} \; + \; E_{\text{mono-dip}} \; + \; E_{\text{mono-quad}} \\ + \; E_{\text{dip-dip}} \; + \; E_{\text{dip-quad}} \; + \; E_{\text{quad-quad}} \end{aligned} \tag{2.4}$$

L'équation 2.4 décrit les interactions entre charge (monopole) d'un atome ou site multipolaire i et charge, dipole et quadrupole d'un site j, ainsi que celles entre dipole de i et dipole et quadrupole de j et enfin, celle entre

quadrupole de i et j. Physiquement, l'approximation multipolaire permet de prendre en compte l'anisotropie de chaque site.

A cela s'ajoute une correction à courte distance de $E_{\text{mono-mono}}$, $E_{\text{mono-dip}}$ et $E_{\text{mono-quad}}$ qui sont les composantes les plus importantes [11, 12]. Cette correction est impérative pour décrire l'électrostatique au niveau quantique. En effet, cette erreur d'évaluation de l'énergie est due à l'approximation multipolaire qui n'est pas capable de reproduire un effet de recouvrement de densité qui tend à favoriser l'interaction : la pénétration de charge (voir section 1.3 du chapitre 1 de la partie 2). $E_{\text{mono-mono}}$ s'exprime sous la forme trois termes représentant les interactions noyau-noyau, électron-noyau et électron-électron. Dans le cas de $E_{\text{mono-dip}}$ et $E_{\text{mono-quad}}$, seule la charge est écrantée.

2.2.2 Répulsion

Le terme d'échange-répulsion, noté $E_{\text{répulsion}}$, est décrit sous trois sommes distinctes. En effet, la répulsion est évaluée entre paires libres du fragment A et celles de B, entre liaisons de chaque fragment et enfin entre liaisons et paires libres (voir equation 2.5).

$$\begin{aligned} E_{\text{répulsion}} = C1 \sum_{\text{paires libres}}^{A} \sum_{\text{paires libres}}^{B \neq A} & \text{rep(paire libre - paire libre)} \\ + C2 \sum_{\text{liaisons}}^{A} \sum_{\text{liaisons}}^{B \neq A} & \text{rep(liaison - liaison)} \\ + C3 \sum_{\text{liaisons}}^{A} \sum_{\text{paires libres}}^{B \neq A} & \text{rep(liaison - paire libre)} \end{aligned} \tag{2.5}$$

Le terme isolé $\text{rep}(i, j)$, avec i et j étant une paire libre et/ou une liaison, évalue le recouvrement des densités électroniques de i et j, noté S_{ij}. $C1$, $C2$

et $C3$ sont des constantes de normalisation. Enfin la répulsion considère ce recouvrement au carré S_{ij}^2 sous forme de deux termes :

$$\text{rep}(i,j) \propto \left(\frac{S_{ij}^2}{r_{ij}} + \frac{S_{ij}^2}{r_{ij}^2} \right) \tag{2.6}$$

2.2.3 Polarisation

La polarisation est une quantité locale dépendant notamment de l'environnement chimique proche. SIBFA utilise les polarisabilités distribuées sur les centroïdes des orbitales localisées de BOYS [89]. Celles-ci sont calculées selon la méthode de GARMER et STEVENS [86]. Selon cette approche, les polarisabilités sont distribuées sur un fragment moléculaire sans interagir ensemble. En effet, leur valeur ne sera modifiée que par le champ électrique global créé par les autres fragments moléculaires.

$$E_{\text{polarisation}} = -\frac{1}{2} \sum_j \Delta\mu(i) E_0(j) \tag{2.7}$$

L'équation 2.7 correspond à l'énergie de polarisation pour un centre polarisable i localisé sur un fragment. $\mu(i)$ est le dipole porté par le centre i. En effet, rappelons que SIBFA évalue la polarisation en considérant les dipoles, sans aller jusqu'aux quadrupoles.

$E(\Delta\mu(i))$ et $E_0(j)$ sont les champs électriques créés par les moments multipolaires induits et permanents respectivement. $\alpha(i)$ représente le tenseur de polarisabilité. Enfin $\Delta\mu(i)$ correspond au dipole induit du centre i et est de la forme suivante :

$$\Delta\mu(i) = \alpha(i) \sum_j^{x,y,z} E(\Delta\mu(i)) + E_0(j) \tag{2.8}$$

De plus, afin d'éviter les problèmes dits de polarisation catastrophe, une correction à courte portée est nécessaire. SIBFA utilise une fonction gaus-

sienne, notée $A(i,j)$ pour réduire la valeur du champ électrique crée par j sur i.

$$E_{j/i}^{final} = (1 - A(i,j))E_{j/i}^{init} \tag{2.9}$$

Cette contribution ne sera pas discutée au cours de cette thèse. Le modèle d'APPLEQUIST bien établi est préféré [85]. (voir chapitre 2 de la partie 1)

2.2.4 Transfert de charge

Le transfert de charge correspond à la mobilité électronique inter-fragment, c'est-à-dire la capacité d'une orbitale d'un fragment A à son état fondamental à partager sa densité électronique avec une orbitale virtuelle d'un fragment B (voir figure 2.2).

2.2.5 Dispersion

Le terme de dispersion se compose aussi de trois sommes distinctes, estimant les interactions atome-atome, atome-paire libre et paire libre-paire libre. L'équation 2.24 présente le cas général du traitement de la dispersion.

$$\begin{aligned} E_{\text{disp}} =& E_{\text{disp}}(\text{atome - atome}) + E_{\text{disp}}(\text{atome - paire libre}) \\ &+ E_{\text{disp}}(\text{atome - paire libre}) \end{aligned} \tag{2.10}$$

$$E_{\text{disp}}(i,j) = E_{\text{disp}}^{ld}(i,j) + E_{\text{échange-disp}}^{cd}(i,j) \tag{2.11}$$

La dispersion étant une interaction à longue distance, sa formulation dans le cadre de SIBFA le prend en compte séparant le traitement total (voir équation 2.24) en deux composantes. L'équation 2.11 présente l'interaction évaluée entre i et j (avec i et j représentant un atome ou une paire libre) et distingue donc un traitement à longue distance, noté ld et un à courte distance, noté cd, correspondant à l'échange-dispersion.

2.3 Calibration

D'un point de vue quantique, il est possible de rationnaliser ces interactions de manière plus précise que ce soit quantitativement ou qualitativement, grâce aux méthodes d'analyse de décomposition de l'énergie intermoléculaire (souvent appelé EDA pour *energy decomposition analysis*). Ces approches représentent un outil précieux dans la calibration de méthodes classiques et plus particulièrement dans le cadre de SIBFA qui tend à reproduire toutes les composantes physiques des interactions intermoléculaires. Dans le cas d'un complexe moléculaire AB formé par deux fragments A et B, l'énergie d'interaction est la différence entre l'énergie du complexe et la somme des énergies des fragments individuels, telle que :

$$E_{\text{intermoléculaire}} = E_{AB} - E_A - E_B \tag{2.12}$$

En effet, lorsque deux fragments moléculaires sont en interaction, ces méthodes permettent de déterminer de manière physiquement séparée les contributions mis en jeu. Il existe plusieurs types d'approches pour décomposer la quantité $E_{\text{intermoléculaire}}$ de l'équation 2.12, selon différents schémas basés sur :

(i) un principe variationnelle : La méthode CSOV (*Constrained Space Orbital Variation*) [90, 91] qui est basée sur la procédure développée par KITAURA et MOROKUMA [92] et RVS (*Reduced Variational Space*) [88] en font partie. Ces deux dernières représentent deux approches théoriquement différentes mais décrivent des termes énergétiques physiquement analogues. Cependant ce schéma variationnel autorise une décomposition telle que :

$$\begin{aligned} E_{\text{intermoléculaire}} = & E_{\text{électrostatique}} + E_{\text{échange-dispersion}} \\ & + E_{\text{polarisation}} + E_{\text{transfert de charge}} \end{aligned} \tag{2.13}$$

Aussi, ces approches sont limitées à des calculs de type Hartree-Fock (HF)

et DFT ou multi-référence SCF (MRSCF) pour RVS et CSOV respectivement. De plus il est nécessaire de tenir compte de l'erreur de superposition de base (BSSE *basis set superposition error*)

(ii) la théorie de la perturbation : SAPT (*Symmetry Adapted Perturbation-Theory*) est la seule méthode de type EDA post-HF, prenant ainsi en compte l'effet de corrélation électronique. De plus, elle est aussi la seule à décomposer l'énergie en cinq composantes physiques intermoléculaires (voir équation 2.3). Rappelons que la méthodologie SIBFA est formulée de cette manière, cela fait de SAPT une approche adaptée à la calibration et la confirmation de ce potentiel. Enfin, un effort d'implémentation de cette approche perturbative permet des calculs sur des dimères d'acides aminés selon des temps de calculs raisonnables via le programme PSI4 [93, 94]. Ceci étant particulièrement intéressant pour le développement et la validation de champs de forces classiques et polarisables qui traitent tout particulièrement de ce type de système moléculaire.

SAPT

Cette approche autorise la description de tout le potentiel intermoléculaire à courte distance, région majoritairement répulsive, jusqu'à longue distance, ayant alors un comportement asymptotique. SAPT est une extension de la théorie de RAYLEIGH-SCHRÖDINGER [95, 96]. En effet, selon RAYLEIGH-SCHRÖDINGER, la perturbation est faible devant l'énergie totale du complexe. LONDON a proposé une approche où le recouvrement des nuages électroniques des fragments moléculaires est négligé, en supposant donc les fragments A et B (voir équation 2.12) suffisamment éloignés [97]. Cette approximation n'est pas valide dans le cadre supramoléculaire et ligand-récepteur puisque les complexes ainsi formés sont à courte distance. D'autres approches ont été développées par la suite [98–100].

Selon le formalisme SAPT, l'hamiltonien H d'un dimère AB s'exprime en une somme de contributions individuelles de chaque monomère A et B (voir équation 2.14).

$$H = F_A + W_A + F_B + W_B + V \tag{2.14}$$

où F_i est l'opérateur de FOCK du monomère i et W_i le potentiel dit de fluctuation défini par MØLLER-PLESSET [100] tel que :

$$W_i = H_i - F_i \tag{2.15}$$

où H_i est l'hamiltonien total du monomère i. Enfin V est le potentiel perturbatif d'interaction du dimère AB.

Finalement le caractère perturbatif n'est évalué que pour W_i et V.
L'énergie d'interaction dans ce cadre s'exprime selon l'équation 2.16 [101].

$$E_{\text{intermoléculaire}} = \sum_{n=1}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(E_{\text{pol}^*}^{nkl} + E_{\text{échange}^*}^{nkl}) \tag{2.16}$$

où n est l'ordre du développement perturbatif, k et l représentent la corrélation électronique intramonomère. Dans la définition des termes énergétiques [94], on définit $v = k + l$, de telle sorte qu'une contribution est décrite par deux indices E^{nv}. E_{pol^*} ne correspond pas uniquement à la polarisation à proprement parler mais au développement de la polarisation dans le cadre SAPT. $E_{\text{échange}^*}$ est un terme répulsif, résultant de l'antisymétrisation des fonctions d'onde des monomères, qui permet de prendre en compte le recouvrement à courte distance. Pour plus de détails, se reporter à la référence [102].

L'approximation des intégrales bi-électroniques est devenue populaire dans le cadre de plusieurs méthodes de chimie quantique, grâce au gain de temps de calcul mais aussi de la réduction du coût de mémoire [103–108]. L'implémentation de la résolution de l'identité RI, aussi appelée *density*

fitted (DF)) permet donc d'abaisser le coût de tels calculs en intervenant dans l'évaluation du terme d'échange de l'équation 2.16. On parle alors de DF-SAPT. Cependant, ceci implique l'utilisation de bases gaussiennes auxiliaires pré-optimisées qui ne sont pas encore disponibles pour tous les atomes. Cette approche est similaire aux pseudopotentiels [109, 110]. Enfin l'approche SAPT dans le contexte des orbitales atomiques développée par KOHN-SHAM existe [111, 112], tout comme DF-DFT-SAPT [113–115].

Ces différentes approches issues du formalisme SAPT permettent la décomposition de l'énergie intermoléculaire en contributions physiques séparées. SAPT est donc un outil précieux et puissant dans la construction d'un potentiel tel que SIBFA reproduisant ces mêmes contributions dans un cadre rigoureux.

2.4 Conclusions & Perspectives

Les interactions non-covalentes sont donc responsables de la formation de complexes biomoléculaires. Ce type d'interaction est régi par le comportement électronique. Par comportement, il faut entendre :

(i) interactions « directes » comme l'électrostatique et VAN DER WAALS comportant la répulsion et dispersion.

(ii) et celles dues à la mobilité électronique, c'est-à-dire l'induction décrivant la polarisation et le transfert de charge.

Leur traitement requiert alors une méthodologie rigoureuse avec l'idée de respecter les principes de séparabilité, d'anistropie, de non-additivité et de transférabilité énoncés dans la section 2.2. Ces derniers se doivent d'être respectés pour rester dans un cadre robuste de la description, aussi

bien quantitative que qualitative, des interactions non-liées. La méthodologie SIBFA s'inscrit dans ce contexte.

La revisite de ce modèle passe aussi par une nouvelle calibration. La méthode choisie, SAPT, permet un raffinement de l'ensemble des contributions intermoléculaires. Cependant, au-delà de ces résultats, la première pierre sur laquelle reposent les approches polarisables, est les multipoles. Il existe plusieurs méthodes de calculs mais aussi plusieurs protocoles. Le niveau de calculs de chimie quantique est aussi à considérer. En effet, jusque là, SIBFA utilisait des sites multipolaires issus de calculs HF ou DFT. En raison des problèmes de délocalisation de la densité électronique connus en DFT, cette méthode n'est pas adaptée dans ce contexte [116]. De nouveaux multipoles au niveau MP2 seront étudiés, établissant ainsi un protocole systématique tenant compte de la corrélation électronique de manière claire.

L'apport scientifique d'une telle approche dans un programme de dynamique moléculaire pourrait grandement bénéficier à la pharmacologie, et plus particulièrement à la conception de médicaments (plus communément appelé *drug design*), ou encore de manière plus générale à la biologie. En effet, au-delà de résultats « bruts » de dynamique moléculaire à analyser ou exploiter, SIBFA permet aussi, grâce à sa formulation, d'interpréter physiquement les effets alors en jeu dans le mécanisme étudié. Cette réflexion interprétative pourrait mener à la proposition de molécules, améliorant ainsi la sélectivité, en compétition avec le ligand « naturel » :

(i) inhibant une étape d'un processus biologique en cascade dans le cas d'un virus.

(ii) activant alors une protéine dans le cas d'une maladie dégénérative comme Alzheimer.

Deuxième partie

Vers la Dynamique Moléculaire
SIBFA

Chapitre 1

Electrostatique

Sommaire

1.1 Introduction 54
1.2 Description charge atomique 58
1.3 Pénétration Electrostatique 61
1.3.1 Point de départ : quelle distribution multipolaire ? . . . 75
1.3.2 Comparaison & discussion de cp-GDMA [1] et cp-MEP . . . 80
1.3.3 Application des corrections de pénétration de charge . . 82
1.4 Conditions périodiques 88
1.5 Conclusions & Perspectives 132

Dans ce chapitre, nous nous intéresserons d'abord à la contextualisation de cette composante énergétique en chimie et plus particulièrement en supramoléculaire d'un point vue classique. Dans un second temps, nous présenterons les théories menant à une meilleure description de la charge atomique. Une troisième section sera dédiée à un problème particulier rencontré avec l'approximation multipolaire à courte distance. Enfin nous étudierons l'applicatibilité de notre traitement en conditions périodiques, indispensable pour réduire le coût de calcul en présence d'un solvant.

1.1 Introduction

Parmi les différents types d'interactions non-covalentes, l'interaction électrostatique est considérée comme la plus importante quantitativement [22, 23]. Elle gouverne l'arrangement d'un bon nombre de complexes supramoléculaires. On parle souvent de force motrice de complexation. Pour ces raisons, son traitement est crucial dans le cas d'étude de ce type de systèmes. Les premières approximations considérant des charges dites partielles sur les atomes et reprennent la formulation, empruntée à la physique, de la loi de Coulomb.

$$E_{coulomb} = \sum_i \sum_j \frac{q_i q_j}{r_{ij}} \tag{1.1}$$

où q_i et q_j sont les charges centrées sur les atomes i et j respectivement et le r_{ij} est la distance entre eux. Ce calcul par paires d'atomes est donc de complexité N^2. Si l'on s'intéresse à des systèmes d'intérêt biologique de plusieurs milliers d'atomes, ce qui est l'essence même de méthodes empiriques telles que les champs de forces, cette complexité devient alors une vraie limite au calcul. Pour s'affranchir de cette dernière, une distance r_{ij} seuil (dit *cutoff*) est utilisée, après laquelle l'électrostatique est nul, baissant ainsi le nombre d'interaction par paires d'atomes à calculer. Cette méthode

pose deux problèmes d'un point vue physique et mathématique :

(i) les effets électrostatiques portent à longue distance et peuvent être coopératif, c'est-à-dire qu'une somme de forces locales participe à un effet global. Dans les études bioinformatiques, le plus souvent classiques, on parle alors d'effets allostériques qui décrivent le même type de situation de phénomènes coopératifs. Au sein d'une protéine, il n'est pas rare d'observer une chaine constituée de molécules d'eau (appelée réseau d'eau) reliant ainsi des parties du système sur plusieurs angströms. Pour cela, j'ai choisi un système réaliste, une glucoamylase (code PDB : 1GAH) représentée figure 1.1, c'est-à-dire qu'il n'est pas le fruit d'une solvatation d'une simulation. Les molécules d'eau ont donc été piégées lors de la cristallisation. Sur la figure 1.1, on peut nettement voir l'eau autour et le long de sites hydrophiles de la protéine.

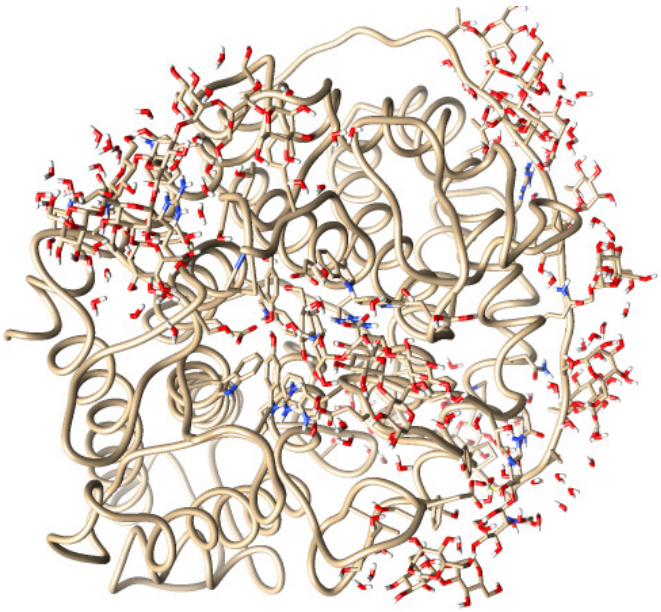
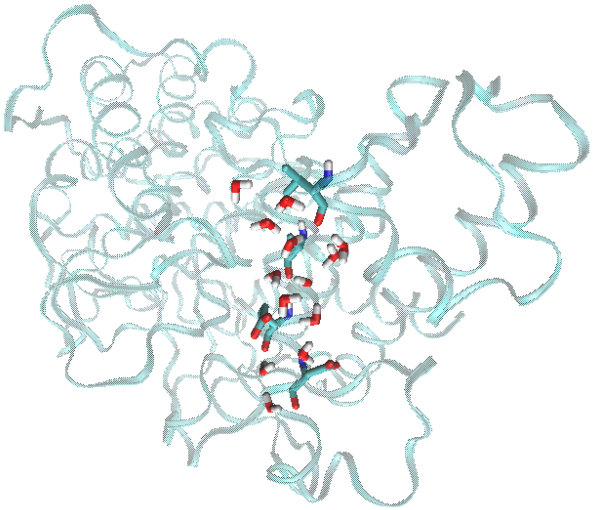


FIGURE 1.1 – exemple de réseau de molécules d'eau au sein de la protéine (code PDB :1GAH [117]).

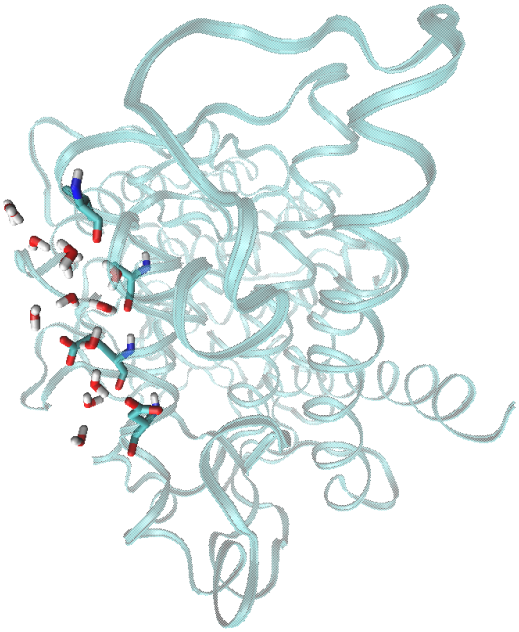
La figure 1.2 reprend le système de la figure 1.1 en mettant en valeur (sous la forme licorice) quatre résidus et des molécules d'eau. Ces résidus correspondent aux acides aminés notés VAL-346, GLU-280, ALA-276 et ASP-283 au sein du système 1GAH. Chacune de ces molécules ou fragments participent à un réseau intermoléculaire stabilisant. Outre la force de la liaison hydrogène entre dimères, les contributions énergétiques et plus particulièrement électrostatiques du voisinage permettent le maintien d'un tel réseau moléculaire par coopérativité. Il est donc nécessaire d'en tenir compte.

(ii) L'utilisation de ce type de valeur seuil crée une discontinuité dans le calcul de l'énergie qui se répercute sur sa non dérivabilité au voisinage de ce seuil. Or, cette propriété est indispensable pour assurer rigoureusement la bonne réalisation de dynamique moléculaire où interviennent les forces à chaque pas de calcul.

La liaison hydrogène est l'exemple typique d'une interaction non-covalente électrostatique (voir figure scan angle). L'anisotropie intrinsèque à ce type d'interaction n'est pas du tout prise en compte par les charges ponctuelles. Dans cette approximation, chaque densité électronique atomique est représentée par un point. Le graphique 1.3 montre le profil énergétique électrostatique en fonction de l'angle formé par la liaison $OH---O$. Il apparait alors clairement qu'il faut aller plus loin que les charges partielles pour traiter au mieux les interactions électrostatiques. Des approximations de chimie quantique permettent une meilleure description de la densité de charge autour de l'atome.

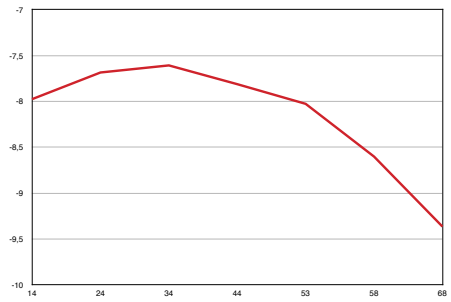


(a) vue de face

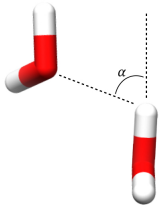


(b) vue de profil

FIGURE 1.2 – sélection réseau intermoléculaire au sein de 1GAH



(a)



(b)

FIGURE 1.3 – Energie électrostatique SAPT en fonction de la variation angulaire de la liaison hydrogène en kcal.mol et schéma moléculaire de la variation angulaire.

1.2 Description charge atomique

Les champs de force de première génération, tels que CHARMM [118] ou AMBER [119] utilisaient et utilisent toujours des charges ponctuelles centrées sur les atomes exploitant ainsi une notion vue dans les premières années d'apprentissage de la chimie : l'électronégativité. En effet, chaque atome du tableau périodique est décrit par un δq, qui représente sa propension à perdre ou gagner un ou plusieurs électrons et donc à former des liaisons hydrogène par exemple. Cet ensemble de charges atomiques est censé décrire une densité électronique moléculaire. La calibration de ces charges est réalisée à partir du potentiel moléculaire électrostatique (plus souvent appelé par l'anglicisme MEP : *Molecular Electrostatic Potential*) calculé par méthodes de chimie quantique (voir figure 2.1)

En 1942, LONDON, bien connu pour ces travaux sur la dispersion, proposa d'aller plus loin que cette description, en approximant cette densité par un ensemble de points chargés. LÖWDIN suggéra, quant à lui, de rajouter deux charges centrés sur les liaisons, formant ainsi un dipole caractéristique d'une liaison, en plus d'une charge atomique inspirée de MULLIKEN. Cette idée tend à rationaliser la notion de liaison chimique par des dipoles, prenant

ainsi en compte l'hybridation. En effet, une double liaison peut interagir de manière non-covalente à sa perpendiculaire avec un élément pauvre en électron. C'est ce qu'on observe avec les systèmes aromatiques notamment, riches en électrons, dans le cas d'interactions π (voir figure 1.4).

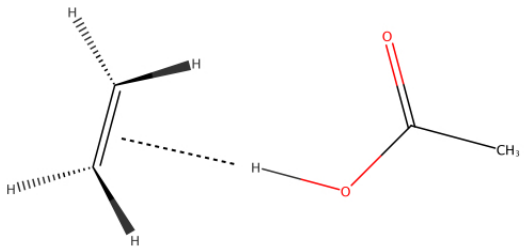


FIGURE 1.4 – exemple d'interaction π entre une double liaison et un atome d'hydrogène.

Ces contributions notables furent les premiers pas vers le développement multipolaire et de champs de force de seconde génération. Les travaux de REIN et CLAVERIE ont permis de rationaliser l'utilisation de moments multipolaires d'un point de vue chimique et mathématique [120, 121]. Le programme GDMA a, par la suite, permis de populariser leur utilisation [84].

$$W = \sum_{i=0} \sum_{j \neq i} W_{ij} \tag{1.2}$$

où W représente la somme des séries des interactions multipolaires. La décomposition des interactions multipolaires est représentée dans les équations de 1.3 à 1.11 :

Monopole - monopole

$$W_{00} = \frac{qq'}{|R|} \tag{1.3}$$

Dipole - monopole

$$W_{01} = (q'\boldsymbol{\mu} \cdot \mathbf{R})|R|^{-3} \tag{1.4}$$

Monopole - dipole

$$W_{10} = -(q\boldsymbol{\mu'} \cdot \mathbf{R})|R|^{-3} \tag{1.5}$$

Dipole - dipole

$$W_{11} = (\boldsymbol{\mu} \cdot \boldsymbol{\mu'}|R|^{-3} - 3(\boldsymbol{\mu} \cdot \mathbf{R})(\boldsymbol{\mu'} \cdot \mathbf{R})|R|^{-5} \tag{1.6}$$

Monopole - quadrupole

$$W_{02} = (q'\mathbf{R} \cdot Q \cdot \mathbf{R})|R|^{-5} \tag{1.7}$$

Quadrupole - monopole

$$W_{20} = (q\mathbf{R} \cdot Q' \cdot \mathbf{R})|R|^{-5} \tag{1.8}$$

Dipole - quadrupole

$$W_{12} = 2(\boldsymbol{\mu'} \cdot Q \cdot \mathbf{R})|R|^{-5} - [5(\boldsymbol{\mu'} \cdot \mathbf{R})(\mathbf{R} \cdot Q \cdot \mathbf{R})]|R|^{-7} \tag{1.9}$$

Quadrupole - dipole

$$W_{21} = -2(\boldsymbol{\mu} \cdot Q' \cdot \mathbf{R})|R|^{-5} + [5(\mathbf{R} \cdot Q' \cdot \mathbf{R})(\boldsymbol{\mu} \cdot \mathbf{R})]|R|^{-7} \tag{1.10}$$

Quadrupole - quadrupole

$$\begin{aligned} W_{22} =& \frac{2}{3}\frac{1}{6}[Q : Q']|R|^{-5} - \frac{20}{3}[\mathbf{R} \cdot Q \cdot \mathbf{R}]|R|^{-7} \\ &+ \frac{35}{3}(\mathbf{R} \cdot Q \cdot \mathbf{R})(\mathbf{R} \cdot Q' \cdot \mathbf{R})|R|^{-9} \end{aligned} \tag{1.11}$$

$\mathbf{R}$ est le vecteur du point non-prime au prime. q, $\boldsymbol{\mu}$ et Q sont la charge, le dipole et le quadrupole. En effet, et c'est là l'une des forces de ce développement, l'analogie avec la théorie des orbitales atomiques est évidente (voir figure 1.5).

Cependant, en chimie quantique, la densité électronique s'étend jusqu'à l'infini et ne peut donc être rigoureusement décrite par une distribution de charge. De plus, le développement multipolaire [122] se calcule autour d'une sphère centrée sur un site (atome ou liaison) dont le rayon est un paramètre

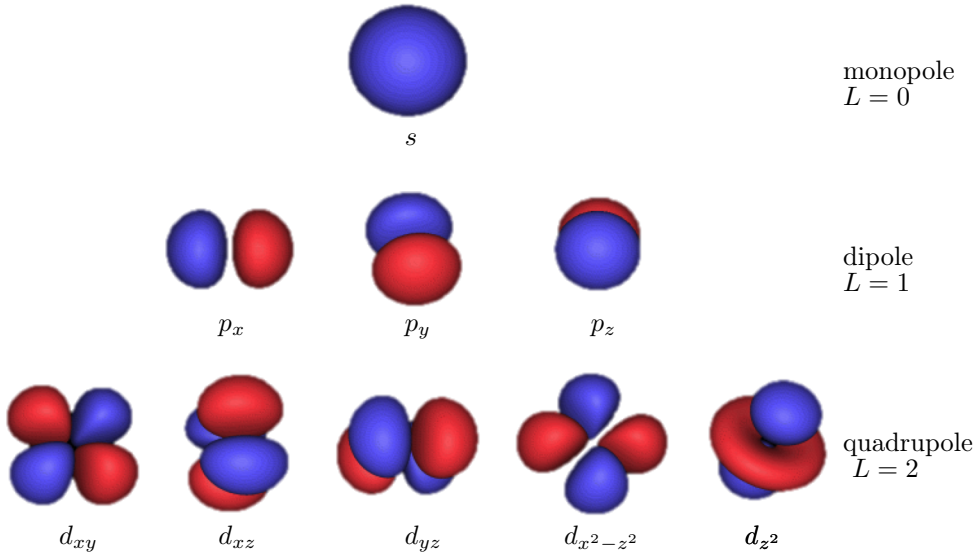


FIGURE 1.5 – Représentation des orbitales atomiques s, p et d ainsi que l'ordre du moment multipolaire associé (réalité avec MOPAC)

du calcul et au-delà de cette sphère, la distribution est donc considérée comme nulle.

Ce problème de description est beaucoup plus important à courte distance jusqu'aux alentours de l'équilibre où les densités atomiques électroniques se recouvrent. En effet, ce recouvrement produit une décroissance rapide de l'énergie, de type exponentielle. Or le développement multipolaire échoue à reproduire directement cet effet : la pénétration électrostatique. Il est alors nécessaire d'introduire un terme de correction.

1.3 Pénétration Electrostatique

La théorie de la perturbation (SAPT) nous permet de calculer chaque composante de l'énergie intermoléculaire. SHERRILL et al. ont démontré que la décroissance de l'électrostatique suit une forme exponentielle [123]. De plus, JEZIORKA et al. ont étudié le dimère de l'hélium afin de reproduire de

manière classique le potentiel intermoléculaire quantique sous la forme d'une série A_n/r^n [101]. Une décroissance à courte portée de type exponentielle est observée et la qualifie même de « phénomène inconnu ». Il est intéressant de noter ce type de décroissance électrostatique sur un dimère de gaz rare alors qu'il ne possède pas de charge partielle et par extension de multipoles. Nous avons étudié ce potentiel sur d'autres dimères de gaz rares le long de l'axe diatomique pour l'Hélium et l'Argon. Le même type de décroissance est clairement observé à courte portée et semble corrélé à la taille de l'atome (voir figure 1.6).

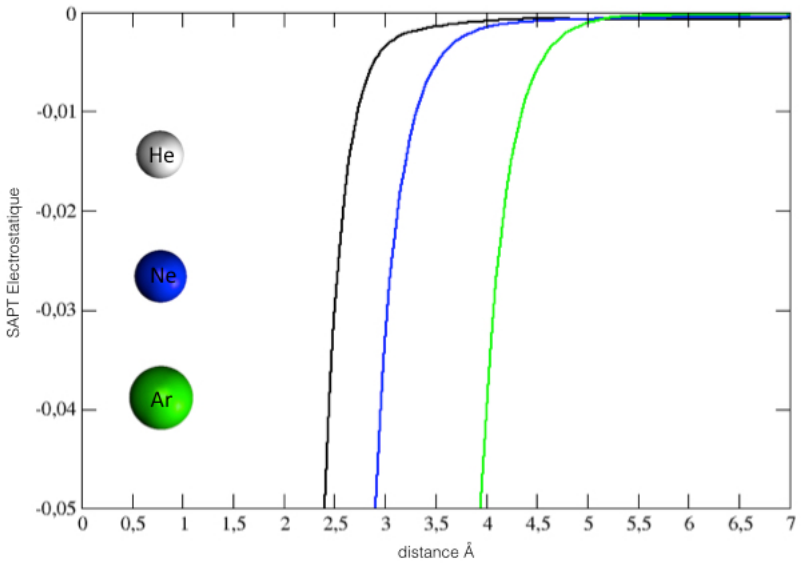


FIGURE 1.6 – Energie électrostatique SAPT (en kcal/mol) : scan intermoléculaire de dimères de gaz rares.

Ceci suggère un traitement au-delà des multipoles, en considérant par exemple la valence électronique, décrivant ainsi la couche électronique externe, première responsable des interactions non-covalentes, ainsi que le rayon atomique.

Avant de discuter d'un terme de correction, rappelons qu'en chimie quan-

tique, l'énergie électrostatique entre deux monomères s'exprime telle que :

$$\begin{aligned} E_{\text{elec}} = & -2\sum_{i}\sum_{\nu} Z_\nu \int \frac{|\varphi_i(1)|^2}{r_{i\nu}} {dr_1}^3 - 2\sum_{j}\sum_{\mu} Z_\mu \int \frac{|\varphi_j(2)|^2}{r_{j\mu}} {dr_2}^3 \\ & + 4\sum_{i}\sum_{j} \int \frac{|\varphi_i(1)|^2 |\varphi_j(2)|^2}{r_{12}} {dr_1}^3 {dr_2}^3 + \sum_{\mu}\sum_{\nu} \frac{Z_\mu Z_\nu}{r_{\mu\nu}} \end{aligned} \tag{1.12}$$

où μ (ν) et φ_i (φ_j) sont le noyau et les orbitales moléculaires non-perturbées du monomère A (B) et Z est le numéro atomique. L'énergie se décompose donc en 3 composantes, décrivant ainsi la répulsion noyau-noyau, l'attraction électron-noyau et la répulsion électron-électron. Afin de reproduire cet effet d'origine purement quantique, et ce malgré un traitement classique, le potentiel SIBFA s'appuie donc sur l'équation 1.12.

D'autres corrections ont été proposées [124–126]. Toutes introduisent un ou plusieurs termes exponentiels, au niveau de l'interaction entre monopoles.

L'article inséré ci-dessous propose un nouveau traitement de la pénétration de charge inspiré de celui de SIBFA mais se limitant, lui aussi, aux interactions entre monopoles. La calibration de cette correction et des multipoles est aussi développée. Ces derniers dérivent du protocole GDMA puis sont ajustés sur le potentiel électrostatique moléculaire (MEP). Certaines composantes sont par la suite réajustées pour se rapprocher des valeurs référence. Cette implémentation a été réalisée dans TINKER, en utilisant donc de nouveaux moments multipolaires. De plus, un nouvelle base de donnée d'énergie intermoléculaire de systèmes, dérivant du S66 [127], mais encore plus étendu jusqu'à cent systèmes à l'équilibre et à son voisinage. Ces calculs SAPT ont réalisé au niveau MP2/aug-cc- pVTZ. Le lien de cette appendice est présent en Annexe A de ce manuscrit. Cette base de donnée saura être utile pour la construction et la validation d'un champ de forces polarisables sur un large panel de systèmes représentatifs d'interactions non-covalentes.

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field

Qiantao Wang,[†,‡] Joshua A. Rackers,[§] Chenfeng He,[†] Rui Qi,[†] Christophe Narth,[∥] Louis Lagardere,[∥] Nohad Gresh,[∥] Jay W. Ponder,[⊥] Jean-Philip Piquemal,*[,∥] and Pengyu Ren*[,†]

[†]Department of Biomedical Engineering and [‡]Division of Medicinal Chemistry, College of Pharmacy, The University of Texas at Austin, Austin, Texas 78712, United States

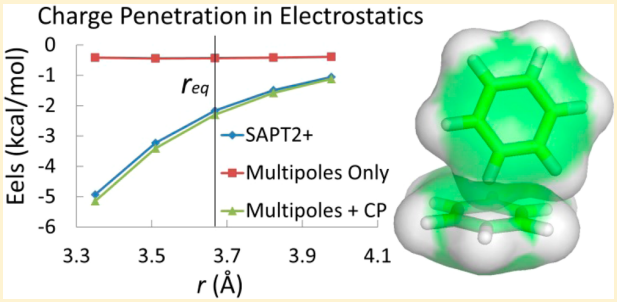
[§]Computational and Molecular Biophysics Program, Division of Biology & Biomedical Sciences, Washington University in St. Louis, St. Louis, Missouri 63110, United States

[∥]Laboratoire de Chimie Théorique, Sorbonne Universités, UPMC Paris 06, UMR 7616, Case Courrier 137, 4 Place Jussieu, F-75005 Paris, France

[⊥]Department of Chemistry, Washington University in St. Louis, St. Louis, Missouri 63130, United States

Supporting Information

ABSTRACT: Classical molecular mechanics force fields typically model interatomic electrostatic interactions with point charges or multipole expansions, which can fail for atoms in close contact due to the lack of a description of penetration effects between their electron clouds. These short-range penetration effects can be significant and are essential for accurate modeling of intermolecular interactions. In this work we report parametrization of an empirical charge−charge function previously reported (Piquemal, J.-P.; et al. *J. Phys. Chem. A* **2003**, 107, 10353) to correct for the missing penetration term in standard molecular mechanics force fields. For this purpose, we have developed a database (S101×7) of 101 unique molecular dimers, each at 7 different intermolecular distances. Electrostatic, induction/polarization, repulsion, and dispersion energies, as well as the total interaction energy for each complex in the database are calculated using the SAPT2+ method (Parker, T. M.; et al. *J. Chem. Phys.* **2014**, *140*, 094106). This empirical penetration model significantly improves agreement between point multipole and quantum mechanical electrostatic energies across the set of dimers and distances, while using only a limited set of parameters for each chemical element. Given the simplicity and effectiveness of the model, we expect the electrostatic penetration correction will become a standard component of future molecular mechanics force fields.



INTRODUCTION

Electrostatic interactions comprise one of the principle interatomic forces, along with exchange-repulsion, dispersion, and polarization or induction. The importance of electrostatic interactions is paramount at long range and for polar molecules. Much development effort has been focused on computational treatment of long-range electrostatics, e.g., the development of particle-meshed Ewald (PME) methods.[1−7] Electrostatic interactions at short range have received less consideration until recently. At close distances, a spherical approximation of atomic charge distributions is insufficiently accurate and use of atomic multipole expansions provides much greater flexibility in modeling complex electrostatic potentials near a molecular surface, an insight which inspired the development of the AMOEBA force field.[6,8−10] Nonetheless, at *very* close interatomic distances, when electron clouds overlap, a point multipole approximation becomes inadequate. The electrostatic potential within a spherical electron cloud no longer behaves as a simple $1/r$ interaction potential at small separation distances. Such deviation from a simple Coulomb potential is referred to as a penetration effect. While the charge penetration effect leads to a negative correction to energy at typical molecular interaction distances, where the electron−electron penetration is dominant, it can be repulsive at very short range.[11] A recent study by Lewis and co-workers reported the counterintuitive result that any ring substitutions of the benzene dimer (parallel) with electron-withdrawing or electron-donating groups yield more favorable electrostatic contributions than the unsubstituted benzene−benzene dimer itself.[12] This result is contrary to the conventional thought that such interactions are correlated with the ability to withdraw or donate electrons to the π cloud as described by the Hunter−Sanders rules.[13] Sherrill and co-workers suggested this is because the electrostatic interactions in such systems at the $\pi-\pi$ stacking distance exhibit a significant charge penetration effect. The multipole

Received: March 20, 2015
Published: April 28, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/acs.jctc.5b00267
J. Chem. Theory Comput. 2015, 11, 2609−2618

model, which the Hunter–Sanders rules are based upon, cannot correctly account for such effects.[14] Moreover, in a recent study of aromatic crystals, a charge penetration corrected AMOEBA-like model predicted better crystal properties than the uncorrected model.[15] It was shown that point atomic multipoles consistently predict positive (repulsive) electrostatic interactions between stacked or T-shaped benzene dimers while symmetry-adapted perturbation theory (SAPT)[16] suggests an opposite trend toward attractive interactions. The current AMOEBA force field seemingly compensates for penetration with a less repulsive van der Waals interaction, so the total interaction energy is reasonable at certain dimer configurations. However, explicit incorporation of the penetration effect provides much better anisotropy in crystal packing and makes the overall force field more transferable. In another study of organochlorine compounds using the AMOEBA model, it was found that the transferability of chlorine van der Waals parameters was unsatisfactory, likely due to lack of an explicit penetration correction.[17]

There have been previous attempts to incorporate the charge penetration effect into implicit solvent models,[18−20] multipole-based electrostatic models,[21−27] charge-density-based (including Gaussian multipole) models,[15,28−30] and combined quantum and molecular mechanics (QM/MM) models.[26,30−34] Generally, the charge penetration correction involves breaking the atom-centered point charge into an effective core and a valence electron density, as suggested by Gordon et al.[21] and Piquemal et al.[22] In this way, the electrostatic interaction between two atoms is described as a sum of interactions between core and valence charge densities, which can be modeled with empirical exponential functions. Alternatively, rigorous integration over the two charge densities can be used to model short-range electrostatic interactions,[15,28,29] with a significantly greater expenditure of computational effort. Others have explored incorporating charge penetration effects into the QM/MM framework, using either screened molecular mechanics (MM) charges[26,31−33] or simple empirical damping corrections.[34] Such screened MM charges are typically parametrized for QM/MM applications and may be not directly applicable in full MM calculations, e.g., to reproduce the attractive SAPT electrostatic energy in a stacked benzene–benzene conformation, unless an explicit term for interactions of the valence charge densities is included, as in the model recently proposed by Wang and Truhlar.[30]

In this study, the charge–charge penetration electrostatic model of Piquemal et al.[22] is revisited, implemented, and extensively tested in the context of the AMOEBA force field and using a new parametrization strategy. The charge penetration corrected AMOEBA point multipole model (multipoles + CP) is developed using a comprehensive set of small molecule complexes, and the parameters are determined for H, C, N, O, P, S, F, Cl, and Br to cover the elements commonly found in organic and biological molecules. To facilitate model development in this and future studies, a new database of SAPT2+[35,36] decomposed quantum mechanical energies constructed for 101 small molecule pairs, each at 7 different intermolecular distances (the S101×7 database), is presented.

METHODS

S101 and S101×7 Databases. In order to systematically examine the electrostatic and other components of intermolecular forces, the S101 and S101×7 databases of homo- and heterodimers of common organic molecules have been constructed. The S101 database contains 101 unique molecule pairs (Figure 1). The first 66 pairs, which cover the majority of

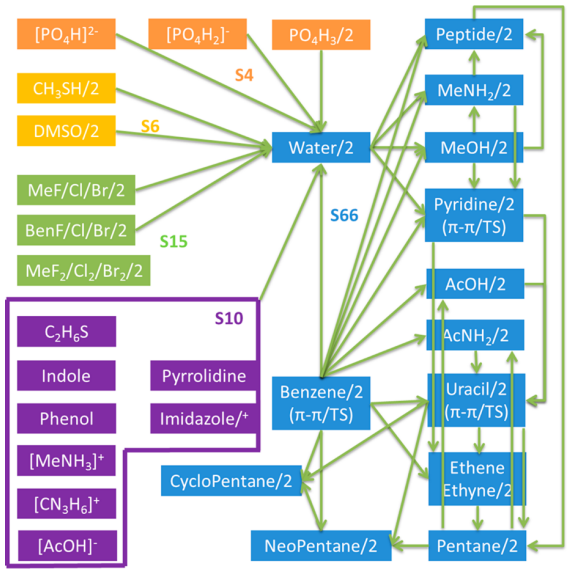


Figure 1. Schematic view of monomers and dimers in the S101 data set. The arrows connect two molecules that form a dimer; "/2" represents the existence of a homo dimer; "/+" indicates both neutral and ionized molecules are included. Different configurations of the same dimers, e.g., $MeNH_2$–water, phenol–water, benzene–benzene, and MeCl–MeCl, are included in the data set to take into account the orientational effect.

the typical organic interactions of H, C, N, and O atoms, are taken from the S66 database from Hobza et al.[37] In addition, 15 complexes containing halogen atoms (F, Cl, and Br), six complexes containing sulfur, and four complexes containing phosphorus have been added. Furthermore, 10 monomer–water complexes, which encompass amino acid side chain analogs (including the charged ones) missing in the S66 data set, have also been added, yielding a total of 101 pairs. To construct the S101×7 database, definitions of the intermolecular distance vectors from the S66×8 database of Hobza et al.[37] were used. Unlike S66×8, each of the 101 model complexes were placed at seven separation distances, corresponding to 0.70, 0.80, 0.90, 0.95, 1.00, 1.05, and 1.10 times the equilibrium intermolecular distances. Compared against the S66×8 database, the S101×7 set includes more dimer configurations at very short separations, which have been rarely investigated but are essential to the study of penetration effects and exchange–repulsion interactions. We have selected 0.7 times the equilibrium distance as the lower bound because the SAPT calculations below show that at this close distance the electrostatic energy is about 50% of the exchange–repulsion energy or higher; i.e., both electrostatic and van der Waals (vdW) components are important in the total interaction energy. As these short distances are being sampled in molecular dynamics simulations at room temperature and even more so at higher temperatures, their contributions to the simulated bulk properties are nonnegligible. Thus, it is essential to ensure the charge penetration model behaves correctly at these short distances.

DOI: 10.1021/acs.jctc.5b00267
J. Chem. Theory Comput. 2015, 11, 2609−2618

The newly added structures among the 101 complexes were optimized at the MP2/cc-pVTZ level of theory with counterpoise correction using the Gaussian09 program.[38] For each of the resulting 707 dimer configurations, the interaction energy has been decomposed using SAPT2+ analysis[35,36] provided by the PSI4 program.[39] The SAPT2+ calculation returns electrostatic, exchange−repulsion, induction, and dispersion energies, all to second order with respect to intramolecular electron correlation. Exact definitions of each component can be found in Figure 1 of Sherrill et al.[36] It should be noted that dispersion energy can only be separated from other effects in long range when two molecules do not overlap. Thus, at van der Waals distances, it may be more appropriately to refer to this as "dispersion-like" or "damped dispersion" energy. This should be kept in mind even though for simplicity the term "dispersion" is used throughout the discussion. All SAPT calculations were carried out using Dunning's correlation consistent basis sets[40,41] at both aug-cc-pVDZ and aug-cc-pVTZ levels. The complete basis set (CBS) limits of the SAPT2+ energies were also estimated. (Data can be found in the Supporting Information)

CBS Extrapolation Scheme. A two-point extrapolation strategy has been used to estimate the complete basis set limit of the exchange−repulsion and dispersion energy at the SAPT2+ level of theory. This is similar to Helgaker's scheme[42] but with an optimized p value (eq 1). Such a protocol has previously been applied to extrapolate the dispersion energy of DFT-SAPT calculation in earlier study.[43]

$$E_{\mathrm{CBS}} = \frac{E_{X+1}(X+1)^p - E_X X^p}{(X+1)^p - X^p} \tag{1}$$

Different p values, 3.0 for exchange−repulsion and 4.3 for dispersion energy, were obtained using the small pairs and subsequently applied for extrapolation over the full S101×7 database.

Scaling of the SAPT2+/CBS Dispersion Energy. Since the truncated terms in the SAPT2+ dispersion energy make a considerable contribution to the total interaction energy, dispersion energies obtained at the SAPT2+/CBS level are scaled by a factor f in order to match the SAPT total interaction energy to those obtained at the CCSD(T)/CBS level of theory (eq 2).

$$S = |E_{\mathrm{total}}^{\mathrm{CCSD(T)/CBS}} - (E_{\mathrm{non\text{-}disp}}^{\mathrm{SAPT2+/CBS}} + E_{\mathrm{disp}}^{\mathrm{SAPT2+/CBS}} f)| \tag{2}$$

By minimizing eq 2 using the 66 pairs in the S66 data set, a scale factor of $f = 0.89$ has been determined and used to construct the S101×7 database.

Modified Charge−Charge Interaction. In order to model the charge penetration effect, the method of Piquemal et al.[22] is revisited. Their original model corrects the charge−charge and charge−dipole interactions. Here, we propose to retain the charge−charge correction only. As a result, each atomic point charge is divided into an effective core and a damped valence electron distribution. Thus, the electrostatic energy between two atomic charges can be written as

$$\begin{aligned} E_{qq}(r) = [&Z_1Z_2 - Z_1(Z_2 - q_2)(1 - \exp(-\alpha_2 r)) \\ &- Z_2(Z_1 - q_1)(1 - \exp(-\alpha_1 r)) \\ &+ (Z_1 - q_1)(Z_2 - q_2)(1 - \exp(-\beta_1 r))(1 - \exp(-\beta_2 r))]/r \end{aligned} \tag{3}$$

where r is the interatomic distance; Z is the positive effective core charge, which is set to be equal to the number of valence electrons of each atom; q is the net charge of the atom, thus $(Z - q)$ can be considered as the magnitude of the (negatively charged) electron cloud; and α and β are two parameters controlling the magnitude of the damping of the electron cloud when the atom is interacting with the core and with electrons from other atoms, respectively. Thus, the total electrostatic energy between two atoms now involves three components, the core−core, core−electron, and electron−electron interactions.

Two methods have been explored to determine the α parameter values. The first method involves fitting the damped potential to the QM electrostatic potential at short range, near or within the molecular surface. By considering a probe charge of $+1e$ as a particle with an effective core charge of $+1e$ and having no valence electrons, $(Z_2 - q_2)$ becomes zero. Thus, the electrostatic potential can be written as

$$V_q(r) = [Z_1 - (Z_1 - q_1)(1 - \exp(-\alpha_1 r))]/r \tag{4}$$

Once Z and q are determined, α can be obtained easily by fitting eq 4 to the QM electrostatic potential.

In the second method, α is intuitively set to be the same as the number of valence electrons (except the hydrogen atom):

$$\alpha = \max\{Z, 2\} \tag{5}$$

When Z and β are fixed in eq 3, the electrostatic energy is more attractive when α is greater. This is in accordance with the intuition that atoms having a larger electron cloud may exhibit a stronger penetration effect. Although the final parameters for H, C, N, O, P, S, F, Cl, and Br were derived based on the second method, the performance of both methods is examined for H, C, N, and O containing molecules in later sections.

As the distance between two atoms increases, eq 3 will reduce to the classical Coulomb charge−charge interaction (q_1q_2/r). Thus, the electrostatic interaction at medium and long distances can still be accurately modeled via a multipole expansion, as the penetration correction diminishes rapidly with distance. As the data will show, the penetration correction is only significant when atomic separation is shorter than the sum of atomic van der Waals radii and thus does not affect the reciprocal space portion of an Ewald summation approach such as particle mesh Ewald (PME). In addition, to ensure the continuity between the real and reciprocal space, a switching function is used near the real space Ewald cutoff distance (typically 7 Å for atomic multipole PME) to ensure the penetration correction completely disappears:

$$f_{\mathrm{switch}}(r) = \begin{cases} 1, & r \le r_{\mathrm{l}} \\ 10\left(\dfrac{r_{\mathrm{u}} - r}{r_{\mathrm{u}} - r_{\mathrm{l}}}\right)^3 - 15\left(\dfrac{r_{\mathrm{u}} - r}{r_{\mathrm{u}} - r_{\mathrm{l}}}\right)^4 + 6\left(\dfrac{r_{\mathrm{u}} - r}{r_{\mathrm{u}} - r_{\mathrm{l}}}\right)^5, & r_{\mathrm{l}} < r < r_{\mathrm{u}} \\ 0, & r \ge r_{\mathrm{u}} \end{cases} \tag{6}$$

where r is the interatomic distance and r_{l} and r_{u} are the lower and upper bounds of the switching function.

Derivation of Atomic Point Multipole Moments. The permanent electrostatic energy in the AMOEBA force field includes through quadrupole−quadrupole interactions. Following previously detailed procedures,[9,10,44] an initial set of atomic multipole moments for each molecule was obtained from distributed multipole analysis (DMA)[45] at the MP2/6-311G** level of theory. Then the dipole and quadrupole moments were further optimized by fitting to the electrostatic potential

DOI: 10.1021/acs.jctc.5b00267
J. Chem. Theory Comput. 2015, 11, 2609−2618

calculated at the MP2/aug-cc-pVTZ level. This procedure is automated for organic molecules by the Poltype program.[44] The same strategy has been used in developing the AMOEBA force field for small molecules,[9] proteins,[10] and organochloroine compounds[17] previously. With the addition of the penetration correction described here, monopole−monopole (charge−charge) interactions are calculated using eq 3 while all other terms in the AMOEBA retain their original form. The SAPT and AMOEBA multipole based intermolecular interaction energies were compared on exactly the same dimer structures.

Parametrization of the Penetration Model. Since H, C, N, and O are the most common elements in organics and biomolecules, their parameters were determined first within the new charge penetration formalism. A training set of 35×7 molecule pairs was used, consisting of 15×7 pairs of hydrogen-bonded complexes, 12×7 pairs of dispersion-dominant complexes, and 8×7 pairs with mixed features of both. The initial parameter searching was done using a divide and conquer approach. For example, an initial set of parameters for sp^3 or sp^2 carbon, nonpolar hydrogen, and hydrogen attached to sp^2 C were obtained by selecting a smaller number of pairs (e.g., 8×7) from the dispersion-dominant and mixed complexes. Similarly, other parameters such as those for sp^3 oxygen and polar hydrogen were obtained initially from water dimers. Systematic scanning is feasible for determination of a small number of parameters. Then, one or more sets of initial parameters for all H, C, N, and O atom types were optimized together using the entire training set. Once the parameters for these four elements were finalized, further parametrization for P, S, F, Cl, and Br was carried out using the subsets in S101×7. In each of these subsets, an 80/20 ratio for the training and testing complexes was maintained to ensure a sufficient amount of data points for each atom type. An optimization program written in Python, using the quasi-Newton and Nelder−Mead simplex methods from the SciPy library, was applied to all of the parametrization work. The first derivative of the sum of unsigned errors with respect to each parameter was calculated numerically.

RESULTS

Convergence of the SAPT2+ Energy toward Basis Set Limit. In order to examine the convergence of the SAPT2+ energy toward the basis set limit, five small molecule pairs were selected and calculated using the aug-cc-pVXZ (X = D, T, Q, or 5, abbreviated as aXZ in the following paragraphs) basis sets. The five pairs are water−water, water−methanol, water−methylamine, ethyne−ethyne (T-shaped), and ethyne−water (CH···O). The energy difference between different aXZ (X = D, T, Q, and 5) basis sets for the total interaction energy and each energy component, including electrostatic, induction, exchange−repulsion, and dispersion energy, are compared. In general, a steady decrease in the energy gaps between aDZ−aTZ, aTZ−aQZ, and aQZ−a5Z can be observed (Table 1). The difference between aQZ and a5Z basis sets of all of the energy components as well as the total interaction energy are already well below 0.05 kcal/mol. In particular, the differences in electrostatic and induction energies are even smaller, at 0.003 and 0.005 kcal/mol, respectively. This implies that the difference between a5Z and a bigger basis set, e.g., a6Z, should be even smaller and negligible. Therefore, the results obtained using the a5Z basis set were used to approximate the complete basis set limit.

Table 1. Average Unsigned Differences between Energy Calculated at Different SAPT2+/aug-cc-pVXZ (X = D, T, Q, and 5) Levels of Theory

	MUE (kcal/mol)				
	electrostatic	induction	exch−repul	dispersion	total
aDZ − aTZ	0.076	0.010	0.213	0.237	0.382
aTZ − aQZ	0.016	0.005	0.091	0.047	0.140
aQZ − a5Z	0.003	0.005	0.036	0.011	0.043

Since SAPT2+ calculations are computationally expensive, the practical size of the basis set has been limited to aug-cc-pVTZ for most molecule pairs in the S101 database. To obtain an estimate of the SAPT2+ energy at the CBS limit, extrapolation thus is necessary. As shown in Figure 2, the electrostatic and induction energy components converge quickly to the CBS limit (approximated by a5Z results). The mean unsigned errors between aTZ and a5Z of five pairs are 0.018 and 0.010 kcal/mol for the two components, respectively. Therefore, for electrostatic and induction energies, the results obtained with the aTZ basis set are considered a reasonable approximation of the CBS limit. For exchange−repulsion and dispersion energy, a two-point scheme was applied to extrapolate the energy calculated at aDZ and aTZ to the CBS limit.

SAPT2+ Estimation of the CBS Limit. As mentioned in the previous section, different energy components converge at different rates with respect to the basis set size. Electrostatic and induction energies calculated using the aTZ basis set are sufficiently converged, while the exchange−repulsion and dispersion energy terms are not. Thus, a two-point extrapolation scheme is used to extrapolate the exchange−repulsion and dispersion energy at aTZ basis set to the complete basis set limit. The extrapolated dispersion energy is further scaled by a factor of 0.89 to compensate for higher order dispersion terms missing in the SAPT2+ approach (refer to Methods for details of the extrapolation and scale factor determination). Then the total SAPT2+ interaction energy is obtained by summing up the individual energy components. Finally, the quality of the SAPT2+ interaction energy at different basis set levels was examined by comparing with the CCSD(T)/CBS interaction energy of the S66 data set (Figure 3).[37] In general, the SAPT2+/CBS estimates with scaled dispersion energy (which will be referred to as SAPT2+/CBS/scaled to distinguish from the SAPT2+/CBS values without scaling) have the smallest mean unsigned error (MUE) among tested combinations, given MUE values of 0.16, 0.68, 0.47, and 0.17 kcal/mol (and RMSE of 0.25, 0.80, 0.56, and 0.21 kcal/mol) for SAPT2+/CBS/scaled, SAPT2+/CBS, SAPT2+/aTZ, and SAPT2+/aDZ, respectively. It is not surprising that the SAPT2+/aDZ combination yields very reasonable total interaction energy as this is consistent with a previous report.[36] However, the small error in the total interaction energy of SAPT2+/aDZ is due to the error cancellation of individual energy components. To best estimate individual energy components, SAPT2+/CBS/scaled results remain the most accurate choice in this study. However, for the larger systems where aug-cc-pVTZ calculations are not practical, SAPT2+/aDZ may be considered as an alternative. For all 707 pairs in the S101×7 database reported in this study, the same strategy (SAPT2+/CBS/scaled) is applied to calculate the individual energy components. All SAPT data can be accessed from Table S4 of the Supporting Information.

DOI: 10.1021/acs.jctc.5b00267
J. Chem. Theory Comput. 2015, 11, 2609−2618

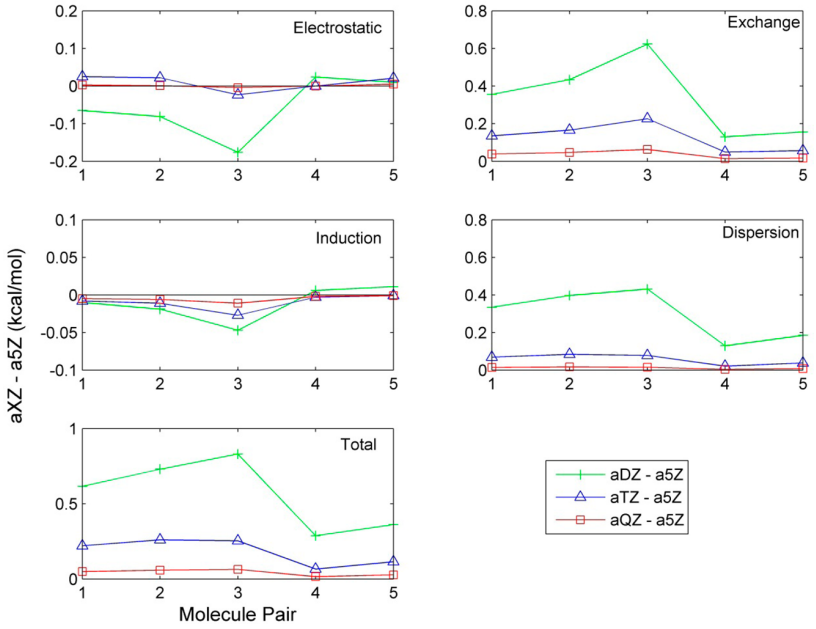


Figure 2. Difference between the SAPT2+ energy components calculated using aug-cc-pVXZ (X = D, T, and Q) basis set with the value obtained using aug-cc-pV5Z.

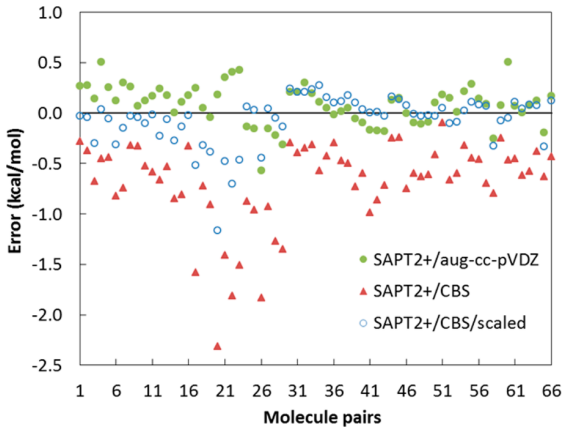


Figure 3. Errors of SAPT2+ interaction energy compared to CCSD(T)/CBS estimation[37] for dimers in the S66 data set. SAPT2+/aug-cc-pVDZ energy is shown in solid circles, SAPT2+/CBS is shown in triangles, and SAPT2+/CBS/scaled is shown in hollow circles.

Short-Range Electrostatic Interactions. The electrostatic interaction energy due to AMOEBA point multipoles as well as the charge penetration correction (multipoles + CP) are calculated and compared with SAPT2+ data for the S101×7 data set (excluding 7×7 complexes containing an ethyne molecule due to a lack of AMOEBA parameters) (Supporting Information Table S3). In our current model, the parameters Z and α for each atom are uniquely determined by the element type. Z is the number of valence electron. α is set equal to Z, or if Z is less than 2, then α is set to 2. The only parameter to be determined for the penetration correction is β in eq 3. For each of the H, C, N, and O elements, three atom types are used for β (Table 2). Hydrogen atoms are divided into nonpolar, aromatic, and polar hydrogens. Carbon, nitrogen, and oxygen all have three β values representing sp^3, sp^2, and aromatic cases.

Table 2. Two Sets of Charge Penetration Parameters for H, C, N, O, P, S, F, Cl, and Br

		valence-α set		fitted-α set	
atom type	Z	α (Å^{-1})	β (Å^{-1})	α (Å^{-1})	β (Å^{-1})
H (nonpolar)	1	2.0	1.999	3.3	2.924
H (aromatic)	1	2.0	2.010	3.3	3.064
H (polar, water)	1	2.0	2.004	3.3	3.178
C (sp^3)	4	4.0	2.646	3.8	2.934
C (aromatic)	4	4.0	2.708	3.8	2.764
C (sp^2)	4	4.0	2.685	3.8	2.673
N (sp^3)	5	5.0	3.097	3.1	2.790
N (aromatic)	5	5.0	3.072	3.1	2.784
N (sp^2)	5	5.0	3.054	3.1	2.761
O (sp^3, hydroxyl, water)	6	6.0	3.661	3.5	3.131
O (aromatic)	6	6.0	4.282	3.5	3.188
O (sp^2, carbonyl)	6	6.0	4.469	3.5	3.213
P (phosphate)	5	5.0	2.360	2.4	2.603
S (sulfide, e.g., R−SH)	6	6.0	2.770	2.6	2.382
S (sulfur IV, e.g., DMSO)	6	6.0	2.381	2.6	2.230
F (organofluorine)	7	7.0	4.275	4.2	4.030
Cl (organochloride)	7	7.0	2.830	3.0	2.594
Br (organobromine)	7	7.0	2.564	2.7	2.336

For sulfur, distinct β values are used for sulfide and sulfur IV, while P, F, Cl, and Br have only a single β value per element in current parametrization.

In general, after fitting of β parameters, the new electrostatic model with charge penetration correction shows excellent agreement with the SAPT2+ results (Figure 4). Taking the valence-α parameter set as an example, for dimers near the equilibrium distances (R_{min}), i.e., 0.90, 0.95, 1.00, 1.05, and 1.10 of R_{min}, the mean unsigned error (MUE) of the original point multipoles is 3.16 kcal/mol, which is reduced about 5-fold to 0.57 kcal/mol after inclusion of the charge penetration correction (Table 3). For the dimers at very short separation, i.e., 0.70 and 0.80 of R_{min}, the MUEs for the corrected and

DOI: 10.1021/acs.jctc.5b00267
J. Chem. Theory Comput. 2015, 11, 2609−2618

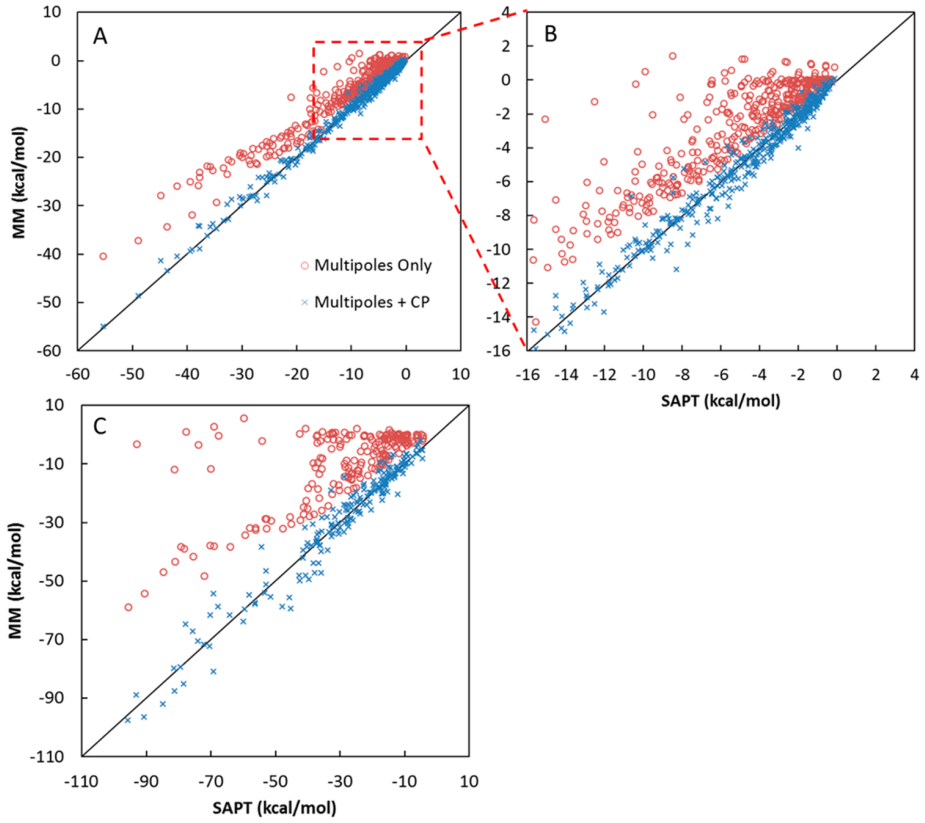


Figure 4. Plots of multipole electrostatic energy (kcal/mol) against the reference SAPT2+/aug-cc-pVTZ calculation for (A) near-equilibrium (0.90, 0.95, 1.00, 1.05, and 1.10) complexes taken from the S101×7 data set, (B) expanded plot of the boxed region in A, and (C) short-range (0.70 and 0.80) complexes in the S101×7 data set. The uncorrected AMOEBA point multipole energy (multipoles only) is shown in red circles, and the charge penetration corrected point multipole energies using the valence-α parameter set (multipoles + CP) are denoted by blue crosses.

Table 3. Differences between AMOEBA Electrostatic Energies, Either with or without Charge Penetration Correction, Compared against SAPT2+/CBS/Scaled Electrostatic Energies for the S101×7 Data Set

S101 set	statistics	multipoles only	multipoles + CP[a]	multipoles + CP[b]
R (0.90−1.10) (94×5 pairs)	MUE	3.16	0.57	0.72
	MSE	3.16	−0.04	−0.17
	RMSE	4.35	0.83	1.04
	% error	52.7%	13.6%	16.5%
R (0.70−0.80) (94×2 pairs)	MUE	19.16	3.28	2.84
	MSE	19.16	−0.15	0.46
	RMSE	24.17	4.63	4.36
	% error	69.3%	13.4%	10.8%
all distance (94×7 pairs)	MUE	7.73	1.35	1.33
	MSE	7.73	−0.07	0.01
	RMSE	13.43	2.57	2.49
	% error	57.4%	13.6%	14.9%

[a]Charge penetration corrected model using the valence-α parameter set. [b]Charge penetration corrected model using the fitted-α parameter set.

uncorrected electrostatic energy are 3.28 and 19.16 kcal/mol, respectively. As shown in Figure 4, it is striking that point-multipole-based electrostatic energy alone yields very large errors for dimers in close contact, and the simple charge penetration correction applied here is able to systematically improve agreement with SAPT-derived electrostatics. Based upon the mean unsigned errors, the charge penetration corrected model results in a percentage error of 13.6% and 13.4% at near-equilibrium and very short separations, respectively. In contrast, the uncorrected model has errors of 53% and 69% for these same two distance ranges. It is clear the charge penetration corrected model not only reduces the magnitude of absolute and relative errors compared to SAPT but also provides consistent performance over a range of distances. In the uncorrected model, the percentage of error at very short distances is larger than at near-equilibrium distances, due to the increased effect of short-ranged charge penetration.

For S66 dimers at near-equilibrium separations and using uncorrected AMOEBA multipoles, the hydrogen-bonded complexes exhibit the largest mean unsigned error of 4.41 kcal/mol, compared to MUEs of 3.08 and 2.08 kcal/mol for dispersion-dominant and mixed complexes (Table 4). This is not surprising since the hydrogen-bonded complexes generally have the strongest electrostatic interactions. However, in terms of relative errors, the dispersion-dominant complexes carry the largest error at 105%, while the hydrogen-bonded and mixed complexes have the mean percentage of errors of 30% and 58%, respectively. It is somewhat surprising the dispersion-dominant complexes have such absolute and relative errors, as they are normally considered to have the weakest electrostatic interaction among the three types.

DOI: 10.1021/acs.jctc.5b00267
J. Chem. Theory Comput. 2015, 11, 2609−2618

Table 4. Electrostatic Energy (kcal/mol) in Different Interaction Types of the S66 Complexes at the Near-Equilibrium (0.90−1.10) Distances in the S101×7 Data Set

	multipoles only		multipoles + CP[a]		multipoles + CP[b]	
S66 set	MUE	% error	MUE	% error	MUE	% error
hydrogen-bonded	4.41	30.5	0.50	3.5	0.47	3.3
dispersion-dominant	3.08	105.3	0.53	23.2	0.69	26.7
mixed	2.08	58.4	0.61	18.7	0.74	23.6

[a]Charge penetration corrected model using the valence-α parameter set. [b]Charge penetration corrected model using the fitted-α parameter set.

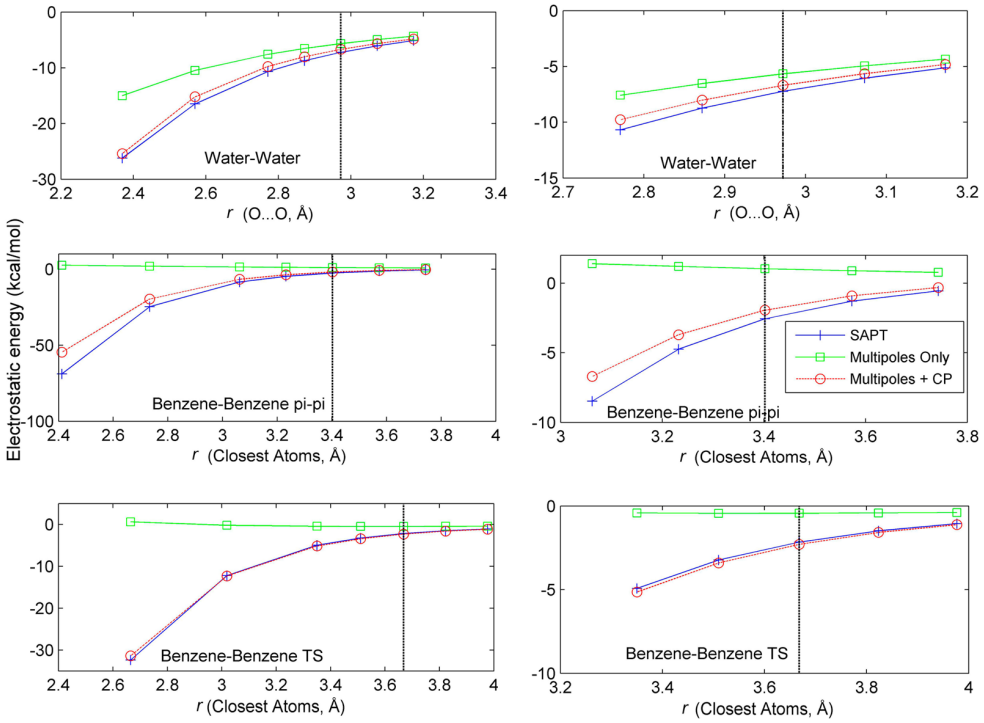


Figure 5. Plots of the electrostatic energy profiles of water−water and benzene−benzene dimer complexes. The valence-α parameter set was used in calculations of the charge penetration corrected model (multipoles + CP). The vertical line indicates the equilibrium distance.

To help understand why dispersion-dominant complexes have such large relative errors, the electrostatic energies of benzene dimers and $\pi-\pi$ stacked and T-shaped complexes, as well as hydrogen-bonded water dimers, are shown in Figure 5. For the uncorrected AMOEBA model, the calculated electrostatic energy is positive for the $\pi-\pi$ benzene pairs yet QM calculations suggest the interaction is attractive with a negative electrostatic energy. Taking the electrostatic energy for this pair at the equilibrium distance as an example, the SAPT2+/CBS/scaled calculation yields a value of −2.6 kcal/mol, while the uncorrected AMOEBA multipoles give +1.0 kcal/mol, an error of 3.6 kcal/mol or 138%. For the T-shaped benzene dimer, the SAPT2+/CBS/scaled and the uncorrected AMOEBA multipoles have values of −2.2 and −0.4 kcal/mol, respectively. The unsigned error is 1.8 kcal/mol or 82% of the SAPT values, both somewhat less than for the $\pi-\pi$ complex. These findings are consistent with the previous study by Tafipolsky and Engels.[15] In contrast, although the hydrogen-bonded water dimer has the larger electrostatic energy of −7.2 kcal/mol, the uncorrected model has an unsigned error of 1.6 kcal/mol and a relative error of only 22%. This trend is in accordance with the averaged errors reported in Table 4 and suggests the electrostatic interaction in dispersion-dominant complexes is the most charge penetration dependent. This might be explained by two effects. First, in the nonpolar molecules, the electron distribution is more "balanced"; i.e., there is more electron density on the hydrogen atoms, hence a stronger penetration effect for hydrogens. Second, in the stacked benzene dimer, interactions between heavier atoms, carbon−carbon for example, suffer stronger charge penetration effect, thus weight more in electrostatic energy. For hydrogen-bonded pairs, although the percentage of error is relatively low for the uncorrected atomic multipoles, the absolute error remains significant. Therefore, a correction is still necessary in order to achieve better accuracy in the force field. It is notable that, after the charge penetration correction, the mean unsigned errors of all three types of complexes are reduced to 0.5−0.6 kcal/mol near the equilibrium distances, which is approaching the possible error of the QM calculation itself.

Alternative Way To Derive the α Parameter. As mentioned in Methods, an alternative way to derive the α parameter is to fit the penetration-damped electrostatic potential (eq 4) to the target QM values. An attempt to use this fitting strategy has been also made, and the resulting parameters have been compared. The parametrization of α is restricted to a single unique value for each element type, as

DOI: 10.1021/acs.jctc.5b00267
J. Chem. Theory Comput. 2015, 11, 2609−2618

Table 5. Comparison of the Two Sets of Parameters, Valence-α and Fitted-α, for H, C, N, and O Containing Molecules in S66 Complexes[a]

S66 set	statistics (kcal/mol)	multipoles only	valence-α	fitted-α	valence-α; single β per element
R (0.90−1.10) (59×5 pairs)	MUE	3.34	0.54	0.62	0.65
	MSE	3.34	−0.04	−0.19	−0.00
	RMSE	4.45	0.74	0.83	0.88
R (0.70−0.80) (59×2 pairs)	MUE	21.34	2.72	2.39	3.52
	MSE	21.34	−0.10	0.41	−0.19
	RMSE	26.84	3.80	3.92	4.40
all distances (59×7 pairs)	MUE	8.48	1.16	1.12	1.47
	MSE	8.48	−0.06	−0.02	−0.06
	RMSE	14.83	2.13	2.21	2.47

[a]An additional set of parameters which has a unique β for each element is also presented.

before. A brute force scanning of the parameter using a grid size of 0.1 $Å^{-1}$ was used to search for the global minimum since the α parameter is less sensitive than β. All 13 monomers (excluding ethyne) in the S66 data set were used in fitting of the α for H, C, N, and O elements. Then β parameters were determined as before with α values fixed to their potential-fitted values. The penetration parameter set obtained this way will be referred to as the fitted-α set, while the parameter set with α based on eq 5 will be referred as the valence-α set. With the fitted-α parameters, the RMSE of the electrostatic potential of the 13 monomers calculated using eq 4 is greatly reduced to 0.07 kcal/mol, compared to an RMSE of 0.95 kcal/mol for the valence-α (see Supporting Information Table S2).

The performance of the two sets of parameters has been compared using the S101 data set. The overall performance of the two parameter sets is very similar to mean unsigned errors of 1.35 and 1.33 kcal/mol for the valence- and fitted-α sets, respectively (Table 3). For near-equilibrium pairs, the valence-α set has a marginally better MUE of 0.57 kcal/mol against 0.72 kcal/mol for the fitted-α set. For short-ranged pairs, the fitted-α set with a MUE of 2.84 kcal/mol yet is slightly better than a MUE of 3.28 kcal/mol of the valence-α set. Similar trends in RMSEs of the two sets of parameters are also observed. However, the valence-α parameter set tends to have more balanced performances for hydrogen-bonded, dispersion-dominant, and mixed complexes, giving the MUEs of 0.50, 0.53, and 0.61 kcal/mol for the three groups, respectively (Table 4). In contrast, the fitted-α set, with a MUE of 0.47 kcal/mol for the hydrogen-bonded complexes, shows subtly better agreement with SAPT results yet has slightly worse performances for the aromatic compounds. The MUEs of the dispersion-dominant and mixed complexes are 0.69 and 0.74 kcal/mol, respectively (electrostatic energy of individual pairs can be found in Table S3 in the Supporting Information). Nonetheless, the two sets of parameters all have excellent agreement with the SAPT results for the whole S101×7 database, while the fitted-α set yields better electrostatic potential than the valence-α set.

The charge penetration model also exhibited good transferability during the fitting of β parameters. Although three atom types are used for H, C, N, and O in the current parametrization, restriction to a single β for each element also results in reasonable accuracy. Simply applying the arithmetic mean of the three β parameters in valence-α parameter set for each element (Supporting Information Table S1) increases the MUE by only 0.1 to 0.65 kcal/mol for the near-equilibrium pairs in the S66 set (Table 5). For pairs with shorter distances, the MUE increases by 0.8 to 3.52 kcal/mol in the same set. Only marginal improvements in MUEs were found after optimizing the β parameters for each element starting from the averaged value. We believe this demonstrates the robustness and transferability of the charge penetration correction and the parametrization strategy. For the purpose of retaining flexibility, we recommend the use of three atom types for each of the H, C, N, and O elements in our final model.

CONCLUSION

The charge penetration effect is usually overlooked in molecular mechanical models and traditional force fields. Our results show that DMA-derived point multipoles systematically underestimate the SAPT electrostatic interaction energy at typical molecular interaction distances based on the 101×7 dimers studied here (see the Supporting Information). An exponential damping function providing a simple charge−charge penetration model suitable for force field incorporation has been revisited, along with a new parametrization strategy. The S101×7 SAPT-decomposed quantum mechanical energy database is developed as a reference for parameter training and for use in future force field comparison. The database is an extension of the S66 and S66×8 data set previously developed by Hobza and co-workers,[37] with additional prototype molecular complexes. The decomposed energies are calculated at the SAPT2+/aug-cc-pVTZ level of theory, with exchange−repulsion and dispersion components extrapolated to the complete basis set limit. The dispersion energy is further scaled to compensate for missing higher order terms in the SAPT2+ method. The total SAPT interaction energy is in excellent agreement with CCSD(T)/CBS results, which are currently considered to be the "gold standard" for estimation of intermolecular interactions, with a mean unsigned error of 0.16 kcal/mol for the S66 data set. Thus, the SAPT results should provide a reliable reference for force field development.

By replacing the idealized charge−charge (Coulomb) interaction with the charge penetration corrected model (eq 3) in the AMOEBA framework, the accuracy of calculated electrostatic energies for the S101×7 database is improved by 5-fold. For the five distance pairs near the equilibrium distances (i.e., 0.90−1.10 times the equilibrium distance), the mean unsigned error of the charge penetration corrected and uncorrected point multipole models are 0.57 and 3.16 kcal/mol, respectively; for the extremely close distance separations (i.e., 0.70 and 0.80 times the equilibrium distance), the mean unsigned errors of the two models are 3.28 and 19.16 kcal/mol, respectively. The improvement for the corrected model is significant and shows a consistent agreement with the quantum mechanics data at both long and short distances. The

DOI: 10.1021/acs.jctc.5b00267
J. Chem. Theory Comput. 2015, 11, 2609−2618

robustness and transferability of this model is also reflected in the use of very limited (element-based) parameters. The charge penetration correction is short-ranged and rapidly converges to the classical Coulomb interaction beyond 6−7 Å. Thus, it can be completely incorporated into the real space of Ewald summation without any additional computational cost in reciprocal space. Because simulations including penetration correction are clearly feasible, there is ongoing work dedicated to the optimization of parallel scaling the coupled penetration/smooth particle mesh Ewald approach. In addition, higher order penetration corrections (charge−dipole and charge−quadrupole penetration) are also possible and have been implemented in models such as SIBFA.[46] Nonetheless, the simple empirical charge penetration model presented in this work provides us with an efficient approach to achieve accurate electrostatic energy that is systematically modeled after SAPT quantum mechanical energy decomposition. The change in the electrostatic component requires re-examination of the van der Waals interaction to arrive at a balanced representation of the total energy. Overall we expect this improvement in the electrostatic component will alleviate the need for error compensation via other components and lead to more balanced and transferable potential energy functions in general. The comprehensive SAPT database we developed in this work will also be useful for many others who are interested in understanding intermolecular forces or evaluating different empirical models.

■ ASSOCIATED CONTENT

Supporting Information

Tables listing unique α and β for each element, RMSEs of the electrostatic potential on the grid compared to MP2/aug-cc-pVTZ calculation, interaction energies, and decomposed interaction energies. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00267.

■ AUTHOR INFORMATION

Corresponding Authors

*(P.R.) E-mail: pren@mail.utexas.edu.
*(J.-P.P.) E-mail: jpp@lct.jussieu.fr.

Funding

We are grateful for the support from the Robert A. Welch Foundation (Grant F-1691 to P.R.), the National Institutes of Health (Grants GM106137 and GM114237 to J.W.P. and P.R.), the National Science Foundation (Grant CHE1152823 to J.W.P.), and CPRIT (Grant RP110532 to Q.W.). The high performance computing resources were provided by TACC and XSEDE (Grant TG-MCB100057 to P.R.). This work was also supported in part by French state funds managed by CALSIMLAB and the ANR within the Investissements d'Avenir program under reference ANR-11-IDEX-0004-02 (to J.-P.P).

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

CBS, complete basis set; DMA, distributed multipole analysis; MSE, mean signed error; MUE, mean unsigned error; PME, particle mesh Ewald; QM/MM, combined quantum and molecular mechanics method; RMSE, root-mean-square error; SAPT, symmetry-adapted perturbation theory

■ REFERENCES

(1) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys. (Berlin, Ger.)* **1921**, *369* (3), 253−287.
(2) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(3) Darden, T.; Perera, L.; Li, L.; Pedersen, L. New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **1999**, *7* (3), R55−R60.
(4) Smith, W. Point multipoles in the Ewald summation (revisited). *CCP5 Newslett.* **1998**, *46*, 18−30.
(5) Nam, K.; Gao, J.; York, D. M. An Efficient Linear-Scaling Ewald Method for Long-Range Electrostatic Interactions in Combined QM/MM Calculations. *J. Chem. Theory Comput.* **2004**, *1* (1), 2−13.
(6) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107* (24), 5933−5947.
(7) Walker, R. C.; Crowley, M. F.; Case, D. A. The implementation of a fast and accurate QM/MM potential method in Amber. *J. Comput. Chem.* **2008**, *29* (7), 1019−1031.
(8) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114* (8), 2549−2564.
(9) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7* (10), 3143−3161.
(10) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046−4063.
(11) Stone, A. J. *The Theory of Intermolecular Forces*. Oxford University Press: New York, 1996.
(12) Watt, M.; Hardebeck, L. K. E.; Kirkpatrick, C. C.; Lewis, M. Face-to-Face Arene−Arene Binding Energies: Dominated by Dispersion but Predicted by Electrostatic and Dispersion/Polarizability Substituent Constants. *J. Am. Chem. Soc.* **2011**, *133* (11), 3854−3862.
(13) Hunter, C. A.; Sanders, J. K. M. The nature of .pi.-.pi. interactions. *J. Am. Chem. Soc.* **1990**, *112* (14), 5525−5534.
(14) Hohenstein, E. G.; Duan, J. N.; Sherrill, C. D. Origin of the Surprising Enhancement of Electrostatic Energies by Electron-Donating Substituents in Substituted Sandwich Benzene Dimers. *J. Am. Chem. Soc.* **2011**, *133* (34), 13244−13247.
(15) Tafipolsky, M.; Engels, B. Accurate Intermolecular Potentials with Physically Grounded Electrostatics. *J. Chem. Theory Comput.* **2011**, *7* (6), 1791−1803.
(16) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94* (7), 1887−1930.
(17) Mu, X.; Wang, Q.; Wang, L.-P.; Fried, S. D.; Piquemal, J.-P.; Dalby, K. N.; Ren, P. Modeling Organochlorine Compounds and the σ-Hole Effect Using a Polarizable Multipole Force Field. *J. Phys. Chem. B* **2014**, *118* (24), 6456−6465.
(18) Chipman, D. M. Charge penetration in dielectric models of solvation. *J. Chem. Phys.* **1997**, *106* (24), 10194−10206.
(19) Chipman, D. M. Reaction field treatment of charge penetration. *J. Chem. Phys.* **2000**, *112* (13), 5558−5565.
(20) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Polarizable dielectric model of solvation with inclusion of charge penetration effects. *J. Chem. Phys.* **2001**, *114* (13), 5691−5701.
(21) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. Evaluation of charge penetration between distributed multipolar expansions. *J. Chem. Phys.* **2000**, *112* (17), 7300−7306.
(22) Piquemal, J. P.; Gresh, N.; Giessner-Prettre, C. Improved formulas for the calculation of the electrostatic contribution to the intermolecular interaction energy from multipolar expansion of the electronic distribution. *J. Phys. Chem. A* **2003**, *107* (48), 10353−10359.

DOI: 10.1021/acs.jctc.5b00267
J. Chem. Theory Comput. 2015, 11, 2609−2618

(23) Gresh, N.; Piquemal, J. P.; Krauss, M. Representation of Zn(II) complexes in polarizable molecular mechanics. Further refinements of the electrostatic and short-range contributions. Comparisons with parallel ab initio computations. *J. Comput. Chem.* **2005**, *26* (11), 1113−1130.
(24) Slipchenko, L. V.; Gordon, M. S. Electrostatic energy in the effective fragment potential method: Theory and application to benzene dimer. *J. Comput. Chem.* **2007**, *28* (1), 276−291.
(25) Werneck, A. S.; Filho, T. M. R.; Dardenne, L. E. General methodology to optimize damping functions to account for charge penetration effects in electrostatic calculations using multicentered multipolar expansions. *J. Phys. Chem. A* **2008**, *112* (2), 268−280.
(26) Kumar, R.; Wang, F. F.; Jenness, G. R.; Jordan, K. D. A second generation distributed point polarizable water model. *J. Chem. Phys.* **2010**, *132* (1), No. 014309.
(27) Stone, A. J. Electrostatic Damping Functions and the Penetration Energy. *J. Phys. Chem. A* **2011**, *115* (25), 7017−7027.
(28) Spackman, M. A. The use of the promolecular charge density to approximate the penetration contribution to intermolecular electrostatic energies. *Chem. Phys. Lett.* **2006**, *418* (1−3), 158−162.
(29) Cisneros, G. A. Application of Gaussian Electrostatic Model (GEM) Distributed Multipoles in the AMOEBA Force Field. *J. Chem. Theory Comput.* **2012**, *8* (12), 5072−5080.
(30) Wang, B.; Truhlar, D. G. Screened Electrostatic Interactions in Molecular Mechanics. *J. Chem. Theory Comput.* **2014**, *10* (10), 4480−4487.
(31) Cisneros, G. A.; Tholander, S. N. I.; Parisel, O.; Darden, T. A.; Elking, D.; Perera, L.; Piquemal, J. P. Simple formulas for improved point-charge electrostatics in classical force fields and hybrid quantum mechanical/molecular mechanical embedding. *Int. J. Quantum Chem.* **2008**, *108* (11), 1905−1912.
(32) Wang, B.; Truhlar, D. G. Including Charge Penetration Effects in Molecular Modeling. *J. Chem. Theory Comput.* **2010**, *6* (11), 3330−3342.
(33) Wang, B.; Truhlar, D. G. Partial Atomic Charges and Screened Charge Models of the Electrostatic Potential. *J. Chem. Theory Comput.* **2012**, *8* (6), 1989−1998.
(34) Wang, Q. T.; Bryce, R. A. Improved Hydrogen Bonding at the NDDO-Type Semiempirical Quantum Mechanical/Molecular Mechanical Interface. *J. Chem. Theory Comput.* **2009**, *5* (9), 2206−2211.
(35) Hohenstein, E. G.; Sherrill, C. D. Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory. *J. Chem. Phys.* **2010**, *133* (1), No. 014101.
(36) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.* **2014**, *140* (9), No. 094106.
(37) Rezac, J.; Riley, K. E.; Hobza, P. S66: A Well-Balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7* (8), 2427−2438.
(38) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian: Wallingford, CT, USA, 2009.
(39) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. Psi4: An open-source ab initio electronic structure program. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2* (4), 556−565.
(40) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007−1023.
(41) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96* (9), 6796−6806.
(42) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N2, and H2O. *Chem. Phys. Lett.* **1998**, *286* (3−4), 243−252.
(43) Rezac, J.; Hobza, P. Extrapolation and Scaling of the DFT-SAPT Interaction Energies toward the Basis Set Limit. *J. Chem. Theory Comput.* **2011**, *7* (3), 685−689.
(44) Wu, J.; Chattree, G.; Ren, P. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, *131* (3), 1−11.
(45) Stone, A. J.; Alderton, M. Distributed multipole analysis. *Mol. Phys.* **1985**, *56* (5), 1047−1064.
(46) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P. Anisotropic, polarizable molecular mechanics studies of inter- and intramoecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *J. Chem. Theory Comput.* **2007**, *3* (6), 1960−1986.

DOI: 10.1021/acs.jctc.5b00267
J. Chem. Theory Comput. 2015, 11, 2609−2618

Au-delà de la qualité de ces corrections présentées, deux points peuvent être discutés :

(i) une correction charge-charge n'est pas forcément adaptée à tous les types de systèmes. Prenons l'exemple d'une molécule de benzène qui possède des charges et un dipole très faible en comparaison à la composante quadrupolaire. Une extension de ce type de correction aux interactions charge-dipole et charge-quadrupole est à envisager.

(ii) enfin la stratégie de calcul des multipoles est primordiale. En effet, dans les travaux présentés ci-dessus, ils ne dérivent pas directement de GDMA mais sont ajustés par la suite sur le MEP. Le schéma de distribution multipolaire est aussi à discuter : seulement atomique ou atomique et milieu de liaison.

La correction introduite dans le potentiel SIBFA par PIQUEMAL *et al.* considère trois composantes du calcul multipolaire de l'électrostatique.

$$\begin{aligned} E^*_{m_i-m_j} =& \frac{1}{r_{ij}}[Z_iZ_j - Z_i(Z_j - q_j)(1 - exp(-\alpha_j r_{ij})) \\ & - Z_j(Z_i - q_i)(1 - exp(-\alpha_i r_{ij})) \\ & + (Z_i - q_i)(Z_j - q_j)(1 - exp(-\beta_i r_{ij})(1 - exp(-\beta_j r_{ij})] \end{aligned} \tag{1.13}$$

où $\alpha_i = \gamma/r^i_{vdw}$ et $\beta_i = \delta/r^i_{vdw}$, γ et δ sont des facteurs de normalisations et r^i_{vdw} des rayons de VAN DER WAALS de l'atome i. Z_i est la valence électronique de l'atome i et q_i la charge (monopole) de l'atome i. Cette première correction tend à reproduire l'équation 1.12.

$$E^*_{m_i-d_j} = \frac{\boldsymbol{\mu}_j.\boldsymbol{r}_{ji}}{r^3_{ij}}(Z_i - (Z_i - q_i)(1 - exp(-\eta_i r_{ij})) \tag{1.14}$$

où $\eta_i = \chi/r^i_{vdw}$, with χ a un facteur de normalisation.

De la même manière que dans l'équation 1.14, la charge est écrantée dans le terme charge-quadrupole.

La transférabilité de ce modèle, limitée dans un premier temps à charge-charge et charge-dipole, a été étudiée en l'appliquant à un autre champ de forces : AMOEBA. Bien que AMOEBA et SIBFA se basent tous les deux sur un développement multipolaire pour le calcul de l'électrostatique. Leur schéma diffère, en plus de multipoles centrés sur les atomes, SIBFA prend en compte ceux sur les liaisons [128]. De plus, le protocole de calcul des multipoles diffère aussi. Au vu de ces traitements quantitativement et qualitativement différents, nous avons décidé d'étendre le modèle de SIBFA et de considérer α_i, β_i et η_i des équations 1.13 et 1.14 comme paramètres de la correction dans le contexte du potentiel AMOEBA.

1.3.1 Point de départ : quelle distribution multipolaire ?

GDMA propose plusieurs options de partitionnement de l'espace et la possibilité d'affecter un rayon aux différents sites autour duquel sera calculée la distribution des moments multipolaires. Ces paramètres ont été testés sur dix dimères d'eau d'orientations différentes et représentatives [129]. Les densités ont été calculées au niveau MP2 avec une base aug-ccpvTZ. Rappelons que jusque là, SIBFA restait au niveau HF, manquant ainsi les effets de corrélations électroniques et sans prendre en compte les orbitales f, jugées trop diffuses. Dans cette étude, en plus des paramètres ci-dessous, la densité électronique de la molécule d'eau a été calculé avec et sans orbitale f pour l'oxygène. Ces résultats ne sont pas présentés, car n'apportant finalement que peu d'information sur le méthodologie à suivre et peu concluant dans le cadre de cette étude. Plusieurs paramètres de ce calcul ont été testés :

1. le schéma des sites multipolaires : (i) atome seulement (tableau 1.1). (ii) atome + milieu de liaison (tableau 1.2).
2. le rayon des atomes d'hydrogène à 0,325 et à 0,650 Å, noté rad325 et rad65.

3. trois options de partitionnement de l'espace, noté « switch » dans GDMA d'où la notation sw$i$ avec i l'option de partitionnement.

Les deux derniers points suivent des protocoles établis dans le cadre de calibration de nouveaux ligand, c'est-à-dire non-défini dans un champ de forces [15], donnant ainsi lieu à la notation sw$i$-rad$j$.

Les tableaux 1.1 et 1.2 reprennent l'énergie électrostatique multipolaire sans correction de pénétration de charge. Enfin le tableau 1.3 reprend pour ces mêmes dimères l'énergie référence SAPT.

index	sw0-rad325	sw0-rad65	sw2-rad325	sw2-rad65	sw4-rad325	sw4-rad65
1	-6.02	-6.54	-6.73	-7.94	-6.75	-7.93
2	-5.12	-5.51	-5.15	-6.92	-5.16	-6.90
3	-5.01	-5.36	-4.85	-6.80	-4.86	-6.77
4	-5.31	-5.01	-4.82	-4.66	-4.81	-4.68
5	-4.81	-4.44	-3.76	-4.05	-3.74	-4.04
6	-4.67	-4.28	-3.44	-3.96	-3.42	-3.95
7	-3.49	-3.44	-3.58	-3.08	-3.58	-3.09
8	-0.93	-1.04	-1.26	-1.10	-1.26	-1.11
9	-3.06	-3.28	-3.79	-3.53	-3.80	-3.56
10	-1.87	-1.92	-2.21	-2.00	-2.21	-2.02

TABLE 1.1 – Energie multipolaire électrostatique à partir de moments atomes en kcal/mol

index	sw0-rad325	sw0-rad65	sw2-rad325	sw2-rad65	sw4-rad325	sw4-rad65
1	-6.18	-6.50	-6.91	-8.14	-6.96	-8.21
2	-5.19	-5.47	-5.22	-7.07	-5.23	-7.11
3	-5.05	-5.33	-4.88	-6.93	-4.88	-6.96
4	-5.13	-4.99	-4.72	-4.63	-4.69	-4.63
5	-4.51	-4.41	-3.55	-3.98	-3.48	-3.95
6	-4.32	-4.26	-3.19	-3.89	-3.11	-3.85
7	-3.59	-3.46	-3.51	-3.06	-3.50	-3.05
8	-1.04	-1.03	-1.27	-1.09	-1.28	-1.10
9	-3.32	-3.29	-3.87	-3.54	-3.89	-3.57
10	-1.96	-1.91	-2.24	-1.97	-2.25	-1.98

TABLE 1.2 – Energie multipolaire électrostatique à partir de moments sur liaisons et atomes en kcal/mol

index	SAPT
1	-7.96
2	-6.72
3	-6.52
4	-6.61
5	-5.71
6	-5.41
7	-4.84
8	-1.49
9	-4.71
10	-2.71

TABLE 1.3 – Energie électrostatique SAPT (3ème ordre - base aug-cc-pvTZ) en kcal/mol.

En observant les résultats « bruts », il est difficile de conclure sur le protocole à suivre. En effet, des variations de près de 2 kcal/mol entre les énergies les plus faibles et les plus élevées sont présentes dans les tableaux 1.1 et 1.2. De plus, en comparant ces résultats, non-corrigés, avec les résultats références (voir tableau 1.3), on s'aperçoit que certaines énergies électrostatiques purement multipolaires sont déjà proches de cette référence. Notons par exemple que les protocoles notés sw2-rad65 et sw4-rad65 des tableaux 1.1 et 1.2 donnent des énergies en très bon accord avec SAPT mais seulement pour certaines orientations des dimères, prenant ainsi en compte la pénétration de charge de manière artificielle.

Finalement, établir la ou les stratégies appropriées passe par une calibration des paramètres de la correction de pénétration de charge appliquée à ses douze jeux de multipoles et une comparaison directe aux résultats SAPT. Le tableau 1.4 reprend les énergies électrostatiques corrigées à courte distance. Le meilleur accord est trouvé avec GDMA [1], c'est-à-dire avec des multipoles centrés sur des atomes uniquement, en utilisant l'option 2 de partitionnement de l'espace de GDMA et avec un rayon de 0,325 Å pour les hydrogènes. GDMA [2] correspond à ce même protocole mais selon un schéma de distribution multipolaire centré sur atome et milieu de liaison. La cin-

quième colonne se réfère aux méthodes ajustant les multipoles sur le potentiel électrostatique, tel que AMOEBA. Enfin la cinquième colonne reprend les résultats d'une correction proposée récemment sur ces mêmes dimères [126]. Le préfixe cp- indique que les termes ont été corrigés à courte portée, jusqu'à charge-quadrupole selon les approches GDMA, jusqu'à charge-dipole pour MEP et limité à charge-charge dans le cas des résultats extraits de la référence 126.

index	cp-GDMA ^{2}	cp-GDMA ^{1}	cp-MEP	Ref. 126	SAPT
1	-8.96	-8.18	-7.87	-6.9	-7.96
2	-7.32	-6.64	-6.94	-6.3	-6.72
3	-7.03	-6.36	-6.86	-6.5	-6.52
4	-6.48	-6.58	-6.10	-6.3	-6.61
5	-5.61	-5.78	-5.65	-6.0	-5.71
6	-5.40	-5.58	-5.56	-6.2	-5.41
7	-4.38	-4.71	-3.96	-4.5	-4.84
8	-1.20	-1.43	-1.19	-1.1	-1.49
9	-5.02	-4.84	-3.90	-4.9	-4.71
10	-2.87	-2.82	-2.48	-3.6	-2.71

TABLE 1.4 – Comparaison de l'énergie multipolaire corrigée à courte portée en kcal/mol sur dix différentes orientations de dimères d'eau. GDMA ^{i} décrit le schéma de la distribution des moments, $i = 1$ seulement atomique et $i = 2$ atome et liaison. Le préfixe cp- désigne les termes corrigés à courte distance.

Entre cp-GDMA ^{1} et cp-GDMA ^{2}, des différences notables d'énergies sont observées, allant de 0,7 à 1 kcal/mol pour les systèmes les plus bas en énergie (index de 1 à 4). Toutefois, cet écart se réduit pour les dimères avec une énergie plus haute. La différence entre ces deux approches sur plusieurs orientations des dimères n'est pas systématique : la relation entre ces deux approches n'est donc pas linéaire et une correction, qui modifie les termes charge-charge, charge-dipole et/ou charge-quadrupole, ne compensera pas forcément ces variations (voir exemple tableau 1.6).

Le tableau 1.5 contient les erreurs moyennes de chaque proposition de correction par rapport aux valeurs référence. cp-GDMA ^{1} se distingue nettement comme meilleure approche avec une erreur moyenne de 0,12 kcal/mol, soit

un tiers par rapport à cp-GDMA [2] et cp-MEP. Enfin la correction proposée dans la référence 126 présente une erreur moyenne quatre fois plus élevée que cp-GDMA [1].

	erreur moyenne
cp-GDMA [2]	0,36
cp-GDMA [1]	0,12
cp-MEP	0,36
Ref. 126	0,47

TABLE 1.5 – Erreur moyenne en kcal/mol par rapport à SAPT.

En se concentrant sur les complexes d'énergie les plus faibles (les cinq premiers dimères), il est intéressant de noter que la différence entre la valeur maximale et minimale de l'évaluation de l'énergie électrostatique est de 0,9 kcal/mol pour la cinquième colonne. cp-GDMA [2] présente un écart de plus de 3 kcal/mol. Enfin, cp-GDMA [1] et cp-MEP montrent, quant à eux, des écarts 2,4 et 2,2 kcal/mol respectivement. Leur variation est en corrélation avec celui de SAPT : 2,3 kcal/mol.

Cette observation tend à démontrer l'adaptabilité ou encore la transférabilité d'un modèle de correction de la pénétration de charge sur différents conformères. En prenant en compte l'erreur moyenne (voir tableau 1.5) et l'écart des valeurs maximales minimales sur les systèmes les plus faibles en énergies (pour les cinq premiers dimères), les approches présentant le meilleur accord avec SAPT sont cp-GDMA [1] suivi de cp-MEP. Rappelons que la première a bénéficié de plusieurs jeux de multipoles et chacun a été calibré pour reproduire au mieux les résultats référence et modifie trois termes du développement multipolaire (charge-charge, charge-dipole et charge-quadrupole) (voir tableau 1.6). La correction charge-quadrupole permet de baisser légèrement l'erreur moyenne. Enfin, la seconde reprend simplement les multipoles présents dans le modèle d'eau d'AMOEBA et cor-

rige les deux premiers termes de ce développement (voir tableau 1.7) avec plus de liberté : en utilisant trois paramètres par atome.

1.3.2 Comparaison & discussion de cp-GDMA [1] et cp-MEP

Les tableaux 1.6 et 1.7 contiennent l'énergie en kcal/mol de la correction de pénétration de charge totale et sa décomposition par termes modifiés. Enfin, le tableau 1.8 reprend les énergies électrostatiques purement multipolaires (donc non-corrigées, de ces deux approches mais aussi de GDMA [2] et les valeurs référence pour ne pas être exhaustif) afin d'avoir une vue complète de l'apport de la correction dans les deux cas.

Au delà de la qualité des multipoles entre ces deux approches, il est intéressant de noter que la correction totale reste dans le même ordre de grandeur, alors que leurs énergies multipolaires (tableau 1.8) sont différentes. Bien que cette correction totale semble décrire le même type de décroissance d'un point de vue global, la décomposition par termes modifiés apporte plus d'éclairage. En effet, dans le cas de GDMA [1], la modification du terme de charge-charge est toujours négative et est la principale responsable de la décroissance énergétique par rapport aux composantes modifiées. Enfin la correction charge-dipole tend à raffiner de manière plus subtile le résultat total de l'énergie électrostatique, tout comme celle apportée à charge-quadrupole. Alors que dans le cas de cp-MEP, seule la correction charge-charge contrôle l'évaluation de l'énergie de pénétration de charge dans le cadre des approches de type AMOEBA. La modification du terme charge-dipole est nulle dans les cas présenté ci-dessous mais l'implémentation laisse tout de même la possibilité de moduler ce terme pour d'autres systèmes. Enfin, le modèle d'eau AMOEBA a été amélioré sur dix géométries représentatives donnant ainsi lieu aux prémices d'un modèle hybride SIBFA/AMOEBA.

Finalement, ces observations du traitement de la pénétration de charge

sur ces deux approches ont pour origine les multipoles utilisés mais donnent des résultats satisfaisant en corrélation avec les calculs références SAPT, en particulier pour l'approche GDMA [1]. Ceci confirmant l'intérêt d'étude préliminaire comme présenté dans les tableaux 1.2 et 1.1 et ce, quelque soit l'approche : purement GDMA ou calibré sur un potentiel électrosatique (MEP). Au-delà d'une discussion sur la qualité des multipoles, il est intéressant de noter que la correction implémentée répond parfaitement aux attentes, c'est-à-dire d'apporter une correction à courte distance en accord avec les valeurs références. Sa tranférabilité est donc confirmée par cette étude, certes sur dix configurations de dimères d'eau. Des tests extensifs de ces deux corrections sont présentés dans la section suivante.

		E_{pen}		
index	total	charge-charge	charge-dipole	charge-quadrupole
1	-1.44	-1.22	-0.11	0.12
2	-1.48	-1.15	-0.14	0.20
3	-1.51	-1.14	-0.15	0.23
4	-1.76	-1.57	-0.03	0.15
5	-2.02	-1.65	-0.08	0.29
6	-2.14	-1.70	-0.10	0.34
7	-1.13	-1.13	0.03	0.03
8	-0.18	-0.25	0.04	-0.04
9	-1.04	-1.00	-0.03	0.02
10	-0.61	-0.58	-0.02	0.01

TABLE 1.6 – Décomposition de la correction de pénétration de charge obtenue via le protocole sw2-rad325 de GDMA avec des moments multipolaires centrés sur les atomes (GDMA [1]) en kcal/mol.

		E_{pen}	
index	total	charge-charge	charge-dipole
1	-1,83	-0,91	-0,93
2	-1,66	-0,93	-0,73
3	-1,64	-0,00	-1,64
4	-1,48	0,18	-1,65
5	-1,39	0,35	-1,73
6	-1,35	0,14	-1,49
7	-0,81	0,71	-1,51
8	-0,25	1,36	-1,61
9	-0,75	0,12	-0,87
10	-0,45	-0,31	-0,14

TABLE 1.7 – Décomposition de la correction de pénétration de charge dans le cadre d'AMOEBA (cp-MEP) en kcal/mol.

index	GDMA [2]	GDMA [1]	MEP	SAPT
1	-6.91	-6.73	-6.04	-7.96
2	-5.22	-5.15	-5.28	-6.72
3	-4.88	-4.85	-5.22	-6.52
4	-4.72	-4.82	-4.63	-6.61
5	-3.55	-3.76	-4.26	-5.71
6	-3.19	-3.44	-4.20	-5.41
7	-3.52	-3.58	-3.15	-4.84
8	-1.28	-1.26	-0.94	-1.49
9	-3.87	-3.79	-3.15	-4.71
10	-2.24	-2.21	-2.03	-2.71

TABLE 1.8 – Comparaison de l'énergie électrostatique purement multipolaire en kcal/mol.

1.3.3 Application des corrections de pénétration de charge

Dans cette section, les deux meilleures corrections ont été étendues à d'autres dimères. Dans un premier temps, le même protocole présenté dans les tableaux 1.1 et 1.2 a été mis en place sur des molécules de benzène afin d'obtenir le jeu de multipoles le plus adapté. Enfin dans un second temps, l'approche développée dans AMOEBA a été appliquée sur des systèmes neutres extraits du S66 [127] (complexes intermoléculaires représentatifs) et cationiques.

En appliquant le même protocole que dans la section 1.3.1, les multipoles les plus adaptés correspondent à la même combinaison que celle de la molécule

d'eau, c'est-à-dire l'option 2 de partitionnement de l'espace et un rayon de 0,325 Å assigné aux hydrogènes. Ce résultat ne tend pas à définir ce protocole comme « universel ». La correction cp-GDMA [1] a été appliquée sur des dimères contenant au moins un benzène. Les paramètres de la correction de pénétration de charge de l'eau restent inchangés, et seuls ceux du benzène ont été ajustés. Le tableau 1.9 reprend ces résultats qui présentent un bon accord avec SAPT, avec une erreur moyenne de 0,27 kcal/mol.

index		SAPT	cp-GDMA [1]
benzène - water	90	-4,37	-4,03
	95	-3,43	-3,36
	100	-2,77	-2,82
	105	-2,29	-2,39
	110	-1,93	-2,05
benzène - benzène *π stacking*	90	-8,48	-8,40
	95	-4,75	-6,27
	100	-2,56	-3,27
	105	-1,29	-1,96
	110	-0,56	-1,13
benzène - benzène *T-shaped*	90	-4,93	-4,66
	95	-3,22	-3 ,35
	100	-2,16	-2,44
	105	-1,49	-1,80
	110	-1,06	-1,34

TABLE 1.9 – Energie électrostatique de dimères composés de benzènes en kcal/mol.

Cette première étape avait pour but de vérifier le protocole de calculs des multipoles et de calibration de la correction proposée.

Enfin, la correction cp-MEP a été appliquée de manière plus extensive. La figure 1.7 représente des profils énergétiques électrostatiques en fonction de la distance intermoléculaire dans le cas de quatre orientations de dimères de l'eau différentes extrait des dix dimères étudiés jusque là.

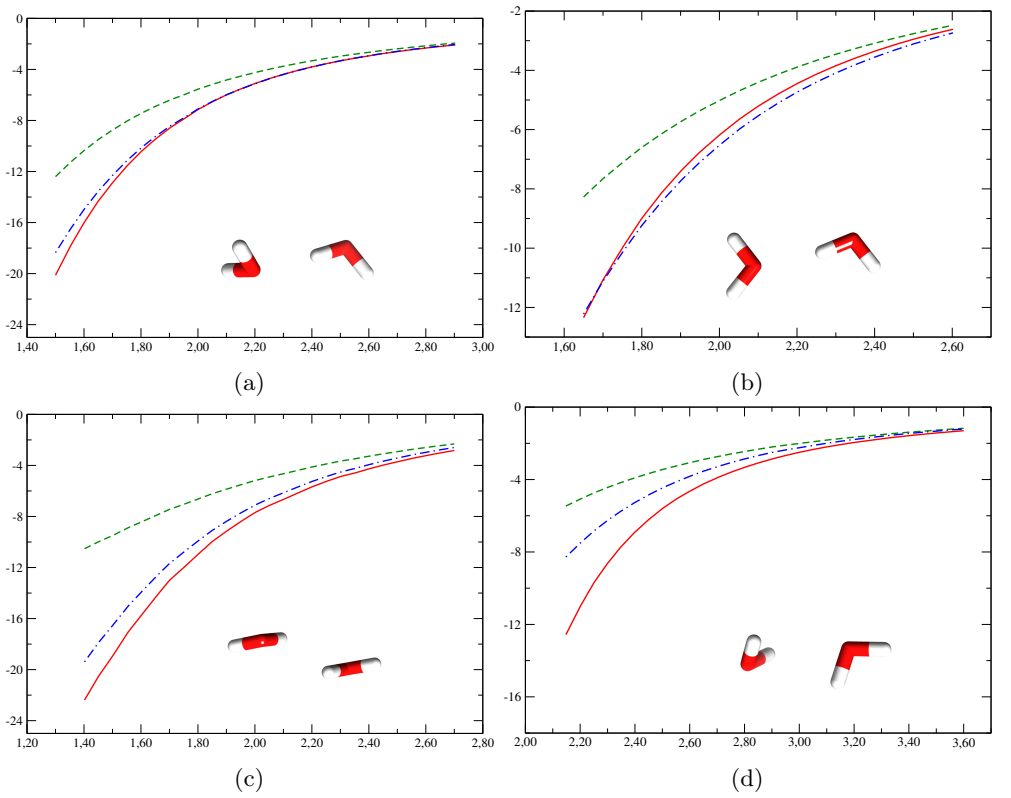


FIGURE 1.7 – Profil de l'énergie électrostatique de dimères de l'eau en fonction de la distance intermoléculaire en kcal/mol. SAPT est représenté par la courbe continue rouge. AMOEBA (MEP) en tiret vert. cp-AMOEBA (cp-MEP) en tiret pointillé bleu.

Les courbes en trait plein rouge correspondent aux calculs références SAPT. Celles vertes en trait discontinu se réfèrent à un calcul de l'énergie électrostatique via des multipoles recalibrés sur le potentiel électrostatique (MEP) réalisé avec le champ de force AMOEBA dans TINKER. Enfin, la correction de pénétration de charge appliquée à AMOEBA (cp-MEP ou cp-AMOEBA) est représentée par les courbes bleues pointillées et en trait discontinu. En considérant les résultats présentés précédemment dans les tableaux 1.4 et 1.5, ainsi que la figure 1.7, la correction cp-AMOEBA (cp-MEP) se confirme transférable le long de la distance intermoléculaire. Le modèle d'eau AMOEBA a été amélioré grâce à cette correction, confirmant ainsi les prémices d'un modèle hybride SIBFA/AMOEBA.

Ce même modèle de correction de l'énergie électrostatique à courte distance a été étendu sur une série de dimères extraite de la base de donnée de systèmes intermoléculaires représentatifs : S66 [127]. Cette série variée contient les dimères d'eau et de benzène précédemment étudiés mais aussi des systèmes présentant des fonctions alcool et amine. Ces complexes ont été évalués au voisinage de l'équilibre et sont représentés sur la figure 1.8. La compilation de l'ensemble de ces résultats montre une bonne corrélation avec les résultats références. Cependant, les écarts significatifs (jusqu'à 5kcal/mol) par rapport à la courbe idéale de corrélation concernent les scans intermoléculaires des dimères de l'eau.

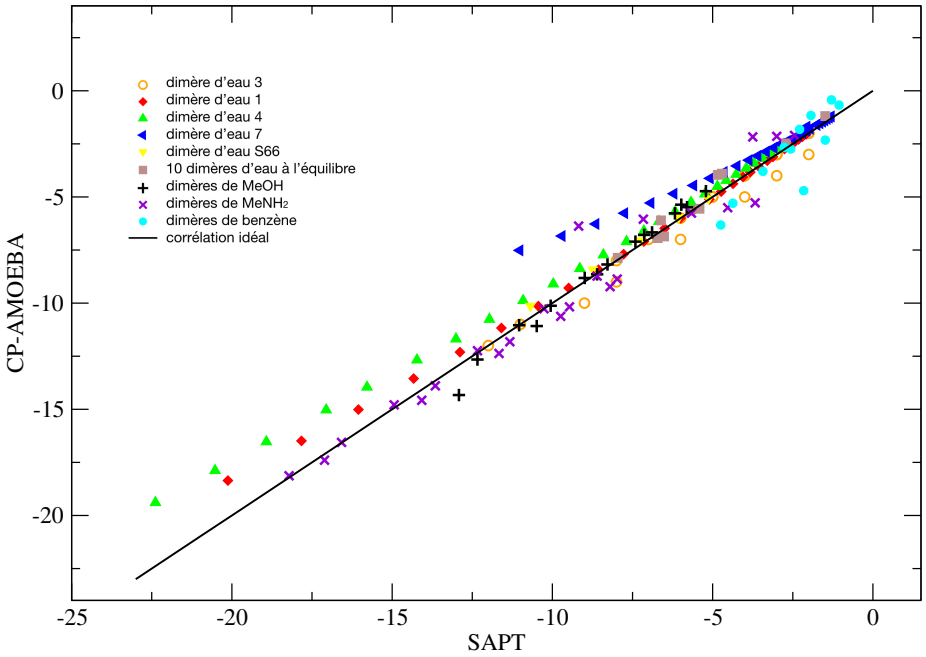


FIGURE 1.8 – Diagramme de corrélation de l'énergie électrostatique entre la référence SAPT et la correction de pénétration (cp-AMOEBA) en kcal/mol sur une série de dimères divers (voir légende)

Sur la figure 1.8, cette observation se rapporte aux triangles verts et bleus et décrivent des distances intermoléculaires très courtes. A cette distance, le caractère répulsif de l'interaction est la composante majoritaire et ne seront pas rencontrés lors de dynamique moléculaire.
Enfin, après avoir confirmé la transférabilité et l'applicabilité de cette correction sur des systèmes neutres, cp-AMOEBA a été appliqué à des systèmes chargés de type complexe eau/cation métallique, incluant les cations K^+, Na^+, Mg^{2+}, Ca^{2+} and Zn^{2+}.

Cette série de cations n'est décrite que par un monopole équivalent à leur charge portée. Ainsi, seule leur composante charge-charge se voit modifiée (équation 1.13). La figure 1.9 rapporte la comparaison des résultats selon la même légende que précédemment (figure 1.7). Cette simple correction de pénétration, charge-charge, améliore nettement le traitement de l'énergie électrostatique et présente un très bon accord avec SAPT. Finalement, la transférabilité du modèle de correction de pénétration de charge est confirmée dans un autre contexte, celui des complexes cations métallique - monomère neutre, très présents en biologie. En effet près d'un tiers des enzymes connues ont besoin d'un cation métallique pour réaliser des catalyses [23].

L'intérêt de méthodes telles que les champs de forces est de traiter des systèmes constitués de milliers d'atomes. Le coût de calcul devient alors limitant. Le calcul de l'énergie électrostatique en condition périodique permet de s'en affranchir mais l'applicabilité de ce type de correction proposée à courte portée reste à étudier.

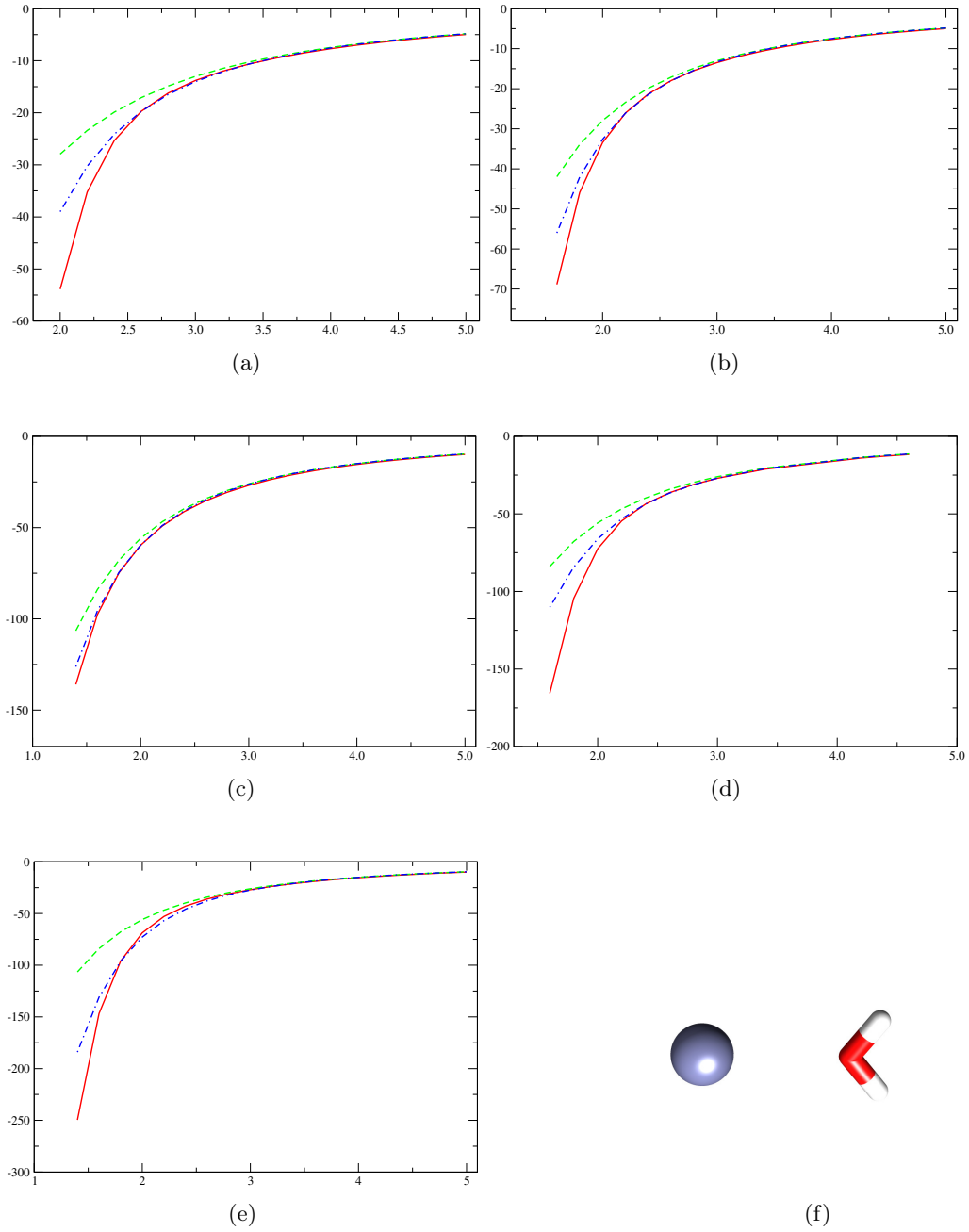


FIGURE 1.9 – Complexes cation métallique - eau : profil de l'énergie électrostatique en fonction de la distance cation - O en kcal/mol. SAPT correspond la la courbe trait plein rouge. AMOEBA en trait discontinu vert. cp-AMOEBA en bleu en trait discontinu et pointillé. a) complexe K^+ eau. b) complexe Na^+ eau. c) complexe Mg^{2+} eau. d) complexe Ca^{2+} eau. e) complexe Zn^{2+} eau. f) illustration de l'approche cation - oxygène (de la molécule d'eau).

1.4 Conditions périodiques

Comme déjà discuté en section 1.1 l'énergie électrostatique est une double sommation sur les atomes, c'est-à-dire un calcul par paires d'atomes non-liés. Pour N atomes, la complexité est N^2. Dès les premiers développements de champs de forces classiques, de nouvelles approches ont rapidement été proposées et plus particulièrement pour le traitement des interactions électrostatiques à longues distances. La première approximation a été l'introduction d'une distance seuil, au-delà de laquelle l'énergie est nulle. Dans le cadre de dynamique moléculaire, la conservation de l'énergie n'est alors plus assurée, tout comme sa dérivabilité au voisinage de ce seuil. Le calcul des forces, indispensable dans ce contexte, n'est donc pas fiable. Pour remédier à cela, plusieurs approches ont été envisagées. La première a été l'utilisation d'une fonction d'amortissement (*switching function*) qui amène l'énergie du système à zero après une distance seuil, appliquée aux interactions non-liées c'est-à-dire électrostatique et VAN DER WAALS. Si d'un point de vue mathématique, les problèmes précédemment cités sont réglés ; d'un point de vue physique, l'énergie totale du système perd en sa qualité par le fait de négliger les interactions au-delà d'une distance seuil. La seconde approche a été introduite par EWALD en 1921. La sommation d'EWALD, initialement développée dans le domaine de la physique théorique, en cristallographie, propose de séparer le calcul en deux termes, tel que :

$$\phi(r) = \phi_{sr}(r) + \phi_{lr}(r) \tag{1.15}$$

où $\phi(r)$ est le potentiel électrostatique selon r. $\phi_{sr}(r)$ est la composante à courte portée (sr pour *short range*) appelée somme directe et $\phi_{lr}(r)$ à longue distance (lr pour *long range*), somme réciproque. Cependant, physiquement, ce dernier terme ne contient pas strictement les interactions à longues distances. Il s'agit d'un abus de langage. En effet, il n'est pas nul à

courte distance. Schématiquement, pour mieux s'en rendre compte, on peut représenter la sommation d'EWALD par l'identité $1 = erf(r) + erfc(r)$ où erf est la fonction d'erreur et erfc son complémentaire (voir figure 1.10). Sur celle-ci, la courbe bleue correspondant au tracé de $erfc(r)/r$ montre bien qu'on ne traite pas uniquement les interactions à longue distance, puisqu'elle n'est pas nulle pour r petit.

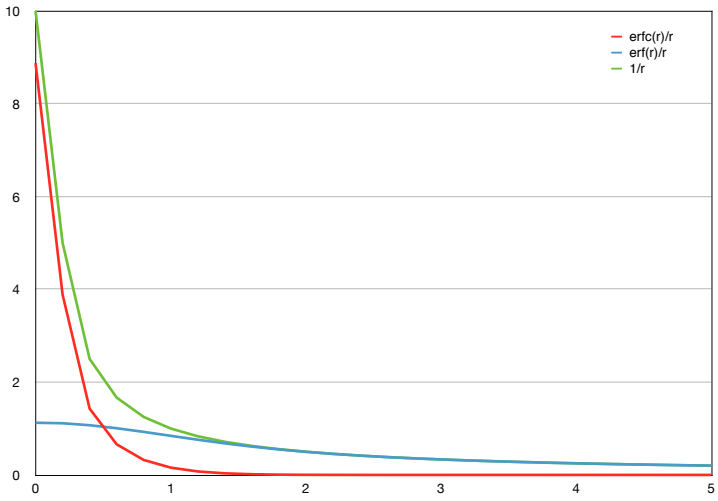


FIGURE 1.10 – Décomposition de $1/r$ en $erf(r)/r$ et $erfc(r)/r$: schéma sommation d'EWALD.

Il est intéressant de noter que cette idée de double somme contrôlée par cette identité a été reprise en développement de la théorie de la fonctionnelle de la densité (DFT), appelée à séparation de portée où un traitement différent est proposé à courte et longue portée [130].

Avant d'aller plus loin, il faut définir le contexte originel d'un tel algorithme et son transfert dans le domaine des champs de forces. Comme énoncé précédemment, cette méthode a été développée en cristallographie où la matière condensée s'organise en réseau symétrique.

Il est alors possible de dégager une structure, dite maille unitaire (voir figure 1.11), qui se répète de manière isotropique : voilà qui définit les conditions périodiques.

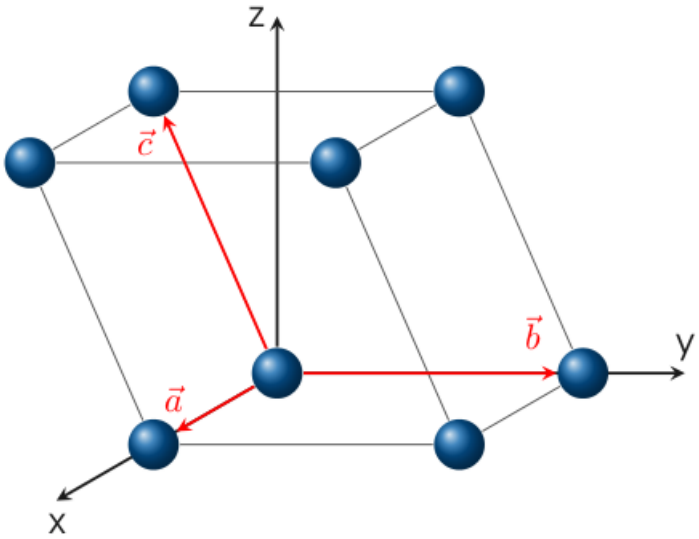


FIGURE 1.11 – exemple de motif : réseau de Bravais monoclinique primitif (paramètre de maille : a $\neq b \neq c, \alpha = \gamma = 90, \beta = 90$)

EWALD a donc exploité cette symétrie d'organisation des particules pour en calculer rigoureusement l'énergie électrostatique et en diminuer le coût. En chimie ou biochimie, on ne peut observer de telles répétitions de motifs moléculaires. Il a donc fallu dans un premier temps rationaliser les conditions périodiques dans le cadre des champs de forces pour profiter pleinement de cette méthode. Pour cela, considérons un système, une protéine par exemple ou un complexe supramoléculaire. Plongeons le dans un bain de milliers de molécules d'eau le plus souvent, permettant ainsi de reproduire les conditions biologiques. Il est alors possible d'étudier l'effet de ce bain sur notre système en conditions périodiques. L'effet de solvatation dit explicite est donc le domaine d'applicabilité de la sommation d'EWALD. Notre système entouré de molécules de solvant forme ainsi le motif qui sera

reproduit (voir zone bleutée sur la figure 1.12), comme le sont ceux en cristallographie à la différence près que dans notre cas, on peut parler de motifs fictifs puisque notre système ne sera pas influencé par son image qui sera à une distance physiquement trop éloigné pour cela (voir figure 1.12). Sur celle-ci, le motif à répliquer est coloré en bleu et correspond au système, au centre, solvaté. C'est cette cellule qui est donc répété de manière isotopique. Sur la figure 1.12, les réplicas ne sont représentés que selon deux directions.

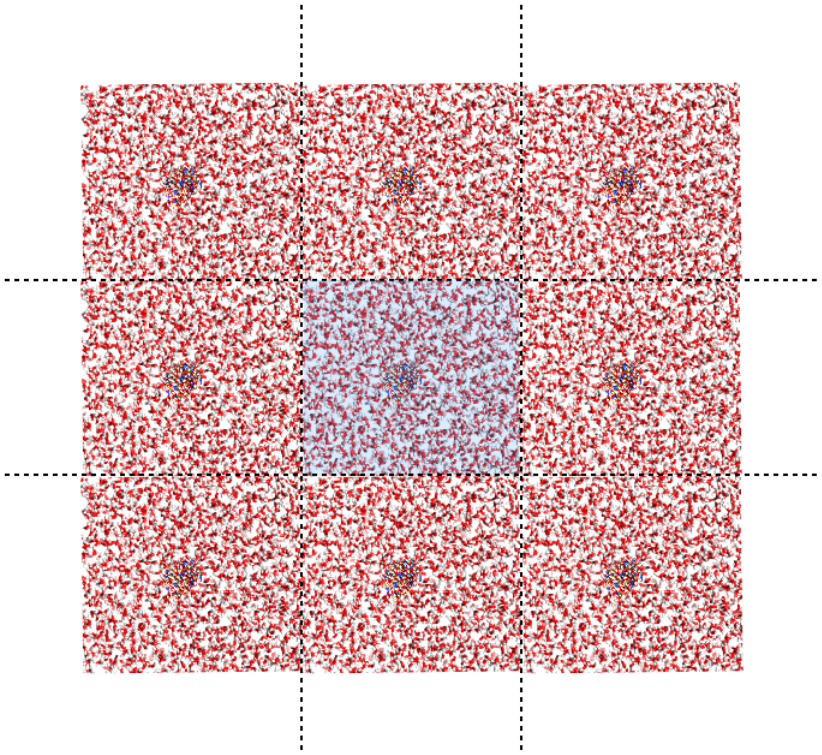


FIGURE 1.12 – Exemple schématique d'un système biologique en conditions périodiques en 2 dimensions.

Revenons maintenant à la formulation de cette sommation, qui divise l'énergie électrostatique sous la forme de trois termes :

$$E_{\text{Ewald}} = E_{\text{directe}} + E_{\text{self}} + E_{\text{réciproque}} \tag{1.16}$$

$$E_{\text{directe}} = \frac{1}{2} \sum_{\mathbf{n}} \sum_{1<i,j<N} q_i q_j \frac{\text{erfc}(\alpha^{1/2} |\mathbf{r_{ijn}}|)}{|\mathbf{r_{ijn}}|} \tag{1.17}$$

$$E_{\text{self}} = -\left(\frac{\alpha}{\pi}\right)^{1/2} \sum_{j=1} q_j^2 \tag{1.18}$$

$$E_{\text{réciproque}} = \frac{1}{2\pi V} \sum_{\mathbf{m}\neq 0} \frac{\exp(-\pi^2 \mathbf{m}^2/\alpha)}{\mathbf{m}^2} S(\mathbf{m}) S(\mathbf{\text{-}m}) \tag{1.19}$$

où $\mathbf{m}$ est le vecteur de la maille unitaire dans sa base duale et V son volume.

$$\mathbf{r_{ijn}} = \mathbf{r_j} - \mathbf{r_i} + \mathbf{n} \tag{1.20}$$

avec $\mathbf{n}$ est la norme du vecteur de maille unitaire. Enfin $S(\mathbf{m})$, appelé facteur de structure, est défini tel que :

$$S(\mathbf{m}) = \sum_{j=1}^{N} q_j \exp(2i\pi \mathbf{m} \cdot \mathbf{rj}) \tag{1.21}$$

α est un paramètre de l'équation, souvent appelé coefficient d'Ewald et permet de contrôler la frontière à partir de laquelle la somme directe tend rapidement vers zéro, où les interactions électrostatiques y sont alors négligées. En pratique, il est souvent confondu avec une distance seuil (*cutoff*). Plus α est grand, plus cette frontière est à courte portée. α varie donc en inverse de la distance seuil. La sommation d'EWALD permet d'abaisser le coût de calcul de N^2 à $N^{3/2}$ tout en conservant l'énergie.

Par la suite dans les années 1990, les travaux combinés de DARDEN et ESSMANN ont permis d'arriver au SPME (*Smooth Particle Mesh Ewald*) [27,131], permettant d'abaisser le coût de calcul en $N \log N$.

Ces méthodes algorithmiques n'ont été présentées ici qu'avec des charges ponctuelles. Qu'en est-t-il avec des multipoles ?
Dans le cadre du développement multipolaire, l'énergie électrostatique est [132, 133] :

$$E_{\text{électrostatique}} = \frac{1}{2}\sum_{\mathbf{n}}\sum_{1\leq i,j\leq N}\hat{L}_i\hat{L}_j\left(\frac{1}{|\mathbf{r_{ijn}}|}\right) \tag{1.22}$$

avec $\hat{L}_i$, l'opérateur multipolaire, s'exprimant :

$$\hat{L}_i = q_i + \boldsymbol{\mu_i}\cdot\nabla_i + \mathbf{Q_i} : \nabla_i\nabla_i \tag{1.23}$$

où ∇_i est le gradient en r_i et μ et Q sont les moments dipolaire et quadrupolaire. Les trois termes de la sommation d'EWALD deviennent alors :

$$E_{directe} = \sum_{\mathbf{n}}\sum_{1\leq i,j\leq N}\hat{L}_i\hat{L}_j\left(\frac{\operatorname{erfc}(\beta|\mathbf{r_{ijn}}|)}{|\mathbf{r_{ijn}}|}\right) \tag{1.24}$$

$$\begin{aligned} E_{\text{self}} = -\frac{\beta}{\sqrt{\pi}}\sum_{j=1}^{N}\Bigg(q_j^2 + \frac{2\beta^2}{3}(\mu_j^2 - 2q_j\operatorname{Tr}(\mathbf{Q_j})) \\ + \frac{4\beta^4}{5}(2\mathbf{Q_j} : \mathbf{Q_j} + \operatorname{Tr}^2(\mathbf{Q_j}))\Bigg) \end{aligned} \tag{1.25}$$

$$E_{\text{réciproque}} = \frac{1}{2\pi V}\sum_{\mathbf{m}\neq 0}\frac{\exp(-\pi^2\mathbf{m}^2/\beta^2)}{\mathbf{m}^2}S(\mathbf{m})S(-\mathbf{m}) \tag{1.26}$$

Maintenant que nous avons vu que l'utilisation du développement multipolaire est assurée en conditions périodiques grâce au SPME, comment y intégrer la correction à courte portée ? Dans un premier temps, reformulons l'énergie due uniquement à la pénétration E_{pen} dans le cadre d'une correction charge-charge et charge-dipole comme cp-AMOEBA :

$$\begin{aligned} E_{pen} &= E^*_{\text{elec}} - E_{\text{elec}} \\ &= \sum\sum_{1\leq i,j\leq N} E^*_{m_i-m_j} - E_{m_i-m_j} + E^*_{m_i-d_j} - E_{m_i-d_j} \end{aligned} \tag{1.27}$$

$$E^*_{\text{elec}} = E_{\text{elec}} + E_{pen} \tag{1.28}$$

avec E^*_{elec} qui est la somme de l'énergie de pénétration (E_{pen}) et l'énergie purement multipolaire (E_{elec}).

E_{pen} n'est finalement que la seule quantité à intégrer dans le SPME multipolaire. Cependant elle ne peut être directement couplée dans le formalisme du SPME à cause de sa forme exponentielle, dans les sommes réciproques et self. D'un point de vue physique, cette correction est à courte distance par nature puis s'atténue comme on peut le voir sur les figures 1.7 et 1.9 où l'énergie électrostatique modifiée tend vers la valeur non-modifiée. L'idée est donc de traiter la quantité corrigée E_{pen} à part, bénéficiant ainsi des avantages computationnels du SPME, restant donc inchangé. La correction apportée tend vers zéro dès 5 Å (voir figures 1.7 et 1.9). Il n'y a pas lieu de la calculer au-delà de ce seuil. L'introduction d'une nouvelle valeur seuil, comme pour le SPME, au-delà de la laquelle E_{pen} sera considérée comme nulle. Afin de rester dans un cadre rigoureux de dynamique moléculaire, plusieurs méthodes dites de *cutoff* ont déjà été proposées [134]. Parmi ces méthodes, nous avons choisi une fonction d'atténuation (*switching function*) afin de traiter E_{pen} à cette valeur seuil [135] :

$$S(r) = \begin{cases} 1 & r \leq R_{int} \\ \frac{(R_{ext}^2 - r^2)^2 (R_{ext}^2 + 2r^2 - 3R_{int}^2)}{(R_{ext}^2 - R_{int}^2)3} & R_{int} < r \leq R_{ext} \\ 0 & r > R_{ext} \end{cases} \tag{1.29}$$

De manière générale, pour toute distance r, l'énergie potentielle (ici électrostatique) est toujours égale à $V(r) \cdot S(r)$ où $V(r)$ est le potentiel non-atténué. Deux valeurs seuil sont définies dans l'équation 1.29 : R_{int} et R_{ext}. En pratique, le potentiel $V(r)$ est amené vers 0 entre ces deux valeurs et la différence entre R_{ext} et R_{int} est de l'ordre de 4 Å. En suivant cette méthodologie, les interactions à courte distance ne sont pas modifiées avant R_{int}.

De plus, la continuité des dérivées premières est assurée, évitant ainsi des artefacts physiques dus à l'instabilité énergétique rencontrée avec une troncature abrupte (voir figure 1.13).

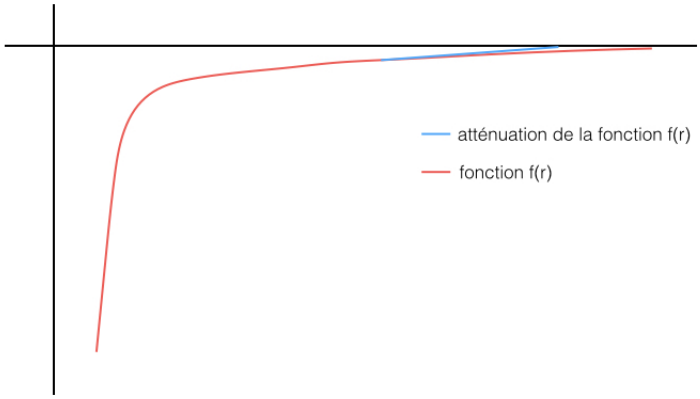


FIGURE 1.13 – Exemple d'application de la fonction d'amortissement sur $f(r)$.

L'implémentation de cette correction du calcul de l'énergie électrostatique, ainsi que ses forces, a été faite dans une version locale de TINKER : TINKER-HP. Les paramètres intrinsèques aux atomes dans le cadre de cette correction ont été ajoutés dans une version modifiée du fichier paramètre du champ de forces polarisable AMOEBA. L'activation de cette correction se fait par un simple mot clef dans le fichier .key où sont répertériorées les instructions pour une simulation numérique.

Ces travaux ont été acceptés dans le Journal of Computational Chemistry (JCC) inséré ci-dessous.

Scalable improvement of SPME multipolar electrostatics in anisotropic polarizable molecular mechanics using a general short-range penetration correction up to quadrupoles

Christophe Narth,[†] Louis Lagardère,[‡] Étienne Polack,[†,¶] Nohad Gresh,[†,§] Qiantao Wang,[‖] David R. Bell,[‖] Joshua A. Rackers,[⊥] Jay W. Ponder,[⊥] Pengyu Y. Ren,[*,‖] and Jean-Philip Piquemal[*,†]

UPMC Univ. Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris, France, UPMC Univ. Paris 06, Institut du Calcul et de la Simulation, F-75005, Paris, France, UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France, Chemistry and Biology Nucleo(s)tides and immunology for Therapy (CBNIT), UMR 8601 CNRS, UFR Biomédicale, Paris 75006, France, Department of Biomedical Engineering, The University of Texas at Austin, Texas 78712, and Department of Biochemistry and Molecular Biophysics, Washington University, St. Louis, Missouri 63110

E-mail: pren@mail.utexas.edu; jpp@lct.jussieu.fr

For Peer Review

Abstract

We propose a general coupling of the Smooth Particle Mesh Ewald (SPME) approach for distributed multipoles to a short-range charge penetration correction modifying to the charge-charge, charge-dipole and charge-quadrupole levels. Such an approach significantly improves electrostatics when compared to ab initio values and has been calibrated on Symmetry-Adapted Perturbation Theory (SAPT) reference data. Various neutral molecular dimers have been tested as results on charged systems (metal cation complexes) are provided. Transferability of the correction is adressed in the context of the implementation of the AMOEBA and SIBFA polarizable force fields in the TINKER-HP software. As the choices of the multipolar distribution are discussed, conclusions are drawn for the future penetration-corrected polarizable force fields highlighting the mandatory need of non-spurious procedures leading to the derivation of well balanced and physically meaningful distributed moments. Finally scalability and parallelism of the short-range corrected SPME approach are addressed, demonstrating that the damping function is computationally affordable and accurate for molecular dynamics simulations of complexe bio- or bioinorganic systems in periodic boundary conditions.

For Peer Review

1 Introduction

The field of molecular modeling is growing exponentially. The increase of available crystallographic structures[3–5] over the years has posed a substantial challenge for molecular force field development. For such systems, classical approaches still encounter successes but also reach their limits when describing non-bonded interactions which can only be reliably computed

*To whom correspondence should be addressed
†UPMC Univ. Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris, France
‡UPMC Univ. Paris 06, Institut du Calcul et de la Simulation, F-75005, Paris, France
¶UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France
§Chemistry and Biology Nucleo(s)tides and immunology for Therapy (CBNIT), UMR 8601 CNRS, UFR Biomédicale, Paris 75006, France
‖Department of Biomedical Engineering, The University of Texas at Austin, Texas 78712
⊥Department of Biochemistry and Molecular Biophysics, Washington University, St. Louis, Missouri 63110

using quantum chemistry. Year after year, it becomes clear that the gap between quantum and classical approaches is too large to be ignored, should one attempt to be predictive. First-generation force fields ignore polarization effects and employ a simple electrostatic term while a van der Waals interaction term takes the role of a buffer contribution including all quantum-like effects. To compute interaction energies of high accuracy that could be compared directly to quantum chemistry and/or experiments, one should build force fields that embody all anisotropic effects and are able to separately reproduce all quantum contributions. Among all terms, electrostatics remains prominent.[6,7] Point charges can appropriately capture some long-range electrostatics but, quantum chemistry clearly shows the need to go beyond such an approximation to handle anisotropic interactions. Usually, distributed multipoles are used to account for polarization, but they are devoid of the short-range electrostatic quantum effects, the so-called charge penetration, which occurs when the molecular densities overlap. Various strategies have been employed to include such effects. They range from damping functions[1,8] to neural networks[9] and to newly developed point-distributed electron densities such as the GEM[10,11] potential. Quantum mechanical force fields[12] also address this topic. Clearly, both short- and long-range electrostatics do matter and one would like to treat them in concert. This was done in the context of GEM that uses continuous electrostatics which resort to electron densities. To do so, a generalized Smooth Particle Mesh Ewald approach was introduced which handles Hermite Gaussians thereby including penetration effects. As shown initially by Darden *et al*,[13,14] the PME and SPME strategies provide a scaling improvement enabling to deal with electrostatics in periodic boundary conditions with a $N\log(N)$ complexity as compared to the usual direct space N^2. In this contribution we propose a simple strategy based on a damping function to include short-range penetration within a distributed multipole SPME framework. Its aim is to be applicable to modern polarizable molecular dynamics using force fields such as AMOEBA[15] and SIBFA.[1,2,16–19] The paper is organized as follows. We first present our choice of charge penetration correction. We then briefly recall the SPME methodology for distributed multipoles before presenting a

For Peer Review

newly developed coupled approach. Finally, we present the parametrization procedure, the numerical results and a parallel implementation.

2 Charge Penetration

The question of the atomic charge description has previously been investigated.[1,10,20–25] In classical force fields, the charge distribution is modeled using point charges calibrated on the Molecular Electrostatic Potential (MEP) computed with QM (Quantum Mechanical) methods.[26] Although the anisotropy, is not directly taken into account even in the case of MEP fitting,[27,28] this method is still widely used among various communities as it allows the simulation of very large systems due to its low computational cost.

The Distributed Multipole Analysis[29] (DMA) has been shown to be a suitable description of the charge distribution to improve electrostatics. Other distributed multipole methods exist. For example, one of the first full-fledged polarizable force fields, SIBFA,[1,2,16–18] (Sum of Interactions Between Fragments Ab-initio computed), originally used the multipole partitioning scheme developed by Vigné-Maeder and Claverie.[30,31] Nevertheless, none of these multipole methods properly account for the overlap of molecular densities at short-range. In physical systems, this overlap gives rise to a mainly attractive energy denoted electrostatic penetration. Hence, distributed multipolescan not be rigorously accurate at all ranges because of their classical approximation of electrostatics: theycan not describe the purely quantum mechanical charge penetration effect.

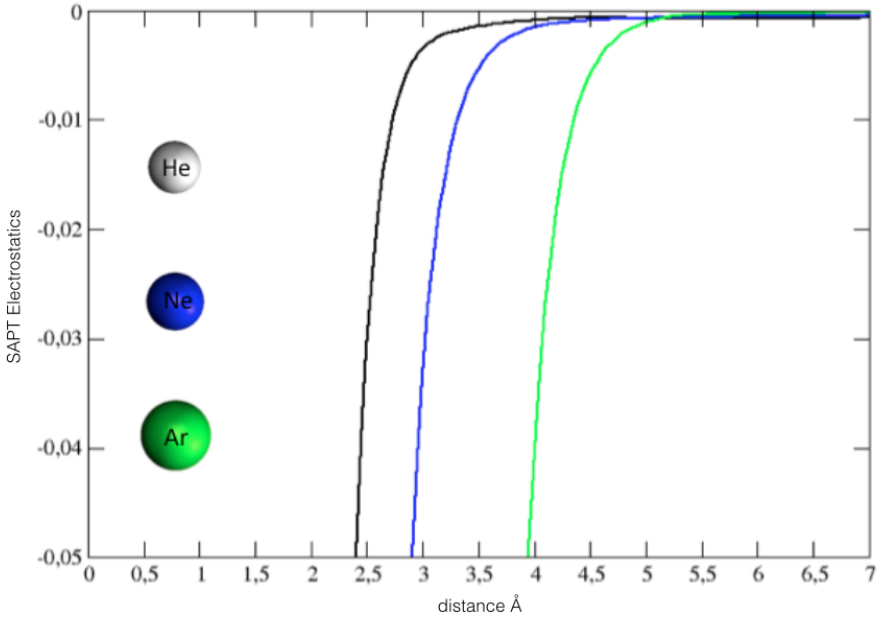
From quantum intermolecular perturbation theory, one can see that charge penetration exhibits an exponential decay being density overlap-dependent.[32]

Charge penetration has been widely studied in QM. For example, Jeziorska *et al* studied Helium dimer intermolecular potential to find a classical description of its potential by means of Energy Decomposition Analysis (EDA), here Symmetry Adapted Perturbation Theory[33,34] (SAPT).

For Peer Review

We extended this SAPT study to Neon and Argon dimers. Figure 1 exhibits the electrostatic component for these three rare gas dimers as a function of their interatomic distances. One can notice i) the existence of a non-zero electrostatic energy component at short-range and ii) the confirmation of the exponential decay. This observation and the fact that the rare gas charge is zero, and therefore does not have any classical electrostatic contribution, suggests that short-range effects should also be included. In that context, Piquemal *et al* have

Figure 1: Electrostatics contribution of rare gas dimers in kcal/mol



proposed a damping correction[1,2] grounded in Quantum Chemistry (QC) (see equation 1). Indeed, the electrostatic energy between two monomers (A and B) can be expressed as :

$$E_{\text{elec}} = -2\sum_{i}\sum_{\nu} Z_\nu \int \frac{|\varphi_i(1)|^2}{r_{i\nu}} {dr_1}^3 - 2\sum_{j}\sum_{\mu} Z_\mu \int \frac{|\varphi_j(2)|^2}{r_{j\mu}} {dr_2}^3 + 4\sum_{i}\sum_{\text{J}} \int \frac{|\varphi_i(1)|^2|\varphi_j(2)|^2}{r_{12}} {dr_1}^3 {dr_2}^3 + \sum_{\mu}\sum_{\nu} \frac{Z_\mu Z_\nu}{r_{\mu\nu}} \quad (1)$$

where μ (ν) and φ_i (φ_j) are, respectively, the nucleus and the unperturbed molecular orbitals of monomer A (B) and Z is the atomic number. The first two terms define nucleus-electron

attraction, the third the electron-electron repulsion and the fourth the nucleus-nucleus repulsion. This strongly suggests that mimicking such a term is a key ingredient to correctly model all the components of electrostatics. The electronic integrals explain the appearance of electrostatic exponential decay. As charge penetration is missing, it seems reasonable to consider the outer-shell electrons (electronic valence) in the construction of an overlap-based short-range correction. Indeed, as shown in Figure 1 the atomic size is important, and is closely linked to the magnitude of the penetration effect. One would like to consider van der Waals-like radii as a component of the parametrization. In that context,[2] Piquemal *et al* proposed a short-range correction to the multipolar electrostatics that includes all these features and was originally implemented in SIBFA. The latter modifies the charge-charge, charge-dipole and charge-quadrupole interactions.

The aim of this paper is to include and to reparametrize this correction in the framework of the Smooth Particle Mesh Ewald. The proposed correction modifies the charge-charge (2) and charge-dipole (3). The latter involves an damped charge that is also appicable to correct the charge-quadrupole interactions.[2] These three terms are the dominant ones.[1]

$$\begin{aligned} E^*_{m_i-m_j} =& \frac{1}{r_{ij}}[Z_iZ_j - Z_i(Z_j - q_j)(1 - exp(-\alpha_j r_{ij})) \\ & - Z_j(Z_i - q_i)(1 - exp(-\alpha_i r_{ij})) \\ & + (Z_i - q_i)(Z_j - q_j)(1 - exp(-\beta_i r_{ij})(1 - exp(-\beta_j r_{ij})] \end{aligned} \tag{2}$$

where $\alpha_i = \gamma/r^i_{vdw}$ and $\beta_i = \delta/r^i_{vdw}$, γ and δ are normalization coefficients and r^i_{vdw} the van der Waals radius of atom i. Z_i is the electronic valence of the atom i, q_i the charge (monopole) of atom i.

$$E^*_{m_i-d_j} = \frac{\boldsymbol{\mu}_j.\boldsymbol{r}_{ji}}{r^3_{ij}}(Z_i - (Z_i - q_i)(1 - exp(-\eta_i r_{ij})) \tag{3}$$

where $\eta_i = \chi/r^i_{vdw}$, with χ a normalization coefficient.

John Wiley & Sons, Inc.
For Peer Review

This charge-charge correction contains the traditional coulombic expression, corresponding to nucleus-nucleus repulsion. The formulation considers two more contributions, nucleus-electron attraction and electron-electron repulsion, where the damping function appears. It describes the depletion of the electronic cloud at short-distance. Following the same idea, the charge-dipole also accounts for this depletion by modulating the charge.[20] It is worth mentioning that at long-range, this term tends to the classical coulombic electrostatic expression. In Molecular Dynamics (MD), conformational exploration of space can lead to very short-interatomic distances where multipoles clearly fail to recover the QM values due to the lack of charge penetration effect. Usually such errors are compensated by parametrization of the van der Waals terms. However, there is no guarantee of a systematic compensation during such MD simulations, the lack of which could lead to errors. In order to apply this methodology to large systems and perform MD, we propose to couple this approach to the Smooth Particle Mesh Ewald[13] (SPME).

2.1 Smooth Particle Mesh Ewald

SPME is based on the Ewald summation[?] which suggests to split the electrostatics into two terms, a short-range E_{dir} and a long-range one E_{rec} and whose complexity is N^2 or even $N^{3/2}$.[36] SPME decreases it to $N \log N$ by using Fast Fourier Transforms (FFT).[?]

SPME can be applied to condensed phase. In that case, the lower complexity leads to particularly large time savings for large systems in Molecular Dynamics. SPME is an efficient and adaptable algorithm, as its formulation for multipoles[37,38] and for the charge penetration correction suggests. In this section, we show how to modify the general SPME formulas in order to include the charge penetration correction.

Let U be a neutral unit cell with N particles in position $(\mathbf{r}_i)_{i\in\{1,\cdots,N\}}$ and respective multipolar moments $(\hat{L}_i)_{i\in\{1,\cdots,N\}}$, $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ its basis and Σ the system made of an infinite number of images of the unit cell.

For Peer Review

For any integers n_1, n_2, n_3, m_1, m_2 and m_3 we will note

$$\mathbf{n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \quad \mathbf{m} = m_1\mathbf{a}_1^* + m_2\mathbf{a}_2^* + m_3\mathbf{a}_3^*,$$

where $(\mathbf{a}_1^*, \mathbf{a}_2^*, \mathbf{a}_3^*)$ is the dual basis of $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ (*i.e.* $\mathbf{a}_i^* \cdot \mathbf{a}_j = \delta_{ij}$).

The electrostatic energy of the system is

$$E_{\text{elec}}(\Sigma) = \frac{1}{2}\sum_{\mathbf{n}}{}' \sum_{1\leq i,j\leq N} \hat{L}_i\hat{L}_j \left(\frac{1}{|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|}\right), \tag{4}$$

where the prime means that the terms $i = j$ are not summed up when $\mathbf{n} = \mathbf{0}$. If we consider multipoles up to quadrupoles, then the operators $\hat{L}_i$ can be written as

$$\hat{L}_i(\mathbf{m}) = q_i + \boldsymbol{\mu}_i \cdot \nabla_i + \boldsymbol{Q}_i : \nabla_i\nabla_i, \tag{5}$$

where ∇_i is the gradient at $\mathbf{r}_i$, $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ are the dipoles and the quadrupole moments.

For any $\mathbf{m} \in \mathbb{R}^3$ we define the structure factor

$$S(\mathbf{m}) = \sum_j \tilde{L}_j(\mathbf{m}) \exp(2\pi\mathbf{m} \cdot \mathbf{r}_j) \tag{6}$$

where $\tilde{L}_j$ is the Fourier transform of the operator $\hat{L}_j$, *i.e.*

$$\tilde{L}_j = q_j + 2i\pi\boldsymbol{\mu}_j \cdot \mathbf{m} - (2\pi)^2\boldsymbol{Q}_j : \mathbf{m}\mathbf{m}, \tag{7}$$

If Σ is supposed to be surrounded by a medium with infinite dielectric constant, then it has been shown that for any $\beta > 0$ the sum can be split as[38]

$$E_{\text{elec}}(\Sigma) = E_{\text{dir}} + E_{\text{self}} + E_{\text{rec}}, \tag{8}$$

where

For Peer Review

$$E_{\text{dir}} = \sum_{\mathbf{n}}{}' \sum_{1\le i,j\le N} \hat{L}_i \hat{L}_j \left(\frac{\operatorname{erfc}\left(\beta|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|\right)}{|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|} \right), \tag{9}$$

$$E_{\text{self}} = -\frac{\beta}{\sqrt{\pi}} \sum_{j=1}^{N} \left(q_j^2 + \frac{2\beta^2}{3} \left(\boldsymbol{\mu}_j^2 - 2q_j \operatorname{Tr}(\boldsymbol{Q}_j) \right) + \frac{4\beta^4}{5} \left(2\boldsymbol{Q}_j \colon \boldsymbol{Q}_j + \operatorname{Tr}^2(\boldsymbol{Q}_j) \right) \right), \tag{10}$$

and (11)

$$E_{\text{rec}} = \frac{1}{2\pi V} \sum_{\mathbf{m}\neq 0} \frac{\exp(-\pi^2 \mathbf{m}^2/\beta^2)}{\mathbf{m}^2} S(\mathbf{m}) S(-\mathbf{m}). \tag{12}$$

The self energy term[39] E_{self} is a bias due to the Ewald summation and does not have any physical meaning. Both the direct and the reciprocal sums converge rapidly. A 9 Å cutoff is common, allowing a $O(N)$ computational complexity for the direct sum. The B-spline approximation of the structure factor on a grid of dimensions $K_1 \times K_2 \times K_3$ enables the reciprocal energy to be approximated by

$$\tilde{E}_{\text{rec}} = \frac{1}{2} \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \sum_{k_3=0}^{K_3-1} Q(k_1,k_2,k_3) \cdot (\theta_{\text{rec}} * Q)(k_1,k_2,k_3), \tag{13}$$

where Q is a derivable function obtained after interpolation of the structure factors at the nodes of the grid; and θ_{rec} a pair-potential function independent of particle position. The regularity of Q implies that the force in position i can be approximated by

$$\tilde{\mathbf{F}}_i = -\nabla_i \tilde{E}_r = \frac{1}{2} \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \sum_{k_3=0}^{K_3-1} \nabla_i Q(k_1,k_2,k_3) \cdot (\theta_{\text{rec}} * Q)(k_1,k_2,k_3). \tag{14}$$

In principle, the short-range correction is only in the direct space and therefore we will show that no modification of the reciprocal sum is required. We will detail our strategy of implementation aimed at an optimal scalability of the corrected electrostatics. As we discussed, the presented implementation in SPME requiries a modification of the direct sum only. To do so, we need to reexpress the damping function by Piquemal et al. in more

For Peer Review

convenient form as a simple correction to the direct sum.

2.2 Coupling Smooth Particle Mesh Ewald and a penetration correction

Let us define E^*_{elec} the total modified/damped electrostatic energy of the system Σ.

$$E^*_{\text{elec}}(\Sigma) = \frac{1}{2}\sum_{\mathbf{n}}{}' \sum_{1\leq i,j\leq N} E^*_{m_i-m_j} + E^*_{m_i-d_j} + E^*_{m_i-q_j} + E_{d_i-d_j} + E_{d_i-q_j} + E_{q_i-q_j} \tag{15}$$

where $E_{d_i-d_j}$ is the interaction between dipoles of site i and site j ; $E_{m_i-q_j}$, $E_{d_i-q_j}$ and $E_{q_i-q_j}$ are respectively the interaction between monopole/dipole/quadrupole of site i and quadrupole of site j. Let us define E_{pen} the total charge penetration energy,

$$E_{pen} = E^*_{\text{elec}} - E_{\text{elec}} = \sum_{\mathbf{n}}{}' \sum_{1\leq i,j\leq N} E^*_{m_i-m_j} - E_{m_i-m_j} + E^*_{m_i-d_j} - E_{m_i-d_j} \tag{16}$$

Let us define the undamped electrostatic interaction between the monopoles of sites i and j $E_{m_i-m_j}$, let us also define the undamped electrostatic interaction between the dipole of site j and the monopole of site i $E_{m_i-d_j}$.

$$\begin{aligned} E^*_{m_i-m_j} - E_{m_i-m_j} &= (q_i - Z_i)(q_i - Z_j)(\exp(-\beta_i r_{ij}) - 1)(\exp(-\beta_j r_{ij}) - 1) \\ &- (q_j - Z_i)Z_j(\exp(-\alpha_i r_{ij}) - 1) - (q_i - Z_j)Z_i(\exp(-\alpha_j r_{ij}) - 1) - q_i q_j + Z_i Z_j \end{aligned} \tag{17}$$

$$E^*_{m_i-d_j} - E_{m_i-d_j} = \frac{\boldsymbol{\mu}_j.\boldsymbol{r}_{ji}}{r^3_{ij}}(Z_i - q_i)exp(-\eta r_{ij}) \tag{18}$$

$$E^*_{\text{elec}} = E_{\text{elec}} + E_{pen} \tag{19}$$

The same strategy has been followed to correct the charge-quadrupole interaction.[2]

As seen in section 2.1, E_{pen} can not be simply included into the SPME framework, as we

For Peer Review

should take a large cutoff for the direct sum to avoid any persistence of penetration outside the direct sum treatment.

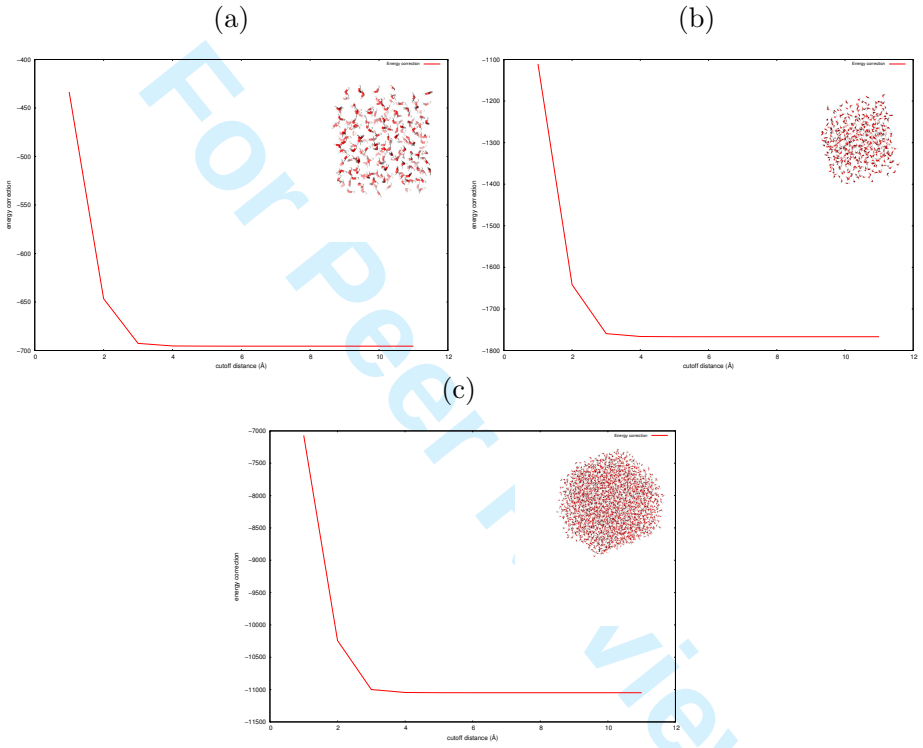
Indeed, the charge penetration energy is prominent at short-range and tends quickly to zero at long-range because of its $\exp(-x)/x^n$ (with $n = 1, 3$) form (see figure 4) but not necessarily fast enough to be maintained within the Ewald direct sum cutoff. As such Ewald cutoff is essential for scalability, one would like to separately handle SPME and penetration issues. This justifies the choice of an additional cutoff method dedicated to penetration. Indeed, for a given precision, one can always find a cutoff for penetration that guarantees an error inferior to this threshold. As such an approach generates a discontinuity and non-derivability in the potential function near the cutoff limit. We will couple it with a switching function that ensures derivability of the energy during a MD. Many smoothing functions have been proposed[40] and we chose to use the one defined in Ref. 41 and 42 (equation (20)). Hence, for every distance r, the potential energy is equal to $V(r) \cdot S(r)$ where $V(r)$ is the unswitched function. The interactions are not modified for distances less than an inner cutoff distance R_{in} and are smoothed to zero at the outer cutoff R_{out}. Besides, the continuity of the first derivatives prevents jumps in the energy during MD or minimization, which could lead to non-physical phenomena such as heating.

Such a technique enables to keep the long-range effects and the computational advantage of SPME using FFTs with an optimal Ewald cutoff, while just adding a computationally cheap correction term with its own cutoff (see Figure 2). Besides, numerical simulations of common systems (up to tens of thousands of atoms) show that a cutoff between 6 and 8 Å guarantees an error inferior to 10^{-3} kcal/mol (see Figure 2). Of course, the smoothing function makes possible to choose a more aggressive cutoff of 4 Å(see Figure 2) which will capture most of the penetration interaction in common systems leading to significant computational speed-up. Overall the final cost in term of computational ressources of the penetration correction is inferior to 10 per cent of the total penetration corrected multipolar electrostatics.

For Peer Review

$$S(r) = \begin{cases} 1 & r \leq R_{in} \\ \frac{(R_{out}^2 - r^2)^2 (R_{out}^2 + 2r^2 - 3R_{in}^2)}{(R_{out}^2 - R_{in}^2)3} & R_{in} < r \leq R_{out} \\ 0 & r > R_{out} \end{cases} \quad (20)$$

Figure 2: Convergence cutoff study of penetration energy. Energy in kcal/mol.



For Peer Review

3 Results

3.1 Parametrization

The electrostatic correction includes four empirical parameters: two for charge-charge interactions (core-electron attraction and electron-electron repulsion), one for charge-dipole interactions, and one for charge-quadrupole. Those were initially adjusted on water dimers.[43–46] As various parametrization strategies are possible, the atomic radii can be used as a pa-

rameter in order to reproduce the charge penetration effect. We apply this methodology to various choices of distributed multipoles. We focused on the GDMA approach by Stone[29] and on the electrostatic potential (ESP)? ren2011polarizable fitted multipoles used in the AMOEBA force field. We first follow the initial parametrization strategy developed for SIBFA and then we choose a simpler exponential functional form for the parametrization that does not take into account any atom radii. In this form, the parameters are the previously defined exponents [2] : α_i, β_i and η_i.

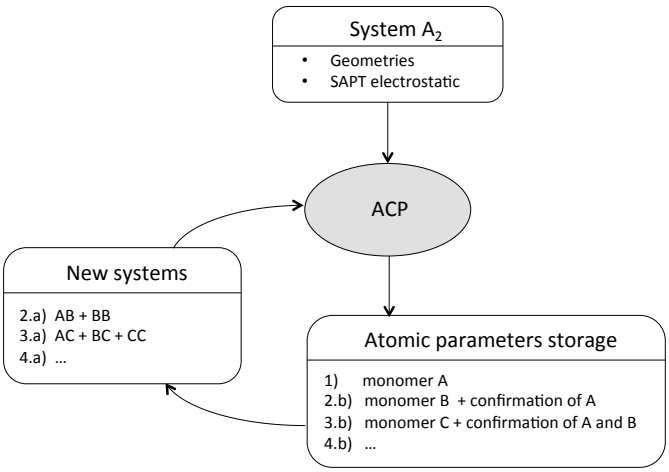
We focused our parametrization on a well established panel of intermolecular systems extracted from the S66 dataset as well as on selected configurations of water dimers obtained from high level ab-initio computation[44,47,48] . Several QM intermolecular energy decomposition techniques are now available such as reduced variational space[49] (RVS) or SAPT.[34] SAPT offers the possibility to include all types of intermolecular interactions, as it allows to take into account electronic correlation. Because our upcoming implementation should reproduce all of them in a new anisotropic polarizable force field based on *ab-initio* calculation, we chose to calibrate the electrostatics on SAPT. Note that the idea is to reproduce electrostatic trends at all ranges, with a special focus at short-ranges where charge penetration matters, which could make a perfect match with QM reference values difficult.

The first step of our calibration was to find the best set of parameters for ten different water dimers associated to stationary point on the ab-initio potential energy surface (denoted A on Figure 3). Four of these dimers were chosen (see Figure 4) to refine our results because they represent different chemical situations. Next, a new dimer (denoted B on Figure 3) as well as monomer-water systems (step 2 on Figure 3) can be calibrated. Thus, new parameters can be tested on dimers involving non-calibrated monomers (denoted C on Figure 3). This also allows to validate the previous calibration.

Each system was considered at its lower energy value configuration and at different intermolecular distances. These kind of variation are likely to happen in MD and accounting for them will make our model more robust, as it would be able to sample density overlapping

For Peer Review

Figure 3: Flowchart Automated Calibration Procedure (ACP)



regions.

3.2 Computational Details

The parallel implementation was first tested on the SGI UV 2000 supercomputer of the ICS (Institut du Calcul et de la Simulation) which is made of 64 nodes each one of them having two Intel Xeon E5-4650L CPUs with eight cores at 2.6 Ghz and 256 GB of DDR3 DIMM RAM. The nodes are able to communicate within a global shared memory system thanks to the SGI NUMALINK interconnect. We also tested our MPI implementation on the Stampede supercomputer of the TACC (Texas Advanced Computing Center) whose architecture is more standard as it consists, for the part we did our tests on, in 6400 compute nodes with two intel Xeon E5-2680 CPUs with eight cores at 2.7Ghz, one Intel Xeon Phi SE10P coprocessor and 32 GB of DDR3 DIMM RAM. These nodes communicate within a 56 GB/s InfiniBand network. The following numerical results are all based on computations made on this supercomputer. SAPT calculations have been performed using PSI4[50] at third order for water dimers and

second for the others, using aug-cc-pTZ basis sets. All DMA multipoles were computed from density at the MP2(full)/aug-cc-pTZ.

3.3 Numerical Results

Table 1: Comparison of the pure multipolar electrostatic energy (in kcal/mol) water dimers. GDMA^i describes the multipolar sites scheme, $i = 1$ atoms only and $i = 2$ atoms and bonds.

index	GDMA^2	GDMA^1	ESP	SAPT
1	-6.913	-6.733	-6.036	-7.960
2	-5.218	-5.154	-5.275	-6.716
3	-4.878	-4.851	-5.220	-6.516
4	-4.724	-4.820	-4.627	-6.610
5	-3.546	-3.761	-4.263	-5.713
6	-3.191	-3.441	-4.204	-5.410
7	-3.517	-3.582	-3.151	-4.838
8	-1.275	-1.256	-0.944	-1.486
9	-3.866	-3.793	-3.152	-4.709
10	-2.237	-2.212	-2.026	-2.712

Table 2: Comparison of the penetration corrected electrostatics (in kcal/mol) water dimers. GDMA^i describes the multipolar sites scheme, $i = 1$ atoms only and $i = 2$ atoms and bonds.

index	cp-GDMA^2	cp-GDMA^1	cp-ESP	Ref. 25	SAPT
1	-8.96	-8.18	-7.87	-6.9	-7.96
2	-7.32	-6.64	-6.94	-6.3	-6.72
3	-7.03	-6.36	-6.86	-6.5	-6.52
4	-6.48	-6.58	-6.10	-6.3	-6.61
5	-5.61	-5.78	-5.65	-6.0	-5.71
6	-5.40	-5.58	-5.56	-6.2	-5.41
7	-4.38	-4.71	-3.96	-4.5	-4.84
8	-1.20	-1.43	-1.19	-1.1	-1.49
9	-2.87	-4.84	-3.90	-4.9	-4.71
10	-5.02	-2.82	-2.48	-3.6	-2.71

Tables 1 and 2 gather various results for ten selected orientation of the water dimer[44] including pure multipolar interactions (Table 1) and damped multipolar ones (Table 2). Such configurations are clearly difficult as they embody various types of orientations involving

For Peer Review

very different anisotropic orbital interactions. The three first columns of table 1 describe the values of the GDMA, atoms and bond midpoints then atoms only and ESP fitted multipoles energies. In the first approximation, compared to reference SAPT and to atoms only GDMA multipoles, one would think that having more centers improve electrostatics. Overall, the ESP fit strategy seems to improve the agreement with SAPT results. The addition of more centers has been already described in the literature by Claverie and Stone[17,29] as a way to capture higher order effects, atoms plus bond midpoints being equivalent to the addition of octupoles within a distributed multipoles analysis. Such an observation might need, however, to be reconsidered. Indeed, one would think first that the derivation of multipoles is non-empirical, however, choices exist within the GDMA algorithm including the choice of the partition approach, the choices of the hydrogen radius etc. Here we tried various settings and kept the following values for the GDMA approach (space partitioning : switching 2, Hydrogen radius 0,325 Å). Choices also exist for the ESP fitted multipoles as, within the procedure, the charge is initially derived from GDMA and is kept frozen whereas dipole and quadrupole moments are reoptimized to minimize the differences with respect to the ab initio electrostatic potential. Overall, with all possible choices, a continuum of multipoles set can be derived with strong energetical differences. Which one should be chosen and are the ESP fitted multipoles really better. The application of the damping function to distributed multipoles casts a new light on such an issue. Clearly, the atoms plus bond midpoints GDMA and ESP fitted multipoles corrected results (see table 2) are improved by damping but surprisingly the best agreement is found for the damping of atoms only GDMA multipoles. A clear-cut explanation can be found by a careful analysis of the multipolar values alone (table 1) compared to SAPT. Both GDMA (Atoms plus midpoints) and ESP multipoles energies are too high. Indeed, even, with the best choice of settings, some GDMA configurations energies are almost equal the full quantum SAPT energies that embodies penetration and therefore that should be larger. In a sense, the apparently good behavior of the multipolar energies is artificial and is a consequence of a spurious fit that only minimizes the ESP potential without

For Peer Review

providing an adequate physical picture of the distribution of moments on centers. It is of course, exacerbated by the use of damping functions that add another correction the GDMA analysis. In other words, if the multipolar energy of a given configuration is, by chance, spuriously correct, the damping correction will only destroy the fictitious agreement. For large basis sets, the advantage of having a larger number of centers tend to generate lower quality distributed multipoles because of the difficult handling of diffuse function.[29] Damping indeed requires physically meaningful multipoles with a consistent representation of the different angular momenta. For ESP fitted multipoles, no charge redistribution is provided for the reoptimization of the dipole and quadrupole moments. It generates a better fit for some configurations but also damages anisotropy. Indeed, moments are related to atomic orbital symmetries. If one component is overfitted, the others are clearly less accurately or spuriously described. Overall, whatever, the strategy to parametrize the damping function, one has to start to analyze the multipolar energy behavior compared to SAPT: both sets of energies should be clearly different as multipoles do not embody penetration energy. Any attempt to damp overfitted energies will lead to a loss of accuracy when damping is added, especially if one studies various anisotropic interactions. If well-balanced multipoles are extracted, the damping is then increasingly effective as the order of correction tend towards charge-quadrupole (see appendix). If overfitted multipoles are chosen, one can still strongly reduce the error with SAPT but a charge-charge correction should be prefered as higher-order corrections will be perturbed by the spurious anisotropy of the charge distribution. It may lead to double countings of the penetration effect for configurations where multipoles are already overestimated. Overall, on the 10 dimer configurations, the best choice is based on GDMA multipoles up to quadrupole fitted on atoms only and corrected by a damping function going up to the quadrupole level.

Besides the application to a SIBFA-like protocol, we also applied this methodology to treat charge penetration within AMOEBA framework, which use only atom centered ESP multipoles on variety of systems. As we discussed, ESP multipoles do take not full advantages of higher

For Peer Review

order penetration correction as the charge-dipole damping appears already weaker than in the case of well behaved multipoles.

In all these cases, we have significantly improved the AMOEBA results without modifying the existing sets of multipoles (amoeba09).

For the previously discussed water dimers, the maximum energy variation between the first six configurations, corresponding to the lowest energies, the charge-charge correction reproduced well the behavior of SAPT energy variations between configurations (energy difference between the most and the less stable configuration : 2.5 for SAPT) whereas other corrections appear strongly less anisotropic (0.9 kcal/mol for Ref. 25). Even for this simplest level of correction, penetration effects are well recovered. Indeed, figure 4 represents the electrostatic energy between two water molecules as a function of their intermolecular distance in Angströms. The first two geometries (fig.4a and 4b) correspond to strong H-bonds while the third (fig. 4c) involves two H-bonds and the fourth (fig. 4d) weak interactions.

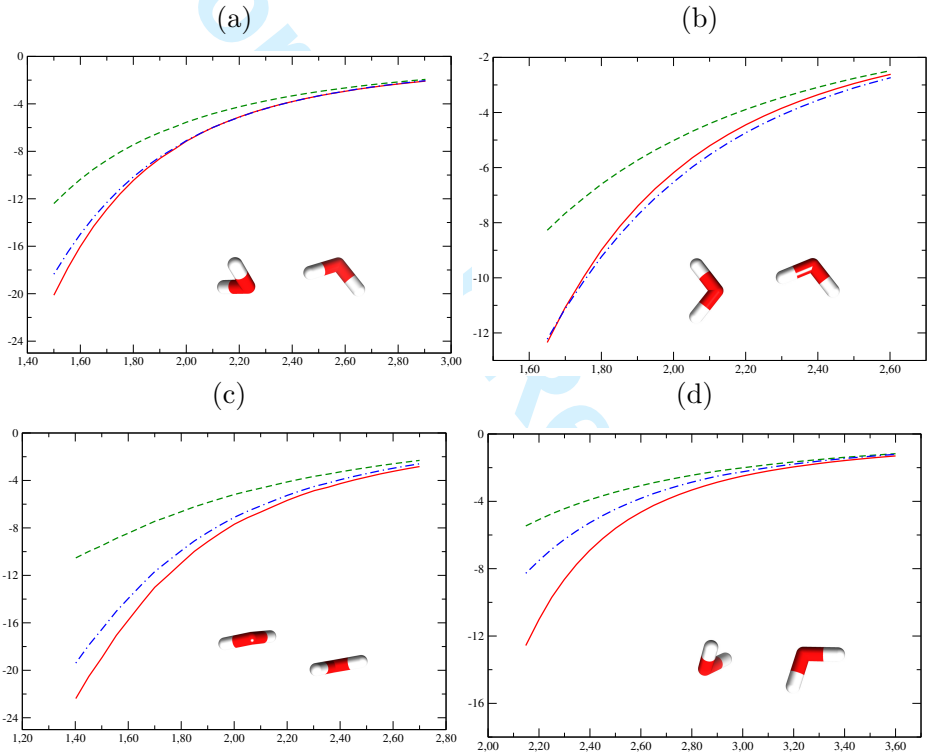
Table 3: RMSE for distance scans of water dimers (see text) in kcal/mol (reference SAPT)

water dimer	CP-AMOEBA	AMOEBA
1	0.51	2.81
3	0.25	1.66
4	1.22	4.86
7	1.39	1.88
S66	0.69	2,21

We also report the root mean square error for intermolecular distance scans (from 1.4 to 2.7 A) of selected water dimers (S66 is an additional water geometry extracted from S66 set) in Table 3 which confirms the improvement of the AMOEBA model in different configurations and the robustness of the correction at very short-range. The small improvement between the RMSE's regarding water dimer 7 (Figure 4d) could be explained by its being a high lying energy dimer obtained only at the CCSDT level[43] Indeed, for this kind of system, the charge penetration correction remains subtle because of the importance of electronic correlation and of the difficulty to model interaction involving the use of very diffuse basis function.

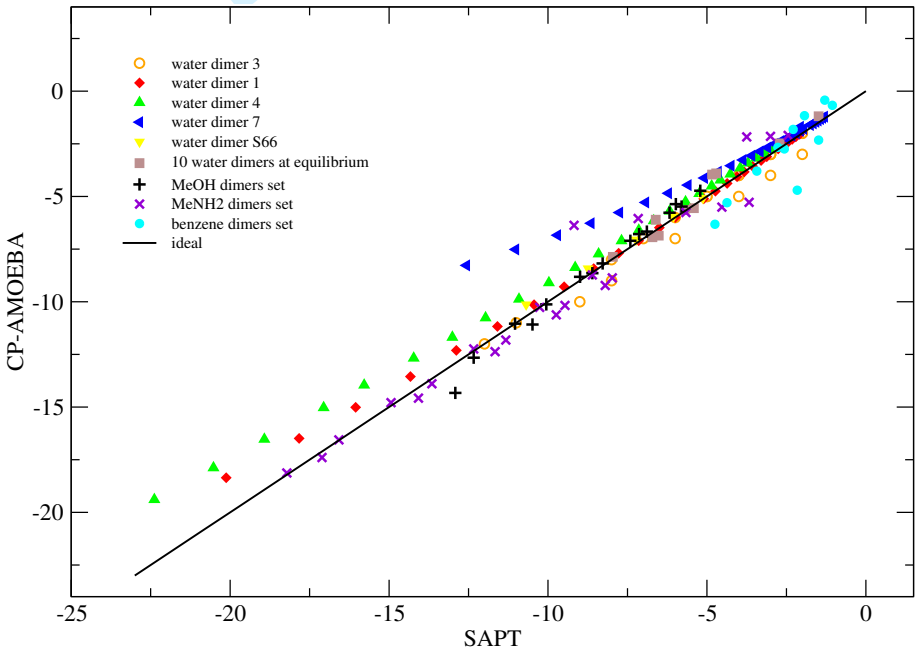
For Peer Review

Figure 4: Water dimers electrostatic profile as function of intermolecular distance in kcal/mol. SAPT is the red line. AMOEBA the green dashed line. cp-AMOEBA the blue dashed dot line.



We continue our investigation on other chemical functions: methanol (MeOH) and methyl amine (MeNH2) (see appendix Tables 5 and 6 respectively). Finally, we investigate again benzene (see appendix table 7, 8 and 9). In all these cases, we improved the electrostatics using AMOEBA multipoles and particularly for benzene dimers. The approximation only breaks at very -unphysical- short-range. We report all our results in Figure 4, a plot of the results obtained with the correction against SAPT reference values. This graph includes all the presented molecules and their distance scans from near equilibrium up to short-distances.

Figure 5: Correlation between SAPT and the penetration correction in kcal/mol within binding range for various dimers (see legend)



The electrostatic interactions energies of some of these complexes are for distances far below equilibrium and for these penetration effects can be very significant. As one can see from tables (appendix), the penetration correction always improved the results compared to SAPT, even for the cases where ESP multipolar energies were really off (see benzene results), with the wrong interaction sign in all cases.

To conclude, we extended our tests to charge system, studying complexes of metal cations

with water including K^+, Na^+, Mg^{2+},Ca^{2+} and Zn^{2+}.

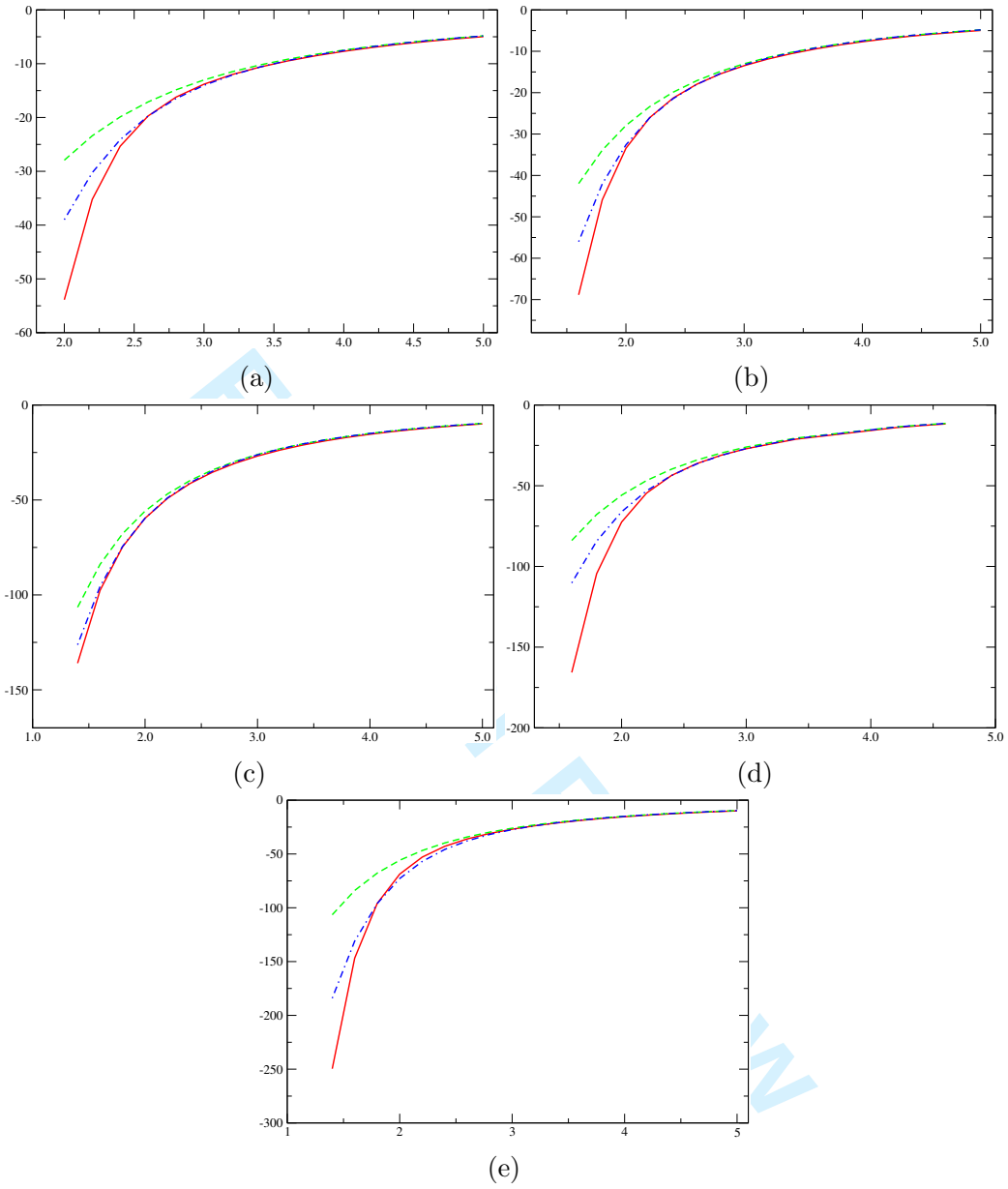


Figure 6: Cation-water complexes : electrostatic variation as function of cation-O distance in kcal/mol. SAPT is the red line, AMOEBA the green dashed line and the penetration correction blue dashed-dot line. a) K^+ water complex. b) Na^+ water complex. c) Mg^{2+} water complex. d) Ca^{2+} water complex. e) Zn^{2+} water complex.

As expected from the SIBFA previous work, penetration effects are a key element to model metals. The simple addition of the charge-charge correction allows a very close agreement with reference SAPT datas.

3.3.1 Parallel Implementation

The loops involved in the computation of the correction term can be easily parallelized with a shared memory (OpenMP), distributed memory (MPI) or hybrid (OpenMP/MPI) paradigm. The additional routines are included in a parallel version of the SPME that we developed in the upcoming TINKER-HP software[51–54] : the atom sites and the corresponding energy and forces are distributed among processes using a spatial decomposition.[55] Furthermore, as the cutoff used in the correction term (between 6 and 8 Å) is usually smaller than the real space cutoff (see Figure 2) used in the SPME, no additional cross-processor communication of the atomic positions between each time step is necessary. The parallel scaling of the computation of the correction using MPI is showed on graph 7 for a water box of 60 000 atoms in a cubic unit cell where each edge is 84 Angströms long. Our tests showed that a pure MPI implementation is faster than a hybrid one before reaching a scaling plateau, but once this plateau is reached, the hybrid implementation allows to make use of more cores and to improve the best timing. The best timings obtained to compute both the electrostatic forces using SPME, and the correction, using the pure MPI implementation are shown in table 4, as well as the best timings obtained with the hybrid implementation (with 2 OpenMP threads per MPI process).

For Peer Review

Figure 7: Parallel scaling for a water box of 60 000 atoms

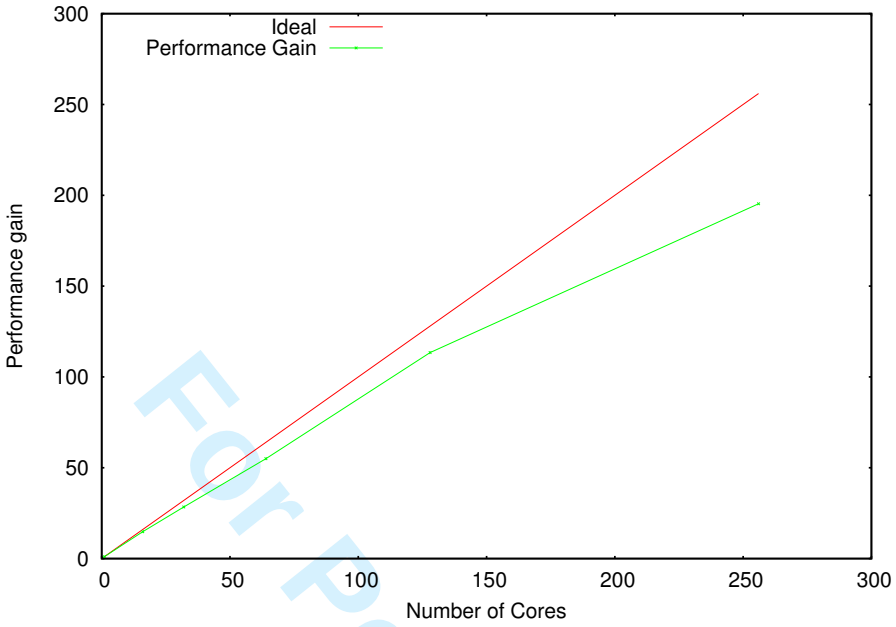


Table 4: Absolute best timings (in seconds) and number of cores used (in parentheses) for the computation of the electrostatic forces with SPME and of the correction in our implementation.

	Pure MPI	Hybrid OpenMP/MPI
SPME	0.07 (256)	0.04 (512)
CP-AMOEBA	0.025 (256)	0.015 (512)

With a cutoff of 8 Å, which is beyond convergence, the total penetration cost is only 25 per cent in both cases. Of course, as we discussed, such a cutoff can be chosen more aggressi vely. A cutoff of 5 angströms captured 99 per cent of penetration interaction and led to strong decrease of the computation cost which became lower than 10 per cent.

4 Conclusions

We have studied the general coupling of an empirical short-range charge penetration correction to the SPME. The corrected SPME approach allows a computationally efficient and accurate improvement of electrostatics in polarizable simulations by correcting the short-

For Peer Review

range charge-charge, charge-dipole and charge-quadrupole terms. The method has been implemented in the TINKER-HP software and the additional cost of the short-range damping function shown to be small as the global electrostatics implementation exhibits a favorable parallel scaling at the MPI and OPEN-MP/MPI hybrid levels. A separate cutoff for the penetration term coupled to a switching function has been implemented in order to maintain the possibility of a performance optimization of SPME thanks to the separate Ewald cutoff. Initial tests have been performed in the context of the AMOEBA and SIBFA force fields (ESP and GDMA approaches respectively). Substantial improvements were noted in the agreement of standard non-corrected multipolar interaction energies (from Stone multipoles[29]) compared to SAPT ab-initio reference data. As we discussed the effects of spurious fit of distributed multipoles, the need of comparing atom-centered GDMA to SAPT is highlighted as one wants to avoid overfitted multipolar energies. When distributed moments are physically meaningful, the correction exhibits a systematic improvement when adding charge-charge, then charge-dipole and charge-quadrupole corrections. For some sets of multipoles, like the ESP one, that exhibit some fictitious inclusion of penetration the correction should be limited to the charge-charge correction only. For these types of multipoles, specific parametrization procedure has been recently proposed in reference[56] based on ESP derivation of the damping parameters.

This opens two possible strategies as one can focus on the sole optimization of the charge-charge term to obtain a significant improvement of atom centered multipolar electrostatics at a very low computational cost.

Strategies in that direction are currently studied in the context of AMOEBA-like potentials.

A second strategy includes higher order damping corrections (charge dipole and beyond), as presented in the SIBFA potential. Such an implementation is therefore the starting point of SIBFA portage into TINKER-HP. Prospective applications in biochemistry and pharmacology should benefit from the inclusion of charge penetration in PBC molecular dynamics with improved representation of short-range contributions exchange-repulsion and charge transfer

For Peer Review

and long-range polarization and dispersion. To conclude, it is important to point out that future inclusion of such correction in polarizable molecular dynamics will require either to reparametrize the van der Waals contribution for the case of AMOEBA-like potentials or to use the penetration-corrected electrostatics in conjunction with an explicit exchange-repulsion contribution like in SIBFA.

Acknowledgements

We wish to thanks Marie-France Couret for technical support. This work was also supported in part by French state funds managed by CALSIMLAB and the ANR within the Investissements d'Avenir program under reference ANR-11-IDEX-0004-02. Qiantao Wang, David R. Bell and Pengyu Ren are grateful for support by the Robert A. Welch Foundation (F-1691) and the National Institutes of Health (GM106137), as well as the high performance computing resources provided by TACC and XSEDE (TG-MCB100057). Funding from French CNRS through PICS grant between UPMC and UT Austin is acknowldeged. Jean-Philip Piquemal and Étienne Polack acknowledge funding from the Délégation Générale de l'Armement (DGA), Maîtrise NRBC, from the French Ministry of Defense.

For Peer Review

References

(1) Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. *The Journal of Physical Chemistry A* **2003**, *107*, 10353–10359.

(2) Piquemal, J.-P.; Chevreau, H.; Gresh, N. *Journal of Chemical Theory and Computation* **2007**, *3*, 824–837.

(3) Allen, F. H. *Acta Crystallographica Section B: Structural Science* **2002**, *58*, 380–388.

(4) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer Jr, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *Archives of biochemistry and biophysics* **1978**, *185*, 584–591.

(5) UniProt Consortium, *Nucleic acids research* **2014**, *42*, D191–D198.

(6) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular cell biology*; WH Freeman, 2000.

(7) Pratt, C. W.; Voet, D.; Voet, J. G. *Fundamentals of biochemistry: life at the molecular level*; John Wiley & Sons, New York, NY, 2005.

(8) Slipchenko, L. V.; Gordon, M. S. *Molecular Physics* **2009**, *107*, 999–1016.

(9) Handley, C. M.; Popelier, P. L. *The Journal of Physical Chemistry A* **2010**, *114*, 3371–3383.

(10) Elking, D. M.; Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A.; Pedersen, L. G. *Journal of chemical theory and computation* **2009**, *6*, 190–202.

(11) Chaudret, R.; Gresh, N.; Narth, C.; Lagardère, L.; Darden, T. A.; Cisneros, G. A.; Piquemal, J.-P. *The Journal of Physical Chemistry A* **2014**,

For Peer Review

(12) Giese, T. J.; Chen, H.; Dissanayake, T.; Giambaşu, G. M.; Heldenbrand, H.; Huang, M.; Kuechler, E. R.; Lee, T.-S.; Panteva, M. T.; Radak, B. K. *Journal of chemical theory and computation* **2013**, *9*, 1417–1427.

(13) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *The Journal of chemical physics* **1995**, *103*, 8577–8593.

(14) Darden, T.; York, D.; Pedersen, L. *The Journal of chemical physics* **1993**, *98*, 10089–10092.

(15) Ren, P.; Wu, C.; Ponder, J. W. *Journal of chemical theory and computation* **2011**, *7*, 3143–3161.

(16) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *Journal of chemical theory and computation* **2007**, *3*, 1960–1986.

(17) Gresh, N.; Claverie, P.; Pullman, A. *Theoretica chimica acta* **1984**, *66*, 1–20.

(18) Piquemal, J.-P.; Williams-Hubbard, B.; Fey, N.; Deeth, R. J.; Gresh, N.; Giessner-Prettre, C. *Journal of computational chemistry* **2003**, *24*, 1963–1970.

(19) Devereux, M.; Gresh, N.; Piquemal, J.-P.; Meuwly, M. *Journal of computational chemistry* **2014**, *35*, 1577–1591.

(20) Cisneros, G.; Tholander, S.; Parisel, O.; Darden, T.; Elking, D.; Perera, L.; Piquemal, J.-P. *International journal of quantum chemistry* **2008**, *108*, 1905–1912.

(21) Werneck, A. S.; Filho, T. M. R.; Dardenne, L. E. *The Journal of Physical Chemistry A* **2008**, *112*, 268–280.

(22) Piquemal, J.-P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *The Journal of chemical physics* **2006**, *124*, 104101.

For Peer Review

(23) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *The Journal of Chemical Physics* **2000**, *112*, 7300–7306.

(24) Stone, A. J. *The Journal of Physical Chemistry A* **2011**, *115*, 7017–7027.

(25) Wang, B.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2014**,

(26) Momany, F. A. *The Journal of Physical Chemistry* **1978**, *82*, 592–601.

(27) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. *Chemical reviews* **2013**, *114*, 779–814.

(28) Wheatley, R. J.; Mitchell, J. B. *Journal of computational chemistry* **1994**, *15*, 1187–1198.

(29) Stone, A. *Chemical Physics Letters* **1981**, *83*, 233–239.

(30) Vigné-Maeder, F.; Claverie, P. *The Journal of chemical physics* **1988**, *88*, 4934–4948.

(31) Claverie, P. Contribution à l'étude des interactions intermoléculaires. Ph.D. thesis, UPMC, 1973.

(32) Hohenstein, E. G.; Duan, J.; Sherrill, C. D. *Journal of the American Chemical Society* **2011**, *133*, 13244–13247.

(33) Jeziorska, M.; Cencek, W.; Patkowski, K.; Jeziorski, B.; Szalewicz, K. *The Journal of chemical physics* **2007**, *127*, 124303–124303.

(34) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chemical Reviews* **1994**, *94*, 1887–1930.

(35) Ewald, P. P. *Annalen der Physik* **1921**, *369*, 253–287.

(36) Perram, J. W.; Petersen, H. G.; De Leeuw, S. W. *Molecular Physics* **1988**, *65*, 875–893.

(37) Sagui, C.; Pedersen, L. G.; Darden, T. A. *The Journal of chemical physics* **2004**, *120*, 73–87.

For Peer Review

(38) Smith, E. R. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **1981**, *375*, 475–505.

(39) Smith, W. *INFORMATION NEWSLETTER FOR COMPUTER SIMULATION OF CONDENSED PHASES* **1998**, 15–25.

(40) Norberg, J.; Nilsson, L. *Biophysical journal* **2000**, *79*, 1537–1553.

(41) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *Journal of computational chemistry* **1983**, *4*, 187–217.

(42) Loncharich, R. J.; Brooks, B. R. *Proteins: Structure, Function, and Bioinformatics* **1989**, *6*, 32–45.

(43) Reinhardt, P.; Piquemal, J.-P. *International Journal of Quantum Chemistry* **2009**, *109*, 3259–3267.

(44) van Duijneveldt-van de Rijdt, J.; Mooij, W.; van Duijneveldt, F. *Physical Chemistry Chemical Physics* **2003**, *5*, 1169–1180.

(45) Smith, B. J.; Swanton, D. J.; Pople, J. A.; Schaefer III, H. F.; Radom, L. *The Journal of chemical physics* **1990**, *92*, 1240–1247.

(46) Tschumper, G. S.; Leininger, M. L.; Hoffman, B. C.; Valeev, E. F.; Schaefer III, H. F.; Quack, M. *The Journal of chemical physics* **2002**, *116*, 690–701.

(47) Rezác, J.; Riley, K. E.; Hobza, P. *Journal of chemical theory and computation* **2011**, *7*, 2427–2438.

(48) Goerigk, L.; Kruse, H.; Grimme, S. *ChemPhysChem* **2011**, *12*, 3421–3433.

(49) Stevens, W. J.; Fink, W. H. *Chemical physics letters* **1987**, *139*, 15–22.

For Peer Review

(50) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 556–565.

(51) Lipparini, F.; Stamm, B.; Cancès, E.; Maday, Y.; Mennucci, B. *Journal of Chemical Theory and Computation* **2013**, *9*, 3637–3648.

(52) Lipparini, F.; Scalmani, G.; Lagardère, L.; Stamm, B.; Cancès, E.; Maday, Y.; Piquemal, J.-P.; Frisch, M. J.; Mennucci, B. *The Journal of chemical physics* **2014**, *141*, 184108.

(53) Lipparini, F.; Lagardère, L.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. *Journal of Chemical Theory and Computation* **2014**, *10*, 1638–1651.

(54) **submitted**,

(55) Lagardère, L.; Lipparini, F.; Polack, E.; Stamm, B.; Schnieders, M.; Cances, E.; Ren, P.; Maday, Y.; Piquemal, J.-P. *Journal of Chemical Theory and Computation* **2015**,

(56) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. *Journal of Chemical Theory and Computation* **2015**,

For Peer Review

A Appendix

A.1 Forces

$$\begin{aligned}\frac{\partial(E^*_{m_i-m_j}-E_{m_i-m_j})}{\partial r_i^\alpha} =& \frac{(r_j-r_i)^\alpha}{r_{ij}^\alpha}(Z_i(Z_j-q_j)exp(-\alpha_j r_{ij})(\alpha_j+\frac{1}{r_{ij}}) \\ &+ Z_j(Z_i-q_i)exp(-\alpha_i r_{ij})(\alpha_i+\frac{1}{r_{ij}}) \\ &+ (Z_i-q_i)(Z_j-q_j)(exp(-(\beta_i+\beta_j)r_{ij})(\beta_i+\beta_j+\frac{1}{r_{ij}}) \\ &- exp(-\beta_i r_{ij})(\beta_i+\frac{1}{r_{ij}}) - exp(-\beta_j r_{ij})(\beta_j+\frac{1}{r_{ij}}))\end{aligned} \tag{21}$$

$$\frac{\partial(E^*_{m_i-d_j}-E_{m_i-d_j})}{\partial r_i^\alpha} = \frac{(Z_i-q_i)exp(-\eta r_{ij})}{r_{ij}^3}(\mu_j^\alpha + (r_j-r_i)^\alpha \frac{\boldsymbol{\mu}_j.\boldsymbol{r}_{ji}}{r_{ij}}(\eta+\frac{3}{r_{ij}})) \tag{22}$$

A.2 Torques

Because the energy correction term contains permanent dipole moments that are usually defined in local frames binded to neighboring atoms, the derivatives of this term with respect to the positions of these atoms have to be taken into account. These are often improperly refered to as 'torques'. Suppose that the k-th atom is used to define the local frame in which the dipole moment $\boldsymbol{\mu}_j$ is written, then the corresponding derivative is the following :

$$\frac{d(E^*_{m_i-d_j}-E_{m_i-d_j})}{dr_k^\alpha} = \frac{\partial(E^*_{m_i-d_j}-E_{m_i-d_j})}{\partial r_k^\alpha} + < \frac{\partial(E^*_{m_i-d_j}-E_{m_i-d_j})}{\partial \boldsymbol{\mu}_j}, \frac{\partial \boldsymbol{\mu}_j}{\partial r_k^\alpha} > \tag{23}$$

$$\frac{d(E^*_{m_i-d_j}-E_{m_i-d_j})}{dr_k^\alpha} = \frac{\partial(E^*_{m_i-d_j}-E_{m_i-d_j})}{\partial r_k^\alpha} + < (Z_i-q_i)exp(-\eta r_{ij})\frac{\boldsymbol{r}_{ji}}{r_{ij}^3}, \frac{\partial \boldsymbol{\mu}_j}{\partial r_k^\alpha} > \tag{24}$$

where the derivatives of $\boldsymbol{\mu}_j$ involve the derivatives of the derivatives of the rotation matrix used to go from the local to the global frame

For Peer Review

A.3 Data

Tables 5 and 6 report electrostatics using two different distributed multipolar sites scheme : (i) atoms plus bonds (table 5). (ii) atoms only (table 6). Space Partitonning and hydrogen radii were also tested. sw$i$-rad$j$ denotes the space partitioning choice (switch i in GDMA) and j is the hydrogen radii.

Table 5: multipolar electrostatics using moments on bonds and atoms in kcal/mol

index	sw0-rad325	sw0-rad65	sw2-rad325	sw2-rad65	sw4-rad325	sw4-rad65
1	-6.18	-6.50	-6.91	-8.14	-6.96	-8.21
2	-5.19	-5.47	-5.22	-7.07	-5.23	-7.11
3	-5.05	-5.33	-4.88	-6.93	-4.88	-6.96
4	-5.13	-4.99	-4.72	-4.63	-4.69	-4.63
5	-4.51	-4.41	-3.55	-3.98	-3.48	-3.95
6	-4.32	-4.26	-3.19	-3.89	-3.11	-3.85
7	-3.59	-3.46	-3.51	-3.06	-3.50	-3.05
8	-1.04	-1.03	-1.27	-1.09	-1.28	-1.10
9	-3.32	-3.29	-3.87	-3.54	-3.89	-3.57
10	-1.96	-1.91	-2.24	-1.97	-2.25	-1.98

Table 6: multipolar electrostatics using moments on atoms only in kcal/mol

index	sw0-rad325	sw0-rad65	sw2-rad325	sw2-rad65	sw4-rad325	sw4-rad65
1	-6.02	-6.54	-6.73	-7.94	-6.75	-7.93
2	-5.12	-5.51	-5.15	-6.92	-5.16	-6.90
3	-5.01	-5.36	-4.85	-6.80	-4.86	-6.77
4	-5.31	-5.01	-4.82	-4.66	-4.81	-4.68
5	-4.81	-4.44	-3.76	-4.05	-3.74	-4.04
6	-4.67	-4.28	-3.44	-3.96	-3.42	-3.95
7	-3.49	-3.44	-3.58	-3.08	-3.58	-3.09
8	-0.93	-1.04	-1.26	-1.10	-1.26	-1.11
9	-3.06	-3.28	-3.79	-3.53	-3.80	-3.56
10	-1.87	-1.92	-2.21	-2.00	-2.21	-2.02

For Peer Review

Table 7: multipolar electrostatics and penetration correction decomposed

SAPT	GDMA[1]	E_{pen}	E_{pen} charge-charge	E_{pen} charge-dipole	E_{pen} charge-quadrupole
-8.18	-6.73	-1.44	-1.22	-0.11	0.12
-6.64	-5.15	-1.48	-1.15	-0.14	0.20
-6.36	-4.85	-1.51	-1.14	-0.15	0.23
-6.58	-4.82	-1.76	-1.57	-0.03	0.15
-5.78	-3.76	-2.02	-1.65	-0.08	0.29
-5.58	-3.44	-2.14	-1.70	-0.10	0.34
-4.71	-3.58	-1.13	-1.13	0.03	0.03
-1.43	-1.26	-0.18	-0.25	0.04	-0.04
-4.84	-3.79	-1.04	-1.00	-0.03	0.02
-2.82	-2.21	-0.61	-0.58	-0.02	0.01

We have retained the S66[47] dimer notations. Tables reports electrostatic energy in kcal/mol.

Table 8: CP-amoeba parameters

fragments	atoms	α	β	η
water	H	3.5	2.9	1.7
	O	4.0	4.9	4.7
methanol	H(-H3C)	3.5	2.6	4.7
	C(-CH3)	5.8	4.9	2.5
	O	6.4	4.5	0.7
	H (-O)	4.8	2.9	1.7
methylamine	H(-H3C)	3.5	2.7	4.7
	C(-CH3)	4.8	5.4	2.5
	N	4.8	4.8	2.7
	H (-N)	2.5	2.9	1.7
benzene	C	2.8	2.5	2.8
	H	3.5	2.3	2.5

For Peer Review

Table 9: Distance scans of S66 water orientations

dimers	SAPT	AMOEBA	CP-AMOEBA
2699_01WaterWater090	-10.687	-7.469	-10.136
2700_01WaterWater095	-8.739	-6.407	-8.423
2701_01WaterWater100	-7.231	-5.542	-7.058
2702_01WaterWater105	-6.052	-4.826	-5.961
2703_01WaterWater110	-5.124	-4.230	-5.078

Table 10: Distance scans of various methanol dimers orientations

dimers	SAPT	AMOEBA	CP-AMOEBA
2707_02WaterMeOH090	-12.337	-7.530	-12.656
2708_02WaterMeOH095	-10.053	-6.282	-10.124
2709_02WaterMeOH100	-8.280	-5.274	-8.180
2710_02WaterMeOH105	-6.891	-4.447	-6.667
2711_02WaterMeOH110	-5.799	-3.764	-5.480
2731_05MeOHMeOH090	-12.915	-6.222	-14.330
2732_05MeOHMeOH095	-10.488	-5.047	-11.082
2733_05MeOHMeOH100	-8.605	-4.096	-8.644
2734_05MeOHMeOH105	-7.133	-3.314	-6.788
2735_05MeOHMeOH110	-5.979	-2.667	-5.364
2755_08MeOHWater090	-11.037	-6.1497	-11.042
2756_08MeOHWater095	-8.991	-5.163	-8.815
2757_08MeOHWater100	-7.410	-4.356	-7.102
2758_08MeOHWater105	-6.177	-3.686	-5.770
2759_08MeOHWater110	-5.211	-3.128	-4.728

For Peer Review

Table 11: Distance scans of various methylamine dimers orientations

dimers	SAPT	AMOEBA	CP-AMOEBA
2715_03WaterMeNH2090	-16.578	-14.156	-16.557
2716_03WaterMeNH2095	-13.653	-12.178	-13.894
2717_03WaterMeNH2100	-11.328	-10.595	-1.817
2718_03WaterMeNH2105	-9.468	-9.311	-10.175
2719_03WaterMeNH2110	-7.978	-8.261	-8.867
2739_06MeOHMeNH2090	-18.216	-12.793	-18.129
2740_06MeOHMeNH2095	-14.938	-10.881	-14.790
2741_06MeOHMeNH2100	-12.337	-9.354	-12.240
2742_06MeOHMeNH2105	-10.262	-8.117	-10.267
2743_06MeOHMeNH2110	-8.605	-7.107	-8.727
2763_09MeNH2MeOH090	-6.139	-3.797	-2.151
2764_09MeNH2MeOH095	-4.758	-3.364	-2.177
2765_09MeNH2MeOH100	-3.748	-3.011	-2.172
2766_09MeNH2MeOH105	-3.002	-2.719	-2.145
2767_09MeNH2MeOH110	-2.444	-2.475	-2.101
2771_10MeNH2MeNH2090	-9.181	-7.864	-6.372
2772_10MeNH2MeNH2095	-7.163	-7.199	-6.048
2773_10MeNH2MeNH2100	-5.661	-6.655	-5.761
2774_10MeNH2MeNH2105	-4.530	-6.207	-5.505
2775_10MeNH2MeNH2110	-3.675	-5.835	-5.278
2787_12MeNH2Water090	-17.110	-13.859	-17.396
2788_12MeNH2Water095	-14.075	-11.953	-14.576
2789_12MeNH2Water100	-11.665	-10.426	-12.369
2790_12MeNH2Water105	-9.740	-9.185	-10.621
2791_12MeNH2Water110	-8.201	-8.169	-9.226

Table 12: Distances scans of Benzene water values

dimers	SAPT	AMOEBA	CP-AMOEBA
3123_54BenzeneWaterOHpi090	-4.370	5.499	-5.296
3124_54BenzeneWaterOHpi095	-3.432	5.486	-3.795
3125_54BenzeneWaterOHpi100	-2.769	5.536	-2.668
3126_54BenzeneWaterOHpi105	-2.288	5.620	-1.815
3127_54BenzeneWaterOHpi110	-1.932	5.721	-1.166

For Peer Review

Table 13: Distances scans of Benzene π stacking

dimers	SAPT	AMOEBA	CP-AMOEBA
2883_24BenzeneBenzenepipi090	-8.480	15.690	-11.851
2884_24BenzeneBenzenepipi095	-4.749	15.556	-6.321
2885_24BenzeneBenzenepipi100	-2.564	15.430	-2.743
2886_24BenzeneBenzenepipi105	-1.292	15.314	-0.428
2887_24BenzeneBenzenepipi110	-0.560	15.208	1.639

Table 14: Benzene T-shaped

dimers	SAPT	AMOEBA	CP-AMOEBA
3067_47BenzeneBenzeneTS090	-4.928	13.934	-13.167
3068_47BenzeneBenzeneTS095	-3.224	13.936	-8.150
3069_47BenzeneBenzeneTS100	-2.160	13.956	-4.706
3070_47BenzeneBenzeneTS105	-1.486	13.985	-2.324
3071_47BenzeneBenzeneTS110	-1.055	14.018	-0.673

For Peer Review

1.5 Conclusions & Perspectives

L'utilisation des multipoles est donc cruciale pour une meilleure évaluation de l'énergie d'interaction d'origine électrostatique, majoritaire dans les complexes supramoléculaires. Les benzènes, très présents dans les systèmes biologiques, ont des moments dipolaire et quadrupolaire relativement important par rapport aux autres systèmes présentés et sont moins bien décrits avec AMOEBA qu'avec SIBFA. L'équilibre avec d'autres interactions, telles que les VAN DER WAALS, lors de la calibration peut aussi l'expliquer. La question du niveau de calcul quantique des multipoles pourrait être discutée, ainsi que le schéma ou nombre de sites puisque SIBFA a la possibilité de faire intervenir des multipoles de liaisons. JOUBERT et POPELIER ont conclu dans leur travaux qu'un plus grand nombre de site est un gage d'une meilleure description de la densité électronique [136], contrairement au conclusion de ces travaux de thèse. Ceci s'explique par les molécules mises en jeu dans le cadre de ces deux études (complexes de VAN DER WAALS et base d'ADN VS eau et benzene). Nous avons donc décidé de rester dans le schéma atomique mais de présenter de nouveaux multipoles calculés au niveau MP2/aug-cc-pvTZ.

D'un point de vue pratique, un calcul de ce niveau est encore très lourd au vu du nombre d'atomes de certains fragments moléculaires. En effet, les programmes de champs de forces reprennent la nomenclature de la Protein Data Bank où une protéine est fragmentée en résidus, des acides aminés par exemple. La même situation est rencontrée pour toute nouvelle introduction, comme par exemple la calibration d'un potentiel ligand. Celle-ci a été réalisée dans une étude en cours sur le VIH où le ligand n'était pas paramétré dans les champs de forces. Cette étape reste donc limitante en terme de temps de calcul mais assure une meilleure description de sa densité électronique. En effet, seul ce niveau de calcul le permet. Un calcul DFT est

exclu puisque la délocalisation électronique est mal traitée. L'autre alternative serait un calcul Hartree Fock, qui ne prend donc pas en compte la corrélation électronique.
D'autant plus que la qualité de la correction à courte portée s'en retrouve aussi affectée au vue des résultats, même si, il est vrai, l'électrostatique est nettement améliorée. De plus, son impact sur de grands systèmes en conditions périodiques permet d'avoir une meilleure estimation de l'énergie en se rapprochant d'une hypothétique valeur QM au vu des résultats sur les dimères. Cette amélioration est donc présente dans le code de simulation numérique TINKER.
Pour aller plus loin et s'affranchir des limites des multipoles, une alternative est envisageable : l'emploi de gaussienne évitant ainsi cet effet à courte portée dont STONE s'est d'ailleurs inspiré dans sa correction de la pénétration. Cette méthode est présente dans GEM (*Gaussian Electrostatic Model*). Son implémentation dans tinker-hp pourrait être une grande avancée dans le domaine des dynamiques moléculaires, proposant ainsi un schéma GEM/SIBFA ou des applications QM/GEM.

De plus, le SPME est donc applicable à tout scalaire du type $1/r^n$ et pourrait donc voir son extension dans d'autres termes du potentiel SIBFA.

Chapitre 2

Interactions de van der Waals et d'Induction : Energies et Forces

Sommaire

2.1 Interactions de van der Waals 136
2.1.1 Introduction 136
2.1.2 Répulsion 138
2.1.3 Dispersion 149
2.2 Traitement de la mobilité électronique : Induction . . . 154
2.2.1 Introduction 154
2.2.2 Polarisation 154
2.2.3 Transfert de charge 160
2.3 Résultats numériques & Discussion 166
2.4 Conclusions & Perspectives 171

2.1 Interactions de VAN DER WAALS

Dans cette première partie du chapitre, nous rappellerons brièvement ce que sont les interactions de type VAN DER WAALS, ainsi que leur traitement dans le cadre des champs de forces classiques. Enfin, dans une deuxième et troisième section, nous présenterons les potentiels SIBFA implémentés pour décrire ce type d'interaction dans le programme de simulation moléculaire numérique TINKER.

2.1.1 Introduction

Bien qu'historiquement, les approximations et théories menant à la simplification de ces interactions sont plus complexes et tiennent compte de l'induction. Cette composante ne sera pas traitée dans cette partie. Les forces dites de VAN DER WAALS sont décrites par des potentiels de type LENNARD-JONES. Celui-ci peut servir de « tampon » de calibration pour représenter la différence entre l'énergie totale et l'électrostatique par exemple mais d'autres stratégies existent [137]. Physiquement, cette énergie perd alors de son sens puisque l'idée est d'avoir une énergie totale en corrélation avec des calculs quantiques.

Ouvrons une parenthèse. Au niveau macroscopique, elles ont été, au début des années 2000, mises en lumière par une étude sur le gecko [138]. En effet, des capillaires, de l'ordre du micromètre, sur ses pattes ont été mis en évidence comme responsable de son aptitude à se fixer sur des surfaces hydrophobes. Les forces de VAN DER WAALS seraient à l'origine de sa capacité à se déplacer sur des surfaces de telle orientation que le commun des mortels ne peut réaliser en raison de la gravité.
Enfin, dans un contexte biologique, il a été montré que les interactions de VAN DER WAALS issues de modèles théoriques ne reproduisent pas les ob-

servations expérimentales [139, 140]. En effet, en chimie quantique et plus particulièrement en DFT, des corrections existent [141–143]. Celles ci sont du même type approximant ces interactions en une fonction, plus étendue qu'un potentiel de LENNARD-JONES, s'exprimant en une série de termes en $1/r^n$, avec des coefficients au numérateur.

Ces potentiels classiques miment ainsi un traitement répulsif à courte distance : la répulsion et attractif à longue distance : la dispersion. De manière générale, par abus de langage, on parle de force de VAN DER WAALS pour expliquer une interaction non-liée faible. Dans ces traitements classiques, seuls les atomes sont considérés. De plus, les paramètres atomiques sont moins extensifs comme le sont ceux sur l'énergie électrostatique, qui dérivent de chaque fragment moléculaire défini (dans le cadre d'une approche basée sur les multipoles).

L'une des particularités du potentiel SIBFA est la prise en compte des paires libres. En effet, les systèmes étudiés en biochimie sont majoritairement constitués d'hydrogène, de carbone, d'oxygène et d'azote. Ces deux derniers atomes possèdent des paires libres. La figure 2.1 représente les bassins de la fonction de localisation électronique (ELF : *electron localization function*) sur un bain de molécules d'eau [144, 145]. Brièvement, cette fonction permet de localiser les zones de fortes densités électroniques, comme les liaisons covalentes ou encore les paires libres.

Intéressons nous à ces volumes (colorés en bleu sur la figure 2.1) autour des atomes d'oxygène (en rouge) : ce sont ses deux paires libres. D'un point de vue physique, on peut aisément observer que les paires libres peuvent entrainer des clashs stériques si on ne les considère pas et mener à des géométries optimisées peu rigoureuses. De plus, étant chacune porteuse de deux électrons, elles peuvent aussi participer aux interactions à plus longue

distance.

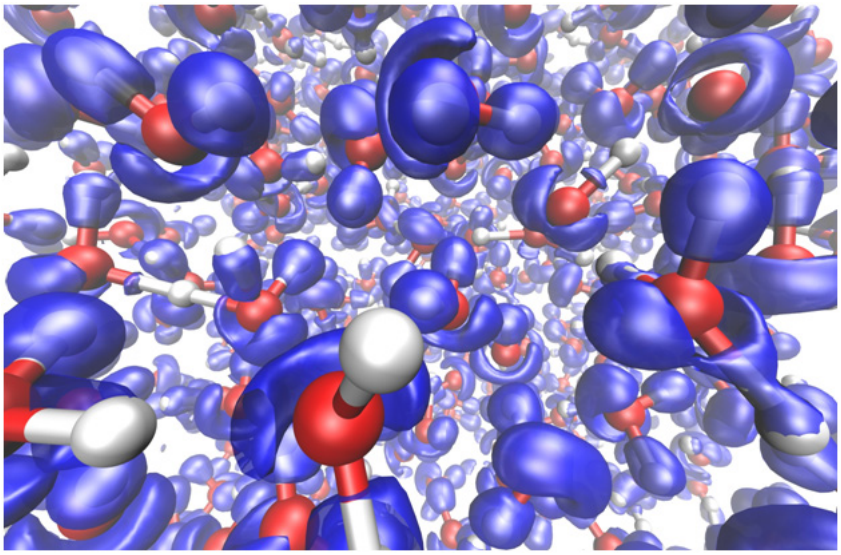


FIGURE 2.1 – représentation des bassins ELF (en bleu) d'un bain de molécules d'eau [146].

En pratique, la prise en compte des paires libres se fait de la même manière que les multipoles. En effet, celles-ci sont stockées dans un fichier les décrivant dans un repère local centré sur l'atome porteur. Ensuite, par une méthode de rotation dite « Z-then-X », ces coordonnées sont transformées dans le repère global cartésien. Leur position dans ce repère local est déterminée par la théorie de GILLESPIE [147] (ou VSEPR : *valence-shell electron-pair repulsion*) ou encore via des calculs ELF.

2.1.2 Répulsion

Cette composante de l'énergie intermoléculaire décrit le recouvrement électronique à courte distance, d'où son caractère répulsif. Dans le chapitre précédent, nous avons justement décrit une correction à courte portée. L'un

des piliers du potentiel SIBFA et la méthodologie sur lequel il s'appuie tend à rationaliser les différentes composantes des forces intermoléculaires en les décomposant de manière non-additive. On peut voir une similarité entre ces traitements mais quelle est donc la différence ?

La correction électrostatique à courte distance ou pénétration de charge est due au caractère classique de la description de la charge : approximation multipolaire. Alors que ce terme de répulsion décrit le recouvrement électronique ou plus précisément le recouvrement des orbitales électroniques selon la théorie des orbitales atomiques - moléculaires menant ainsi à des clashs stériques. En champ de forces classique, elle est traitée le plus souvent par un seul terme en r^{-12} ou r^{-14}. Notre méthode se veut plus raffinée en reproduisant la composante d'échange dans la théorie SAPT. En pratique, ce recouvrement est évalué entre les régions riches électroniquement comme les paires libres ou liaisons aromatiques. Pour cela, le terme d'échange-répulsion se décompose en trois doubles sommations distinctes sur, liaison - liaison, liaison - paire libre et paire libre - paire libre, représentées respectivement sur les figures 2.3, 2.4 et 2.7. A la différence de la figure 2.1, le codes couleurs des bassins ELF distinguent les paires libres (en bleu) des liaisons covalentes (en vert). Reprenons les termes de l'équation 2.5 du chapitre 2 de la partie 1. Les notations introduites par la figure 2.2 seront reprises dans les sous-sections suivantes.

Répulsion liaison - liaison

Dans un premier temps, définissons ce terme de recouvrement orbitalaire dans le cadre de SIBFA (voir équation 2.1). Entre deux atomes A et C appartenant à des fragments moléculaires non-liés (voir figure 2.2), le recouvrement est évalué en considérant les orbitales des atomes. Dans l'équation 2.1, X_i décrit l'orbitale i de l'atome X.

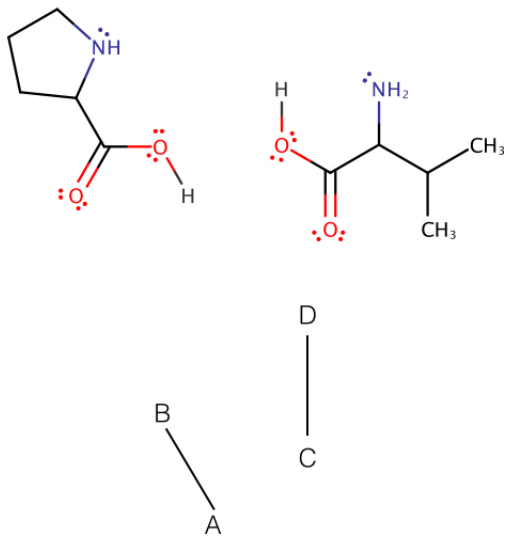


FIGURE 2.2 – schéma répulsion proline (gauche) valine (droite). AB et CD représentent les liaisons OH de la proline et la valine respectivement, en gardant leur orientation.

$$\begin{aligned} \text{rec(A,C)} = \text{rec}(A_S, C_S) + \text{rec}(A_P, C_S) \\ + \text{rec}(A_S, C_P) + \text{rec}(A_P, C_P) \end{aligned} \tag{2.1}$$

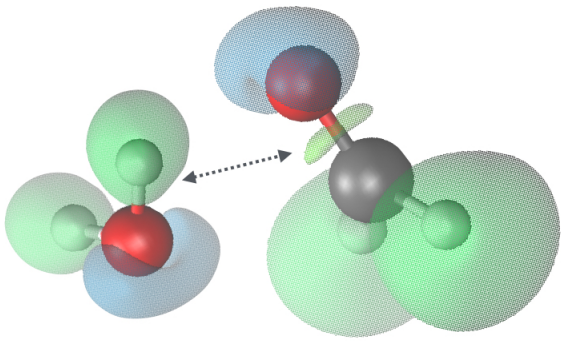


FIGURE 2.3 – illustration répulsion liaison - liaison : cas de l'eau (gauche) et du formaldéhyde (droite)

Les recouvrements orbitalaires se déterminent à partir du recouvrement

des orbitales s :

$$\begin{aligned} &\mathrm{rec}(A_P, C_S) = \\ &c_p(A)c_s(C)f^P_{orb}(A)f^S_{orb}(C)\cos(\mathbf{AB},\mathbf{AC})\mathrm{rec}(A_S, C_S) \end{aligned} \tag{2.2}$$

L'équation 2.2 correspond à l'évaluation du recouvrement entre l'orbitale p de l'atome A et l'orbitale s de l'atome C. L'angle ainsi formé par **AB** et **AC** permet d'introduire l'anisotropie dans le calcul de l'énergie de répulsion. $c_s(A)$ et $c_p(C)$, ainsi que $\mathrm{f}^P_{orb}(A)$ et $f^S_{orb}(C)$ correspondent aux coefficients d'hybridations et facteurs orbitalaires respectivement [148]. Enfin, les équations 2.3 et 2.4 décrivent les autres recouvrements possibles pour être complet.

$$\begin{aligned} &\mathrm{rec}(A_S, C_P) = \\ &c_s(A)c_p(C)f^S_{orb}(A)f^P_{orb}(C)\cos(\mathbf{CA},\mathbf{CD})\mathrm{rec}(A_S, C_S) \end{aligned} \tag{2.3}$$

$$\begin{aligned} &\mathrm{rec}(A_P, C_P) = \\ &c_p(A)c_p(C)f^P_{orb}(A)f^P_{orb}(C)\cos(\mathbf{AB},\mathbf{AC})\cos(\mathbf{CA},\mathbf{CD})\mathrm{rec}(A_S, C_S) \end{aligned} \tag{2.4}$$

Reste donc à exprimer le terme de recouvrement entre orbitales s.

$$\mathrm{rec}(A_S, C_S) = M_{AC}\exp(-\alpha\rho_{AC}) \tag{2.5}$$

$$\rho_{AC} = \frac{r_{AC}}{4\sqrt{R_{rep}(A)R_{rep}(C)}} \tag{2.6}$$

où $R_{rep}(A)$ et $R_{rep}(C)$ sont les paramètres intrinsèques aux atomes de ce terme énergétique. r_{AC} est la distance entre les atomes A et C.

$$M_{AC} = \sqrt{K_{AC}\left(1 - \frac{q_A}{val(A)}\right)\left(1 - \frac{q_C}{val(C)}\right)} \tag{2.7}$$

où q_A (q_C) et $val(A)$ ($val(C)$ sont la charge et la valence électronique de l'atome A (C). Enfin K_{AC} est le facteur de proportionnalité des orbitales atomiques.

Le recouvrement entre orbitales s est défini, le terme de recouvrement $rec(A, C)$ l'est donc aussi. La même procédure est à suivre pour $rec(A, D)$, $rec(B, C)$ et $rec(B, D)$. Il y a donc quatre interactions prises en compte dans ce terme décrivant la répulsion entre liaisons et est représenté sur la figure 2.3. Leur somme formant ainsi $rec(AB, CD)$.

$$\mathrm{rec}(AB, CD) = \mathrm{rec}(A, C) + \mathrm{rec}(A, D) + \mathrm{rec}(B, C) + \mathrm{rec}(B, D) \tag{2.8}$$

Finalement, le terme de répulsion entre liaisons de fragments n'appartenant pas à la même molécule s'exprime :

$$\begin{aligned}\mathrm{rep}(AB, CD) =& N_{occ}(AB)N_{occ}(CD)\mathrm{rec}(AB, CD)^2/d_{AB-CD} \\ &+ N_{occ}(AB)N_{occ}(CD)\mathrm{rec}(AB, CD)^2/d^2_{AB-CD}\end{aligned} \tag{2.9}$$

où $N_{occ}(AB)$ ($N_{occ}(CD)$) est le nombre d'occupation de la liaison AB (CD), c'est-à-dire le nombre d'électrons de la liaison considérée. d_{AB-CD} est la distance entre les barycentres des liaisons AB et CD.

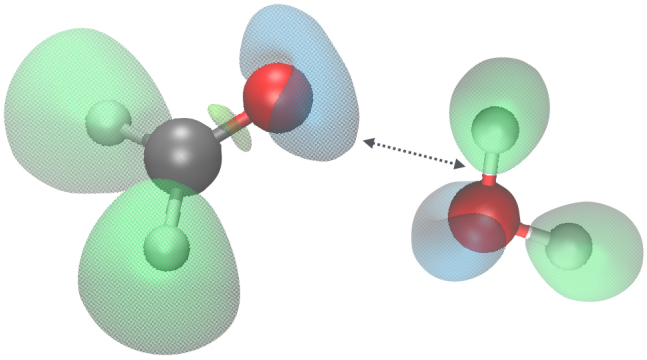
Répulsion paire libre - liaison

Pour ce deuxième terme, la méthodologie reste similaire. Sur la figure 2.2, B et C sont porteurs de paires libres, que l'on notera i et j. Le terme extrait de l'équation 2.5 du chapitre 2 de la partie 1, s'exprime dans ce cas :

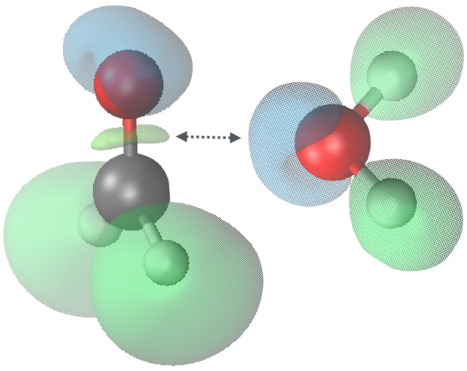
$$\begin{aligned}\mathrm{rep}(\text{liaison - paire libre}) = &\ \mathrm{rep}(AB, j_1) + \mathrm{rep}(AB, j_2) \\ &+ \mathrm{rep}(CD, i_1) + \mathrm{rep}(CD, i_2)\end{aligned} \tag{2.10}$$

où j_1 et j_2 (i_1 et i_2) désignent chacun une paire libre de l'atome d'oxygène C (B). Enfin, le porteur de la paire libre est considéré dans les équations ci-dessus, de 2.2 jusqu'à 2.9. Seuls le dénominateur, le nombre d'occupation et les angles sont décrits directement par la paire libre dans l'équation 2.9. Dans notre cas (voir figure 2.2), l'interaction entre la liaison AB et la paire libre j_1 portée par l'atome C, le terme d_{AB-j_1} correspond à la distance

entre la paire libre et le barycentre de la liaison. La prise en compte de l'anisotropie via les cosinus dans les équations 2.2, 2.3 et 2.4, dans le cadre des interactions de répulsion liaisons - paire libre est détaillée en figure 2.5, en simplifiant : l'atome C ne porte qu'une paire libre.



(a) interaction liaison de l'eau - paire libre du formaldéhyde



(b) interaction liaison du formaldéhyde - paire libre de l'eau

FIGURE 2.4 – illustration répulsion liaison - paire libre : cas formaldéhyde (gauche) et de l'eau (droite)

(a) recouvrement de l'atome A sur la paire libre j porté par C.

(b) recouvrement de la paire libre j porté par C sur l'atome A.

FIGURE 2.5 – schéma interaction rep(AB, j).

Finalement, dans le calcul du type rep(liaison - paire libre) représenté sur la figure 2.4, deux interactions sont estimées.

Répulsion paire libre - paire libre

Enfin, ce troisième terme estime les interactions de répulsion entre paires libres portées par des atomes appartenant à des molécules différentes et est représenté sur la figure 2.7. Dans la section précédente, les termes décrits par le porteur ou par la paire libre ont déjà été discutés. Finalement, la répulsion entre deux paires libres a et b, rep(a,b) ne représente qu'une interaction à calculer. En conclusion, l'expression du calcul de l'énergie de répulsion de SIBFA comprend les interactions présentées dans l'équation 2.11 et la figure 2.6 dans le contexte de l'exemple de la figure 2.2.

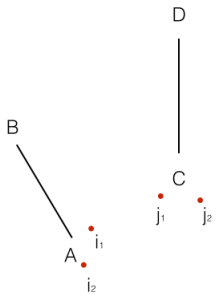


FIGURE 2.6 – schéma interaction de répulsion complète

$$\begin{aligned} E_{\text{rep}} = & E_{rep}(A,C) + E_{rep}(A,D) + E_{rep}(B,C) + E_{rep}(B,D) \\ & + E_{rep}(A,j_1) + E_{rep}(A,j_2) + E_{rep}(B,j_1) + E_{rep}(B,j_2) \\ & + E_{rep}(C,i_1) + E_{rep}(C,i_2) + E_{rep}(D,i_1) + E_{rep}(D,i_2) \\ & + E_{rep}(i_1,j_1) + E_{rep}(i_1,j_2) + E_{rep}(i_2,j_1) + E_{rep}(i_2,j_2) \end{aligned} \tag{2.11}$$

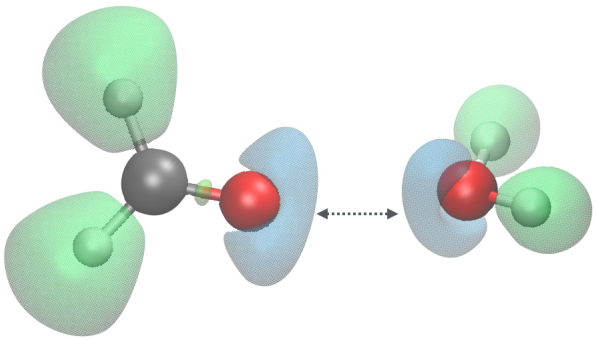


FIGURE 2.7 – illustration répulsion paire libre - paire libre : cas de l'eau (gauche) et du formaldéhyde (droite)

Forces

Dans le terme de l'énergie de répulsion présentée dans les équations 2.1 à 2.11, rappelons l'expression de la répulsion dans le cadre de deux sites (liaison ou paire libre) i et j :

$$\begin{aligned} \text{rep}(i,j) = & N_{occ}(i)N_{occ}(j)\text{rec}(i,j)^2/d_{i-j} \\ & + N_{occ}(i)N_{occ}(j)\text{rec}(i,j)^2/d_{i-j}^2 \end{aligned} \tag{2.12}$$

Finalement, la dépendance spatiale des gradients se concentre sur deux termes de cette équation : i) au dénominateur : d_{i-j}, la distance entre les deux sites considérés. ii) au numérateur : dans le terme de recouvrement

$\mathrm{rec}(i,j)$.

$$\begin{aligned}\frac{\partial \mathrm{rep}(i,j)}{\partial r_k^\alpha} =& N_{occ}(i)N_{occ}(j)2\frac{\partial \mathrm{rec}(i,j)}{\partial r_k^\alpha}\mathrm{rec}(i,j)/d_{i-j} \\ &+ N_{occ}(i)N_{occ}(j)\mathrm{rec}(i,j)^2(-1)\frac{\partial d_{i-j}}{\partial r_k^\alpha}/d_{i-j}^2 \\ &+ N_{occ}(i)N_{occ}(j)2\frac{\partial \mathrm{rec}(i,j)}{\partial r_k^\alpha}\mathrm{rec}(i,j)/d_{i-j}^2 \\ &+ N_{occ}(i)N_{occ}(j)\mathrm{rec}(i,j)^2(-2)\frac{\partial d_{i-j}}{\partial r_k^\alpha}/d_{i-j}^3\end{aligned} \tag{2.13}$$

avec k, le k-ième atome et α = x, y, z.

Recouvrement entre orbitales s

Considérons le recouvrement entre orbitale s des sites i et j

$$\mathrm{rec}(i_s, j_s) = M_{ij}\exp(-\alpha\rho_{ij}) \tag{2.14}$$

avec M_{ij} ne dépendant pas des coordonnées. Rappelons que ρ_{ij} est défini tel que :

$$\rho_{ij} = \frac{r_{ij}}{4\sqrt{R_{rep}(i)R_{rep}(j)}} \tag{2.15}$$

Reformulons ainsi $\mathrm{rec}(i_s, j_s)$:

$$\mathrm{rec}(i_s, j_s) = M_{ij}\exp(\beta r_{ij}) \tag{2.16}$$

avec $\beta = \frac{-\alpha}{4\sqrt{R_{rep}(i)R_{rep}(j)}}$

$$\frac{\partial \mathrm{rec}(i_s, j_s)}{\partial r_k^\alpha} = M_{ij}\beta\frac{\partial (r_{ij})}{\partial r_k^\alpha}\exp(\beta r_{ij}) \tag{2.17}$$

Recouvrement entre orbitales s et p

Finalement dans le cas de l'évaluation d'un recouvrement entre orbitale s de deux sites, les forces de la répulsion s'expriment telles que :

$$\frac{\partial \text{rep}(i,j)}{\partial r_k^\alpha} = \frac{N_{occ}(i)N_{occ}(j)\text{rec}(i,j)}{d_{i-j}} \left(2\frac{\partial \text{rec}(i_s,j_s)}{\partial r_k^\alpha} - \frac{\partial (d_{i-j})}{\partial r_k^\alpha}/d_{i-j} + 2\frac{\partial \text{rec}(i_s,j_s)}{\partial r_k^\alpha}/d_{i-j} - 2\frac{\partial (d_{i-j})}{\partial r_k^\alpha}/d_{i-j}^2 \right) \tag{2.18}$$

Le recouvrement entre orbitale s ainsi défini, le cas de l'évaluation d'un recouvrement s et p n'est modifié que par l'introduction de l'anisotropie via les cosinus. Simplifions la formule en regroupant les constantes en G. La figure 2.8 reprend ainsi l'exemple du recouvrement entre une orbitale p de l'atome A appartenant à la liaison AB et l'orbitale s de l'atome C appartenant à la liaison CD.

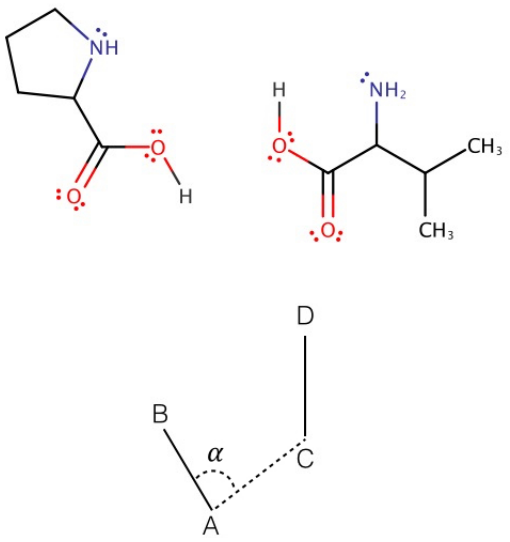


FIGURE 2.8 – schéma répulsion proline (gauche) valine (droite). AB et CD représentent les liaisons OH de la proline et la valine respectivement, en gardant leur orientation.

$$\mathrm{rec}(A_P, C_S) = G\cos(\mathbf{AB}, \mathbf{AC})\mathrm{rec}(A_S, C_S) \tag{2.19}$$

$$\begin{aligned}\cos(\mathbf{AB}, \mathbf{AC}) &= \frac{\mathbf{AB}\cdot\mathbf{AC}}{|AB||AC|} \\ &= \frac{(r^x_{AB}r^x_{AC} + r^y_{AB}r^y_{AC} + r^z_{AB}r^z_{AC})}{\sqrt{({r^x_{AB}}^2 + {r^y_{AB}}^2 + {r^z_{AB}}^2)({r^x_{AC}}^2 + {r^y_{AC}}^2 + {r^z_{AC}}^2)}}\end{aligned} \tag{2.20}$$

Par dérivation d'une fonction composée (*chain rule*), on a :

$$\frac{\partial(\cos(\mathbf{AB}, \mathbf{AC}))}{\partial r^\alpha_A} = \frac{\partial(\cos(\mathbf{AB}, \mathbf{AC}))}{\partial r^\alpha_{AB}}\ \frac{\partial(r^\alpha_{AB})}{\partial r^\alpha_A} \tag{2.21}$$

Le premier terme défini tel que :

$$\begin{aligned}&\frac{\partial(\cos(\mathbf{AB}, \mathbf{AC}))}{\partial r^\alpha_{AB}} = \\ &\Bigg[(r^\alpha_{AC}\sqrt{({r^x_{AB}}^2 + {r^y_{AB}}^2 + {r^z_{AB}}^2)({r^x_{AC}}^2 + {r^y_{AC}}^2 + {r^z_{AC}}^2)} \\ &-r^\alpha_{AB}(r^\alpha_{AC})^2\frac{(r^x_{AB}r^x_{AC} + r^y_{AB}r^y_{AC} + r^z_{AB}r^z_{AC})}{\sqrt{({r^x_{AB}}^2 + {r^y_{AB}}^2 + {r^z_{AB}}^2)({r^x_{AC}}^2 + {r^y_{AC}}^2 + {r^z_{AC}}^2)}}\Bigg] \\ &\frac{1}{({r^x_{AB}}^2 + {r^y_{AB}}^2 + {r^z_{AB}}^2)({r^x_{AC}}^2 + {r^y_{AC}}^2 + {r^z_{AC}}^2)}\end{aligned} \tag{2.22}$$

Finalement, le second terme de l'équation 2.21 donne le signe du gradient total du cosinus.

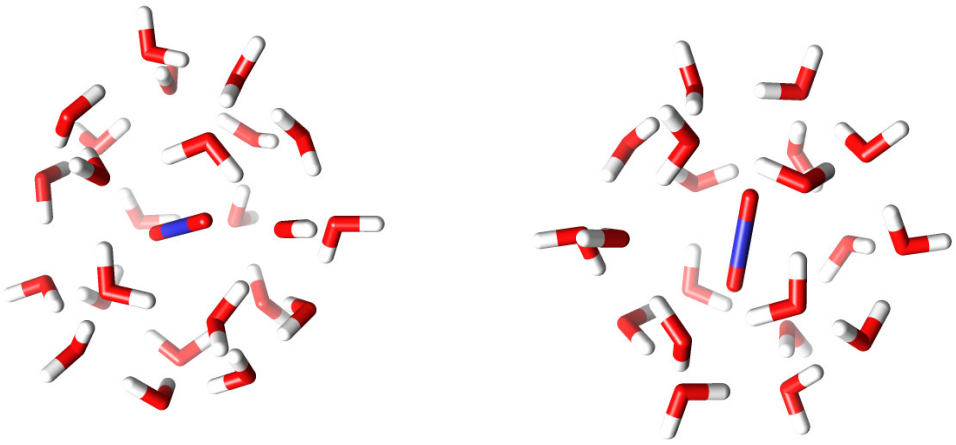
Le gradient du recouvrement, défini en équation 2.19, peut alors s'écrire :

$$\begin{aligned}\frac{\partial\mathrm{rec}(A_P, C_S)}{\partial r^\alpha_A} =& G\frac{\partial(\cos(\mathbf{AB}, \mathbf{AC}))}{\partial r^\alpha_{AB}}\ \frac{\partial(r^\alpha_{AB})}{\partial r^\alpha_A}\mathrm{rec}(A_S, C_S) \\ &+ G\cos(\mathbf{AB}, \mathbf{AC})\frac{\partial\mathrm{rec}(A_S, C_S)}{\partial r^\alpha_A}\end{aligned} \tag{2.23}$$

2.1.3 Dispersion

L'énergie de dispersion peut être vue comme un effet global, dans le sens où les complexes gouvernés par cette composante ne sont pas formés de deux monomères. D'autant plus que ces systèmes se présentent souvent dans un contexte supramoléculaire, c'est-à-dire qu'ils piègent une molécule ou un ion en leur centre. Ce genre d'étude se heurte souvent à la complexité de l'interaction maintenant un tel arrangement moléculaire mais aussi aux effets chélates.

Les clathrates ou plus généralement les complexes formés par l'eau sont représentatifs de ce type d'interactions non-covalentes. Cet effet moléculaire que l'on peut donc qualifier de global présente une énergie attractive qui permet le maintien de tels édifices moléculaires. La figure 2.9 représente deux types de structures de clathrates, se différenciant par l'organisation des molécules d'eau constituant des faces hexagonales et pentagonales. Les clathrates correspondent à des agglomérats de molécules d'eau, formant ainsi des cages, capable de piéger des molécules de dioxyde de carbone, de méthane ou encore de dihydrogène [149, 150]. Ce type de système se rencontre dans le contexte de l'astrochimie [151]. Ces glaces sont présentes dans l'espace et sur Titan, un satellite de Saturne. Ajoutons que même sur Terre, il existe aussi différentes structures de la glace [152]. Enfin, dans une étude récente [153], une nouvelle approche développée en DFT puis adaptée à l'état solide [154, 155] a été utilisée sur des clathrates, ciblant ainsi le problème de traitement de la dispersion lors de l'étude de ce type de complexe supramoléculaire. Rajoutons que cette correction énergétique entre dans le même cadre que celle de SIBFA en considérant non seulement un terme « classique » en $\frac{1}{r}$ mais aussi un second à plus courte distance.



(a) complexe CO_2 - eau (à face hexagonale)

(b) complexe CO_2 - eau (à face pentagonale))

FIGURE 2.9 – Complexes CO_2 - eau.

Formulation

Le terme de dispersion se compose aussi de trois sommes distinctes, estimant les interactions atome-atome, atome-paire libre et paire libre-paire libre décrites dans l'équation 2.24 du chapitre 2 de la partie 1.

$$E_{\text{disp}}(\text{i,j}) = E^{ld}_{\text{disp}}(\text{i,j}) + E^{cd}_{\text{éch-disp}}(\text{i,j}) \tag{2.24}$$

où i et j sont un atome ou une paire libre.

Dispersion longue distance E^{ld}_{disp}

L'interaction de dispersion à longue distance, notée ld se décompose sous la forme d'un potentiel en $\frac{1}{r^n}$ avec $n = 6, 8, 10$:

$$\begin{aligned} E^{ld}_{\text{disp}}(\text{atome - atome}) = \; & (\text{disp}(1/r^6) + \text{disp}(1/r^8) \\ & + \text{disp}(1/r^{10}))C_{disp} \end{aligned} \tag{2.25}$$

avec C_{disp}, facteur de normalisation de l'énergie totale.

$$\text{disp}(1/r^n) = C^n_{disp} D(n)/{m_{ij}}^n \tag{2.26}$$

avec C^n_{disp}, un coefficient dépendant de n et i et j le couple impliqué dans l'évaluation de l'énergie de dispersion.

$$m_{ij} = \frac{r_{ij}}{2\sqrt{R_{disp}(i) R_{disp}(j)}} \tag{2.27}$$

$R_{disp}(i)$ et $R_{disp}(j)$ sont les paramètres atomiques du terme de dispersion.

$$D(n) = B_{ij} \exp(-|\alpha(n) d_{ij}|) \tag{2.28}$$

avec $\alpha(n)$, comme C^n_{disp}, un préfacteur dépendant de n. B_{ij} une constante dépendante du couple i et j. Ajoutons qu'en pratique,

$$d_{ij} = \beta \left(\frac{R_{disp}(i) + R_{disp}(i)}{r_{ij}} \right) - 1 \tag{2.29}$$

avec r_{ij} la distance entre les atomes i et j et β une constante.
Le traitement de la dispersion longue distance, E^{ld}_{disp}, est ainsi défini.

Dispersion courte distance $E^{cd}_{\text{éch-disp}}$

Enfin, comme la dispersion est une interaction à longue distance, il est nécessaire d'ajouter un terme à courte distance : l'échange dispersion afin de respecter le cadre non-additif.

$$E^{cd}_{\text{éch-disp}}(i,j) = t_{ij} C_{xd} \exp(-\alpha_{xd} m_{ij}) \tag{2.30}$$

où α_{xd} est une constante associée à ce terme et différent de celui défini dans l'équation 2.28. m_{ij} est défini dans l'équation 2.27 et C_{xd} une constante associée à l'échange dispersion.

$$t_{ij} = B_{ij} \left(1 - \frac{q_i}{val(i)} \right) \left(1 - \frac{q_j}{val(j)} \right) \tag{2.31}$$

où q_i (q_j) et $val(i)$ ($val(j)$) sont respectivement la charge et la valence électronique de l'atome i (j).

Forces

L'énergie de dispersion SIBFA se décompose en deux termes, le premier décrivant les interactions à « longue distance » et enfin le second à courte portée.

Dispersion à longue portée

$$E_{disp}^{ld} = \sum_{n=6,8,10} \frac{C_{disp}^{n} D(n)}{{m_{ij}}^{n}} \tag{2.32}$$

Seul $D(n)$ et m_{ij} présentent une dépendance spatiale et sont présentés en équations et 2.28 et 2.27 respectivement.
m_{ij} est égale à r_{ij} à une constante près.

$$m_{ij} = A\ r_{ij} \tag{2.33}$$

avec $A = [4R_{disp}(i)R_{disp}(j)]^{-\frac{1}{2}}$.

$$\frac{\partial (m_{ij})^n}{\partial_{r_k^\alpha}} = An \frac{\partial r_{ij}}{\partial r_k^\alpha} (r_{ij})^{(n-1)} \tag{2.34}$$

Dans l'équation 2.28, seul d_{ij} est une variable, fonction de r_{ij}.

$$d_{ij} = \frac{B}{r_{ij}} - 1 \tag{2.35}$$

où $B = \beta[R_{disp}(i) + R_{disp}(j)]$
Deux cas sont donc à considérer : $d_{ij} > 0$ et $d_{ij} < 0$. Enfin le cas où $d_{ij} = 0$ n'est pas rencontré. Ceci impliquerait que :

$$\begin{aligned} &\frac{B}{r_{ij}} - 1 = 0 \\ &\Rightarrow B = r_{ij} \end{aligned} \tag{2.36}$$

Or B est supérieur à des valeurs de distances interatomiques en Å.

$$\frac{\partial d_{ij}}{\partial r_k^\alpha} = \begin{cases} -\dfrac{B}{(r_{ij})^2}\dfrac{\partial r_{ij}}{\partial r_k^\alpha} & d_{ij} > 0 \\ \dfrac{B}{(r_{ij})^2}\dfrac{\partial r_{ij}}{\partial r_k^\alpha} & d_{ij} < 0 \end{cases} \tag{2.37}$$

Le gradient de $D(n)$ s'exprime alors :

$$\frac{\partial D(n)}{\partial r_k^\alpha} = -B_{ij}\alpha(n)\frac{\partial d_{ij}}{\partial r_k^\alpha}\exp\left(-|\alpha n)d_{ij}|\right)) \tag{2.38}$$

Finalement on obtient :

$$\frac{\partial E_{disp}^{ld}}{\partial r_k^\alpha} = \sum_{n=6,8,10} C_{disp}^n \left(\frac{1}{(m_{ij})^n}\frac{\partial D(n)}{\partial r_k^\alpha} + D(n)\frac{\partial (m_{ij})^n}{\partial_{r_k^\alpha}}\right) \tag{2.39}$$

Dispersion à courte portée : échange-dispersion Reformulons l'énergie d'échange dispersion, en regroupant les constantes en P :

$$E_{ech-disp}^{cd} = P\exp(-\alpha_{cd}m_{ij}) \tag{2.40}$$

Seul m_{ij} est une variable du système, dont la dérivée partielle est définie en equation 2.34.

$$\frac{\partial E_{ech-disp}^{cd}}{\partial r_k^\alpha} = -P\alpha_{cd}\frac{\partial m_{ij}}{\partial r_k^\alpha}\exp(-\alpha_{cd}m_{ij}) \tag{2.41}$$

2.2 Traitement de la mobilité électronique : Induction

2.2.1 Introduction

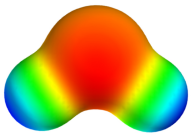
L'introduction d'un terme explicite de polarisation au sein des programmes de dynamique moléculaire peut être considérée comme une des plus grandes avancées de ces dernières années d'un point de vue méthodologique, au-delà de l'accélération des codes.
Physiquement, la polarisation correspond à une modification de la densité de charge d'une molécule influencée par son environnement, c'est-à-dire l'ensemble des molécules voisines de la molécule. La figure 2.10 représente cet effet du à l'approche du cation K^+ sur une molécule d'eau. Le code couleur se défini de la manière suivante : du rouge au bleu équivaut à du plus positif au plus négatif. La figure 2.10a se réfère à une molécule d'eau isolée, qui n'est pas polarisée par son environnement. Enfin les figures 2.10b à 2.10e montre l'effet de polarisation sur la densité de charge de l'eau par l'approche de K^+. Elles révèlent nettement une redistribution de la densité de charge du système. En effet, la molécule d'eau de la figure 2.10d ou 2.10e montre une nette différence avec la figure 2.10a. D'un point de vue physique, sa distribution de charge est plus uniforme.

Enfin, le transfert de charge correspond aussi à un effet de l'environnement mais proche plus particulièrement (voir figure 2.2 du chapitre 2 de la partie 1).

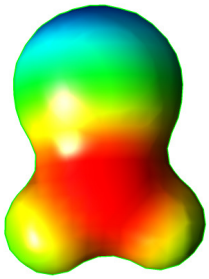
2.2.2 Polarisation

Considérons un système dont la distribution de charge est décrite par des sites multipolaires. Ces sites peuvent être atomiques ou sur le milieu de liaison ou même moléculaires. Le schéma de distribution importe peu. La distribution multipolaire à un site i s'exprime telle que :

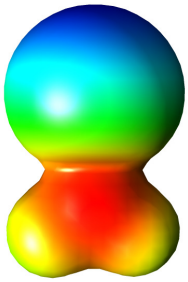
$$M_i = M_i^{perm} + M_i^{ind} \tag{2.42}$$



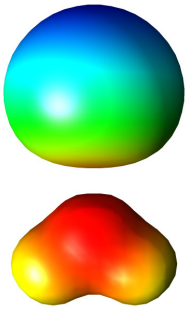
(a) H_20 isolé



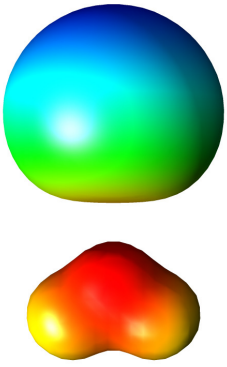
(b) K^+ - H_20 : 1,6 Å



(c) K^+ - H_20 : 2,2 Å



(d) K^+ - H_20 : 2,8 Å



(e) K^+ - H_20 : 3,4 Å

FIGURE 2.10 – Potentiel électrostatique de l'approche du cation K^+ sur l'oxygène de l'eau.

où M_i^{perm} et M_i^{ind} correspondent aux multipoles, au point i, permanents et induits respectivement. Ajoutons toutefois que ce modèle autorise une description de la charge par des charges atomiques partielles, sans pour autant considérer des moments multipolaires d'ordre plus élevé. L'énergie électrostatique entre deux sites multipolaires limitée à l'ordre 1, s'écrit :

$$E_{\text{électrostatique}} = (q_i + \mu_{\mathbf{i}} \cdot \nabla_i)(q_j + \mu_{\mathbf{j}} \cdot \nabla_j)\frac{1}{r_{ij}} \tag{2.43}$$

Ainsi, l'énergie totale d'origine électrostatique incluant donc l'induction du système s'exprime telle que :

$$\begin{aligned} E_{\text{électrostatique}}^{\text{total}} = \sum_i \sum_{i \neq j} \Bigg[& \frac{1}{2} q_i \Omega_0(\mathbf{r_{ij}}) q_j + \boldsymbol{\mu_i} \Omega_1(\mathbf{r_{ij}}) q_j \\ & + \frac{1}{2} \boldsymbol{\mu_i} \Omega_2(\mathbf{r_{ij}}) \boldsymbol{\mu_j} + \frac{1}{2} \boldsymbol{\mu_i} T_\alpha^{-1} \boldsymbol{\mu_j} \Bigg] \end{aligned} \tag{2.44}$$

où Ω_0, Ω_1 et Ω_2 correspondent aux matrices d'interactions charge-charge, charge-dipole et dipole-dipole respectivement. T_α^{-1} est une matrice diagonale comportant les polarisabilités de chaque atome. Enfin le dernier terme est l'énergie de polarisation dans lequel les dipoles notés $\boldsymbol{\mu_i}$ et $\boldsymbol{\mu_j}$ doivent être non nuls. Cette énergie $E_{\text{électrostatique}}^{\text{total}}$ doit être minimisée par les dipoles induits du dernier terme.

$$\begin{aligned} & \frac{\partial}{\partial \boldsymbol{\mu}} E_{\text{électrostatique}}^{\text{total}} = 0 \\ & \Rightarrow (T_\alpha^{-1} + \Omega_2)\boldsymbol{\mu} = -\Omega_1 q \end{aligned} \tag{2.45}$$

Finalement, une fois que le dipole est connu, l'énergie et les forces peuvent alors être calculées. La nature non-additive de l'énergie de polarisation est reflétée dans l'équation 2.45. En effet, ajouter un site (ou un atome) de plus change la valeur de $E_{\text{électrostatique}}^{\text{total}}$ et donc la valeur des dipoles.

Lorsque deux atomes sont trop proches, l'interaction dipole-dipole peut mener au phénomène de « catastrophe de polarisation » où l'énergie atteint

des valeurs non-physiques [76, 156]. Pour éviter cela, un facteur d'atténuation est utilisé à courte distance [39,76,157]. Une autre solution est d'utiliser des polarisabilités faibles mais la précision de l'énergie de polarisation en pâtit [156, 158].

L'effort actuel se concentre sur la résolution rapide de l'équation 2.45, un large système linéaire, dans le cadre de dynamique moléculaire. En effet, l'énergie électrostatique en conditions périodiques grâce au SPME permet un gain de temps de calcul sans perte de précision (voir section conditions périodique du chapitre 1 de la partie 2) et n'est plus une limite computationnelle. Pour autant, peu de travaux se concentrent sur la précision de l'énergie de polarisation ainsi obtenue, impliquant des études sur les polarisabilités ou encore le facteur d'atténuation. Dans le cadre du projet TINKER-HP, une implémentation du calcul de l'énergie de polarisation a été réalisée en conditions périodiques [159].

Physiquement, la polarisation peut apporter des éclairages sur des systèmes membranaires. Le schéma de la figure 2.11 décrit ce que l'on peut observer dans ce type de systèmes [160]. En effet, ils présentent une cavité en forme de tube en leur centre autorisant une libre circulation de certaines particules (eau ou cations). D'un point de vue biologique, cette « circulation » permet à des ligands d'être transportés vers leur potentielle cible thérapeutique passant à travers des membranes.
Le système représenté sur la figure 2.12 correspond à une aquaporine, qui est de taille raisonnable dans le cadre de simulations numériques et est représentatif des systèmes membranaires composés de plusieurs centaines de milliers d'atomes. Lorsqu'une molécule d'eau s'approche de la cavité de l'aquaporine, elle est attirée en son centre pour la traverser. Une molécule d'eau est soumise aux effets d'environnement et la polarisation aurait un rôle important à jouer dans la description de cette trajectoire. Une étude

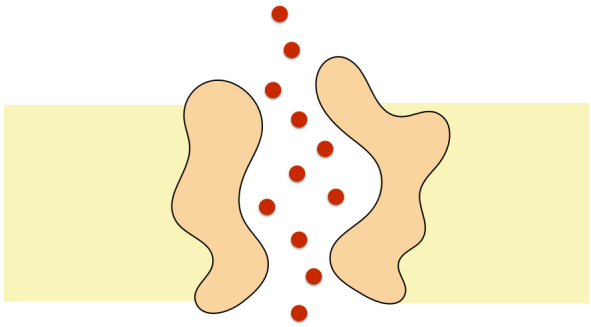
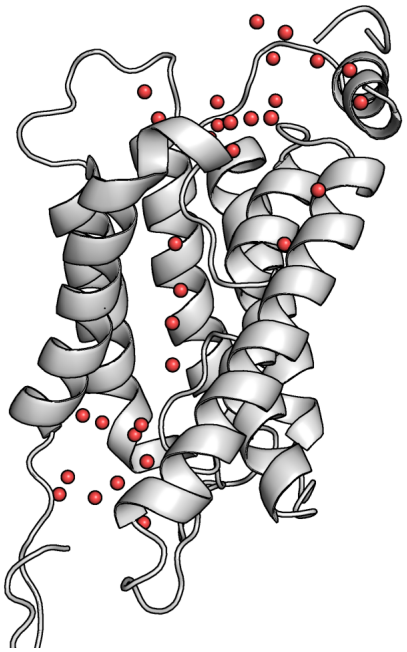


FIGURE 2.11 – schéma d'une membrane traversée par des molécules d'eau représentées par des sphères rouges.

comparative d'approche classique et polarisable de la vitesse de diffusion de l'eau à travers une telle membrane permettrait de répondre à l'apport de la polarisation dans le contexte des transports membranaires.



(a) vue de profil

(b) vue de dessus

FIGURE 2.12 – sélection de réseaux d'eau traversant l'aquaporine 1J4N [161]. Les molécules d'eau sont représentées par les sphères rouges.

2.2.3 Transfert de charge

D'un point de physico-chimique, l'énergie d'induction est composée de la polarisation et du transfert de charge. Le transfert de charge correspond au transfert électronique entre l'orbitale d'un fragment moléculaire à l'état fondamental et une orbitale excitée d'un autre fragment (voir figure 2.13).

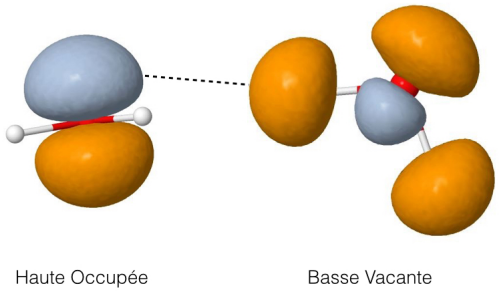


FIGURE 2.13 – schéma transfert de charge entre deux molécules d'eau. L'orbitale moléculaire la plus haute occupée et la plus basse vacante sont représentées [162].

La formulation SIBFA du transfert de charge est basé sur l'approche développée par MURREL *et al* [163]. Cette énergie est dérivée du recouvrement entre les orbitales localisées d'un fragment électron-donneur et les orbitales virtuelles d'un fragment électron-accepteur [164, 165]. Prenant en compte la proximité spatiale de ces deux fragments, la description du transfert de charge se réduit à la prise à compte des paires libres (dont le porteur est donc riche électroniquement : electron-donneur) et aux liaisons impliquant un atome considéré lourd et un hydrogène (electron-accepteur).

$$E_{\text{transfert de charge}} = -2C_{TC} \sum_{\text{paires libres}} \sum_{\text{accepteurs}} N_{occ}^{\alpha} \frac{(T_{\alpha\beta^*})^2}{\Delta E_{\alpha\beta^*}} \tag{2.46}$$

où C_{TC} est une constante de normalisation raffinée par des méthodes de type EDA et plus particulièrement CSOV. Notons que cette valeur s'est mon-

trée transférable dans l'étude de systèmes impliquant des cations métalliques [166,167]. N_{occ}^{α} est le nombre d'occupation de la paire libre considérée lors du calcul. $T_{\alpha\beta^*}$ est fonction : i) du recouvrement de densité de transition entre l'électron-donneur et l'accepteur. L'anisotropie du recouvrement est prise en compte de la même manière quand le terme de répulsion (voir équation 2.2, 2.3 et 2.4). ii) du potentiel électrostatique est ressenti par le porteur de la paire libre. Ce champ comprend la composante électrostatique et la polarisation. Ceci introduit une relation explicite entre le terme de polarisation et celui de transfert de charge en les différenciant explicitement. Enfin le dénominateur de l'équation 2.46 : $\Delta E_{\alpha\beta^*}$ décrit l'énergie nécessaire pour autoriser un tel transfert d'electron entre donneur (paire libre) et accepteur (porteur d'atome hydrogène).

$$\begin{aligned}\Delta E_{\alpha\beta^*} = &\left(I_{\text{paire libre}} + \sum_{i \neq A} V(A) \right) \\ &+ \left(\text{Aff}_{\beta^*} + \sum_{j \neq B} V(B) \right)\end{aligned} \tag{2.47}$$

$I_{\text{paire libre}}$ et Aff_{β^*} correspondent respectivement au potentiel d'ionisation de la paire libre et à l'affinité électronique de l'accepteur. i (j) représente les porteurs de paires libres (porteurs d'atomes d'hydrogène). $V(A)$ ($V(B)$) est le potentiel électrostatique en A le porteur de la paire libre considéré (B, porteur d'un atome d'hydrogène considéré) dans l'interaction.

Formulation

Intéressons nous au numérateur de la formule 2.46 :

$$\left(T_{\alpha\beta^*}\right)^2 = \left(T_{\alpha} - T_{\beta^*}\right)^2 \tag{2.48}$$

$$T_\alpha = \big[(D^1_{tc}V_1)(c_s(A)f_{orb}(A) + c_p(A)f_{orb}(A))M_{AH}\cos(\mathbf{AK},\mathbf{AH})\big] \exp(-\eta_{tc}\ \rho_{AH}) \tag{2.49}$$

où K représente la paire libre. D^1_{tc} et η_{tc} sont des constantes. V_1 indique le potentiel entre l'accepteur (B) et la paire libre. V_1 correspond à la différence entre le potentiel exercé en H et le potentiel entre l'accepteur et la paire libre.

$$T_{\beta^*} = \big[(D^2_{tc}V_2)(c_s(A)f_{orb}(A) + c_p(A)f_{orb}(A))M_{AB}\cos(\mathbf{AK},\mathbf{AB})\big] \exp(-\eta_{tc}\ \rho_{AB}) \tag{2.50}$$

D^2_{tc} est une constante. V_2 correspond à la différence entre le potentiel exercé en A et le potentiel entre l'accepteur et la paire libre.
Les coefficients d'hybridation C_S et C_P, les facteur de recouvrement orbitalaires f_{orb} sont déjà définis dans la formulation de l'énergie de répulsion. Il en va de même pour M_{AB} et M_{AH}.

Enfin intéressons nous au dénominateur de l'équation 2.46.

$$\Delta E_{\alpha\beta^*} = I_K + V - (\mathrm{Aff}(H) + V(A) + V(B)) \tag{2.51}$$

où I_K est le potentiel d'ionisation de la paire libre. Aff(H) est l'affinité électronique de l'hydrogène porté par l'atome lourd. V est le potentiel entre accepteur (A) et donneur (B). Enfin $V(A)$ et $V(B)$ sont les potentiels exercé en A et B.

Les paramètres intrinsèques aux atomes dans le cadre du transfert de charge apparaissent aux dénominateurs du terme ρ_{ij} défini dans l'équation 2.6

Forces

Dans l'expression du transfert de charge, la dépendance spatiale ne touche que le terme exponentiel, le cosinus et les potentiels électrostatiques. Le gradient du terme exponentiel est déjà traité dans les équations 2.16 et 2.17. Enfin les potentiels sont simplement en $1/r$.

L'anisotropie de l'interaction dû au transfert de charge est modélisée par la dépendance angulaire exprimée par le cosinus. Celle-ci sera détaillée dans le cas particulier de la gestion des paires libres, intervenant aussi dans les forces de dispersion et de répulsion.

Les paires libres sont définis dans un repère local, à partir de trois atomes du fragments moléculaires auxquelles leur atome porteur appartient, formant ainsi un repère non-orthogonal finalement. Leur rotation dans le repère global conserve cette dépendance à ces trois atomes. Les coordonnées globale d'une paire libre sont fonction de celles de ces trois atomes. Cette dépendance est donc à prendre en compte dans le calcul des forces.

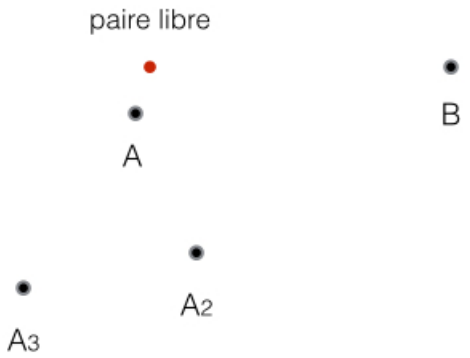


FIGURE 2.14 – schéma dépendance spatiale d'une paire libre du repère local au global. Les atomes sont représentés par les sphères noires et la paire libre en rouge.

La figure 2.14 décrit la dépendance de la paire libre portée par l'atome A mais aussi sa dépendance aux atomes A, A_1, A_2. Enfin B est l'accepteur. Dans le cas de l'évaluation de l'anisotropie du transfert de charge, défini entre A, la paire libre (K) et B, il convient donc de calculer les forces tel que :

$$\begin{aligned}\cos(\mathbf{AK},\mathbf{AB}) &= \frac{\mathbf{AK}\cdot\mathbf{AB}}{|AK||AB|} \\ &= \frac{(r^x_{AK}r^x_{AB} + r^y_{AK}r^y_{AB} + r^z_{AK}r^z_{AB})}{\sqrt{({r^x_{AK}}^2 + {r^y_{AK}}^2 + {r^z_{AK}}^2)({r^x_{AB}}^2 + {r^y_{AB}}^2 + {r^z_{AB}}^2)}}\end{aligned} \tag{2.52}$$

De la même manière que dans l'équation 2.21, par décomposition de fonction composée :

$$\begin{aligned}\nabla(\cos(\mathbf{AK},\mathbf{AB})) =& \frac{\partial\cos(\mathbf{AK},\mathbf{AB})}{\partial r^\alpha_{AK}}\left(\frac{\partial r_{AK}}{\partial r^\alpha_A} + \frac{\partial r_{AK}}{\partial r^\alpha_K}\right) \\ &+ \frac{\partial\cos(\mathbf{AK},\mathbf{AB})}{\partial r^\alpha_{AB}}\left(\frac{\partial r_{AB}}{\partial r^\alpha_A} + \frac{\partial r_{AB}}{\partial r^\alpha_B}\right)\end{aligned} \tag{2.53}$$

Cette dépendance des paires libres selon trois atomes se reflète alors dans le deuxième membre de second terme de la première ligne de l'équation ci-dessus :

$$\frac{\partial r_{AK}}{\partial r^\alpha_K} = \frac{\partial r_{AK}}{\partial r^\alpha_K}\frac{\partial r_K}{\partial r^\alpha_A} + \frac{\partial r_{AK}}{\partial r^\alpha_K}\frac{\partial r_K}{\partial {r_{A_2}}^\alpha} + \frac{\partial r_{AK}}{\partial r^\alpha_K}\frac{\partial r_K}{\partial {r_{A_3}}^\alpha} \tag{2.54}$$

Le gradient du numérateur s'exprime donc :

$$\frac{\partial T_{\alpha\beta^*}}{\partial r^\alpha_k} = \frac{\partial T_\alpha}{\partial r^\alpha_k} - \frac{\partial T_{\beta^*}}{\partial r^\alpha_k} \tag{2.55}$$

$$
\begin{aligned}
\frac{\partial T_\alpha}{\partial r_k^\alpha} =& \Bigg[\left(D_{tc}^1 \frac{\partial V_1}{\partial r_k^\alpha}\right)(c_s(A) f_{orb}(A) \\
& + c_p(A) f_{orb}(A)) M_{AH} \cos(\mathbf{AK}, \mathbf{AH})\Bigg] \\
& + \Bigg[(D_{tc}^1 V_1)(c_s(A) f_{orb}(A) \\
& + c_p(A) f_{orb}(A)) M_{AH} \frac{\partial \cos(\mathbf{AK}, \mathbf{AH})}{\partial r_k^\alpha}\Bigg] \\
& + \Bigg[(D_{tc}^1 V_1)(c_s(A) f_{orb}(A) \\
& + c_p(A) f_{orb}(A)) M_{AH} \cos(\mathbf{AK}, \mathbf{AH})\Bigg]\left(-\eta \frac{\partial \rho_{AH}}{\partial r_k^\alpha}\right) \\
& \exp(-\eta_{tc}\ \rho_{AH})
\end{aligned}
\tag{2.56}
$$

$$
\begin{aligned}
\frac{\partial T_{\beta^*}}{\partial r_k^\alpha} =& \Bigg[\left(D_{tc}^2 \frac{\partial V_2}{\partial r_k^\alpha}\right)(c_s(A) f_{orb}(A) \\
& + c_p(A) f_{orb}(A)) M_{AH} \cos(\mathbf{AK}, \mathbf{AB})\Bigg] \\
& + \Bigg[(D_{tc}^1 V_2)(c_s(A) f_{orb}(A) \\
& + c_p(A) f_{orb}(A)) M_{AH} \frac{\partial \cos(\mathbf{AK}, \mathbf{AB})}{\partial r_k^\alpha}\Bigg] \\
& + \Bigg[(D_{tc}^2 V_2)(c_s(A) f_{orb}(A) \\
& + c_p(A) f_{orb}(A)) M_{AH} \cos(\mathbf{AK}, \mathbf{AB})\Bigg]\left(-\eta \frac{\partial \rho_{AB}}{\partial r_k^\alpha}\right) \\
& \exp(-\eta_{tc}\ \rho_{AB})
\end{aligned}
\tag{2.57}
$$

Le gradient du dénominateur peut s'écrire :

$$
\frac{\partial \Delta E_{\alpha\beta^*}}{\partial r_k^\alpha} = \frac{\partial V}{\partial r_k^\alpha} - \frac{\partial V(A)}{\partial r_k^\alpha} - \frac{\partial V(B)}{\partial r_k^\alpha} \tag{2.58}
$$

Finalement, les forces du transfert de charges sont :

$$
\begin{aligned}
\frac{\partial E_{tc}}{\partial r_k^\alpha} =& -2 C_{tc} N_{occ}^\alpha \frac{T_{\alpha\beta^*}}{\Delta E_{\alpha\beta^*}} \\
& \left[2 \frac{\partial T_{\alpha\beta^*}}{\partial r_k^\alpha} - \frac{\partial \Delta E_{\alpha\beta^*}}{\partial r_k^\alpha} \frac{1}{\Delta E_{\alpha\beta^*}}\right]
\end{aligned}
\tag{2.59}
$$

2.3 Résultats numériques & Discussion

Interactions de VAN DER WAALS

Suivant le fil conducteur d'un modèle d'eau hybride AMOEBA/SIBFA calibré sur SAPT (voir chapitre précédent), les premiers résultats sont sur le jeu de molécule de dix dimères. La première étape de cette implémentation a été évidemment la reproductibilité de ces termes énergétiques par rapport aux résultats SIBFA puis une étude de la somme de l'énergie de répulsion et de dispersion.

Les tableau 2.1, 2.2 et 2.3 reprennent les résultats SAPT, de l'implémentation de SIBFA dans TINKER-HP, de SIBFA originel (noté SIBFA *) et du potentiel 14-7 décomposé d'AMOEBA pour la répulsion, la dispersion et la somme des deux respectivement.

index	SAPT	SIBFA	SIBFA *	AMOEBA
1	-2,20	-2,08	-1,85	-1,78
2	-2,01	-1,93	-1,73	-1,65
3	-1,97	-1,88	-1,69	-1,64
4	-2,11	-2,15	-1,82	-1,91
5	-2,01	-2,08	-1,77	-1,91
6	-1,98	-2,09	-1,79	-1,93
7	-1,68	-1,57	-1,39	-1,36
8	-0,80	-0,74	-0,67	-0,50
9	-1,61	-1,48	-1,29	-1,22
10	-1,12	-1,04	-0,93	-0,80

TABLE 2.1 – Comparaison des énergies de dispersion sur dix dimères d'eau représentatifs à l'équilibre en kcal/mol.

L'érreur moyenne par rapport aux résultats SAPT est de 0,09 kcal/mol pour SIBFA et 0,28 kcal/mol pour AMOEBA dans le cas du traitement de la dispersion. Au-delà d'un meilleur traitement systématique de ce terme énergétique, que ce soit par SIBFA ou SIBFA *, la calibration des paramètres a permis d'améliorer grandement les résultats en s'approchant des valeurs SAPT.

index	SAPT	SIBFA	SIBFA *	AMOEBA
1	7,64	7,30	6,40	4,06
2	6,36	6,18	5,55	3,65
3	6,10	5,82	5,29	3,62
4	5,77	5,96	5,42	3,59
5	4,87	4,67	4,48	3,63
6	4,53	4,70	4,87	3,67
7	3,96	3,53	3,03	1,77
8	1,18	0,75	0,53	0,24
9	4,03	3,48	2,89	1,67
10	2,07	1,82	1,44	0,78

TABLE 2.2 – Comparaison des énergies de répulsion sur dix dimères d'eau représentatifs à l'équilibre en kcal/mol.

Le traitement de la répulsion s'avère plus complexe mais SIBFA * montre déjà des résultats en accord avec SAPT. La nouvelle implémentation calibré sur ces résultats permet d'améliorer cet accord avec une erreur moyenne de 0,30 kcal/mol contre 1,98 kcal/mol pour AMOEBA.

index	SAPT	SIBFA	SIBFA *	AMOEBA
1	5,44	5,22	4,55	2,28
2	4,35	4,25	3,82	2,00
3	4,12	3,94	3,60	1,98
4	3,66	3,81	3,60	1,68
5	2,86	2,59	2,71	1,72
6	2,54	2,61	3,08	1,74
7	2,28	1,96	1,64	0,41
8	0,39	0,01	-0,14	-0,26
9	2,42	2,00	1,60	0,45
10	0,95	0,78	0,51	-0,02

TABLE 2.3 – Comparaison des énergies de VAN DER WAALS sur dix dimères d'eau représentatifs à l'équilibre en kcal/mol.

Un potentiel de type LENNARD-JONES montre nettement ses limites, avec même des valeurs de mauvais signe. Les interactions de répulsion et dispersion nécessitent une formulation plus raffinée comme SIBFA propose dans le cadre d'une méthode non-additive et séparative.

La figure 2.15 reprend les résultats du tableau 2.3 sous la forme d'un diagramme de corrélation par rapport aux résultats référence SAPT. Le décrochage du potentiel de VAN DER WAALS d'AMOEBA est assez net par rapport à la courbe idéale. Enfin, on observe l'amélioration de l'accord entre les résultats SIBFA et SIBFA * qui confirme donc son approche.

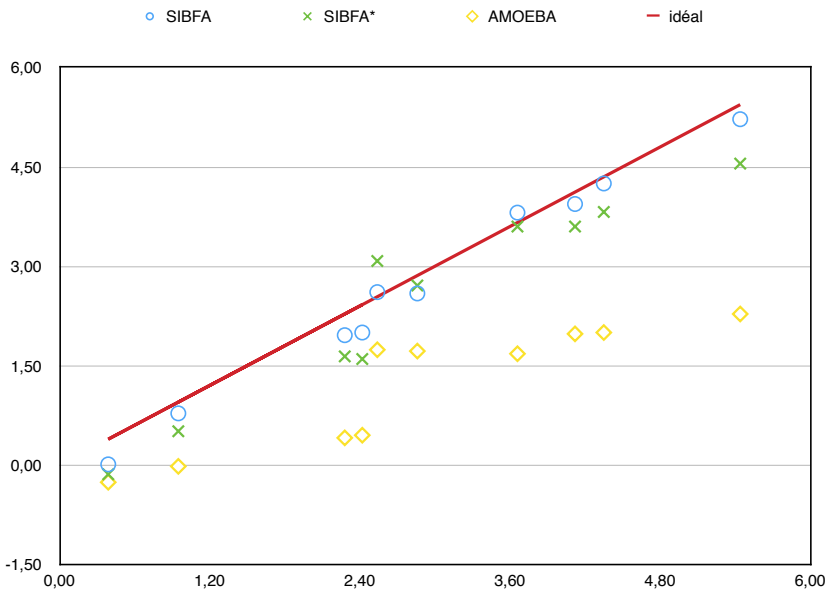


FIGURE 2.15 – Diagramme de corrélation de la somme des énergies de répulsion et dispersion sur les dix dimères d'eau en kcal/mol en lien avec le tableau 2.3. La courbe rouge représente une corrélation idéale. Les cercles bleus les résultats SIBFA. Les croix vertes SIBFA *. Les losanges jaunes AMOEBA.

Transfert de charge

L'implémentation de l'approche SAPT, dans le cadre du transfert de charge, dans le programme Psi4 ne suit pas la même méthodologie que SIBFA mais celle développée par Stone et Misquitta [168] qui présente une dépendance en base. Une approche a été proposée plus récemment par Misquitta qui s'affranchit de cette dépendance [169]. C'est la raison pour laquelle j'ai choisi de rajouter la composante énergétique du transfert de charge obtenue selon la méthode RVS dans le tableau 2.4.

index	RVS	SIBFA	SAPT
1	-0,91	-0,8	-0,74
2	-0,76	-0,72	-0,61
3	-0,73	-0,7	-0,59
4	-0,41	-0,43	-0,36
5	-0,34	-0,33	-0,28
6	-0,30	-0,35	-0,24
7	-0,21	-0,24	-0,18
8	-0,06	-0,06	-0,04
9	-0,19	-0,31	-0,22
10	-0,09	-0,18	-0,11

Table 2.4 – Comparaison des énergies de transfert de charge sur dix dimères d'eau représentatifs à l'équilibre en kcal/mol.

La figure 2.16 représente le diagramme de corrélation par rapport à SAPT, reprenant ainsi les résultats du tableau 2.4. D'un point de vue théorique, il est intéressant de noter que les résultats RVS présentent un écart moyen de 0,07 kcal/mol par rapport à SAPT. Aux regards de cette observation, et au-delà d'une recalibration, une reflexion sur le transfert de charge serait pertinente. D'autant plus que SIBFA se place dans le cadre d'une méthode non-additive et séparitive. Chaque terme énergétique doit donc être indépendant afin de reproduire une énergie totale « réaliste », en corrélation avec des résultats de chimie quantique.

Une alternative serait de considérer l'induction, c'est-à-dire la somme de

la polarisation et du transfert de charge, comme valeur à comparer avec des méthodes de type EDA et plus particulièrement SAPT. Pour cela, l'énergie de polarisation doit être calibrée et s'affranchir des problèmes d'atténuation et polarisabilités en conclusion de ce chapitre. Cette alternative assurerait un traitement énergétique non-additif de la mobilité électronique.

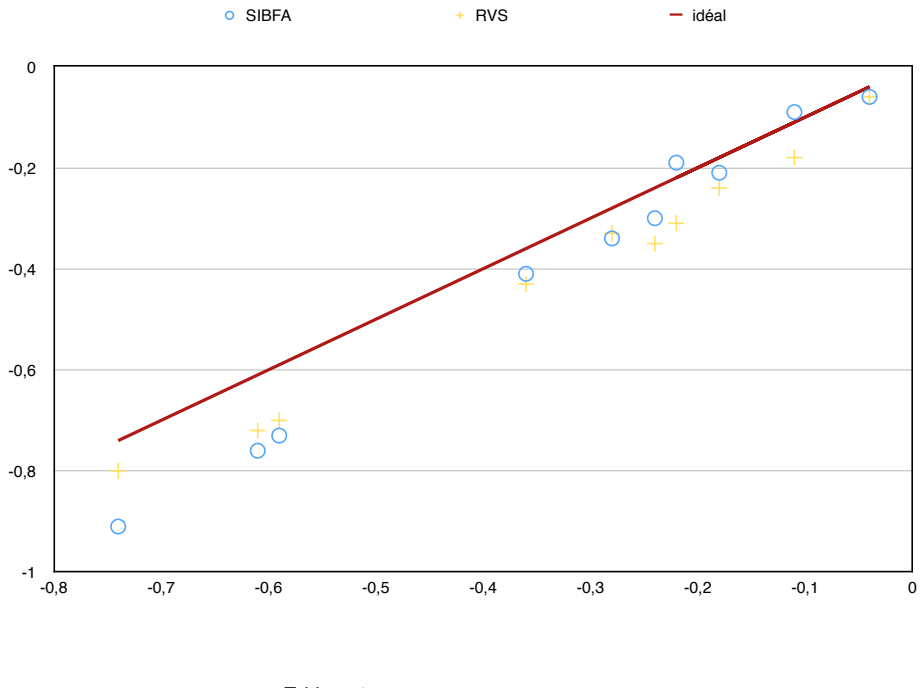


FIGURE 2.16 – Diagramme de corrélation de l'énergie de transfert de charge sur les dix dimères d'eau en kcal/mol en lien avec le tableau 2.4. La courbe rouge représente une corrélation idéale. Les cercles bleus les résultats SIBFA. Les croix jaunes RVS.

2.4 Conclusions & Perspectives

Interaction de van der Waals

Une meilleure description de ces deux composantes, que sont la répulsion et la dispersion, permettra aux termes de van der Waals de jouer un vrai rôle dans l'évaluation des interactions d'un système. En effet, jusque là, le potentiel décrivant les interactions de van der Waals avait un rôle de tampon dans le cadre d'une description de type Lennard-Jones. Rajoutons toutefois qu'un potentiel de Buckimgham parfois utilisé ou tout au moins disponible selon les implémentations se rapproche dans sa forme exponentielle de la description du type sibfa dans son traitement de l'échange-dispersion [170].

Une nouvelle voie d'application s'ouvrirait alors, comme l'astrochimie avec l'étude notamment des clathrates d'eau piégeant de petites molécules. L'étude de la structure de la glace reste hors de portée des approches de type eda. Seule une méthode classique pourrait donner des réponses. Les protéines antigel seraient alors l'étape suivante, très stimulante [171, 172].

Induction

L'inclusion d'un terme de polarisation explicite dans un programme de dynamique moléculaire repose sur deux efforts de travail : i) l'optimisation computationnelle. ii) la reproductibilité des résultats référence.

Jusqu'à l'introduction des approches exploitant les conditions périodiques, telle que le spme, le calcul de l'énergie électrostatique restait limitant en temps de calcul. Cette limite affranchie, le calcul de l'énergie de polarisation se retrouve dans la même situation. L'intégration de ce terme dans le cadre

du SPME ouvre de nouveaux horizons d'un point de vue méthodologie des interactions non-covalentes. En effet, l'importance de la polarisation dans les systèmes comportant des cations métalliques, très présents dans les systèmes biologiques puisque primordial dans les catalyses enzymatiques [23]. Ces nouveaux horizons peuvent être notamment représentés par des simulations de transports d'espèces chimiques (cations par exemple) à travers des membranes. Ce type d'étude souvent réalisé en conditions « trop » réaliste dans le sens où la taille du système se révèle trop importante pour de longues simulations atomiques de dynamiques moléculaires dans le cadre BORN-OPPENHEIMER. Le coût mémoire devient alors limitant. Des approches du type gros grains (*coarse grain*) sont préférées, se différenciant du modèle tout-atome. Selon le niveau d'approximation, le système est simplifié par des points (gros grains) décrivant des fragments moléculaires de plusieurs atomes. Ces modèles gros grains n'incluent des effets de polarisation que pour la description approximée de modèle d'eau ou encore de liquide ionique [173–176]. Ajoutons toutefois que ces approximations peuvent être calibrées à partir de simulations atomiques comprenant un terme explicite de polarisation.

Concernant le second point soulevé, le raffinement de ces résultats passe par des études de la polarisation à courte distance. En effet, celle-ci implique une fonction d'atténuation de THOLE mais aussi les polarisabilités atomiques qui sont réduites [76,156,158]. L'effet de polarisation peut être vu comme un effet global dû à l'environnement chimique. Or comme on peut l'observer sur la figure 2.10, la mobilité électronique et plus précisément la distribution de charge des deux fragments se voit nettement modifiée entre le dimère et le monomère isolé. La polarisation n'est donc pas seulement un effet global mais aussi local. Cela justifie une comparaison systématique avec les résultats SAPT.

L'apport d'un terme énergétique de transfert de charge dans un programme de mécanique moléculaire permettrait une meilleure compréhension des interactions non-covalentes impliquées dans certains complexes supramoléculaires. Au-delà d'un calcul statique, l'évaluation de l'énergie de transfert de charge au cours d'une dynamique moléculaire, dans le cadre d'un échantillonnage conformationnelle par exemple, apporterait des réponses interprétatives quant aux interactions mises en jeu au sein du système étudié. De plus, le couplage d'une telle approche, reproduisant les autres contributions énergétiques intermoléculaires, aux méthodes interprétatives permettrait des analyses complètes des interactions physiques. Parmi ces méthodes, citons NCIplot et ELF, l'approche empirique de la première associée à SIBFA semble la plus réalisable à courte portée [144, 145, 177, 178]. En effet, l'analyse des interactions non-covalentes, via le programme NCIplot, permet déjà de s'intéresser aux différents types d'interactions non-liées. Citons notamment les interactions intermoléculaires des halogènes, présentant un effet physique unique, avec une polarité différente selon une approche radiale ou perpendiculaire *sigma-hole* [179,180]. En effet, cette propriété unique est un sujet de recherche en médecine et plus particulièrement dans la conception médicamenteuse, participant fortement à la stabilisation de ligand au sein de la cavité du récepteur. Une étude est présentée, sous forme de chapitre de livre, sur les différentes types d'interactions non-covalentes en annexe.

En conclusion de ce chapitre, le modèle d'eau hybride AMOEBA/SIBFA dont les prémisses avaient déjà été présentés dans le chapitre précédent, continue son raffinement. En effet, les composantes énergétiques de la répulsion et la dispersion ont été reproduite en très bons accord avec SAPT. De plus, le transfert de charge présente, lui aussi, une bonne corrélation avec les résultats référence. Enfin la polarisation AMOEBA, non-traité dans ces travaux, est déjà bien calibré pour reproduire une valeur proche des résultats

référence. Rappelons que la polarisation est isotrope, on ne se heurterait donc pas à un modèle sur-paramétré qui « s'écroulerait » selon différentes orientations spatiales. Ces travaux de thèse ont permis d'établir, dans un programme de dynamique moléculaire, incluant donc forces et énergies, un modèle d'eau reproduisant avec une très bonne corrélation chaque composante intermoléculaire.

Conclusions & Perspectives

Le développement de champs de forces n'a eu de cesse de proposer de nouvelles méthodes ou formulations ces dernières décennies, que ce soit d'un point de vue classique ou polarisable. Cette constance, et ce depuis la première distribution d'un programme de mécanique moléculaire dans les années 80, montre l'intérêt de l'utilisation de tels programmes dans le domaine de la reconnaissance moléculaire mais aussi les efforts méthodologique à fournir. Ces modèles permettent l'étude de systèmes macromoléculaires ou encore de ligands composés de quelques dizaines d'atomes.

Dans le cadre des approches purement classiques, ces efforts se sont concentrés, ces dernières années, sur le traitement de la polarisation tout en restant dans la description de la densité électronique avec des charges partielles. Ceux-ci donnant lieu au modèle de Drude, ajoutant une charge fictive relié à l'atome porteur, excepté les hydrogènes, par un oscillateur harmonique mais aussi aux charges fluctuantes. Rappelons que le formalisme des dipoles induits est applicable aux charges partielles. Cependant, l'anisotropie électrostatique n'est pas reproduit, biaisant ainsi les résultats obtenus selon l'orientation de l'interaction. Ajoutons que cette observation prend tout son sens lorsque les atomes engagés dans l'interaction ne sont pas alignés, comme dans le cas d'une liaison hydrogène où la liaison covalente $H - O$ forme un angle non-nul avec la liaison non-covalente de ce même atome d'oxygène et l'hydrogène de l'autre fragment $O - - - H$. Ces deux méthodes modifient la composante électrostatique, induisant ainsi le non respect de la non-additivité des termes énergétiques. Néanmoins elles sont toutes disponibles dans des programmes massivement utilisés en dynamique moléculaire, comme CHARMM.

Les champs de forces polarisables ou dits de seconde génération autorisent un meilleur traitement des interactions non-covalentes. Distinguons tout de même les approches AMOEBA et SIBFA qui n'ont pas la même philo-

sophie mais qui partagent une certaine méthodologie dans leur formulation de l'énergie électrostatique et de polarisation. En effet, ces deux champs de forces utilisent des multipoles afin de mieux décrire la densité électronique de fragments moléculaires à partir de calculs *ab-initio.*. En pratique, ces sites multipolaires interviennent dans le calcul de l'énergie électrostatique et de polarisation selon des schémas différents. En effet, le protocole de calcul des multipoles diffère, ainsi que le formalisme de la polarisation.

L'emploi de multipoles est donc crucial dans le traitement des interactions d'origine électrostatiques et polarisables pour reproduire des valeurs référence. Toutefois la pertinence de leur valeur via le programme GDMA mérite réflexion. Ce programme propose plusieurs partitionnement de l'espace, c'est-à-dire de la densité électronique et la possibilité de sélectionner la valeur du rayon assigné par type d'atomes formant ainsi une sphère sur laquelle sont calculés les multipoles. Les différentes combinaisons de ces options constituent un ensemble de jeux de multipoles qui ne sont pas tous adaptés par rapport aux valeurs énergétiques références. Le niveau de calcul de la densité est aussi un élément à prendre en compte. En effet, jusque là, la procédure de SIBFA évaluait les multipoles au niveau HARTREE-FOCK, c'est-à-dire sans corrélation électronique. Dans le cadre de ces travaux de thèse, elle a été calculé au niveau MP2, en testant différents protocoles afin de s'assurer de la meilleure description de la densité électronique, modélisant ainsi l'électrostatique en corrélation avec les résultats références.

La qualité des multipoles est donc un aspect primordial dans la construction et la validation d'un champ de forces polarisables. En effet, il s'agit tout simplement du point de départ avant une simulation. D'autant plus que pour reproduire les résultats issus de la chimie quantique, il est nécessaire d'apporter une correction à courte portée. Rappelons que cette dernière est dû à l'approximation multipolaire qui entraine une surestimation de l'énergie à

courte distance et autour de l'équilibre. Une analyse des résultats brutes permet déjà d'éliminer certains jeux de multipoles en raison de surestimations énergétiques. Il est alors nécessaire de déterminer le jeu de multipoles le plus adapté à la correction de pénétration de charge. La correction présenté dans le chapitre Electrostatique s'est montré transférable selon les modèles (charge-charge seulement ou charge-charge et charge-dipole et au-delà) et selon des multipoles ayant un protocole de calculs différents.

Une alternative serait d'exploiter le potentiel GEM (*Gaussian Electrostatic Model*). En effet, cette approche plus raffinée qui va au delà des multipoles distribués, en employant des gaussiennes modulées par des polynômes d'Hermite qui permettent de décrire des effets électroniques contractés ou diffus. Celles-ci décrivent de manière plus précise la densité électronique. Il est possible d'extraire des multipoles à partir de ces gaussiennes. Ce protocole est moins permissif que le précédent. Cette option pourrait permettre d'établir une méthodologie efficace et robuste afin de construire une base de donnée de multipoles corrélés.

Cependant, le potentiel GEM n'est pas uniquement pertinent quant au calcul des multipoles. En effet, il s'agit d'un champ de force à part entière qui permet de s'affranchir d'une correction pour la pénétration de charge puisque le recouvrement entre gaussiennes est évalué et compris dans la composante électrostatique. En pratique, cela signifie des paramètres en moins mais un calcul plus couteux en raison du modèle gaussien.

De plus, il est tout à fait envisageable d'étendre la décomposition intermoléculaire proposé par le potentiel SIBFA à GEM. En effet, la formulation non-additive et explicite le permet. Le reproductibilité des résultats de décomposition d'énergie intermoléculaire avec un très bon accord par l'implémentation de SIBFA dans TINKER ne peut se limiter à l'un ou l'autre terme de manière robuste. En effet, SIBFA s'inscrit dans un cadre de non-additivité et sépartibilité des termes énergétiques, par conséquent chaque

contribution doit suivre ce principe. Il n'est donc pas suffisant de se rapprocher de valeurs référence pour l'énergie électrostatique, même si elle est la contribution souvent considéré comme force motrice dans l'association de complexe de type ligand-récepteur. C'est aussi pour cette raison que la composante électrostatique a mérité le plus de temps passé au cours de ces trois années de thèse, élaborant ainsi une solide méthodologie quant au calcul des multipoles, dans la validation de cette contribution. Les résultats présentés ont permis d'élaborer un modèle d'eau avec une énergie électrostatique en excellente corrélation avec les résultats référence que sont SAPT.

Suivant ainsi ces principes de non-additivité et de sépartibilité, l'implémentation des termes décrivant les interactions de VAN DER WAALS et l'induction et l'accord avec SAPT a été validé, étendant ainsi la qualité du modèle d'eau hybride AMOEBA/SIBFA. En effet, les interactions de VAN DER WAALS sont principalement traité par un potentiel de type LENNARD-JONES ou dérivé, qui ne reproduit pas une énergie, même totale, en accord avec des calculs de chimie quantique. L'approche SIBFA, recalibrée dans le cadre de SAPT, donne de bons résultats, que ce soit séparément ou sommés. Ces résultats ont été obtenus selon différentes orientations de dimères d'eau reproduisant ainsi l'anisotropie des énergies de répulsion et de dispersion. Ce modèle d'eau peut déjà ambitionner l'étude de complexes tels que les clathrates d'eau.

De plus une interprétation plus raffinée de l'induction, en différenciant la polarisation et le transfert de charge, c'est-à-dire mimant la mobilité électronique intra- et interfragment moléculaire, est vérifié par l'approche SIBFA du transfert de charge. Un effet, certes plus fin, mais qui autorise, au-delà de la reproductibilité des résultats SAPT, une meilleur compréhension des phénomènes physiques en jeu, en rationnalisant les interactions intermolé-

culaires. Pour rester dans ce cadre, l'étude de l'énergie d'induction apparait nécessaire quant au respect de la non-additivité et sépartibilité énergétique. Celui-ci passe par une réflexion sur les modèles employés aujourd'hui : le facteur d'atténuation de THOLE et les polarisabilité, voire même au-delà, avec les hyperpolarisabilités exploitant ainsi les quadrupoles. Ceci rajouterait un coût au calcul de l'énergie de polarisation mais qui serait pertinente selon les atomes pris en compte. En effet, les cations métallique ou les éléments du bloc f en bénéficieraient avec une évaluation de l'énergie de polarisation plus raffinée.

Cependant, ces approches ne sont pas encore intensivement utilisés comme le peuvent être les champs de forces classiques. Ceci est dû à l'actuel implémentation de ces méthodes qui n'est pas massivement parallèle. TINKER, le programme dans lequel a été implémentée la méthodologie SIBFA, était à l'origine en OPENMP, limitant ainsi ses performances. Le projet TINKER-HP permet de passer outre cette limite purement computationnelle avec une implémentation MPI et/ou hybride (MPI / OPENMP). L'inclusion de la correction de pénétration de charge et de la polarisation dans le cadre du SPME en conditions périodiques, autorisant alors des applications des conditions proches du vivant.

La revisite du modèle SIBFA ne s'apparente pas à un simple « copier-coller » du programme actuel. L'objectif étant son implémentation dans un programme de dynamique moléculaire et son optimisation. Celle-ci passe par des efforts de parallélisation mais aussi de simplification de son utilisation, en ayant l'ambition d'une large distribution. Cette revisite s'accompagne aussi d'une réflexion quant aux résultats référence. En effet, plusieurs approches de décomposition de l'énergie intermoléculaire existent. Seul SAPT est capable de reproduire toutes les composantes non-covalentes. Son implémentation dans PSI4 permet d'accéder à des applications sur des dimères de

plus en plus grand, ce qui a longtemps été une limite à cette approche. Enfin cette revisite de SIBFA passe aussi par une mise à plat du modèle complet dans le cadre du traitement des interactions intermoléculaires, s'étendant jusqu'à l'implémentation des forces. Ceci autorisant alors les premières dynamiques moléculaires SIBFA et ce, en conditions périodiques.

En fil conducteur de ces travaux de thèses, la molécule d'eau a souvent été le sujet d'étude et de validation de l'implémentation de SIBFA. L'eau regroupe beaucoup de propriétés que l'on retrouve dans les autres fragments moléculaire. Son oxygène est riche électronique, de part ses doublets non-liants, proposant ainsi un large site attracteur de liaison hydrogène et polarisable. Elle possède aussi deux hydrogènes, capable de réaliser des liaison du même type et eux aussi sont polarisables. L'eau est tout simplement essentielle afin de se placer dans un environnement réaliste aux yeux de la biologie. L'ensemble de ces résultats montre les prémisses d'un modèle d'eau SIBFA. Prémisse, car pour l'heure, il s'agit d'un modèle d'eau hybride AMOEBA/SIBFA, puisque seuls les composantes intermoléculaires ont été modifiées reproduisant avec de très bon accords les résultats référence. L'intramoléculaire est la prochaine étape de ce modèle, ainsi que la thermodynamique, laissant entrevoir des applications stimulantes. Pour l'heure, ces travaux de thèse ont permis de réaliser le premier modèle d'eau énergétique d'inspiration SIBFA, calibré sur une méthode robuste de décomposition d'énergie SAPT.

L'apport d'un tel programme, regroupant performance et précision, devait bénéficier à de nombreux domaines tels que la biochimie et la pharmacologie.

Enfin, étant donné que les résultats *ab-initio* sont reproduits avec de très bons accords, il serait intéressant de coupler cette méthodologie à une

approche dites réactives, simulant la réactivité de systèmes complexes en modélisant les formations et cassures de liaisons covalentes [181, 182]. Cette perspective étendrait le champs d'application de SIBFA aux catalyses en simulant le chemin réactionnel en lieu et place de la DFT. Plus raisonnablement, l'association d'un programme tel que TINKER-HP, mêlant efficacité et méthodologie, avec un programme de chimie quantique autoriserait des applications hybrides QM/MM, très proches d'un traitement complet quantique au niveau énergétique avec notamment la polarisation entre les deux zones assurée par de bons multipoles.

Table des figures

1 schéma association ligand-récepteur (« clef - serrure ») . . . 2
2 diagramme secteur de la composition atomique du corps humain en pourcentage relatif [7]. 5

1.1 Schéma élongation liaison. Cas de l'éthane. 11
1.2 Schéma déformation angulaire. Cas d'un acide. 11
1.3 Schéma déformation diédrale. Cas du dipeptide de l'alanine. 12
1.4 schéma dièdre impropre 13
1.5 Schéma correction d'Urey-Bradley. Cas de l'acide acétique. 14
1.6 tracé schéma d'un potentiel du type 12-6 16
1.7 Exemple de solvatation explicite d'un système. 18
1.8 schéma modèle d'eau en lien avec le tableau 1.2 19
1.9 Exemple de solvatation implicite d'un système. 20
1.10 exemple d'application d'un schéma deux couches de solvatation. 21
1.11 schéma des deux états redox du cofacteur TTQ 26
1.12 schéma simplifié illustrant un diagramme de Jablonski . . 28
1.13 Illustration de spectres de différentes structures secondaires obtenus par dichroïsme circulaire 30

2.1 Potentiel électrostatique moléculaire. 35
2.2 schéma RVS (*Reduced Variational Space*) illustrant la décomposition des interactions intermoléculaires [88]. 40

1.1 exemple de réseau de molécules d'eau au sein de la protéine (code PDB :1GAH [117]). 55

1.2 sélection réseau intermoléculaire au sein de 1GAH 57

1.3 Energie électrostatique SAPT en fonction de la variation angulaire de la liaison hydrogène en kcal.mol et schéma moléculaire de la variation angulaire. 58

1.4 exemple d'interaction π entre une double liaison et un atome d'hydrogène. 59

1.5 Représentation des orbitales atomiques s, p et d ainsi que l'ordre du moment multipolaire associé (réalité avec MOPAC) 61

1.6 Energie électrostatique SAPT (en kcal/mol) : scan intermoléculaire de dimères de gaz rares. 62

1.7 Profil de l'énergie électrostatique de dimères de l'eau en fonction de la distance intermoléculaire en kcal/mol. SAPT est représenté par la courbe continue rouge. AMOEBA (MEP) en tiret vert. cp-AMOEBA (cp-MEP) en tiret pointillé bleu. . . . 84

1.8 Diagramme de corrélation de l'énergie électrostatique entre la référence SAPT et la correction de pénétration (cp-AMOEBA) en kcal/mol sur une série de dimères divers (voir légende) . . 85

1.9 Complexes cation métallique - eau : profil de l'énergie électrostatique en fonction de la distance cation - O en kcal/mol. SAPT correspond la la courbe trait plein rouge. AMOEBA en trait discontinu vert. cp-AMOEBA en bleu en trait discontinu et pointillé. a) complexe K^+ eau. b) complexe Na^+ eau. c) complexe Mg^{2+} eau. d) complexe Ca^{2+} eau. e) complexe Zn^{2+} eau. f) illustration de l'approche cation - oxygène (de la molécule d'eau). 87

1.10 Décomposition de $1/r$ en $erf(r)/r$ et $erfc(r)/r$: schéma sommation d'EWALD. 89
1.11 exemple de motif : réseau de Bravais monoclinique primitif (paramètre de maille : a $\neq b \neq c, \alpha = \gamma = 90, \beta = 90$) . . . 90
1.12 Exemple schématique d'un système biologique en conditions périodiques en 2 dimensions. 91
1.13 Exemple d'application de la fonction d'amortissement sur $f(r)$. 95

2.1 représentation des bassins ELF (en bleu) d'un bain de molécules d'eau [146]. 138
2.2 schéma répulsion proline (gauche) valine (droite). AB et CD représentent les liaisons OH de la proline et la valine respectivement, en gardant leur orientation. 140
2.3 illustration répulsion liaison - liaison : cas de l'eau (gauche) et du formaldéhyde (droite) 140
2.4 illustration répulsion liaison - paire libre : cas formaldéhyde (gauche) et de l'eau (droite) 143
2.5 schéma interaction rep(AB, j). 144
2.6 schéma interaction de répulsion complète 144
2.7 illustration répulsion paire libre - paire libre : cas de l'eau (gauche) et du formaldéhyde (droite) 145
2.8 schéma répulsion proline (gauche) valine (droite). AB et CD représentent les liaisons OH de la proline et la valine respectivement, en gardant leur orientation. 147
2.9 Complexes CO_2 - eau. 150
2.10 Potentiel électrostatique de l'approche du cation K^+ sur l'oxygène de l'eau. 155
2.11 schéma d'une membrane traversée par des molécules d'eau représentées par des sphères rouges. 158

2.12 sélection de réseaux d'eau traversant l'aquaporine 1J4N [161]. Les molécules d'eau sont représentées par les sphères rouges. 159
2.13 schéma transfert de charge entre deux molécules d'eau. L'orbitale moléculaire la plus haute occupée et la plus basse vacante sont représentées [162]. 160
2.14 schéma dépendance spatiale d'une paire libre du repère local au global. Les atomes sont représentés par les sphères noires et la paire libre en rouge. 163
2.15 Diagramme de corrélation de la somme des énergies de répulsion et dispersion sur les dix dimères d'eau en kcal/mol en lien avec le tableau 2.3. La courbe rouge représente une corrélation idéale. Les cercles bleus les résultats SIBFA. Les croix vertes SIBFA *. Les losanges jaunes AMOEBA. 168
2.16 Diagramme de corrélation de l'énergie de transfert de charge sur les dix dimères d'eau en kcal/mol en lien avec le tableau 2.4. La courbe rouge représente une corrélation idéale. Les cercles bleus les résultats SIBFA. Les croix jaunes RVS. 170

Liste des tableaux

1.1 Composition du plasma sanguin en g/dl [25] 17
1.2 Tableau de paramètres des modèles d'eau. Distance en Angströms, angle en degrés. A et C sont les constantes au numérateur de l'expression du potentiel de LENNARD-JONES. [41] 20
1.3 Longueurs de liaisons et écart-types en Angström du TTQ moyennés sur l'ensemble des trajectoires. 27

1.1 Energie multipolaire électrostatique à partir de moments atomes en kcal/mol 76
1.2 Energie multipolaire électrostatique à partir de moments sur liaisons et atomes en kcal/mol 76
1.3 Energie électrostatique SAPT (3ème ordre - base aug-cc-pvTZ) en kcal/mol. 77
1.4 Comparaison de l'énergie multipolaire corrigée à courte portée en kcal/mol sur dix différentes orientations de dimères d'eau. GDMA ^{i} décrit le schéma de la distribution des moments, $i = 1$ seulement atomique et $i = 2$ atome et liaison. Le préfixe cp- désigne les termes corrigés à courte distance. . 78
1.5 Erreur moyenne en kcal/mol par rapport à SAPT. 79
1.6 Décomposition de la correction de pénétration de charge obtenue via le protocole sw2-rad325 de GDMA avec des moments multipolaires centrés sur les atomes (GDMA ^{1}) en kcal/mol. . 81

1.7 Décomposition de la correction de pénétration de charge dans le cadre d'AMOEBA (cp-MEP) en kcal/mol. 82
1.8 Comparaison de l'énergie électrostatique purement multipolaire en kcal/mol. 82
1.9 Energie électrostatique de dimères composés de benzènes en kcal/mol. 83

2.1 Comparaison des énergies de dispersion sur dix dimères d'eau représentatifs à l'équilibre en kcal/mol. 166
2.2 Comparaison des énergies de répulsion sur dix dimères d'eau représentatifs à l'équilibre en kcal/mol. 167
2.3 Comparaison des énergies de VAN DER WAALS sur dix dimères d'eau représentatifs à l'équilibre en kcal/mol. 167
2.4 Comparaison des énergies de transfert de charge sur dix dimères d'eau représentatifs à l'équilibre en kcal/mol. 169

C.5 paramètres atomiques de différents fragments moléculaires de la correction de pénétration de charge dans le cadre de cp-AMOEBA (cp-MEP) 238
C.6 paramètres atomiques de différents fragments moléculaires de la correction de pénétration de charge dans le cadre de cp-GDMA [1] 238
C.7 paramètres de la correction de pénétration de charge dans le cadre de cp-GDMA [1] 238
C.8 paramètres et constantes de l'énergie de répulsion de SIBFA 239
C.9 constantes de l'énergie de dispersion SIBFA à longue portée . 239
C.10 constantes de l'énergie de dispersion SIBFA à courte portée . 239

Annexe A

Résultats SAPT

Les tableaux récapitulatifs des résultats SAPT sur le S66 étendu de l'article inséré en section 1.3 du chapitre 1 de la partie 2 sont regroupés sur plus de 100 pages. Ci-dessous le lien pour les consulter :
http ://pubs.acs.org/doi/abs/10.1021/acs.jctc.5b00267

Annexe B

Chapitre NCIplot : Etude des différents types d'interactions non-covalentes

A complete NCI perspective: from new bonds to reactivity

Christophe Narth, Zeina Maroun, Roberto A. Boto, Robin Chaudret, Marie-Laure Bonnet, Jean-Philip Piquemal and Julia Contreras-García

Abstract The Non-Covalent Interaction (NCI) index is a new topological tool that has recently been added to the theoretical chemist's arsenal. NCI fills a gap that existed within topological methods for the visualization of non-covalent interactions. Based on the electron density and its derivatives, it is able to reveal both attractive and repulsive interactions in the shape of isosurfaces, whose color code reveals the

Christophe Narth
UPMC Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris,France,
CNRS, UMR 7616, Laboratoire de Chimie Théorique, F. 75005, Paris, France, e-mail: christophe.narth@lct.jussieu.fr

Zeina Maroun
Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark, e-mail: zeinamaroun89@gmail.com

Roberto A. Boto
UPMC Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris,France,
CNRS, UMR 7616, Laboratoire de Chimie Théorique, F. 75005, Paris, France, e-mail: alboto@lct.jussieu.fr

Robin Chaudret
Scienomics, 16 rue de l'Arcade, F. 75008, Paris, France, e-mail: Robin.Chaudret@scienomics.com

Marie-Laure Bonnet
ENSCP, Ecole Nationale Superieure de Chimie de Paris Chimie ParisTech, UMR 8247 Institut de Recherche de Chimie Paris, F-75231, Paris, France,
CNRS, UMR 8247, Institut de Recherche de Chimie Paris, F-75231, Paris, France, e-mail: marie-laure.bonnet@chimie-paristech.fr

Jean-Philip Piquemal
UPMC Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris,France,
CNRS, UMR 7616, Laboratoire de Chimie Théorique, F. 75005, Paris, France,e-mail: jpp@lct.jussieu.fr

Julia Contreras-García
UPMC Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris,France,
CNRS, UMR 7616, Laboratoire de Chimie Théorique, F. 75005, Paris, France, e-mail: contrera@lct.jussieu.fr

nature of the interaction. It is interesting to note that NCI can even be calculated at the promolecular level, making it a suitable tool for big systems, such as proteins or DNA. Within this chapter we will review the main characteristics of NCI, its similarities with and differences from previous approaches. Special attention will be paid to the visualization of new interaction types. Being based on the electron density, NCI is not only very stable with respect to the calculation method, but it is also a suitable tool for detecting new bonding mechanisms, since all such mechanisms should have a detectable effect on the electron density. This type of approach overcomes the limitations of bond definition, revealing all interaction types, irrespective of whether they have a name or have previously been identified. Finally, we will show how this tool can be used to understand chemical change along a chemical reaction. We will show several examples of *torquoselectivity* and put forward an explanation of selectivity based on secondary interactions which is complementary to the historical orbital approach.

1 Introduction

As defined by Linus Pauling, “Chemistry is the science of substances: their structure, their properties, and the reactions that change them into other substances”. [1] The first aspects, structure and properties, are clearly associated with the arrangement of atoms in a molecule, i.e. the chemical bond. These bonds determine Pauling’s third aspect, chemical reactivity. In other words, chemical bonds are the undisputed foundation of chemistry and their visualization should allow chemists to understand how molecules behave at the most fundamental level.

Achievement of a mechanistic understanding of chemical and biological functions as well as the structure of solid materials depends on knowing the geometric structure and the nature of bonds. But, despite the ”chemical bond being a fundamental concept in chemistry, **“what is a chemical bond?” still remains a critical question** for the chemical community because of the lack of a unique definition as well as an unclear understanding of its physical nature.

Successful numerical solution of the Schrödinger equation has yielded energies and properties of atoms and molecules, but not yet a clear physical explanation of chemical bonding. There is even a controversy on the mechanistic origin of the most “simple” chemical bond, covalent bonding, as it was remarked by Burdett in his classical book. [2]

As recently as in 2007, a special issue was devoted to the “90 Years of Chemical Bonding”. [3] In this issue, the chemical bond was compared to a unicorn, “a mythical but useful creature, which brings law and order [· · ·] in an otherwise chaotic and disordered world”. [4] Everyone knows how it looks despite nobody ever having seen one. [4–6] This line of reasoning is similar to Coulson’s comment: “Sometimes it seems to me that a bond between two atoms has become so real, so tangible, so friendly, that I can almost see it. Then I awake with a little shock, for a chemical bond is not a real thing. It does not exist. No one has ever seen one. No one ever can. It is a figment of our own imagination.” [7] Even chemical bonds have been described as “noumena” rather than “phenomena”. [8–10]

Chemical bonds together with other concepts such as atomic orbitals, electron shells, lone pairs, aromaticity, atomic charges, (hyper-) conjugation, strain, etc. do not correspond to physical observables. Such concepts therefore cannot be unambiguously defined in pure quantum theory, but constitute a rich set of “fuzzy”, yet invaluably useful concepts. [11–14] They lead to constructive ideas and developments when appropriately used and defined.

In chemistry as well as in physics, advanced theories are held by two milestones: i) a mathematical structure/formalism disclosing the basic entities of the theory and their mathematical relationships, and ii) an “interpretative” recipe of basic entities of the theory. The latter discloses the qualitative meaning of the basic entities and their relation to other known entities within and beyond the theory. It is important to highlight that the connection between the mathematical formalism and its interpretation is always subtle. This problem can be traced back to the lack of a clear and unambiguous definition of a bond in quantum mechanics and the plethora of inter-

pretations that have been introduced with various “meanings” of the “mathematical symbols/entities” of the theory. [15]

In front of this quandary, two opposite attitudes can be envisaged. On the one hand, the old and negative statement of the French mathemathician R. Thom: “Il me faut cependant avouer que la chimie proprement dite ne m’a jamais beaucoup intéressée. Pourquoi? Peut-être parce que des notions telles que celles de valence, de liaison chimique etc., m’ont toujours semblé peu claires du point de vue conceptuel.” (I should admit that chemistry never really interested me. Why? Perhaps because notions such as those of valence, chemical bond, etc., always appeared unclear to me from the conceptual point of view). On the other hand, the more actual and pragmatic comment of Alvarez et al.: “Chemistry has done more than well in a universe of structure and function on the molecular level with just this imperfectly defined concept of a chemical bond. Or maybe it has done so well precisely because the concept is flexible and fuzzy”. [16]

However, it is important to note that scientific arguments, debates, and controversies are at the heart of chemistry. This situation has been clearly stated in the very recent paper entitled “The Nature of the Fourth Bond in the Ground State of C_2: The Quadruple Bond Conundrum” by Danovich et al., [17] in which these authors recongnize that they are in front of a “Rashomon effect”, in which the bonding picture is becoming too fuzzy to be constructive anymore.

In trying to overcome this dichotomy, topological analysis has become one of the most useful tools to characterize interactions. Visualization of bonding interactions between atoms and molecules is a long-standing quest in theoretical and computational chemistry. The main interest lies in creating a tool that enables not only to see the interaction, but also to interpret its character and properties. Different types of bonding can be revealed by various topological methods, each based on different scalar fields. [18] This chapter deals with a new interpretative tool, NCI (standing for Non-Covalent Interactions), for revealing non-covalent interactions, which tries to brind some intuitive order into this fuzzy set of ideas.

1.1 Historical framework

Over the years, different approaches have been developed to reveal chemical bonds. Covalent bonds are intuitively represented using conventional Lewis structures. [19] Molecular Orbital (MO) theory has been very useful and successful for the theoretical analysis of chemical reactions and chemical reactivity. The frontier orbital theory [20] and the orbital symmetry rules of Woodward and Hoffman [21] are paradigmatic examples of the possibilities of quantum chemistry within the MO theory.

To reduce the dimensionality of the problem, three-dimensional interpretative approaches have been introduced. The conceptual density functional theory pioneered by Parr et al. [22] has been at the origin of very useful reactivity descriptors. Another low dimensional approach has originally been developed by Bader: [23–25]

the topological approach. Within these approaches, 3D space is divided following the gradient of a scalar function into mutually disjointed regions. Bader's QTAIM (Quantum Theory of Atoms In Molecules) theory is based on the topological analysis of the density and provides an atomic picture of the system. Chemical bonds and the underlying molecular graph may be traced by the analysis of its bond critical points (BCPs, first order saddle points). Following the same philosophy, ELF [26,27] (Electron Localization Function) topological analysis divides the space into chemically intuitive regions associated with electron pairs so that electronic shells, bonds, and lone pairs are revealed. Also, purely electrostatic interactions can be analyzed using electrostatic potential maps. [28]

1.2 Weak interactions

Chemical interactions between a protein and a drug, or a catalyst and its substrate, self-assembly of nanomaterials, [29, 30], and even some chemical reactions, [31, 32] are dominated by non-covalent interactions. This class of interactions spans a wide range of binding energies, and encompasses hydrogen bonding, dipole-dipole interactions and London dispersion [33] as well as more up to date interactions such as halogen bonds, CH$\cdots\pi$ and $\pi\cdots\pi$ interactions. Repulsive interactions, also known as steric clashes, should not be disregarded either.

More specifically, non-covalent interactions are of paramount importance in chemistry and especially in bio-disciplines, [34, 35] since they set up the force field scenario through which chemical species interact with each other without a significant electron sharing between them. They represent, in fact, the machinery through which molecules recognize themselves and establish how molecules will approach and eventually pack together.

During the last decade, non-covalent interactions have also raised a great deal of interest in the context of self-assembly [36] and crystallization, [37] whose underlying general rules are at the moment too faraway to be fully rationalized and understood. [38] Knowledge of such rules would in principle allow to build from scratch (even complex) materials exhibiting the desired properties. [29, 39, 40] Although it can not be ignored that a given observed structure is generally the outcome of a "drawing" among a plethora of energetically similar, but structurally dissimilar options, [41] understanding intermolecular non-covalent interactions and their mutual interplay in the supramolecular assemblies is nonetheless a fundamental step in making progress in structural prediction and evolution.

1.2.1 Weak interactions: the need for a new approach

Non-covalent interactions are frequently visualized using distance-dependent contacts, generally without consideration of hydrogen atoms. [42–44] Hydrogen-bonds can be identified from the molecular geometry [45] and from ELF, [46] while grid-

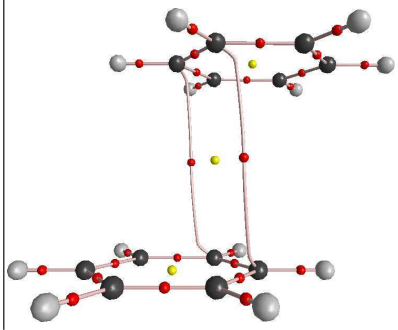
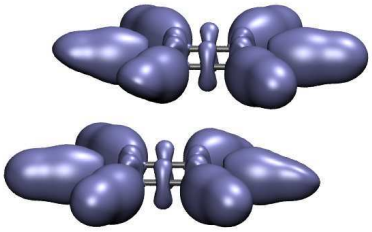
based calculations from classical force fields are used to model other van der Waals interactions. [47] The crucial role of weak interactions can also be analyzed in an indirect manner through property computations (from population to electrostatic moments). [48] However, these fluctuations are not easily visualized. In other words, a visual quantum chemical approach was conspicuously missing in this scenario.

Let's look at an example to clearly pinpoint the state of the art of topological approaches to weak interactions. The image provided by QTAIM and ELF of benzene dimer is provided in Table 1. Let's first focus on a given benzene molecule: The electron density shows maxima (cusps) for the C and H atoms, whereas C-H and C-C bonds are represented by BCPs (in red). ELF, instead, provides a picture based on electron localization, so that isosurfaces appear around atoms and C-H bonds. In both cases the chemical structure is revealed as expected from chemical intuition. However, the ELF picture is obviously more intuitive thanks to the isosurfaces.

Now, if we take a look at the complete system, the stacking dimer, we can see that intermolecular BCPs appear along with a ring critical point (second order saddle point, in yellow). The fact that interactions are related to saddle points, locates them on the interatomic surfaces, so that they highlight interatomic contact, but they do not have an associated region within this approach. Moreover, the critical points unite pairs of C atoms, which is not the chemical picture we have of a stacking interaction: it should appear as a benzene to benzene interaction. The VSEPR regions of benzene dimer are clearly identified by ELF, but nothing is seen for the inter-benzene stacking interaction.

In other words, both approaches fail to correctly provide a picture of delocalized interactions. Thus, it is the aim of this chapter to introduce such a tool, show its advantages over previous theories and its ability to provide a complete and holistic vision of non-covalent interactions and their change directly from the electron density.

Table 1 Comparison between QTAIM and ELF topologies in benzene dimer

Method	**QTAIM**	**ELF**
Function	Electron density	Pauli kinetic energy density
Partition	Atoms	Lewis' pairs
Example		 ELF=0.9

2 NCI: Non-Covalent Interaction index

The electron density has a fundamental advantage over MO-based descriptors because it is an experimentally accessible scalar field and it is a local function defined within the exact many-body theory, also supported by the Hohenberg-Kohn theorem. [49] The relationship between electron density topology and physical/chemical properties can be understood from the Hohenberg-Kohn theorem, which asserts that a system's ground-state properties are a consequence of its electron density. Furthermore since chemical reactions proceed by $\rho(\mathbf{r})$ redistributions, methods that analyze $\rho(\mathbf{r})$ distributions should help to understand the electron structure of molecules and thus chemical reactivity (see Section 8).

Our approach, introduced in the coming sections, uses the density and its derivatives, allowing simultaneous analysis and visualization of all non-covalent interaction types as real-space surfaces, thus adding an important tool to the chemist's arsenal. [50–52]

2.1 The reduced density gradient

The reduced density gradient, *s* or RDG, is a fundamental dimensionless quantity in DFT used to describe the deviation from a homogeneous electron distribution. [49, 53, 54] Properties of the reduced gradient have been investigated in depth in the process of developing increasingly accurate functionals. [55]

The origin of the reduced density gradient can be traced to the generalized gradient contribution to the GGA exchange energy, E_X^{GGA}, from density-functional theory:

$$E_X^{GGA} - E_X^{LDA} = -\sum \int F(s)\rho^{4/3}(\mathbf{r})d\mathbf{r}, \tag{1}$$

where $F(s)$ is a function of *s* for a given spin with

$$s = \frac{1}{C_F}\frac{|\nabla\rho|}{\rho^{4/3}}, \tag{2}$$

c_F being the Fermi constant, $c_F = 2(3\pi^2)^{1/3}$ and the $4/3$ exponent of the density ensuring that *s* is a dimensionless quantity.

s accounts for local density inhomogeneities due to its differential behavior depending on the chemical region of the molecule. The reduced density gradient assumes large values in the exponentially-decaying density tails far from the nuclei, where the density denominator approaches zero more rapidly than the gradient numerator. Small values of *s* occur close to the nuclei, due to the combination of large densities and small density gradients. The lower bound on the reduced density gradient is zero, as occurs throughout a homogeneous electron gas and at bond critical points.

The effect of bonding on the reduced density gradient is especially easy to visualize when s is plotted as a function of the density. Graphs of $s(\rho)$ assume the form $f(x) = ax^{-1/3}$, where a is a constant (see Appendix I). This can easily be proven from a STO model density. For a single atomic orbital $\psi = e^{-\alpha r}$, the density is $\rho = e^{-2\alpha r}$ and the gradient is $\nabla\rho = -2\alpha\rho$, such that

$$s(\rho) = \frac{1}{C_F}\frac{2\alpha\rho}{\rho^{4/3}} = \frac{2\alpha}{C_F}\rho^{-1/3}. \tag{3}$$

When there is overlap between atomic orbitals, a spike in the $s(\rho)$ diagram appears (Figure 1). The points forming this spike can identify the interaction when they are mapped back to real space. This procedure is able to reveal non-covalent interactions.

To explore the features associated with small reduced gradients, we first examine plots of s versus ρ (Figure 1). These plots were generated by evaluating the B3LYP [56, 57] density and reduced gradients on cuboid grids, with a 0.1 au step size, for each molecule or dimer. To provide even more sampling of the small low-density, low-gradient regions in hydrogen-bonded complexes, additional calculations were performed for water and formic acid dimers with a much denser 0.025 au grid.

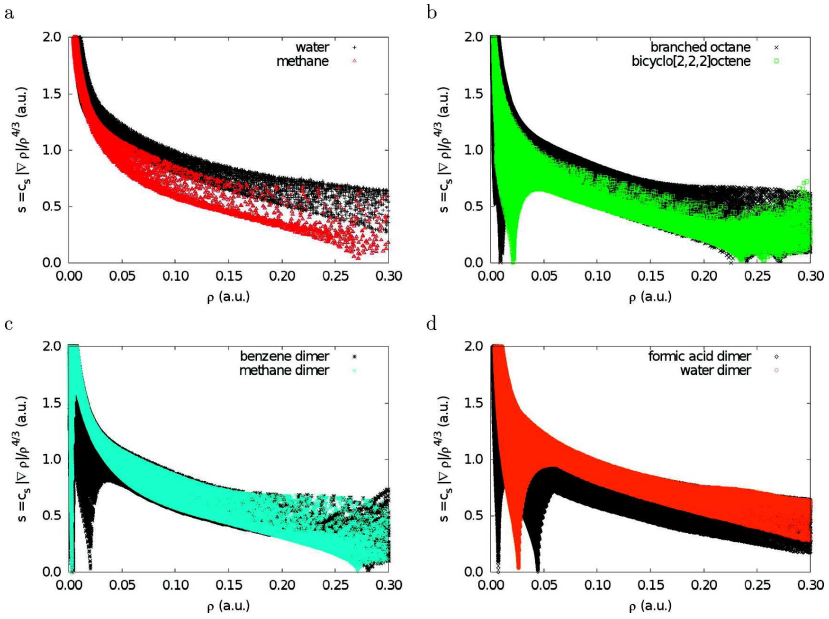


Fig. 1 Plots of the electron density and its reduced gradient for methane, water, branched octane, bicyclo [2,2,2]octene, and the homomolecular dimers of methane, benzene, water, and formic acid. The data was obtained by evaluating B3LYP/6-31G* density and gradient values on cuboid grids. Reprinted with permission from Ref. [50]. Copyright 2010 American Chemical Society.

Plotting s versus ρ, as in Figure 1, reveals the basic pattern of intramolecular interactions. Methane (Figure 1a) illustrates the typical covalent bond pattern. The top left-side points (small density and large reduced gradient) correspond to the exponentially-decaying tail regions of the density, far from the nuclei. The points on the bottom right side (density values of ca. 0.25 au and low reduced gradient) correspond to the C-H covalent bonds. Covalent bonds have their characteristic BCP in the electron density, corresponding to $s = 0$. Regions near the nuclei have larger density values and appear beyond the right edge of the plot. The plot has an overall shape of the form $a\rho^{-1/3}$ because atomic and molecular densities are piecewise exponential. The results for water are very similar, the only difference being that the covalent bonds lie at higher density values, past the edge of the plot. In Figure 1b-d, we consider six examples of chemical systems displaying various types of non-covalent interactions. Plots of s versus ρ for these systems all exhibit a new feature: one or more spikes in the low-density, low-gradient region, a signature of non-covalent interactions. This is the basis of NCI.

2.2 The density second eigenvalue

According to the divergence theorem, [58] the sign of the Laplacian of the density, $\nabla^2\rho$, indicates whether the net gradient flux is entering, $\nabla^2\rho < 0$, or leaving, $\nabla^2\rho > 0$, an infinitesimal volume around a reference point. Hence, it highlights whether the density is concentrated or depleted at that point, relative to the surrounding environment. To differentiate between different types of weak interactions one cannot resort to the sign of the Laplacian itself (as is common within AIM theory) since it is dominated by the principle axis of variation and is positive for all closed-shell interactions. [59]

To understand bonding in more detail, it is often useful to decompose the Laplacian into the contributions along the three particular axes of maximal variation. These components are the three eigenvalues λ_i of the electron-density Hessian matrix, such that, $\nabla^2\rho = \lambda_1 + \lambda_2 + \lambda_3$, ($\lambda_1 < \lambda_2 < \lambda_3$). At points with zero gradient, analysis of the Hessian eigenvalues is analogous to determining the signature of the critical point. Thus, at nuclei (cusps/maxima of ρ), all the eigenvalues are negative, while at the center of cages or holes (minima of ρ) all the eigenvalues are positive. In the remaining points of space $\lambda_3 > 0$, $\lambda_1 < 0$, and λ_2 can be either positive or negative. Within the NCI framework, the sign of λ_2 (i.e. the perpendicular plane) enables identification of the interaction type. Attractive interactions appear at $\lambda_2 < 0$ whereas in the cases where λ_2 is positive (as in rings or cages), usually several atoms interact but are not bonded, which corresponds to steric crowding according to classical chemistry.

Both van der Waals interactions and hydrogen bonds show negatives value of λ_2 at the critical point (with $\lambda_2 \simeq 0$ for van der Waals interactions). This can be attributed to the homomorphic virial path associated with the bonding direction, which defines a line along which the potential-energy density is maximally negative. Con-

versely, non-bonding interactions, such as steric crowding, result in density depletion, such that $\lambda_2 > 0$. Analogously, the homeomophism ensures that these critical points (both ring and cages points) identify lines of minimally-negative potential-energy density.

3 Interpreting NCI

3.1 The 2D plot

Once the second eigenvalue has been introduced to separate attractive from repulsive interactions, it is necessary to categorize the interactions by their strength. This can be done thanks to the properties of the corresponding critical points. Characteristic densities of van der Waals interactions are much smaller than densities at which hydrogen bonds appear. However, steric clashes and hydrogen bonds span similar density ranges and overlap in plots of $s(\rho)$. This is illustrated for the phenol dimer in Figures 2a-b. This is a hydrogen-bonded complex that exhibits non-bonding interactions within each benzene ring and a stacking interaction between the benzene rings. We thus have the three main types of interactions: Van der Waals, HB and steric clashes. Whereas van der Waals is well differenciated because it appears at smaller densities, the steric clash and the hydrogen bond spikes overlap in Figure 2a.

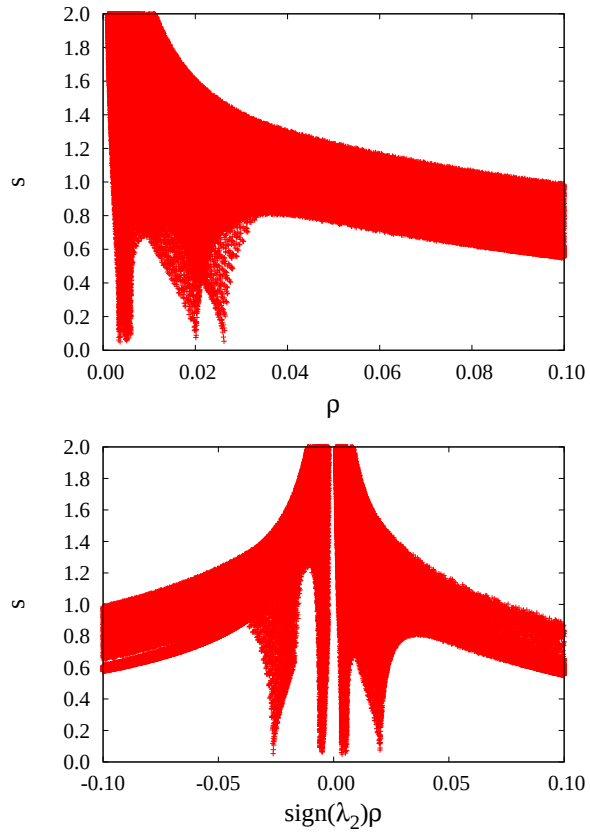
The interaction type can be distinguished if the $s(\rho)$ diagrams are modified by plotting $\text{sign}(\lambda_2)\rho$ as the ordinate. Analysis of the sign of λ_2 thus helps to discern between the different types of weak interactions, whereas the density itself provides information about their strength; both are combined in the value of $\text{sign}(\lambda_2)\rho$. When the Hessian eigenvalues are considered, the different nature of these interactions is made clear: the benzene-ring interactions remain at positive values, whereas the hydrogen bond now lies at negative values, within the attractive regime. The NCI spikes nearest zero density correspond to weakly-attractive dispersion interactions between the phenyl rings.

3.2 3D plot

3.2.1 3D visualization

The 3D spatial visualisation of the non-covalent interactions as defined above is done using the data from the 2D plots as input to construct 3D plots composed of reduced density gradient isosurfaces.

In a nutshell, a cut-off value of s close to zero, typically $s < 0.5$, is chosen in order to recover all the non-covalent interactions in the system, i.e. all the spikes



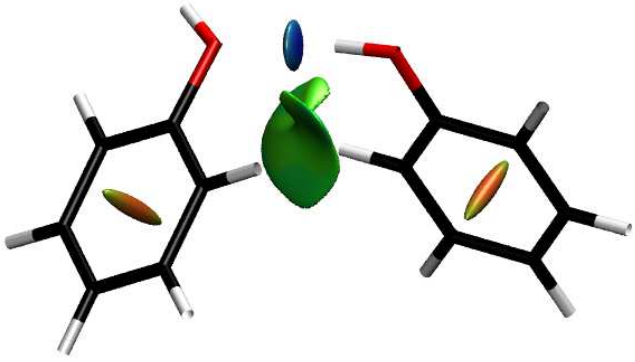


Fig. 2 Plots of (a) $s(\rho)$, (b) NCI isosurface, for the phenol dimer. The $s = 0.6$ isosurface is colored over the range -0.03 $<$ sign(λ_2)ρ $<$ +0.03 au.

of the 2D plots. The corresponding reduced density gradient isosurfaces give rise to closed domains in the molecular space which highlight the spatial localization of the interactions within the system (see Figure 2c). Since 3D isosurfaces are, by definition, regions of low reduced gradient, the density is nearly constant within these.

At this stage, however, the types of interaction corresponding to the several isosurfaces are not apparent. In order to discriminate between them, the density oriented by the sign of λ_2 is further used (as in the 2D plot). A RGB (red-green-blue) coloring scheme is chosen to rank interactions, where red is used for destabilizing interactions, blue for stabilizing interactions and green for delocalized weak interactions. The intensity of these colors (i.e. the deepness of the color) is associated with a higher local density and therefore to a stronger interaction.

The isovalue (or cut-off) of s(ρ) chosen for plotting the 3D isosurface determines which features will appear in the NCI plot as well as their spatial extension. On the one hand, all NCI spikes do not strictly achieve s=0, so that too low a value might miss some of the interactions of interest. [60] On the other hand, too high a value would disclose atomic tails of the density. [61] The cut-off is therefore chosen from the 2D plot so that all spikes, but only spikes, are captured to render a meaningful picture which recovers both attractive and repulsive interactions.

3.2.2 2D and 3D shapes

Topological features of the electron density are very stable with respect to the calculation method. The main effect of different methods on the s vs. ρ diagram is a shift of spikes. The only rule of thumb seems that the same s value should be used when comparing to each other the various NCI, both in the same or in different systems, provided a single method was employed to obtain the various electron densities. This not being the case, different s values are seemingly required to compare on similar grounds the s-based results for differently computed electron density (e.g. from wavefunction, multipolar). In other words, a shift of cut-offs is needed to obtain comparable images This can be related to the fact that s roughly behaves like $\rho^{-1/3}$ (see Eq. 3), so that the effect of the method on the density is directly followed by the s-value. This information is crucial when moving towards bigger systems (see Section 6).

It emerges that a one-to-one inverse correlation seems to exist among the directionality (and the strength) of specific non-covalent interactions and the surface/volume ratio of the corresponding s isosurface. In particular, the stronger the interaction, the smaller and more disc-shaped the s surface appears in real space (and the more negative the ρsign(λ_2) values are).

In some cases, noticeably in ring closings, bicolored isosurfaces appear (see Figure 3). They result from stabilizing features (revealed by the blue color), counterbalanced by destabilising interactions due to steric crowding (revealed by the red color), such as ring closure. [62]

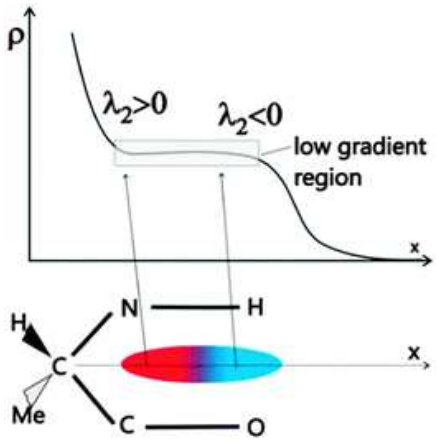


Fig. 3 Scheme of a mixed color isosurface in a C5 ring formed in the groove between a carbonyl and a NH group within a peptide residue. Within this isosurface, the curvature varies sign, leading to two well differenciated parts. A blue external part which illustrates the directional NH···O interaction and a red part which is indicative of the strain in the 5-membered ring resulting from a multicentric density in the inner region close to the backbone. In the top panel the variation of the density and the sign of λ_2 is depicted. Reproduced from Ref. [62] with permission from the PCCP Owner Societies.

4 Small molecules

To explore the features associated with small reduced gradients, we first examine 3D NCI plots in representative small molecules.

Figure 4 displays NCI isosurfaces for branched octane, bicyclo[2,2,2]octene, and the homomolecular dimers of methane, benzene, water, and formic acid. These isosurfaces provide a rich visualization of non-covalent interactions as broad areas of real space, rather than simple pair-wise contacts between atoms. We first consider the sterically-crowded molecules bicyclo[2,2,2]octene (Figure 4a) and the branched octane isomer (Figure 4b). In the first case, the low-density, low-gradient region corresponds to the center of the cage, where steric repulsion between the bridgehead carbons is expected. For the branched octane isomer, the isosurface lies between the closely-interacting methyl groups on opposite sides of the central C-C bond. The interactions are repulsive nearer the C-C bond and weakly attractive between the hydrogen atoms. Dispersion and hydrogen bonding can also be clearly detected. In the dispersion-bound methane dimer (Figure 4d), the isosurface forms a disc between the individual monomers. For the water dimer (Figure 4e), the isosurface lies between a hydrogen donor and oxygen acceptor, characteristic of hydrogen bonds. The formic acid dimer (Figure 4f) reveals stronger HBs than in the water dimer, and also weak van der Waals interactions between the two closely-interacting acidic hydrogens.

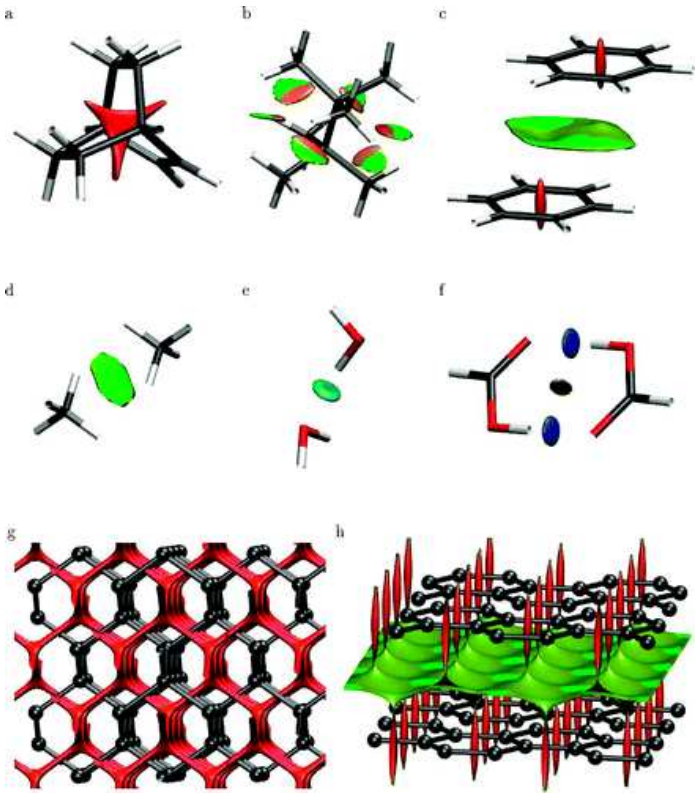


Fig. 4 NCI isosurfaces ($s = 0.5$) for a) branched octane, b) bicyclo[2,2,2]octene, and the homomolecular dimers of c) benzene, d) methane, e) water, and f) formic acid. Gradient isosurfaces are also shown for cuboid sections of g) diamond and h) graphite. Color coding in the range -0.04 $<$ sign(λ_2)ρ $<$ +0.02 au. Reprinted with permission from Ref. [50]. Copyright 2010 American Chemical Society.

Finally, in the benzene dimer (Figure 4c), there is an area of non-bonded overlap located at the center of each benzene ring, resembling the isosurface for bicyclo[2,2,2]octene. There is another lower-density surface between the overlapping portions of the benzenes, where π-stacking is expected. It is important to compare this image to the ones displayed in Table 1. The intermolecular interaction in benzene dimer appears very clearly with NCI as a surface that highlights the benzene to benzene stacking interaction, well beyond the pair interactions found with AIM.

5 Solids

As we saw in the Introduction, non-covalent interactions in solid state have raised a great deal of interest lately, due to their relevance for self-assembly [36] and crystallization [37] processes.

Indeed, crystalline solids exhibit rich and challenging bonding patterns. We consider the prototypical examples of carbon in the diamond (Figure 4g) and graphite (Figure 4h) phases at their equilibrium geometries. In diamond, the carbon atoms are sp^3 hybridized and are connected by strong covalent bonds that form a tridimensional, tetrahedral network. Figure 4g shows a NCI isosurface for a cuboid section of the diamond crystal. The non-covalent surface extends through the voids of the structure, creating a network similar to that of the covalent bonds.

Graphite in its α form (Figure 4h) has a bidimensional, hexagonal lattice, with the carbon atoms sp^2 hybridized and covalently bonded to their three nearest neighbors. The NCI isosurface shows areas of nonbonded overlap at the center of the hexagonal rings, as seen previously in benzene. $\pi - \pi$ stacking interactions between the graphene sheets are clearly manifested by the isosurfaces filling the interlayer spaces.

6 Biological Systems

Understanding of non-covalent interactions is crucial for the comprehension of the 3D structure and, thus, of the activity of biosystems. [65, 66] However, the calculation of the electron density in these systems is totally unbearable. Approximations need thus to be sought.

6.1 Promolecular densities

Densities are stable (see Section 3.2.2) to such an extent that NCI characteristics are already contained in the sum of atomic densities, ρ_i^{at}. [63, 64] The resulting molecular density, also known as promolecular density, ρ^{pro}, is then given by:

$$\rho^{pro} = \sum_i \rho_i^{at} \tag{4}$$

A promolecular density obtained from simple exponential atomic pieces is able to predict low-density, low-reduced-gradient regions similar to density-functional results. The free atomic densities used in these calculations consist of one Slater-type function for each electron shell, fit to closely reproduce spherically-averaged, density-functional atomic densities (see Appendix I).

Approximate promolecular densities were constructed by summing exponential atomic densities for bicyclo[2,2,2]octene, and the homomolecular dimers of methane and water.

Resultant plots of s^{pro} versus ρ^{pro} for these species show the same 2D features seen in Figure 1. Also, 3D isosurfaces generated from the promolecular density (Figure 5) are very similar to those obtained previously with self-consistent DFT and even MP2 densities (Figure 4). For all cases considered, results at the self-consistent and promolecular level are qualitatively equivalent. Only slight quantitative differences are introduced by density relaxation that, as expected, shift the s versus ρ spikes to more bonding regimes. Specifically, a large shift toward smaller density values is observed in the spikes corresponding to non-bonded overlap, introducing less repulsion and greater stability. Taking this shift into account in the choice of isosurfaces, results at the self-consistent and promolecular level are qualitatively equivalent for all cases considered (see Figure 5 bottom). For example, lower cut-offs on the gradient (0.25-0.35) and higher cut-offs on the density (0.08-0.09 au) were required in order to generate the isosurface for bicyclo[2,2,2]octene.

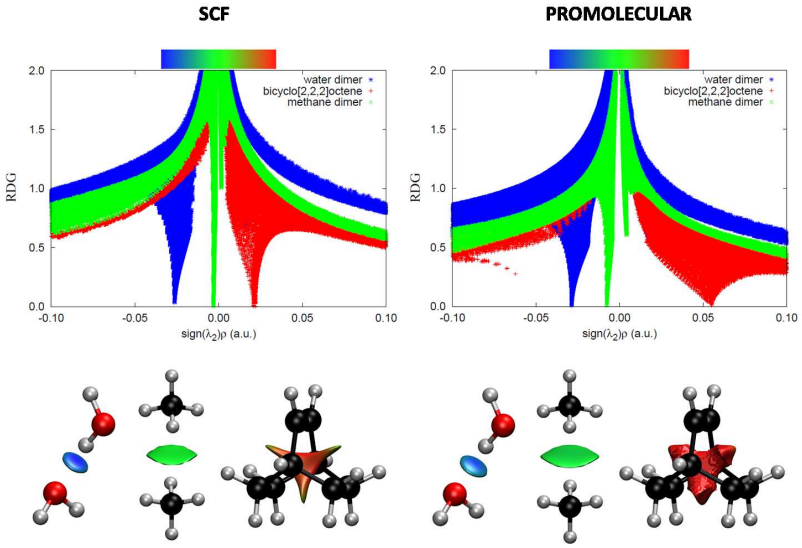


Fig. 5 Comparison between SCF and promolecular NCI results. The same $s(\rho)$ features are obtained using self-consistent (left) and promolecular (right) calculations, with a shift toward negative (stabilizing) regimes. Bottom: Taking the shift in spikes into account (i.e., changing the cut-off), the isosurface shapes remain qualitatively unaltered for selected small molecules. Figures are shown for both SCF (left) and promolecular densities (right). NCI surfaces correspond to s = 0.6 and a color scale of -0.03 $< \rho$ sign(λ_2) $<$ +0.03 au for SCF densities. For promolecular densities, s = 0.5 (water and methane dimers) or s = 0.35 (bicyclo[2,2,2]octene), and the color scale is -0.04 $< \rho$ sign(λ_2) $<$ +0.04 au. Reprinted with permission from Ref. Ref. [51]. Copyright 2011 American Chemical Society.

6.2 Examples

Promolecular densities obviously lack relaxation; however, the promolecular densities are extremely useful in biomolecular systems, such as proteins or DNA. Because the calculation of the electron density in these systems becomes extremely computationally expensive, the promolecular density becomes an attractive option: non-covalent interactions can be analyzed with only the molecular geometry required as input.

We first consider two model polypeptides: an α-helix consisting of 15 alanine residues and an anti-parallel β-sheet consisting of 17 glycine residues. Geometries of the polypeptides were obtained with the MMFF force field using the spartan program. [67] Both were capped with $COCH_3$ and $NHCH_3$ groups. Figure 6 displays low-gradient isosurfaces for cuboid regions at the center of these polypeptides, colored according to the corresponding density values. For the β-sheet, the lowest-density portions of the gradient isosurface arise from hydrophobic, dispersion-dominated interactions, primarily involving the CH_2 groups of the glycines. The higher-density regions correspond to inter-residue hydrogen-bonds and repulsive interactions between the adjacent C=O and N-H groups. For the α-helix, the isosurface has a large, low-density region within the helix and between the side-chain methyl groups. The higher-density portions of the isosurface correspond to inter-residue hydrogen-bonds along the helix and repulsive interactions between adjacent N-H groups.

We also considered the non-covalent interactions between nucleobases in the B-form of double-strand, six-base-pair (TGTGTG) DNA. The structure was obtained using the X3DNA program [68] with ideal geometry parameters. [69] Figure 6c displays the low-gradient isosurface for a cuboid section in the center of the DNA helix, colored according to the $\text{sign}(\lambda_2)\rho$ values. The calculated isosurface resembles that of graphite, with broad, low-density regions indicative of π-stacking between base-steps. The interactions between individual deoxyadenosine-deoxythymidine and deoxycytidine-deoxyguanosine pairs are shown in Figure 6d,e. The isosurfaces show non-bonded overlap within the nucleobase rings, as in benzene and graphite, and hydrogen-bonding motifs similar to the formic acid dimer. The strong N-H$\cdots$O and N-H$\cdots$N hydrogen bonds can be clearly distinguished from the weaker, attractive C-H$\cdots$O interaction by the density values, as shown in different colors.

The hydrogen-bonding surfaces in the DNA model have density values of ca. 0.065 au, compared to density values of ca. 0.035 au for the polypeptide hydrogen bonds. This is evident from the degree of blue shading for the hydrogen bonds in Figure 6. Since density values at hydrogen-bond critical points correlate with the interaction strength, [25, 70] our results indicate that the hydrogen bonds between nucleobase pairs are substantially stronger than between amino acids, in agreement with literature data. [71–73]

Let us now consider the interaction between a ligand and a protein active site. The low-gradient isosurface for a tetracycline inhibitor bound to the tetR protein in Figure 7 shows a complex web of non-covalent interactions between the ligand and active site. When analyzing non-covalent interactions in protein-ligand complexes,

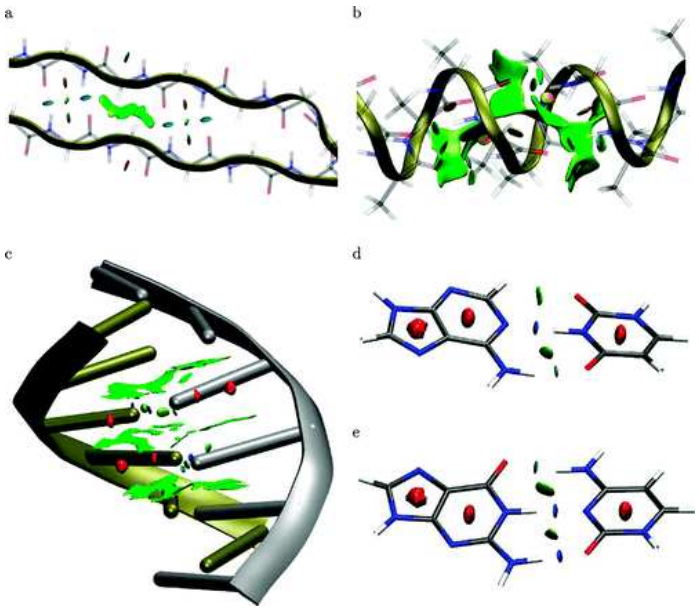


Fig. 6 Gradient isosurfaces ($s^{pro} = 0.35$) for cuboid sections of the (a) β-sheet and (b) α-helix polypeptides. Gradient isosurfaces ($s^{pro} = 0.25$) are also shown for the (c) B-form of DNA, and the (d) A-T and (e) C-G base pairs. The surfaces are colored according to values of sign(λ_2)ρ, ranging from -0.06 to +0.05 au. Reprinted with permission from Ref. [50]. Copyright 2010 American Chemical Society.

it is usually assumed that these interactions are due to a specific contact between two atoms. [44] However, it is clearly seen in Figure 7 that this assumption is only partly correct. Hydrogen bonds, such as those between the tetracycline amine groups and two water molecules (shown in orange), are directional and specific. Conversely, van der Waals, dipole-dipole, and hydrophobic interactions, such as those between the tetracycline and the Leu61, Val91, Ile136, and Val166 residues (shown in yellow), are not atom-specific and occupy broader regions in space. The figure reveals some steric clashes (orange and red regions of the isosurface) that must be offset by stronger, attractive interactions to give binding in this crystal structure. A ligand "fits" the geometry of the active site, and the interaction energy between the ligand and protein is comprised of many small contributions. When trying to design a new ligand to fit a specific active site, one should consider all such interactions.

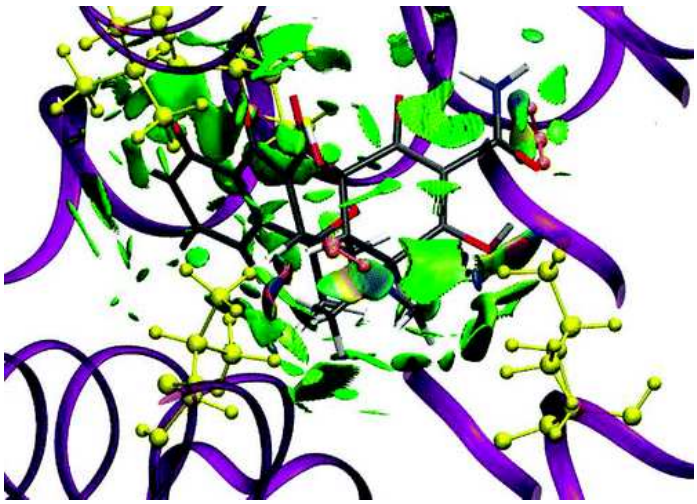


Fig. 7 Gradient isosurfaces (s_{pro} = 0.35) for interaction between the tetR protein and tetracycline inhibitor. Surfaces colored in the sign(λ_2)ρ range from -0.06 to +0.05 au. Reprinted with permission from Ref. [50]. Copyright 2010 American Chemical Society.

7 New bonds: do we really need to name them all?

Hydrogen bonding was postulated in the early 20^{th} century based on the stunning macroscopic differences between the first and second row hydrides, i.e. water is a high boiling liquid without, which there would be no life and hydrogen sulphide is a stinking gas, under ambient conditions. [1]

With the advent of molecular beam and cryogenic experimental methods as well as the ever advancing theoretical methods, HBs have been proved to exist in H_2S as well. [74] But things go even further. About a century later, chemists have identified a wealth of new bonding types along the periodic table. Halogen bonds (XBs) (group 17) are frequently exploited for crystal engineering. [75] Recently, similar bonding mechanisms have been proposed for adjacent main-group elements, and non-covalent "chalcogen bonds" (group 16) [76] and "pnictogen bonds" (group 15) [77] have also been identified in crystal structures. Recently, even carbon bonding (group 14) [78] has been proposed as a stabilizing interaction.

One of the most interesting features of NCI is that it is based on the density. Thus, it is expected to be able to reveal any type of bonding. Before and beyond its corresponding identification and characterization, the signature on the electron density will be present. In the coming sections we briefly show some representatives of the new bonds series, highlighting the possibility to reveal interactions all along the periodic table from just the fast analysis of NCI.

7.1 Halogen bonding

Halogen bonds (XBs) occur between a halogen atom, playing the role of Lewis acid, and a Lewis base. This non-covalent interaction is analog to hydrogen bonding in the sense that in both cases, an atom or group of atoms with high electron density donates charge to an acceptor which is electron poor. Similar to HBs, XBs are also anisotropic and involved in various fields, such as supramolecular chemistry or even materials chemistry. Moreover, halogenated coumpounds are often encountered in medicinal chemistry [79] and drug discovery. [80] Quantum Chemistry approaches have revealed a sigma-hole along the axis defined by the halogen atom and the acceptor. This can be studied thanks to the Molecular Electrostatic Potentials, showing the charge distribution as well as the nature of the interaction as electrostatics and charge transfer driven.

We have analyzed NCI in a series of trifluoromethyl halides, CF_3X, where X = Cl,Br,I with dimethyl ether (DME), dimethyl sulfide (DMS), trimethyl phosphine (TMP) and imidazolin-2-ylidene (NHC). All systems were optimized using second-order Møller-Plesset perturbation theory with the aug-cc-pVDZ(-PP) basis set. [81] The wavefunctions were obtained at DFT level using the B3LYP functional and the 6-31++G** basis set except for iodine, where the pseudopotential LANL2DZ was used.

In the first four cases, the 2D plots (Figure 8) clearly show a spike at vert low density, corresponding to a typical van der Waals interaction. In Figure 8b, a second spike arises, exhibiting non bonding interactions between the chlorine atom and DMS hydrogens. They are closer to the halogen than the sulfide atom and benefit from the high electron density at the halogen.

In agreement with previous studies, the prominent spike is shifted to larger electron densities (left) for iodine, reflecting a stronger non bonding interaction (see Figures 8e-f) as X molecular weight increases.

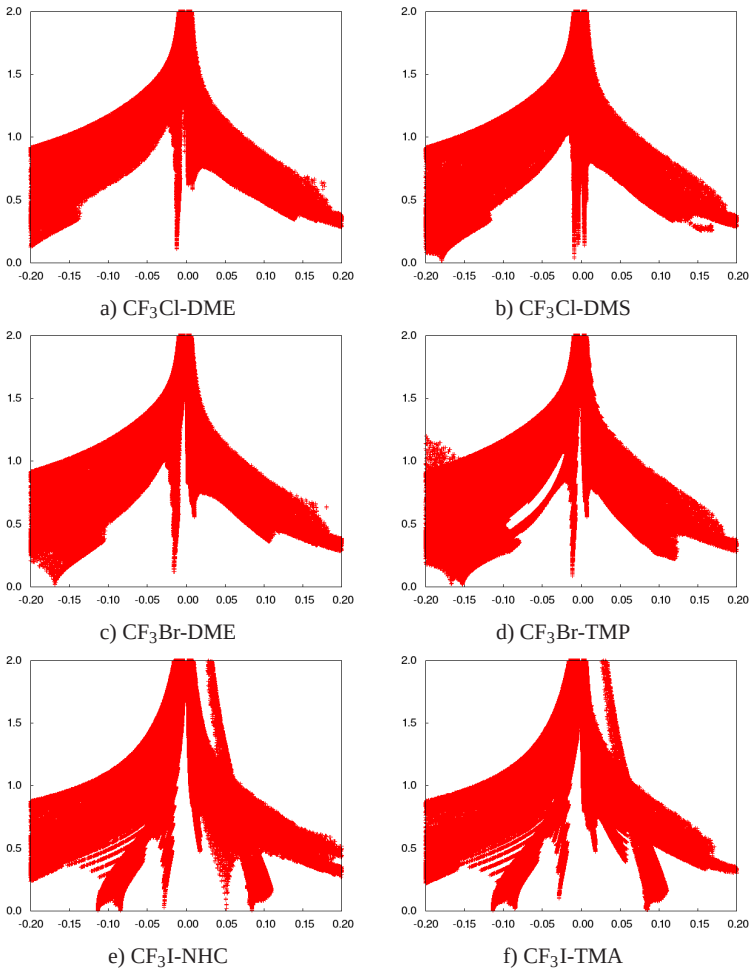
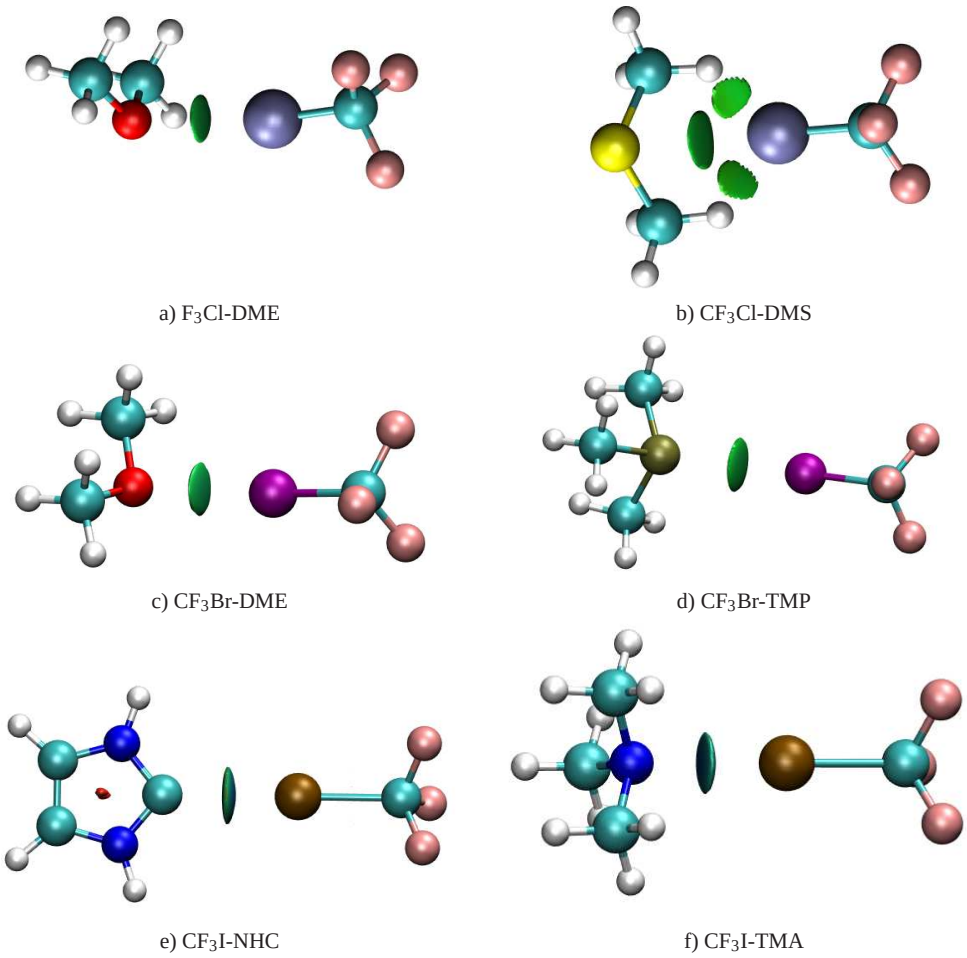


Fig. 8 2D plots for halogen bonded complexes: a) CF_3Cl-DME b) CF_3Cl-DMS c) CF_3Br-DME d) CF_3Br-TMP e) CF_3I-NHC f) CF_3I-TMA. See Section 7.5.2 for an interpretation of iodine compounds extra peaks.



a) F_3Cl-DME b) CF_3Cl-DMS

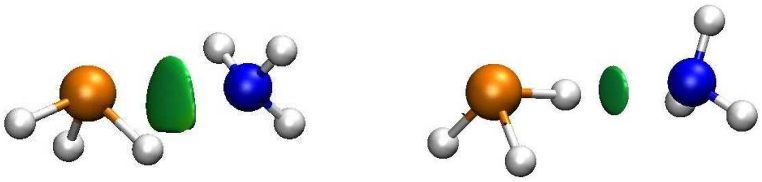
c) CF_3Br-DME d) CF_3Br-TMP

e) CF_3I-NHC f) CF_3I-TMA

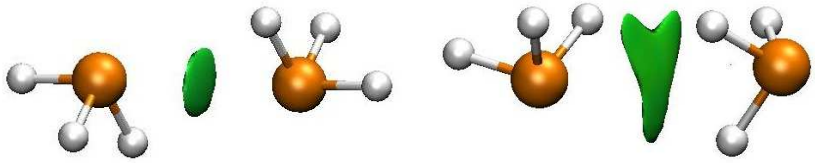
Fig. 9 3D plots for halogen bonded complexes: a) CF_3Cl-DME b) CF_3Cl-DMS c) CF_3Br-DME d) CF_3Br-TMP e) CF_3I-NHC f) CF_3I-TMA. NCI isosurfaces correspond to s=0.5 au and a colorscale of -0.04 $<$ sign(λ_2)ρ $<$ +0.04 au.

7.2 Pnictogen bonding

Pnictogen bonding is a weak non-covalent bonding involving group 15 elements as electron density acceptors. Similiar to halogens, the pnictogen atoms possess a sigma-hole: a region of positive electrostatic potential in the direction of the bond,



NH_3-PH_3



PH_3-PH_3

Fig. 10 Pnictogen bond examples: a-b) NH_3-PH_3, c-d) PH_3-PH_3. 3D isosurface was produced using the cut-off values of $s = 0.5$ and $\rho < 0.05$.

which is attracted to a lone pair on a nucleophile with an outer negative electrostatic potential. [82, 83]

Pnictogen bonding is present in the complex between NH_3 and PH_3 (Figure 10a) [77] where the N atom is the donor of electron density. The sigma-hole gives rise to the equilibrium geometries: the two molecules are oriented such that the P and N atoms face one another directly, without the intermediacy of a H atom. This attraction is due in part to the transfer of electron density from the lone pair of the N atom to the σ^* antibond of a P-H covalent bond. Unlike in hydrogen bonds, the pertinent hydrogen is oriented about 180° away from the N (instead of toward), and the N lone pair overlaps with the lobe of the P-H σ^* orbital that is closest to the P. The calculated binding energy of this pnictogen-bonded complex is more stable than the hydrogen-bonded complex that is formed between the same two molecules where the N atom is the proton acceptor (Figure 10b).

We have optimized the NH_3-PH_3 and PH_3-PH_3 pairs at the ωB97X-D/6-31+G* level of theory. The NCI analysis of the NH_3-PH_3 complexes shows the presence of non-covalent bonding and this is illustrated by the 3D isosurfaces of both complexes (Figure 10a-b). The green color of both interactions indicates that the two types of bonding have similar bonding strengths corresponding to that of van der Waals interactions. Whereas in the H-mediated complex, a typical picture of HB is obtained (Figure 10b), a thick surface is obtained in the case of the pnictogen bonding (Figure 10a), which is extended like in the case of van der Waals, but thick like HB ones.

It is interesting to note that in contrast to halogen bonds, there is no requirement of a sigma-hole of positive electrostatic potential on the P atom, nor is it necessary for the two interacting atoms to be of differing potential. In fact, the two atoms can be identical, as the global minimum of the PH_3 homodimer has the same structure,

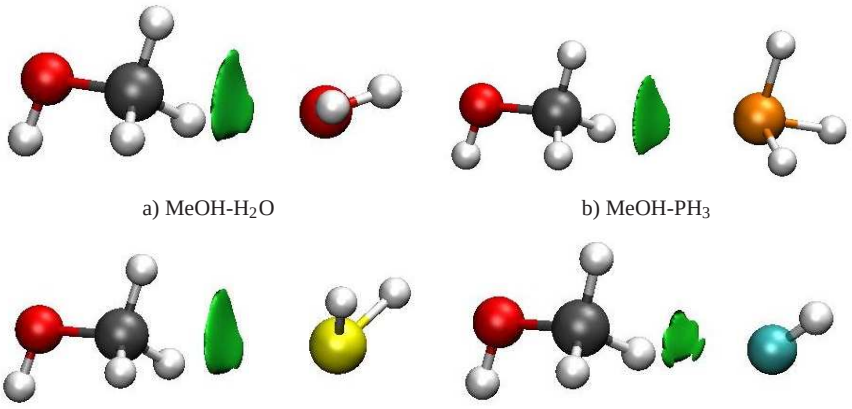
characterized by a P···P attraction. Indeed, for the complex between PH_3 and PH_3, the P atoms possess a partial positive charge and none of the located minima found on the potential energy surface correspond to a hydrogen-bonded complex. [77] The two minima that were located correspond to complexes where the P atoms approach one another (Figure 10c-d). The complex with the symmetric geometry (Figure 10c) was found to be dominated by electrostatic interactions, corresponding to pnictogen bonding, whereas the second structure (Figure 10d) was found to be dominated by dispersion. This shows in the NCI isosurfaces where the interaction region in Figure 10d occupies a larger volume than that of Figure 10c. This is in agreement with the more diffuse character of the dispersion interaction compared to the pnictogen bond which is more concentrated along the bonding direction.

7.3 Carbon bonding

Recently, inspired by the identification of halogen, chalcogen and pnictogen bonding, Mani et al. [78] investigated whether carbon, being a positive centre, can accept electron density. Indeed, both experimental and theoretical studies agree that the tetrahedral face of methane can act as a hydrogen bond acceptor. Rotational spectra of complexes like $CH_4\cdots HF$, $CH_4\cdots HCl$, $CH_4\cdots HCN$ and $CH_4\cdots H_2O$ further confirm this fact. [84] While the tetrahedral face of methane has an electron rich centre and can act as a hydrogen bond acceptor, substitution of one of its hydrogens with some electron withdrawing group (such as -F/OH) can make the opposite face electron deficient. The complex between CH_3F and H_2O has a potential energy minimum with water oxygen pointing towards the tetrahedral face of CH_3. Similar interactions are also found for several methanol complexes in which the electron deficient atom (oxygen) interacts with one of the water's lone pair.

Four examples of complexes, which represent minima on the potential energy surface, are shown in Figures 11a-d. All molecules and bimolecular complexes were optimized at the ωB97X-D/6-31+G* level of theory. In these complexes, the electron density donors (O, P, S and F atoms) are oriented towards the CH_3 face of methanol. Through NCI analysis, the presence of intermolecular interaction is evident in all the complexes. The weak nature of this type of interaction is indicated by the green color of the NCI-isosurfaces that corresponds to van der Waals interactions.

It should be noted that, in spite of the weakness of these type of interactions, they are extremely relevant, since these and similar interactions could give an enthalpic contribution to the so called "hydrophobic interactions". [78]



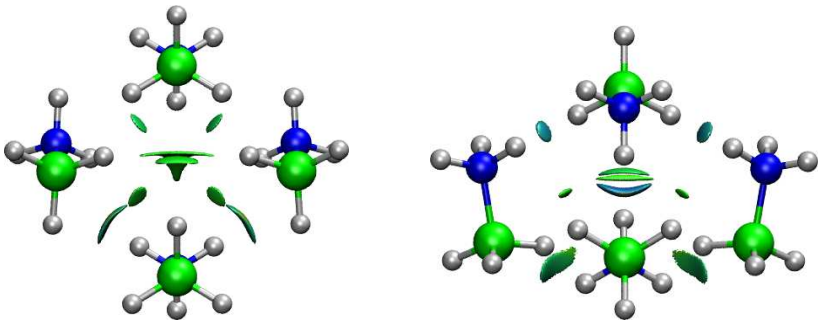
a) MeOH-H_2O b) MeOH-PH_3

c) MeOH-SH_2 d) MeOH-FH

Fig. 11 Carbon bond examples for several electron donors (O, P, S and F): a) MeOH-H_2O b) MeOH-PH_3 c) MeOH-SH_2 d) MeOH-FH. 3D isosurface was produced using the cut-off values of $s = 0.5$ and $\rho < 0.05$.

7.4 The di-hydrogen bond

Very recently the new term di-hydrogen bond was coined to describe an interaction of the type D-H···H-E, where D is a typical hydrogen donor such as N or O. The interesting thing about this type of bond, is that the acceptor atom is also a hydrogen.



a) Gaseous phase b) Crystal structure

Fig. 12 Dihydrogen interactions in a BH_3NH_3 tetramer in a) the fully optimized gas-phase geometry and b) the solid-state geometry. NCI surfaces correspond to $s = 0.4$ au and a color scale of -0.03 $<$ sign(λ_2)ρ $<$ +0.03 au. Reprinted with permission from Ref. [51]. Copyright 2011 American Chemical Society.

Thus, the accepting hydrogen atom must be negatively charged and E has to be an atom capable of accommodating a hydridic hydrogen. Transition metals and boron are some known examples of atoms occurring at position E. Within di-hydrogen bonded complexes, BH_3NH_3 is perhaps the most widely studied. [86–88] We have analyzed the tetramer $(BH_3NH_3)_4$, whose geometry has been derived from the solid state. Figure 12 shows the NCI results for $(BH_3NH_3)_4$. It can be seen that each BH_3NH_3 molecule interacts with the surrounding ones establishing one di-hydrogen bond with each, and numerous van der Waals contacts. The surfaces obtained are, in all cases, completely analogous to those obtained in previous examples for hydrogen bonds.

We have also studied a series of nine complexes presenting di-hydrogen bonds (LiH-HCCH, LiH-HCN, LiH-HCF_3, NaH-HCCH, NaH-HCN, NaH-HCF_3, HBeH-HCCH, HBeH-HCN, HBeH-HCF_3) to check the ability of NCI to detect new types of bonds even at the promolecular level. After using the MAPS Platform [89] to set up the initial systems, the complexes were optimized at the MP2/aug-cc-pVDZ level with NWChem. [90] The optimized coordinates were used to perform NCI promolecular analysis. The results are displayed in Figure 13.

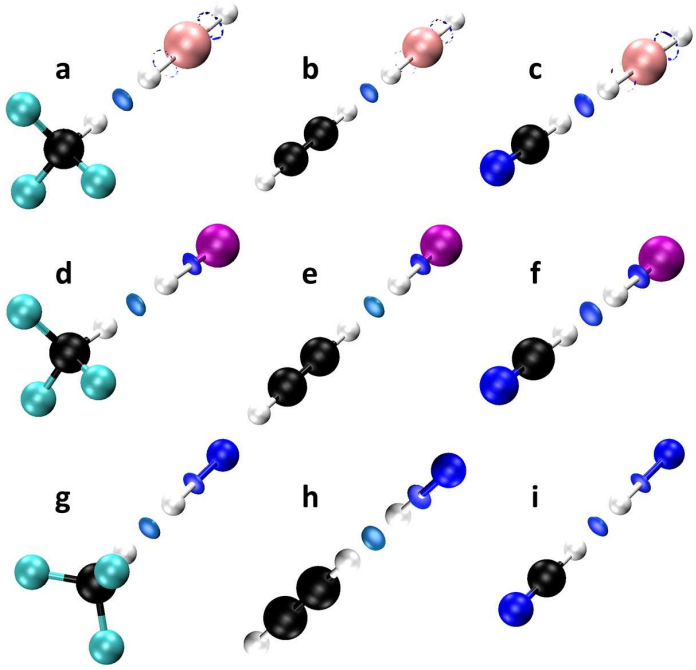


Fig. 13 NCI surfaces for several di-hydrogen complexes: a) HBeH-HCF_3 b) HBeH-HCCH c) HBeH-HCN d) LiH-HCF_3 e) LiH-HCCH f) LiH-HCN g) NaH-HCF_3 h) NaH-HCCH i) NaH-HCN. The NCI isosurfaces were plotted for s = 0.3 and a color scale of -0.03 $<$ sign(λ_2)ρ $<$ +0.03.

Even at this rough level, NCI allows to follow the evolution of the interaction strength for the different systems. The interaction basins appear similar to those from strong hydrogen bonds. They also seem to be stronger than the di-hydrogen bond in the BH_3NH_3 complex. It should be noted that this agrees with the fact that di-hydrogen bonds have been attributed very variable strengths. Crabtree *et al.* [91] have placed the NH···HB contact at the upper end of the energy range quoted for hydrogen bonds. Popelier instead, has found it to be in the range of normal HB strengths [86] and Morrison and Siddick [87] assigned it towards the lower end of the hydrogen bond strength spectrum. Our results show that the range of energies covered go (at least) from the strong to the medium HB-type of strength.

7.5 Metal driven interactions

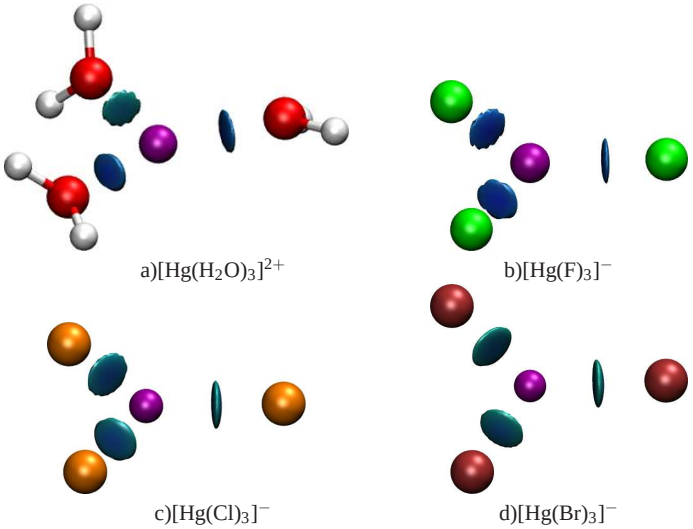


Fig. 14 Complexation of Hg. a)$[Hg(H_2O)_3]^{2+}$ b)$[Hg(F)_3]^-$ c)$[Hg(Cl)_3]^-$ d)$[Hg(Br)_3]^-$. Reprinted with permission from Ref. [51]. Copyright 2011 American Chemical Society.

Although NCI is usually used for weak interactions, its basis does not limit the tool to only weak interactions. Indeed, it can be used for covalent and ionic interactions, as well (see S.I. in Ref. [50] and Ref. [61]). Most commonly, it is also used within metallorganic frameworks to detect interactions with metals. We will analyze several model examples relelvant to solvation and protein structure.

7.5.1 Hg^{2+} complexation

Understanding the complexation of ions and their preferential ligands is of prime relevance when studying their bioactivity. As an example, their absorbance and transport through the body is extremely dependent on their complexation. It would then be extremely interesting to be able to identify the main series as well as the potential substitution sites. In addition Mercury(II) is a heavy metal cation which is specially challenging for quantum mechanical treatment as both correlation and relativistic effects play a crucial role in modeling its complexes.

Figures 14a-d illustrate the ability of NCI to visualize in a fast and efficient manner the complexation sites of Hg and to discriminate the strength of the binding energies between the cation and its ligands. Figure 14a shows the $[Hg(H_2O)_3]^{2+}$ complex and suggests that the three waters bound to Hg are not equivalent, one of the water molecules being more weakly bonded than the others. In order to corroborate this observation, it is possible to perform a decomposition of the interaction energy of such a complex using the RVS [85] (Reduced Variational Space) procedure. Both polarization and charge transfer are significantly weaker for one water molecule: whereas two of the water molecules show a polarization energy of ca. -15 kcal/mol, and a charge transfer energy of -10 kcal/mol, the third water shows a stabilization due to polarization and charge transfer by only -12.7 kcal/mol and -6.2 kcal/mol, respectively.

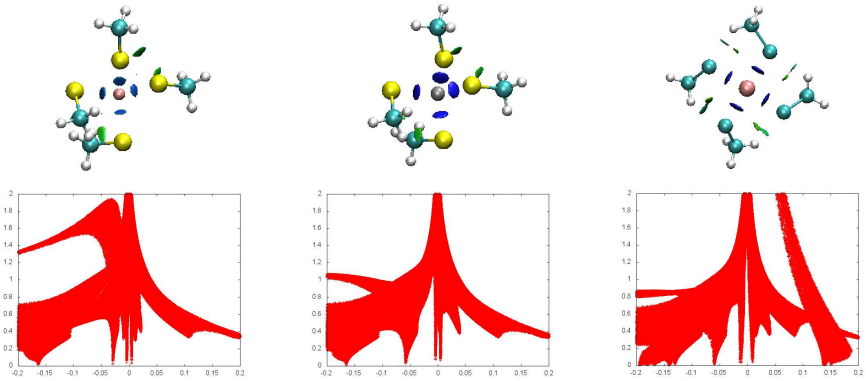
The ability of NCI to recover the ordering of ligands is also applicable when different ligand series are analyzed. Figure 14b-d show $[Hg(X)_3]$ complexes (X standing for F, Cl and Br). It is clearly seen that F is more strongly bonded than Cl and Br. This result is in agreement with their binding energies, which are -632.8, -571.0 and -562.6 kcal/mol, respectively.

7.5.2 Zn^{2+} fingers

Metals play a decisive role in many protein active sites as cofactors. The zinc finger is a small protein structural motif that can be found in many biological systems. It is characterised by the fact than one or more zinc ions can stabilize the fold of a protein. One or two Zn(II) cations [100] are often tetrahedrally coordinated to four or six amino acids such as cysteine (Cys) or histidine (Hys) forming four major cores: $ZnCys_4$, $ZnCys_3His$, $ZnCys_2His_2$ and $ZnCys_6$. [101]

Starting from the $ZnCys_4$ core, its modeling can be carried out by substitution of Cys by methyl thiolate, CH_3S^-. [101] In order to analyse the effect of the environement within the HSAB theory, we have analyzed the series $M^{2+}[SCH_3]_4^{2-}$, with M=Mg,Zn,Pd (from hard to soft Lewis acids). All complexes were optimized at the ωB3LYP/6-31++G** level of theory except for Pd, which was optimized with the pseudopotential LANL2DZ.

In all cases, the tetracoordination of the metal to the XMe ligands clearly stands out as a strong interaction (deep blue in Figure 15). It is interesting to note that the interaction spike shifts from ca. $\text{sign}(\lambda_2)\rho$=-0.03 au. to $\text{sign}(\lambda_2)\rho$=-0.06 au. when



a) $[Mg(SMe)_4]^{2-}$ b) $[Zn(SMe)_4]^{2-}$ c) $[Pd(SMe)_4]^{2-}$

Fig. 15 NCI plots for Zn finger model: a) $[Mg(SMe)_4]^{2-}$, b) $[Zn(SMe)_4]^{2-}$, c) $[Pd(SMe)_4]^{2-}$. Top: 3D NCI plot. Bottom: 2D NCI plot.

passing from Mg(II), which is a hard cation, to Zn(II) and Pd(II), which are intermediate and soft cations. This is in agreement to what is expected from the nature of the S bridge (soft) to the ligands within the HSAB theory.

Spikes appear at very low densities which correspond to secondary interactions in between the ligands, which stabilize the whole structure. As a final note, the big spikes appearing for Pd at ca. ρ=0.1, along with the peaks appearing for the iodine compounds in Section 7.1, are an artifact from the pseudopotential, and as such, should be disregarded.

8 Reactivity

Understanding and predicting chemical reactivity are some of the achievements of quantum chemistry. In this regard, the Woodward-Hoffmann rules [21] for pericyclic reactions have become a classical reference. By definition, pericyclic reactions evolve via a cyclic aromatic transition state of delocalized electrons where bond making and bond breaking occur simultaneously in a cyclic array. Using the orbital symmetry conservation, Woodward-Hoffmann proposed a list of rules of thumb able to predict the mechanism and, hence the stereoselectivity of pericyclic reactions. Examples include cycloadditions, electrocyclizations, sigmatropic rearrangements, and chelotropic reactions. Much work has been devoted to show that electron circulation on the pericyclic transition states may be smartly characterized by the topology of the electron localization function (ELF). [92] Recently, it was shown how the combined analysis of the NCI method and ELF may be employed as a visual tool to understand the electron reorganization along an intrinsic reaction path (IRC). [93] Contrary to ELF, the reduced density gradient does no suffer from catastrophes (sudden creation and/or destruction of critical points), being possible to preclude the bonding formation from the first stages of the reaction.

One example of application of NCI to predict the outcome of pericyclic reactions is provided by the thermal ring-opening of *cis*- and *trans*- 1,2,3,4-tetrafluoro-1,4-bis(pentafluorsulfanyl)cyclobutene (see Figure 16). As a thermal, 4*n* electron process, the Woodward-Hoffmann rules predict that the conrotatory opening is more favorable than the disrotatory one. [21, 94] Additionally, a given terminal substituent may either rotate "outward", leading to (*E*,*E*)-1,2,3,4-tetrafluoro-1,4-bis(pentafluorosulfanyl)butadiene or "inward" to yield (*Z*,*Z*)-1,2,3,4-tetrafluoro-1,4-bis(pentafluorosulfanyl)butadiene. Activation energies obtained at the ωB97X-D/6-31G** level for (*E*,*E*) and (*Z*,*Z*) transition states are 41.55 kcal/mol and 21.12 kcal/mol, respectively. Because this kind of stereoselectivity is related to the direction of the twist, it was named *torquoselectivity* by Houk and co-workers. [95]

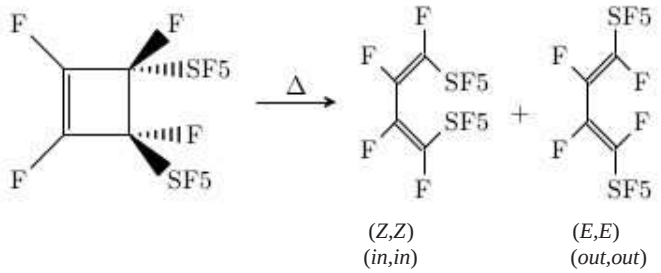


Fig. 16 "Outward" (out) and "inward" (in) conrotatory processes for the thermal ring opening of *trans*-1,2,3,4-tetrafluoro-1,4- bis(pentafluorsulfanyl)cyclobutene.

Rondan and Houk proposed in 1984 a widely accepted orbital model able to explain *torquoselectivity*. [96, 97] In a nutshell, this model states that electron donor substituents at C_3 and C_4 preferentially rotate outward in order to maximize the stabilizing interaction with the HOMO of the breaking C_3-C_4 bond and to minimize the repulsive interaction with the LUMO of the same bond. Electron acceptor substituents undergo the opposite effect, and, consequently, inward rotation is preferred. Since only certain orbitals are included in the model, a wrong selection of the interacting orbitals leads to wrong predictions. This disadvantage is common for all theories based on a selected group of orbitals, such as the frontier orbital theory. [98] To avoid this flaw, Ponec decided to reinvestigate the problem in terms of an electron density based indicator, such as the molecular similarity approach. [99] He showed that the origin of the *torquoselectivity* comes from the low electron reorganization required to convert reactants into products.

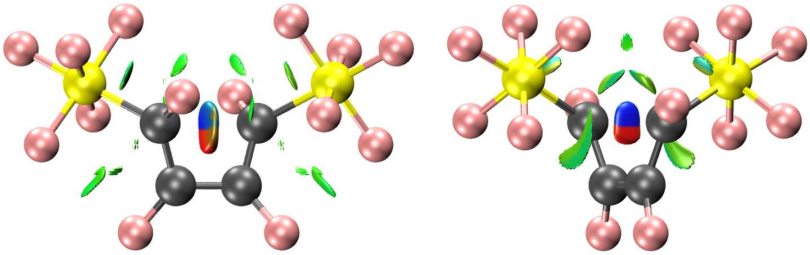


Fig. 17 NCI isosurfaces of (*E*,*E*)(left) and (*Z*,*Z*)(right) transition states.

Additionally, NCI analyses of both (*Z*,*Z*) and (*E*,*E*) transition states provide us with topological arguments to understand this differential selectivity. As seen in Figure 17, out of the breaking carbon-carbon covalent interaction (blue isosurface) and its repuslive counterpart ring tension (red isosurface), we can differentiate three types of noncovalent interactions (green isosurfaces):

Type 1 Fluor-fluor interaction between pentafluorosulfanyl groups
Type 2 Pentafluorosulfanyl-carbon interaction
Type 3 Fluor-fluor interaction between pentafluorosulfanyl and fluoro groups

All of them are present in the (*Z*,*Z*) transition state, whereas only interactions of type 3 are found in the (*E*,*E*) one (see Figure 18). Thus, dispersion interactions between pentafluorosulfanyl groups and those with the carbon cycle should be the driving force of the process. Thus, *torquoselectivity* can also be understood in terms of secondary interactions as revealed by NCI: within this approach products are driven by the accumulation of noncovalent interactions in the transition state.

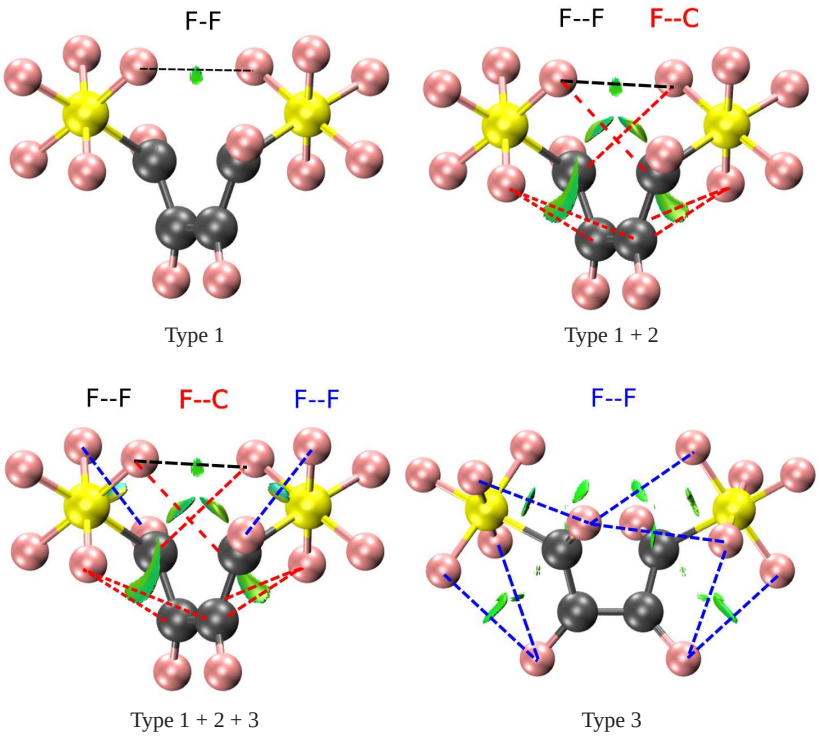


Fig. 18 Noncovalent interactions types in (*Z*,*Z*)(top and bottom left) and (*E*,*E*) (bottom, right) transition states. Black, red and blue dashed lines represent type 1, type 2 and type 3 interactions respectively.

9 Summary and conclusions

In conclusion, non-covalent interactions have a unique signature and their presence can be revealed solely from the electron density. Noncovalent interactions are highly nonlocal and manifest in real space thanks to the NCI analysis: in other words, as low-gradient isosurfaces with low densities. The sign of the second Hessian eigenvalue is used to identify the interaction type, and its strength can be derived from the density on the non-covalent interaction surface.

NCI provides a rapid and rich representation of van der Waals interactions, hydrogen bonds, and steric clashes. If obliged by the size of the system, NCI can be approximated from promolecular densities, so that it only requires the atomic coordinates as input. Thus, it is applicable to large systems, such as proteins or DNA.

Since it is based on the electron density, it is applicable to all types of chemical bonds. We have reviewed here several such examples along the periodic table: halogen bonds, pnicogenic bonds, di-hydrogen bonds. We have even gone to higher densities to show that NCI is also able to reveal and characterize interactions in organometallic systems. Finally, we have looked at the change of chemical interactions along a reaction path, reformulating orbital rules in the *torquoselectivity.*

In summary, we have shown that the electron density and its derivatives contain all the information needed to characterize all chemical bonds and their change, making NCI a holistic tool in the analysis of weak (and not so weak) interactions.

Acknowledgements This work undertaken(partially) in the framework of CALSIMLAB is supported by the public grant ANR-11-LABX-0037-01 overseen by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (reference: ANR-11-IDEX-0004-02)

Part I
Appendix

Appendix

Behavior of Model densities

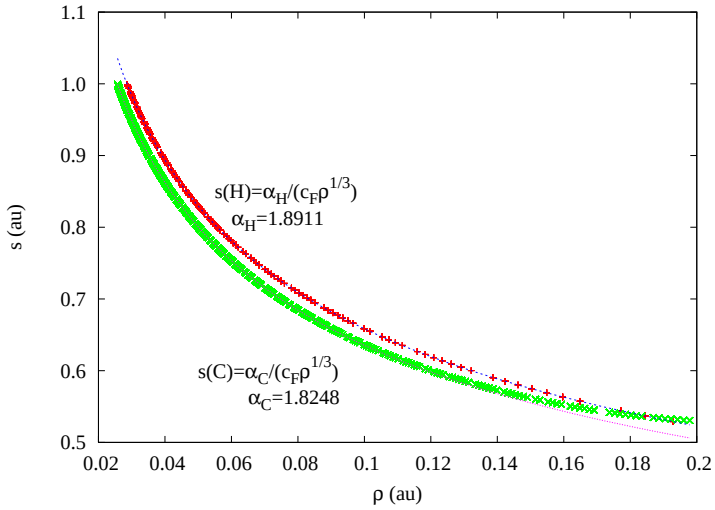


Fig. 19 Behavior of $s(\rho)$ for model densities $\rho = e^{-\alpha r}$ for hydrogen and carbon.

Parameters for promolecular calculations

Table 2 Parameterized exponents (η_i) and coefficients (c_i) of sphericalized atomic densities, in atomic units.

Atom	c_1	η_1	c_2	η_2	c_3	η_3
H	0.2815	0.5288	–	–	–	–
He	2.437	0.3379	–	–	–	–
Li	11.84	0.1912	0.06332	0.9992	–	–
Be	31.34	0.1390	0.3694	0.6945	–	–
B	67.82	0.1059	0.8527	0.5300	–	–
C	120.2	0.0884	1.172	0.5480	–	–
N	190.9	0.0767	2.247	0.4532	–	–
O	289.5	0.0669	2.879	0.3974	–	–
F	406.3	0.0608	3.049	0.3994	–	–
Ne	561.3	0.0549	6.984	0.3447	–	–
Na	760.8	0.0496	22.42	0.2511	0.06358	1.0236
Mg	1016	0.0449	37.17	0.2150	0.3331	0.7753
Al	1319	0.0411	57.95	0.1874	0.8878	0.5962
Si	1658	0.0382	87.16	0.1654	0.7888	0.6995
P	2042	0.0358	115.7	0.1509	1.465	0.5851
S	2501	0.0335	158.0	0.1369	2.170	0.5149
Cl	3024	0.0315	205.5	0.1259	3.369	0.4974
Ar	3625	0.0296	260.0	0.1168	5.211	0.4412

References

1. L. Pauling. 1947. *General Chemistry*, Dover Publications, Inc.
2. J. K. Burdett. 1997. *Chemical Bonds-A Dialog*. Wiley: Chichester, England.
3. G. Frenking, and S. Shaik (editors). 2007. Special Issue: 90 Years of Chemical Bonding. *Journal of Computational Chemistry* 28:1-466.
4. G. Frenking, and A. Krapp. 2007. Unicorns in the world of chemical bonding models. *Journal of Computational Chemistry* 28:15-24.
5. S. Shaik, H. S. Rzepa, and R. Hoffmann. 2013. One molecule, two atoms, three views, four bonds? *Angewandte Chemie International Edition* 52:3020-33.
6. G. Frenking, and M. Hermann. 2013 *Angewandte Chemie International Edition* 52:5922-25.
7. C. A. Coulson. 1953. *The Spirit of Applied Mathematics*, Clarendon Press 20-21.
8. H. Jacobsen. 2010. Kinetic energy density and covalent bonding - a complementary analysis at the border of bond and no bond. *Dalton Transactions* 39:5426-28.
9. R. G. Parr, P. W. Ayers, and R. F. Nalewajski. 2005. What is an atom in a molecule? *The Journal of Physical Chemistry A* 109:3957-59.
10. H. Jacobsen. 2013. Topology maps of bond descriptors based on the kinetic energy density and the essence of chemical bonding. *Physical Chemistry Chemical Physics* 15:5057-66.
11. P. v. R. Schleyer. 2005. Introduction: Delocalization Pi and Sigma. *Chemical Reviews* 105:3433-35.
12. P. L. A. Popelier. 2007. Preface. *Faraday Discussions*135:1-3.
13. I. V. Alabugin, K. M. Gilmore, and P. W. Peterson, P. W. 2011. Hyperconjugation. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1:109141.
14. J. F. Gonthier, S. N. Steinmann, M. D. Wodrich, and C. Corminboeuf. 2012. Quantification of "fuzzy" chemical concepts: a computational perspective. *Chemical Society Reviews* 41:467187.
15. A. Whitaker. *Einstein, Bohr and the Quantum Dilemma*. 1996. Cambridge. Cambridge University Press.
16. S. Alvarez, R. Hoffmann, and C. Mealli. 2009. A bonding quandary – or – a demonstration of the fact that scientists are not born with logic. *Chemistry-A European Journal* 15:8358-73.
17. D. Danovich, P. C. Hiberty, W. Wu, H. S. Rzepa, ,and S. Shaik. 2014. The Nature of the Fourth Bond in the Ground State of C-2 : The Quadruple Bond Conundrum. *Chemistry-A European Journal* 20:6214-20.
18. P. L. Ayers, R. J. Boyd *et al*. 2015. Six questions on topology in theoretical chemistry. *Computational and Theoretical Chemistry* 1053: 2-16.
19. G. N. Lewis. 1916. The atom and the molecule *J. Am. Chem. Soc.* 38:762-786.
20. K. Fukui, T. Yonezawa, and H. Shingu. 1952. A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *Journal of Chemical Physics* 20:722-725.
21. R. B. Woodward, and R. Hoffmann. 1969. The Conservation of Orbital Symmetry *Angew. Chem., Int. Ed. Engl.* 8:781-932.
22. R. G. Parr, R. A. Donnelly, M. Levy, and W. E. Palke. 1978. Electronegativity: The density functional viewpoint *Journal of Chemical Physics* 68:3801.
23. R. F. W. Bader. 1991. A quantum theory of molecular structure and its applications. *Chem. Rev.* 91:893-928.
24. R. F. W. Bader. *Atoms in Molecules: A Quantum Theory.* 1990. Oxford. International Series of Monographs on Chemistry 22; Oxford Science Publications.
25. C. F. Matta, and R. J. Boyd. *In The Quantum Theory of Atoms in Molecules. 2007. New York. Wiley-VCH.* 1-34.
26. A. D. Becke, and K. E. J. Edgecombe. 1990. A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.* 92: 5397-5403.
27. B. Silvi, and A. Savin. 1994. Classification of chemical bonds based on topological analysis of electron localization functions. *Nature* 371:683-686.
28. B. Honig, and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science* 268:1144-1149.

29. H. Fenniri, M. Packiarajan, K. L. Vidale, D. M. Sherman, K. Hallenga, K. V. Wood, J. G. Stowell. 2001. Helical Rosette Nanotubes: Design, Self-Assembly, and Characterization. *J. Am. Chem. Soc.* 123:3854-55.
30. P. Kruse, E. R. Johnson, G. A. DiLabio, R. A. Wolkow. 2002. Patterning of Vinylferrocene on H-Si(100) via Self-Directed Growth of Molecular Lines and STM-Induced Decomposition. *Nano Lett.* 2:807-810.
31. S. S. Sheiko, F. C. Sun, A. Randall, D. Shirvanyants, M. Rubinstein, H. Lee, K. Matyjaszewski. 2006. Adsorption-induced scission of carbon-carbon bonds. *Nature.* 440:191-194.
32. G. A. DiLabio, P. G. Piva, P. Kruse, R. A. Wolkow. 2004. Dispersion interactions enable the self-directed growth of linear alkane nanostructures covalently bound to silicon. *J. Am. Chem. Soc.* 126:16048-50.
33. P. A. Kollman. 1977. Noncovalent interactions *Chem. Rev.* 10:365-371.
34. J. Cerniý, P. Hobza. 2007. Non-covalent interactions in biomacromolecules. *Phys. Chem. Chem. Phys.* 9:5291-5303.
35. A. L. Lehninger, D. L. Nelson, M. M. Cox. 1993. *Principles of Biochemistry. Second edition*. Worth Publishers, Inc.
36. N. Krishnamoorthy, M. H. Yacoub, S. N. Yaliraki. 2011. A computational modeling approach for enhancing self-assembly and biofunctionalisation of collagen biomimetic peptides. *Biomaterials* 32:7275-85.
37. A. Dutta, A. D. Jana, S. Gangopadhyay, K. Kumar Das, J. Marek, R. Marek, J. Brus, M. Ali. 2011. Unprecedented $\pi\pi$ interaction between an aromatic ring and a pseudo-aromatic ring formed through intramolecular H-bonding in a bidentate Schiff base ligand: crystal structure and DFT calculations. *Phys. Chem. Chem. Phys.* 13:15845-53.
38. A. Gavezzotti. *Molecular Aggregation. Structure Analysis and Molecular Simulation of Crystals and Liquids.* 2007. Oxford. IUCr monographs on Crystallography n. 19, Oxford University Press.
39. S. Keinan, M. A. Ratner, T. J. Marks. 2004. Molecular zippers–designing a supramolecular system. *J. Chem. Phys. Lett.* 392:291-296.
40. G. R. Desiraju. *Crystal Engeneering. The design of organic Solids* 1989. Amsterdam. Elsevier
41. G. M. Day *et al.* 2009. *Acta Cryst. B.* 65:107-125.
42. J. M. Word, S. C. Lovell, T. H. LaBean, H. C. Taylor, M. E. Zalis, B. K. Presley, J. S. Richardson, and D. C. Richardson. 1999. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* 285:1735-1747.
43. I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall III, J. Snoeyink, J. S. Richardson, and D. C. Richardson. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35:W375-W383.
44. V. Sobolev, A. Sorokine, J. Prilusky, E. E. Abola, and M. Edelman. 1999. Automated analysis of interatomic contacts in proteins. *Bioinformatics* 15:327-332.
45. I. K. McDonald, and J. M. Thornton. 1994. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* 238:777-793.
46. M.E. Alikhani, F. Fuster, and B. Silvi. 2005. What can tell the topological analysis of ELF on hydrogen bonding? *Struct. Chem.* 16:203-210.
47. R. D. Cramer III, D. E. Patterson, and J. D. Bunce. 1988. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 110:5959-5967.
48. J. Pilmé, J.-P. Piquemal. 2008. Advancing beyond charge analysis using the electronic localization function: Chemically intuitive distribution of electrostatic moments *J. Comput. Chem.* 29:1440-49.
49. P. Hohenberg, and W. Kohn. 1964. Inhomogeneous electron gas *Phys. Rev. B* 136:B864.
50. E. R. Johnson, S. Keinan, P. Mori-Sanchez, J. Contreras-García, A. J. Cohen, and W. Yang. 2010. Revealing Noncovalent Interactions *J. Am. Chem. Soc.* 132:6498-6506.
51. J. Contreras-García, E. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. Beratan, and W. Yang. 2011. NCIPLOT: a program for plotting non-covalent interaction regions *J. Chem. Theor. Comp.* 7:625-632.

52. J. Contreras-García, E. R. Johnson, W. Yang. 2011. Analysis of Hydrogen-Bond Interaction Potentials from the Electron Density: Integration of Noncovalent Interaction Regions *J. Phys. Chem. A* 115:12983-12990.
53. A. D. Becke. 1995. *Modern Electronic Structure Theory*. World Scientific, Yarkony.
54. A. J. Cohen, P. Mori-Sánchez, and W. Yang. 2008. Insights into current limitations of density functional theory. *Science* 321:792-794.
55. A. Zupan, K. Burke, M. Ernzerhof, , J. P. Perdew. 1997. Distributions and averages of electron density parameters: Explaining the effects of gradient corrections. *J. Chem. Phys.* 106:10184-193.
56. A.D. Becke. 1993. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98:5648-52.
57. C. Lee, W. Yang, and R.G. Parr. 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B.* 37:785.
58. G. Arfken, G. *Mathematical Methods for Physicists*. 1985. Orlando Academic Press.
59. R. F. W. Bader, H. Essén. 1984. The characterization of atomic interactions. *J. Chem. Phys.* 80:1943-1960.
60. J. R. Lane, J. Contreras-García, J.-P. Piquemal, B. J. Miller, and H. G. Kjaergaard. 2013. Are Bond Critical Points Really Critical for Hydrogen Bonding? *J. Chem. Theory Comp.* 9:3263-3266.
61. J. Contreras-García, M. Calatayud, J.-P. Piquemal, and J. M. Recio. 2012. Ionic interactions: Comparative topological approach *Comp. Theo. Chem.* 998:193-201.
62. R. Chaudret, B. de Courcy, J. Contreras-García, E. Gloaguen, A. Zehnacker-Rentien, M. Mons, and J.-P. Piquemal. 2013. Unraveling Non Covalent Interactions within Flexible Biomolecules: from electron density topology to gas phase spectroscopy *Phys. Chem. Chem. Phys.* 16:9876-9891.
63. M. A. Spackman, E. N. Maslen. 1986. Chemical properties from the promolecule. *J. Phys. Chem.* 90:2020-2027.
64. A. Martin Pendás, V. Luaña, L. Pueyo, E. Francisco, P. Mori-Sánchez. 2002. Hirshfeld surfaces as approximations to interatomic surfaces. *J. Chem. Phys.* 117:1017-1023.
65. S. Fiedler, J. Broecker, , S. Keller. 2010. Protein folding in membranes. *Cellular and Molecular Life Sciences*, 67:1779-98.
66. K. A. Dill. 1990. Dominant forces in protein folding. *Biochemistry*, 29:7133-55.
67. B. J. Deppmeier, A. J. Driessen, W. Hehre, J. A. Johnson, P. E. Klunzinger, and M. Watanabe. 2002. *Spartan ES 1.0.2*. Wavefunction Inc., Irvine.
68. X.-J. Lu , and W. K. Olson. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 31:5108-121.
69. http://rutchem.rutgers.edu/ olson/Tsukuba/.
70. E. Espinosa, E. Molins, , C. Lecomte. 1988. Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chem. Phys. Lett.* 285:170-173.
71. R. Wieczorek , J. J. Dannenberg. 2004. Comparison of fully optimized α-and 310-helices with extended β-strands. An oniom density functional theory study. *J. Am. Chem. Soc.* 126:14198-205.
72. R. Viswanathan, A. Asensio, and J. J. Dannenberg. 2004. Cooperative hydrogen-bonding in models of antiparallel β-sheets. *J. Phys. Chem. A.* 108:9205-212.
73. P. Jurecka , and P. Hobza. 2003. True Stabilization Energies for the Optimal Planar Hydrogen-Bonded and Stacked Structures of Guanine···Cytosine, Adenine···Thymine, and Their 9-and 1-Methyl Derivatives: Complete Basis Set Calculations at the MP2 and CCSD (T) Levels and Comparison with Experiment. *J. Am. Chem. Soc.* 125:15608-613.
74. E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt. 2011. Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure Appl. Chem.* 83:1637-1641.
75. P. Metrangolo and G. Resnati. 2001. Halogen bonding: a paradigm in supramolecular chemistry. *Chemistry - A European Journal* 7:2511-2519.

76. D. Manna and G. Mugesh. 2012 Regioselective deiodination of thyroxine by iodothyronine deiodinase mimics: an unusual mechanistic pathway involving cooperative chalcogen and halogen bonding. *J. Am. Chem. Soc.* 134:4269-4279.
77. S. Scheiner. A New Noncovalent Force: Comparison of P···N Interaction with Hydrogen and Halogen Bonds *J. Chem. Phys.* 134:094315-094319.
78. D. Mani and E. Arunan. 2013. The X-C···Y (X=O/F, Y=O/S/F/Cl/Br/N/P) 'Carbon Bond' and Hydrophobic Interactions. *Phys. Chem. Chem. Phys.* 15:14377-14383.
79. H. Matter, M. Nazaré, S. Güssregen, D. W. Will, H. Schreuder, A. Bauer, M. Urmann, K. Ritter, M. Wagner, and V. Wehner. 2009. Evidence for C-Cl/C-Br π interactions as an important contribution to protein-ligand binding affinity. *Angewandte Chemie*, 121:2955-2960.
80. L. A. Hardegger, B. Kuhn, B. Spinnler, L. Anselm, R. Ecabert, M. Stihle, B. Gsell, R. Thoma, J. Diez, J. Benz, et al. 2011. Systematic investigation of halogen bonding in protein-ligand interactions. *Angewandte Chemie International Edition* 50:314-318.
81. K. A. Peterson, B. C. Shepler, D. Figgen, and H. Stoll. 2006. On the spectroscopic and thermochemical properties of ClO, BrO, IO, and their anions. *The Journal of Physical Chemistry A* 110:13877-13883.
82. J. Murray, P. Lane and P. Politzer. 2007. A Predicted New Type of Directional Noncovalent Interaction. *Int. J. Quantum Chem.* 107:2286-2292.
83. J. George, V. L. Deringer and R. Dronskowski. Cooperativity of Halogen, Chalcogen, and Pnictogen Bonds in Infinite Molecular Chains by Electronic Structure Theory. *J. Phys. Chem. A* 118:3193-3200.
84. A.C. Legon, B.P. Roberts, and A.L. Wallwork. 1990. Rotational spectra and geometries of the gas-phase dimers (CH_4,HF) and (CH_4,HCl) *Chem. Phys. Lett.* 173:107-114
85. W. J. Stevens , W. H. Fink. 1987. Frozen fragment reduced variational space analysis of hydrogen bonding interactions. Application to the water dimer. *Chem. Phys. Lett.*, 139:15-22.
86. P. L. A. Popelier. 1998. Characterization of a dihydrogen bond on the basis of the electron density. *J. Phys. Chem. A*, 102:1873-78.
87. C. A. Morrison , M. M. Siddick. 2004. Dihydrogen bonds in solid BH_3NH_3. *Angew. Chem. Int. Ed.*, 116:4780-4782.
88. C. F. Matta, J. Hernndez-Trujillo, T. Tang, and R. F. W. Bader. 2003. Hydrogen–hydrogen bonding: a stabilizing interaction in molecules and crystals. *Chem. Eur. J.*, 9:1940-1951.
89. Scienomics; MAPS platform, version 3.4, 2014, France
90. M. Valiev, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. van Dam, D. Wang, J. Nieplocha, E. Apra, T.L. Windus, and W.A. de Jong. 2010. NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* 181:1477-89.
91. T. B. Richardson, de Gala S., Crabtree, , P. E. M. R. H.; Siegbahn. 1995. Unconventional Hydrogen Bonds: Intermolecular B-H···H-N Interactions. *J. Am. Chem. Soc.*, 117:12875-76.
92. E. Matito, J. Poater, M. Duran, and M. Solà. 2006. Electron Fluctuation in Pericyclic and Pseudopericiclyc Reations. *A European Journal of Chemical Physics and Physical Chemistry* 7:111-13.
93. N. Gillet, R. Chaudret, J. Contreras-Garcia, W. Yang, B. Silvi, and J.-P. Piquemal. 2012.Coupling Quantum Interpretative Techniques: Another Look at Chemical Mechanism in Organic Reactions. *Journal of Chemical Theory and Computation* 8:3993-97.
94. R. B. Woodward, R. J. Hoffmann. 1965. Selection Rules for Concerted Cycloaddition Reactions. *Journal of the American Chemical Society* 87:2046-48.
95. K. N. Houk, Y. Li, and J. D. Evanseck. 1992. Transition Structures of Hydrocarbon Pericyclic Reactions. *Angewandte Chimie International Edition* 31:682-708
96. W. Kirmse, N. G. Rondan, and K. N. Houk. 1984. Stereoselective substituent effects on conrotatory electrocyclic reaction of cyclobutenes. *Journal of the American Chemical Society* 106:7989-91-
97. N. G. Rondan, K. N. Houk. 1984. Theory of setereoselection in conrotatory electrocyclic reactions of substitued cyclobutenes. *Journal of the American Chemical Society* 107:2099-111.

98. M. J. S. Deward. 1989. A Critique of Frontier Orbital Theory. *Journal of Molecular Structure (Theochem)* 200:301-23.
99. R. Ponec, G. Yuzhakov, and J. Pecka. 1996. Similarity approach to chemical reactivity. Torquoselectivity in pericyclic reactions. *Journal of Mathematical Chemistry* 20:301-10.
100. B. De Courcy, J.-P. Dognon, C. Clavaguéra, N. Gresh, J.-P. Piquemal. 2011. Interactions within the Alcohol Dehydrogenase (ADH) Zn(II)-metalloenzyme active site: interplay between subvalence, electron correlation/dispersion and charge transfer/induction effects. *Int. J. Quantum Chem.* 111:1213-1221.
101. Y.-M. Lee, C. Lim. 2011. Factors controlling the reactivity of zinc finger cores. *J. Am. Chem. Soc.* 133:8691-8703.

Annexe C

Paramètres et constantes du potentiel SIBFA

C.1 Electrostatique

cp-AMOEBA

fragments	atoms	α	β	η
eau	H	3.5	2.9	1.7
	O	4.0	4.9	4.7
methanol	H(-H3C)	3.5	2.6	4.7
	C(-CH3)	5.8	4.9	2.5
	O	6.4	4.5	0.7
	H (-O)	4.8	2.9	1.7
methylamine	H(-H3C)	3.5	2.7	4.7
	C(-CH3)	4.8	5.4	2.5
	N	4.8	4.8	2.7
	H (-N)	2.5	2.9	1.7
benzène	C	2.8	2.5	2.8
	H	3.5	2.3	2.5

TABLE C.5 – paramètres atomiques de différents fragments moléculaires de la correction de pénétration de charge dans le cadre de cp-AMOEBA (cp-MEP)

cp-GDMA [1]

fragments	atoms	r_{pen}
eau	H	0,650
	O	1,480
benzène	C	1,075
	H	1,625

TABLE C.6 – paramètres atomiques de différents fragments moléculaires de la correction de pénétration de charge dans le cadre de cp-GDMA [1]

γ	δ	χ	Ω
2,44	2,30	2,25	102,59

TABLE C.7 – paramètres de la correction de pénétration de charge dans le cadre de cp-GDMA [1]

Ω est un paramètre fixé constant, tel que γ, δ et χ, intervenant dans la correction charge-quadrupole.

C.2 Répulsion & Dispersion

Répulsion

	c_s	c_p	f^s_{orb}	f^p_{orb}	K_i	R_{rep}
O	1,000	0,000	1,03	1,49	1,36	1,46
H	0,500	0,866	1,00	0,00	1,00	1,17

TABLE C.8 – paramètres et constantes de l'énergie de répulsion de SIBFA

La variable K_{AC} défini dans l'équation 2.7 du chapitre 2 de la partie 2 est défini par le produit de K_A et K_C. Enfin α est égale à 14.

Dispersion

Le terme B_{ij} défini dans l'équation 2.38 et 2.31 du chapitre 2 de la partie 2 est équivalent à K_{AC} défini ci-dessus.

$\alpha(6)$	$\alpha(8)$	$\alpha(10)$	C^6_{disp}	C^8_{disp}	C^{10}_{disp}
1,23	1,28	1,36	1,00	0,318	0,180

TABLE C.9 – constantes de l'énergie de dispersion SIBFA à longue portée

α_{xd}	C_{xd}
8,80	1,37

TABLE C.10 – constantes de l'énergie de dispersion SIBFA à courte portée

Enfin, les paramètres intrinsèques aux atomes du traitement de la dispersion SIBFA sont 1,48 et 1,21 pour l'oxygène et l'hydrogène respectivement.

Bibliographie

[1] J.-M. Lehn. *La chimie supramoléculaire : concepts et perspectives* (De Boeck Supérieur, 1997).

[2] F. Wöhler. *Sur la formation artificielle de l'urée*. In *Annales de chimie et de physique*, vol. 37, 330–334 (1828).

[3] R. B. Woodward. *The total synthesis of vitamin b12*. Pure and Applied Chemistry, **33**(1), 145–178 (1973).

[4] E. Fischer. *Synthesen in der zuckergruppe ii*. Berichte der deutschen chemischen Gesellschaft, **27**(3), 3189–3232 (1894).

[5] P. Hohenberg & W. Kohn. *Inhomogeneous electron gas*. Phys. Rev., **136**, B864–B871 (Nov 1964).

[6] R. G. Parr & W. Yang. *Density-functional theory of atoms and molecules*, vol. 16 (Oxford university press, 1989).

[7] W. M. Haynes. *CRC handbook of chemistry and physics* (CRC press, 2014).

[8] N. Gresh, G. A. Cisneros, T. A. Darden & J.-P. Piquemal. *Anisotropic, polarizable molecular mechanics studies of inter- and intramolecular interactions and ligand-macromolecule complexes. a bottom-up strategy*. Journal of chemical theory and computation, **3**(6), 1960–1986 (2007).

[9] N. Gresh, P. Claverie & A. Pullman. *Theoretical studies of molecular conformation. derivation of an additive procedure for the computation of intramolecular interaction energies. comparison withab initio scf computations*. Theoretica chimica acta, **66**(1), 1–20 (1984).

[10] J.-P. Piquemal, B. Williams-Hubbard, N. Fey, R. J. Deeth, N. Gresh & C. Giessner-Prettre. *Inclusion of the ligand field contribution in a polarizable molecular mechanics : Sibfa-lf*. Journal of computational chemistry, **24**(16), 1963–1970 (2003).

[11] J.-P. Piquemal, H. Chevreau & N. Gresh. *Toward a separate reproduction of the contributions to the hartree-fock and dft intermolecular interaction energies by polarizable molecular mechanics with the sibfa potential*. Journal of Chemical Theory and Computation, **3**(3), 824–837 (2007).

[12] J.-P. Piquemal, N. Gresh & C. Giessner-Prettre. *Improved formulas for the calculation of the electrostatic contribution to the intermolecular interaction energy from multipolar expansion of the electronic distribution*. The Journal of Physical Chemistry A, **107**(48), 10353–10359 (2003).

[13] M. Devereux, N. Gresh, J.-P. Piquemal & M. Meuwly. *A supervised fitting approach to force field parametrization with application to the sibfa polarizable force field*. Journal of computational chemistry, **35**(21), 1577–1591 (2014).

[14] J. W. Ponder. *Tinker molecular modeling package v 6.2*.

[15] P. Ren, C. Wu & J. W. Ponder. *Polarizable atomic multipole-based molecular mechanics for organic molecules*. Journal of chemical theory and computation, **7**(10), 3143–3161 (2011).

[16] W. L. Jorgensen, D. S. Maxwell & J. Tirado-Rives. *Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids*. Journal of the American Chemical Society, **118**(45), 11225–11236 (1996).

[17] C. Oostenbrink, A. Villa, A. E. Mark & W. F. Van Gunsteren. *A biomolecular force field based on the free enthalpy of hydration and solvation : The gromos force-field parameter sets 53a5 and 53a6*. Journal of computational chemistry, **25**(13), 1656–1676 (2004).

[18] A. Warshel, M. Levitt & S. Lifson. *Consistent force field for calculation of vibrational spectra and conformations of some amides*

and lactam rings. Journal of Molecular Spectroscopy, **33**(1), 84–99 (1970).

[19] A. Warshel & M. Levitt. *Program for the consistent force field evaluation of equilibrium geometries and vibrational frequencies of molecules*. Quantum Chemistry Program Exchange, **247** (1974).

[20] J. W. Ponder & D. A. Case. *Force fields for protein simulations*. Advances in protein chemistry, **66**, 27–85 (2003).

[21] M. Born & R. Oppenheimer. *Zur quantentheorie der molekeln*. Annalen der Physik, **389**(20), 457–484 (1927).

[22] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore & J. Darnell. *Molecular cell biology* (WH Freeman, 2000).

[23] C. W. Pratt, D. Voet & J. G. Voet. *Fundamentals of biochemistry : life at the molecular level* (John Wiley & Sons, New York, NY, 2005).

[24] J. E. Jones. *On the determination of molecular fields. ii. from the equation of state of a gas*. Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences, **106**(738), 463–477 (1924).

[25] N. A. Christie, P. O. Buenaventura, H. C. Fernando, N. T. Nguyen, T. L. Weigel, P. F. Ferson & J. D. Luketich. *Results of expandable metal stents for malignant esophageal obstruction in 100 patients : short-term and long-term follow-up*. The Annals of thoracic surgery, **71**(6), 1797–1802 (2001).

[26] B. Xia, V. Tsui, D. A. Case, H. J. Dyson & P. E. Wright. *Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized born model, and with explicit water*. Journal of biomolecular NMR, **22**(4), 317–331 (2002).

[27] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee & L. G. Pedersen. *A smooth particle mesh ewald method*. The Journal of chemical physics, **103**(19), 8577–8593 (1995).

[28] H. J. Berendsen, J. P. Postma, W. F. van Gunsteren & J. Hermans. *Interaction models for water in relation to protein hydration*. In *Intermolecular forces*, 331–342 (Springer, 1981).

[29] H. Berendsen, J. Grigera & T. Straatsma. *The missing term in effective pair potentials*. Journal of Physical Chemistry, **91**(24), 6269–6271 (1987).

[30] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey & M. L. Klein. *Comparison of simple potential functions for simulating liquid water*. The Journal of chemical physics, **79**(2), 926–935 (1983).

[31] A. Glättli, C. Oostenbrink, X. Daura, D. P. Geerke, H. Yu & W. F. van Gunsteren. *On the transferability of the spc/l water model to biomolecular simulation*. Brazilian journal of physics, **34**(1), 116–125 (2004).

[32] M. W. Mahoney & W. L. Jorgensen. *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions*. The Journal of Chemical Physics, **112**(20), 8910–8922 (2000).

[33] B. Chen, J. Xing & J. I. Siepmann. *Development of polarizable water force fields for phase equilibrium calculations*. The Journal of Physical Chemistry B, **104**(10), 2391–2401 (2000).

[34] S. W. Rick, S. J. Stuart & B. J. Berne. *Dynamical fluctuating charge force fields : Application to liquid water*. The Journal of chemical physics, **101**(7), 6141–6156 (1994).

[35] C. J. Burnham & S. S. Xantheas. *Development of transferable interaction models for water. i. prominent features of the water dimer potential energy surface*. The Journal of chemical physics, **116**(4), 1479–1492 (2002).

[36] H. Yu, T. Hansson & W. F. van Gunsteren. *Development of a simple, self-consistent polarizable model for liquid water*. The Journal of chemical physics, **118**(1), 221–234 (2003).

[37] J. Jeon, A. E. Lefohn & G. A. Voth. *An improved polarflex water model*. The Journal of chemical physics, **118**(16), 7504–7518 (2003).

[38] G. Lamoureux, A. D. MacKerell Jr & B. Roux. *A simple polarizable model of water based on classical drude oscillators*. The Journal of chemical physics, **119**(10), 5185–5197 (2003).

[39] P. Ren & J. W. Ponder. *Polarizable atomic multipole water model for molecular mechanics simulation*. The Journal of Physical Chemistry B, **107**(24), 5933–5947 (2003).

[40] K. A. Beauchamp, Y.-S. Lin, R. Das & V. S. Pande. *Are protein force fields getting better ? a systematic benchmark on 524 diverse nmr measurements*. Journal of chemical theory and computation, **8**(4), 1409–1414 (2012).

[41] A. R. Leach. *Molecular modelling : principles and applications* (Pearson education, 2001).

[42] A. Klamt & G. Schuurmann. *Cosmo : a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient*. J. Chem. Soc., Perkin Trans. 2, 799–805 (1993).

[43] M. Schaefer & M. Karplus. *A comprehensive analytical treatment of continuum electrostatics*. The Journal of Physical Chemistry, **100**(5), 1578–1599 (1996).

[44] D. Qiu, P. S. Shenkin, F. P. Hollinger & W. C. Still. *The gb/sa continuum model for solvation. a fast analytical method for the calculation of approximate born radii*. The Journal of Physical Chemistry A, **101**(16), 3005–3014 (1997).

[45] G. D. Hawkins, C. J. Cramer & D. G. Truhlar. *Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium*. The Journal of Physical Chemistry, **100**(51), 19824–19839 (1996).

[46] J. C. Maxwell. *V. illustrations of the dynamical theory of gases. ?part i. on the motions and collisions of perfectly elastic spheres*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, **19**(124), 19–32 (1860).

[47] J. C. Maxwell. *Ii. illustrations of the dynamical theory of gases*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, **20**(130), 21–37 (1860).

[48] L. Boltzmann. *Weitere studien über das wärmegleichgewicht unter gasmolekülen* (Springer, 1970).

[49] L. Verlet. *Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules*. Physical review, **159**(1), 98 (1967).

[50] W. Van Gunsteren & H. Berendsen. *A leap-frog algorithm for stochastic dynamics*. Molecular Simulation, **1**(3), 173–185 (1988).

[51] W. C. Swope, H. C. Andersen, P. H. Berens & 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*. The Journal of Chemical Physics, **76**(1), 637–649 (1982).

[52] D. Beeman. *Some multistep methods for use in molecular dynamics calculations*. Journal of Computational Physics, **20**(2), 130–139 (1976).

[53] S. Nosé. *A unified formulation of the constant temperature molecular dynamics methods*. The Journal of chemical physics, **81**(1), 511–519 (1984).

[54] W. G. Hoover. *Canonical dynamics : equilibrium phase-space distributions*. Physical Review A, **31**(3), 1695 (1985).

[55] H. C. Andersen. *Molecular dynamics simulations at constant pressure and/or temperature*. The Journal of chemical physics, **72**(4), 2384–2393 (1980).

[56] H. J. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola & J. Haak. *Molecular dynamics with coupling to an external bath*. The Journal of chemical physics, **81**(8), 3684–3690 (1984).

[57] J. Richard. *Molecular simulation of fluids (theory, algorithms and object-orientation)* (1999).

[58] C. Narth, N. Gillet, B. Lévy, I. Demachy & A. de la Lande. *Investigation of the molecular mechanisms of electronic decoherence within a quinone cofactor*. Canadian Journal of Chemistry, **91**(7), 628–636 (2013).

[59] C. Narth, N. Gillet, F. Cailliez, B. LÈvy & A. de la Lande. *Electron transfer, decoherence, and protein dynamics : Insights from atomistic simulations*. Accounts of chemical research (2015).

[60] A. De la Lande, N. S. Babcock, J. Řezáč, B. C. Sanders & D. R. Salahub. *Surface residues dynamically organize water bridges to enhance electron transfer between proteins*. Proceedings of the National Academy of Sciences, **107**(26), 11799–11804 (2010).

[61] R. G. Parr & W. Yang. *Density-functional theory of atoms and molecules*, vol. 16 (Oxford university press, 1989).

[62] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai & G. Seifert. *Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties*. Physical Review B, **58**(11), 7260 (1998).

[63] D. M. Jameson & S. E. Seifried. *Quantification of protein–protein interactions using fluorescence polarization*. Methods, **19**(2), 222–233 (1999).

[64] T. Heyduk & J. C. Lee. *Application of fluorescence energy transfer and polarization to monitor escherichia coli camp receptor protein and lac promoter interaction*. Proceedings of the National Academy of Sciences, **87**(5), 1744–1748 (1990).

[65] J. J. Hill & C. A. Royer. *[19] fluorescence approaches to study of protein-nucleic acid complexation*. Methods in enzymology, **278**, 390–416 (1997).

[66] M. S. Nasir & M. E. Jolley. *Fluorescence polarization : an analytical tool for immunoassay and drug discovery*. Combinatorial Chemistry and High Throughput Screening, **2**, 177–190 (1999).

[67] A. I. S. Holm, L. M. Nielsen, S. V. Hoffmann & S. B. Nielsen. *Vacuum-ultraviolet circular dichroism spectroscopy of dna : a valuable tool to elucidate topology and electronic coupling in dna*. Physical Chemistry Chemical Physics, **12**(33), 9581–9596 (2010).

[68] D. G. Lewis & W. C. Johnson. *Circular dichroism of dna in the vacuum ultraviolet*. Journal of molecular biology, **86**(1), 91–96 (1974).

[69] L. Zhong & W. C. Johnson. *Environment affects amino acid preference for secondary structure.*. Proceedings of the National Academy of Sciences, **89**(10), 4462–4465 (1992).

[70] G. Lamoureux, E. Harder, I. V. Vorobyov, B. Roux & A. D. MacKerell. *A polarizable model of water for molecular dynamics simulations of biomolecules*. Chemical Physics Letters, **418**(1), 245–249 (2006).

[71] S. Patel, A. D. Mackerell & C. L. Brooks. *Charmm fluctuating charge force field for proteins : Ii protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model*. Journal of computational chemistry, **25**(12), 1504–1514 (2004).

[72] J. L. Banks, G. A. Kaminski, R. Zhou, D. T. Mainz, B. J. Berne & R. A. Friesner. *Parametrizing a polarizable force field from ab initio data. i. the fluctuating point charge model*. The Journal of chemical physics, **110**(2), 741–754 (1999).

[73] S. Patel & C. L. Brooks. *Charmm fluctuating charge force field for proteins : I parameterization and application to bulk organic liquid simulations*. Journal of computational chemistry, **25**(1), 1–16 (2004).

[74] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey & M. L. Klein. *Comparison of simple potential functions for simulating liquid water*. The Journal of chemical physics, **79**(2), 926–935 (1983).

[75] E. Neria, S. Fischer & M. Karplus. *Simulation of activation free energies in molecular systems*. The Journal of chemical physics, **105**(5), 1902–1921 (1996).

[76] B. T. Thole. *Molecular polarizabilities calculated with a modified dipole interaction*. Chemical Physics, **59**(3), 341–350 (1981).

[77] J.-F. Truchon, A. Nicholls, R. I. Iftimie, B. Roux & C. I. Bayly. *Accurate molecular polarizabilities based on continuum electrostatics*. Journal of chemical theory and computation, **4**(9), 1480–1493 (2008).

[78] S. Brdarski & G. Karlström. *Modeling of the exchange repulsion energy*. The Journal of Physical Chemistry A, **102**(42), 8182–8192 (1998).

[79] M. Carignano, G. Karlström & P. Linse. *Polarizable ions in polarizable water : a molecular dynamics study*. The Journal of Physical Chemistry B, **101**(7), 1142–1147 (1997).

[80] A. Holt & G. Karlström. *Improvement of the nemo potential by inclusion of intramolecular polarization*. International Journal of Quantum Chemistry, **109**(6), 1255–1266 (2009).

[81] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio Jr et al. *Current status of the amoeba polarizable force field*. The journal of physical chemistry B, **114**(8), 2549–2564 (2010).

[82] P. Ren & J. W. Ponder. *Consistent treatment of inter-and intramolecular polarization in molecular mechanics calculations*. Journal of computational chemistry, **23**(16), 1497–1506 (2002).

[83] D. Hagberg, G. Karlström, B. O. Roos & L. Gagliardi. *The coordination of uranyl in water : A combined quantum chemical and molecular simulation study*. Journal of the American Chemical Society, **127**(41), 14250–14256 (2005).

[84] A. Stone. *Distributed multipole analysis, or how to describe a molecular charge distribution*. Chemical Physics Letters, **83**(2), 233–239 (1981).

[85] J. Applequist, J. R. Carl & K.-K. Fung. *Atom dipole interaction model for molecular polarizability. application to polyatomic molecules*

and determination of atom polarizabilities. Journal of the American Chemical Society, **94**(9), 2952–2960 (1972).

[86] D. R. Garmer & W. J. Stevens. *Transferability of molecular distributed polarizabilities from a simple localized orbital based method*. The Journal of Physical Chemistry, **93**(25), 8263–8270 (1989).

[87] N. Gresh, K. El Hage, E. Goldwaser, B. De Courcy, R. Chaudret, D. Perahia, C. Narth, L. Lagardere, F. Lipparini & J.-P. Piquemal. *Addressing the issues of non-isotropy and non-additivity in the development of quantum chemistry-grounded polarizable molecular mechanics.*. Quantum Modeling of Complex Molecular Systems (2015).

[88] W. J. Stevens & W. H. Fink. *Frozen fragment reduced variational space analysis of hydrogen bonding interactions. application to the water dimer*. Chemical physics letters, **139**(1), 15–22 (1987).

[89] J. Foster & S. Boys. *Canonical configurational interaction procedure*. Reviews of Modern Physics, **32**(2), 300 (1960).

[90] P. S. Bagus, K. Hermann & C. W. Bauschlicher Jr. *A new analysis of charge transfer and polarization for ligand–metal bonding : Model studies of al4co and al4nh3*. The Journal of chemical physics, **80**(9), 4378–4386 (1984).

[91] P. S. Bagus & F. Illas. *Decomposition of the chemisorption bond by constrained variations : order of the variations and construction of the variational spaces*. The Journal of chemical physics, **96**(12), 8962–8970 (1992).

[92] K. Kitaura & K. Morokuma. *A new energy decomposition scheme for molecular interactions within the hartree-fock approximation*. International Journal of Quantum Chemistry, **10**(2), 325–340 (1976).

[93] J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke & M. L. Abrams. *Psi4 : an open-source ab initio electronic structure program*. Wiley Interdisciplinary Reviews : Computational Molecular Science, **2**(4), 556–565 (2012).

[94] T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno & C. D. Sherrill. *Levels of symmetry adapted perturbation theory (sapt). i. efficiency and performance for interaction energies*. The Journal of chemical physics, **140**(9), 094106 (2014).

[95] J. W. S. B. Rayleigh. *The theory of sound*, vol. 2 (Macmillan, 1896).

[96] E. Schrödinger. *Quantisierung als eigenwertproblem*. Annalen der physik, **385**(13), 437–490 (1926).

[97] F. London. *The general theory of molecular forces*. Trans. Faraday Soc., **33**, 8b–26 (1937).

[98] J. Hirschfelder. *Perturbation theory for exchange forces, i*. Chemical Physics Letters, **1**(8), 325–329 (1967).

[99] J. Hirschfelder. *Perturbation theory for exchange forces, ii*. Chemical Physics Letters, **1**(9), 363–368 (1967).

[100] C. Møller & M. S. Plesset. *Note on an approximation treatment for many-electron systems*. Physical Review, **46**(7), 618 (1934).

[101] B. Jeziorski, R. Moszynski & K. Szalewicz. *Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes*. Chemical Reviews, **94**(7), 1887–1930 (1994).

[102] E. G. Hohenstein. *Implementation and applications of density-fitted symmetry-adapted perturbation theory*. Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA (2011).

[103] B. I. Dunlap, J. Connolly & J. Sabin. *On some approximations in applications of xα theory*. The Journal of Chemical Physics, **71**(8), 3396–3402 (1979).

[104] M. Feyereisen, G. Fitzgerald & A. Komornicki. *Use of approximate integrals in ab initio theory. an application in mp2 energy calculations*. Chemical Physics Letters, **208**(5), 359–363 (1993).

[105] A. P. Rendell & T. J. Lee. *Coupled-cluster theory employing approximate integrals : An approach to avoid the input/output and*

storage bottlenecks. The Journal of chemical physics, **101**(1), 400–408 (1994).

[106] O. Vahtras, J. Almlöf & M. Feyereisen. *Integral approximations for lcao-scf calculations*. Chemical Physics Letters, **213**(5), 514–518 (1993).

[107] F. Weigend. *A fully direct ri-hf algorithm : Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency*. Physical Chemistry Chemical Physics, **4**(18), 4285–4291 (2002).

[108] H.-J. Werner, F. R. Manby & P. J. Knowles. *Fast linear scaling second-order møller-plesset perturbation theory (mp2) using local and density fitting approximations*. The Journal of chemical physics, **118**(18), 8149–8160 (2003).

[109] R. A. Friesner, R. B. Murphy, M. D. Beachy, M. N. Ringnalda, W. T. Pollard, B. D. Dunietz & Y. Cao. *Correlated ab initio electronic structure calculations for large molecules*. The Journal of Physical Chemistry A, **103**(13), 1913–1928 (1999).

[110] T. J. Martinez & E. A. Carter. *Pseudospectral Methods Applied to the Electron Correlation Problem*. Modern Electronic Structure Theory. Series : Advanced Series in Physical Chemistry, ISBN : <ISBN>978-981-02-1960-4</ISBN>. World Scientific Publishing Company, Edited by D R Yarkony, vol. 2, pp. 1132-1165, **2**, 1132–1165 (Sep. 1995).

[111] W. Kohn & L. J. Sham. *Self-consistent equations including exchange and correlation effects*. Phys. Rev., **140**, A1133–A1138 (Nov 1965).

[112] A. J. Misquitta, R. Podeszwa, B. Jeziorski & K. Szalewicz. *Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations*. The Journal of chemical physics, **123**(21), 214103 (2005).

[113] R. Bukowski, R. Podeszwa & K. Szalewicz. *Efficient calculation of coupled kohn–sham dynamic susceptibility functions and dis-*

persion energies with density fitting. Chemical physics letters, **414**(1), 111–116 (2005).

[114] A. Hesselmann & G. Jansen. *The helium dimer potential from a combined density functional theory and symmetry-adapted perturbation theory approach using an exact exchange–correlation potential*. Physical Chemistry Chemical Physics, **5**(22), 5010–5014 (2003).

[115] R. Podeszwa, R. Bukowski & K. Szalewicz. *Density-fitting method in symmetry-adapted perturbation theory based on kohn-sham description of monomers*. Journal of Chemical Theory and Computation, **2**(2), 400–412 (2006).

[116] A. J. Cohen, P. Mori-Sánchez & W. Yang. *Challenges for density functional theory*. Chemical Reviews, **112**(1), 289–320 (2011).

[117] A. E. Aleshin, B. Stoffer, L. M. Firsov, B. Svensson & R. B. Honzatko. *Crystallographic complexes of glucoamylase with maltooligosaccharide analogs : relationship of stereochemical distortions at the nonreducing end to the catalytic mechanism*. Biochemistry, **35**(25), 8319–8328 (1996).

[118] A. D. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. a. Ha et al. *All-atom empirical potential for molecular modeling and dynamics studies of proteins*. The journal of physical chemistry B, **102**(18), 3586–3616 (1998).

[119] W. Cornell, P. Cieplak, I. Gould, K. Merz, J. Caldwell, D. Spellmeyer & P. Kollman. *A 2nd-generation force-field for the simulation of proteins, nucleic-acids, and small molecules*. In *ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY*, vol. 204, 40–COMP. AMER CHEMICAL SOC 1155 16TH ST, NW, WASHINGTON, DC 20036 (1992).

[120] P. Claverie. *Contribution à l'étude des interactions intermoléculaires*. Ph.D. thesis, UPMC (1973).

[121] R. Rein. *On physical properties and interactions of polyatomic molecules with application to molecular recognition in biology*. Adv. Quantum Chem, **7**, 335–397 (1973).

[122] A. Stone. *Distributed multipole analysis, or how to describe a molecular charge distribution*. Chemical Physics Letters, **83**(2), 233–239 (1981).

[123] C. D. Sherrill. *Computations of Noncovalent Interactions*, vol. 26, chap. 1, 1–38 (John Wiley & Sons, Inc., 2009).

[124] A. Stone. *The theory of intermolecular forces* (Oxford University Press, 2013).

[125] A. J. Stone. *Electrostatic damping functions and the penetration energy*. The Journal of Physical Chemistry A, **115**(25), 7017–7027 (2011).

[126] B. Wang & D. G. Truhlar. *Screened electrostatic interactions in molecular mechanics*. Journal of Chemical Theory and Computation (2014).

[127] J. Rezác, K. E. Riley & P. Hobza. *S66 : A well-balanced database of benchmark interaction energies relevant to biomolecular structures*. Journal of chemical theory and computation, **7**(8), 2427–2438 (2011).

[128] F. Vigné-Maeder & P. Claverie. *The exact multicenter multipolar part of a molecular charge distribution and its simplified representations*. The Journal of chemical physics, **88**(8), 4934–4948 (1988).

[129] J. van Duijneveldt-van de Rijdt, W. Mooij & F. van Duijneveldt. *Testing the quality of some recent water–water potentials*. Physical Chemistry Chemical Physics, **5**(6), 1169–1180 (2003).

[130] A. V. Krukau, G. E. Scuseria, J. P. Perdew & A. Savin. *Hybrid functionals with local range separation*. The Journal of chemical physics, **129**(12), 124103 (2008).

[131] T. Darden, D. York & L. Pedersen. *Particle mesh ewald : An $n\log(n)$ method for ewald sums in large systems*. The Journal of chemical physics, **98**(12), 10089–10092 (1993).

[132] C. Sagui, L. G. Pedersen & T. A. Darden. *Towards an accurate representation of electrostatics in classical force fields : Efficient implementation of multipolar interactions in biomolecular simulations*. The Journal of chemical physics, **120**(1), 73–87 (2004).

[133] E. R. Smith. *Electrostatic energy in ionic crystals*. Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, **375**(1763), 475–505 (1981).

[134] J. Norberg & L. Nilsson. *On the truncation of long-range electrostatic interactions in dna*. Biophysical journal, **79**(3), 1537–1553 (2000).

[135] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan & M. Karplus. *Charmm : A program for macromolecular energy, minimization, and dynamics calculations*. Journal of computational chemistry, **4**(2), 187–217 (1983).

[136] L. Joubert & P. Popelier. *Improved convergence of the 'atoms in molecules' multipole expansion of electrostatic interaction*. Molecular Physics, **100**(21), 3357–3365 (2002).

[137] J. Wang, P. Cieplak, J. Li, Q. Cai, M.-J. Hsieh, R. Luo & Y. Duan. *Development of polarizable models for molecular mechanical calculations. 4. van der waals parametrization*. The Journal of Physical Chemistry B, **116**(24), 7088–7101 (2012).

[138] K. Autumn, M. Sitti, Y. A. Liang, A. M. Peattie, W. R. Hansen, S. Sponberg, T. W. Kenny, R. Fearing, J. N. Israelachvili & R. J. Full. *Evidence for van der waals adhesion in gecko setae*. Proceedings of the National Academy of Sciences, **99**(19), 12252–12256 (2002).

[139] L. Yang, C. Adam, G. S. Nichol & S. L. Cockroft. *How much do van der waals dispersion forces contribute to molecular recognition in solution ?*. Nature chemistry, **5**(12), 1006–1010 (2013).

[140] K. D. Shimizu. *Intermolecular forces : A solution to dispersion interactions*. Nature chemistry, **5**(12), 989–990 (2013).

[141] M. Jeziorska, W. Cencek, K. Patkowski, B. Jeziorski & K. Szalewicz. *Pair potential for helium from symmetry-adapted perturbation theory calculations and from supermolecular data.*. The Journal of chemical physics, **127**(12), 124303–124303 (2007).

[142] S. Grimme. *Semiempirical gga-type density functional constructed with a long-range dispersion correction*. Journal of computational chemistry, **27**(15), 1787–1799 (2006).

[143] S. Grimme. *Accurate description of van der waals complexes by density functional theory including empirical corrections*. Journal of computational chemistry, **25**(12), 1463–1473 (2004).

[144] B. Silvi & A. Savin. *Classification of chemical bonds based on topological analysis of electron localization functions*. Nature, **371**(6499), 683–686 (1994).

[145] A. Savin, R. Nesper, S. Wengert & T. F. Fässler. *Elf : The electron localization function*. Angewandte Chemie International Edition in English, **36**(17), 1808–1832 (1997).

[146] B. Bishop. *Lab researcher clarifies water's ionic conductivity*.

[147] a. R. Gillespie & R. Nyholm. *Inorganic stereochemistry*. Quarterly Reviews, Chemical Society, **11**(4), 339–380 (1957).

[148] R. S. Mulliken. *Electronic population analysis on lcao–mo molecular wave functions. i*. The Journal of Chemical Physics, **23**(10), 1833–1840 (1955).

[149] G. Román-Pérez, M. Moaied, J. M. Soler & F. Yndurain. *Stability, adsorption, and diffusion of ch 4, co 2, and h 2 in clathrate hydrates*. Physical review letters, **105**(14), 145901 (2010).

[150] M. J. Deible, O. Tuguldur & K. D. Jordan. *Theoretical study of the binding energy of a methane molecule in a (h2o) 20 dodecahedral cage*. The Journal of Physical Chemistry B, **118**(28), 8257–8263 (2014).

[151] B. A. Buffett. *Clathrate hydrates*. Annual Review of Earth and Planetary Sciences, **28**(1), 477–507 (2000).

[152] L. Smeller. *Pressure–temperature phase diagrams of biomolecules*. Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology, **1595**(1), 11–29 (2002).

[153] F. Izquierdo-Ruiz, A. Otero-de-la Roza, J. Contreras-García, J. Menéndez, O. Prieto-Ballesteros & J. Recio. *Guest–host interactions in gas clathrate hydrates under pressure*. High Pressure Research, **35**(1), 49–56 (2015).

[154] A. D. Becke & E. R. Johnson. *Exchange-hole dipole moment and the dispersion interaction revisited*. The Journal of chemical physics, **127**(15), 154108 (2007).

[155] A. Otero-de-la Roza & E. R. Johnson. *Van der waals interactions in solids using the exchange-hole dipole moment model*. The Journal of chemical physics, **136**(17), 174109 (2012).

[156] J. Applequist, J. R. Carl & K.-K. Fung. *Atom dipole interaction model for molecular polarizability. application to polyatomic molecules and determination of atom polarizabilities*. Journal of the American Chemical Society, **94**(9), 2952–2960 (1972).

[157] G. A. Kaminski, H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. Cao, R. B. Murphy, R. Zhou & T. A. Halgren. *Development of a polarizable force field for proteins via ab initio quantum chemistry : first generation model and gas phase tests*. Journal of computational chemistry, **23**(16), 1515–1531 (2002).

[158] J. Applequist. *An atom dipole interaction model for molecular optical properties*. Accounts of Chemical Research, **10**(3), 79–85 (1977).

[159] L. Lagardère, F. Lipparini, E. Polack, B. Stamm, M. Schnieders, E. Cances, P. Ren, Y. Maday & J.-P. Piquemal. *Scalable evaluation of polarization energy and associated forces in polarizable molecular dynamics : Ii. towards massively parallel computations using smooth particle mesh ewald*. Journal of Chemical Theory and Computation (2015).

[160] U. K. Eriksson, G. Fischer, R. Friemann, G. Enkavi, E. Tajkhorshid & R. Neutze. *Subangstrom resolution x-ray structure*

details aquaporin-water interactions. Science, **340**(6138), 1346–1349 (2013).

[161] H. Sui, B.-G. Han, J. K. Lee, P. Walian & B. K. Jap. *Structural basis of water-specific transport through the aqp1 water channel*. Nature, **414**(6866), 872–878 (2001).

[162] P. Chaquin & F. Fuster. *Orbimol*.

[163] J. Murrell, M. Randic & D. Williams. *The theory of intermolecular forces in the region of small orbital overlap*. In *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*, vol. 284, 566–581. The Royal Society (1965).

[164] N. Gresh, P. Claverie & A. Pullman. *Computations of intermolecular interactions : Expansion of a charge-transfer energy contribution in the framework of an additive procedure. applications to hydrogen-bonded systems*. International Journal of Quantum Chemistry, **22**(1), 199–215 (1982).

[165] N. Gresh, P. Claverie & A. Pullman. *Intermolecular interactions : Elaboration on an additive procedure including an explicit charge-transfer contribution*. International journal of quantum chemistry, **29**(1), 101–118 (1986).

[166] N. Gresh. *Energetics of zn2+ binding to a series of biologically relevant ligands : A molecular mechanics investigation grounded on ab initio self-consistent field supermolecular computations*. Journal of computational chemistry, **16**(7), 856–882 (1995).

[167] N. Gresh, J. E. Šponer, N. Špa ?ková, J. Leszczynski & J. Šponer. *Theoretical study of binding of hydrated zn (ii) and mg (ii) cations to 5'-guanosine monophosphate. toward polarizable molecular mechanics for dna and rna*. The Journal of Physical Chemistry B, **107**(33), 8669–8681 (2003).

[168] A. J. Stone & A. J. Misquitta. *Charge-transfer in symmetry-adapted perturbation theory*. Chemical Physics Letters, **473**(1), 201–205 (2009).

[169] A. J. Misquitta. *Charge transfer from regularized symmetry-adapted perturbation theory*. Journal of Chemical Theory and Computation, **9**(12), 5313–5326 (2013).

[170] R. A. Buckingham. *The classical equation of state of gaseous helium, neon and argon*. In *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*, vol. 168, 264–283. The Royal Society (1938).

[171] F. Sicheri & D. Yang. *Ice-binding structure and mechanism of an antifreeze protein from winter flounder* (1995).

[172] C. A. Knight, A. L. De Vries & L. D. Oolman. *Fish antifreeze protein and the freezing and recrystallization of ice* (1984).

[173] T. Ha-Duong, N. Basdevant & D. Borgis. *A polarizable coarse-grained water model for coarse-grained proteins simulations*. Chemical Physics Letters, **468**(1), 79–82 (2009).

[174] S. O. Yesylevskyy, L. V. Schäfer, D. Sengupta & S. J. Marrink. *Polarizable water model for the coarse-grained martini force field*. PLoS Comput Biol, **6**(6), e1000810 (2010).

[175] S. Riniker & W. F. van Gunsteren. *A simple, efficient polarizable coarse-grained water model for molecular dynamics simulations*. The Journal of chemical physics, **134**(8), 084110 (2011).

[176] Y. Wang, S. Izvekov, T. Yan & G. A. Voth. *Multiscale coarse-graining of ionic liquids*. The Journal of Physical Chemistry B, **110**(8), 3564–3575 (2006).

[177] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan & W. Yang. *Nciplot : a program for plotting noncovalent interaction regions*. Journal of chemical theory and computation, **7**(3), 625–632 (2011).

[178] S. Noury, X. Krokidis, F. Fuster & B. Silvi. *Topmod package*. Universite Pierre et Marie Curie (1997).

[179] H. Matter, M. Nazaré, S. Güssregen, D. W. Will, H. Schreuder, A. Bauer, M. Urmann, K. Ritter, M. Wagner

& V. Wehner. *Evidence for c ? cl/c ? br ? ? ? π interactions as an important contribution to protein–ligand binding affinity*. Angewandte Chemie, **121**(16), 2955–2960 (2009).

[180] L. A. Hardegger, B. Kuhn, B. Spinnler, L. Anselm, R. Ecabert, M. Stihle, B. Gsell, R. Thoma, J. Diez, J. Benz et al. *Systematic investigation of halogen bonding in protein–ligand interactions*. Angewandte Chemie International Edition, **50**(1), 314–318 (2011).

[181] A. C. Van Duin, S. Dasgupta, F. Lorant & W. A. Goddard. *Reaxff : a reactive force field for hydrocarbons*. The Journal of Physical Chemistry A, **105**(41), 9396–9409 (2001).

[182] W. A. Goddard III, A. Van Duin, K. Chenoweth, M.-J. Cheng, S. Pudar, J. Oxgaard, B. Merinov, Y. H. Jang & P. Persson. *Development of the reaxff reactive force field for mechanistic studies of catalytic selective oxidation processes on bimoo x*. Topics in Catalysis, **38**(1-3), 93–103 (2006).

Résumé

Le but de cette thèse est une revisite du potentiel SIBFA. Ceci inclue un travail et une réflexion sur la méthodologie de cette approche avec une implémentation proposant une utilisation plus large. De plus, une nouvelle calibration de champ de forces raffiné est permise aujourd'hui. En effet, la décomposition d'énergie intermoléculaire SAPT donne accès à toutes les composantes avec rigueur. La reproduction des résultats *ab-initio* par un potentiel analytique laisse entrevoir des applications prometteuse. Au delà du coup de calcul considérablement réduit par rapport aux méthodes de chimie quantique, son intégration dans un code de dynamique moléculaire ouvre les portes à de nombreuses études encore plus prometteuses hors de portée de la chimie quantique. Enfin l'optimisation de ce code, avec une parallélisation bien étudiée, en feront un outil majeur de la biochimie. Dans une première partie, nous introduirons les notions et principes essentiels à la dynamique moléculaire. Un premier chapitre exposera la mécanique classique utilisé dans les programmes les plus distribués et utilisés. Un second chapitre introduira les méthodes permettant un meilleur traitement des interactions non-covalentes essentielles dans les études de complexes ligand-recepteur. Une seconde partie abordera de manière plus concrète la stratégie d'implémentation de SIBFA dans Tinker. Celle-ci s'organisera autour de trois chapitre, traitant chaque composante énergétique intermoléculaire.
L'objectif de cette thèse est de proposer un socle solide autour du traitement des interactions non covalentes dans le cadre des champs de forces polarisables de dernières générations et de présenter le modèle d'eau hybride AMOEBA/SIBFA.

Mots-clés: Champs de Forces Polarisables, Dynamique Moléculaire, Décomposition d'énergie, Interactions intermoléculaires.