Université de Montréal

Direction des bibliothèques

AVIS

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

This document was digitized by the Records Management & Archives Division of Université de Montréal.

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Modéliser la polarisation électronique par un continuum diélectrique intramoléculaire
Vers un champ de force polarisable pour la chimie bioorganique

par
Jean-François Truchon

Département de chimie
Faculté des Arts et des Sciences

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

décembre, 2008

Faculté des arts et des sciences
Grade conféré à compter du
06 AOUT 2009
Université de Montréal

© Jean-François Truchon, 2008

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée :

Modéliser la polarisation électronique par un continuum diélectrique intramoléculaire

Vers un champ de force polarisable pour la chimie bioorganique

présentée par :

Jean-François Truchon

a été évaluée par un jury composé des personnes suivantes :

Matthias Ernzerhof, président-rapporteur

Radu I. Iftimie, directeur de recherche

Christopher I. Bayly, codirecteur

Benoît Roux, codirecteur

Michel Lafleur, membre du jury

Enrico Purisima, examinateur externe

Normand Mousseau, représentant du doyen de la FES

Résumé

Cette thèse présente l'étude d'une nouvelle méthodologie appelée EPIC (*Electronic Polarization from the Internal Continuum*) pour inclure la polarisation électronique dans un champ de force de mécanique moléculaire. Un continuum diélectrique intramoléculaire permet de modéliser l'induction électronique avec précision. L'outil mathématique employé repose sur l'équation de Poisson, issue de l'électrostatique classique, et sur une fonction de diélectrique qui définit un volume moléculaire polarisable. La fonction de diélectrique moléculaire, élément principal du modèle, est construite à l'aide de rayons atomiques et d'un diélectrique intérieur, des paramètres empiriques ajustables.

Une nouvelle formule pour calculer le tenseur de polarisabilité d'un volume diélectrique de forme quelconque permet d'optimiser les paramètres empiriques choisis. L'accord entre les polarisabilités moléculaires du modèle EPIC avec l'expérience ou la mécanique quantique dépasse les précédents de la littérature, en particulier au niveau du nombre inférieur de paramètres ajustables requis. L'anisotropie de la polarisabilité est obtenue sans la complexité des modèles polarisables déjà utilisés en mécanique moléculaire. La validation est effectuée sur un ensemble de 707 tenseurs de polarisabilité calculés avec B3LYP aux fins de la présente étude. Le modèle obtenu couvre ainsi une grande partie des fonctionnalités chimiques propres aux biomolécules et à la chimie bioorganique.

Des calculs du potentiel électrostatique induit par une perturbation électrique locale indiquent que le modèle EPIC et la mécanique quantique sont en excellent accord. De plus, un protocole général pour l'optimisation des charges atomiques partielles, baignant dans un volume diélectrique moléculaire, est dérivé. Ces avancements sont mis à l'épreuve par le calcul de la courbe d'énergie potentielle d'interaction d'un complexe cation-π et du pont-H formé par un dimère de 4-pyridone – applications où la polarisabilité est essentielle.

Le modèle EPIC est ajusté sur des molécules isolées. Alors, la réutilisation de ces paramètres pour la phase condensée est vérifiée en comparant des indices de réfraction expérimentaux avec la constante diélectrique optique calculée en appliquant EPIC sur les molécules de gouttes liquides obtenues par dynamique moléculaire. La corrélation montre une pente unitaire et un coefficient de corrélation de 0.95.

Des calculs d'énergies libres d'hydratation avec 485 solutés montrent que le potentiel électrostatique permanent, l'induction électronique et la polarisation d'un solvant implicite peuvent être obtenus avec le même ensemble de paramètres empiriques. Cela démontre la justesse des fondements physiques du modèle présenté. Nos travaux sur les solvants implicites nous ont amenés à revoir certaines idées préconçues sur la signification des paramètres traditionnellement utilisés et à proposer une fonction de diélectrique à trois zones qui reflète mieux les différents phénomènes physiques en présence. Le découplage de chacune des étapes de paramétrisation est un avantage considérable pour généraliser le modèle EPIC.

Un champ de force polarisable général et fiable a le pouvoir d'améliorer la prédictibilité des simulations moléculaires. Le modèle électrostatique EPIC est précis et requiert peu de paramètres ajustables, ce qui pourra en faire la pierre angulaire pour la mise au point d'un champ de force polarisable général et précis. Cela pourrait avoir des retombées importantes pour le design de médicaments, l'étude des processus biologiques, etc.

Mots-clés : Champ de force, mécanique moléculaire, polarisabilité, polarisation électronique, potentiel électrostatique, diélectrique, Poisson-Boltzmann, paramétrisation, DRESP, EPIC.

Abstract

This thesis presents the study of a new approach, called EPIC (Electronic polarization from the Internal Continuum), to include electronic polarization in molecular mechanical force fields. An intramolecular dielectric continuum is shown to accurately model the electronic induction. EPIC is based on Poisson's equation, from classical electrostatics, and makes use of a dielectric function that varies in space, defining a polarizable molecular volume. The obtained partial differential equation is solved with a finite difference algorithm. The molecular dielectric function, central to EPIC, is built with atomic radii and an internal dielectric constant: empirical and adjustable parameters.

The chosen empirical parameters are optimized with a new numerical procedure for the calculation of the polarizability tensor resulting from a dielectric volume. The agreement between the EPIC molecular polarizabilities and those obtained with quantum mechanics or experiment exceeds literature precedents, especially because of the much smaller number of fitted parameters required. The polarizability anisotropy is accurately modeled without the extra complexity necessary in previous polarizable models. EPIC is validated on a dataset containing 707 polarizability tensors calculated with B3LYP for this study. Thereby, the presented optimized parameters can account for the polarizability of a wide variety of functional chemical groups found in biomolecules and bioorganic chemistry.

The assessment of the electrostatic potential induced by a local electric perturbation shows excellent agreement between EPIC and quantum mechanics. Furthermore, a new general approach for the calculation of atomic partial charges placed in a dielectric volume is derived. These progress are tested with the calculation of the potential energy dissociation curve for a cation- π system and a H-bonded 4-pyridone dimer, both shown to be strongly dependent on polarizability.

The EPIC model is fitted uniquely on isolated molecules. Hence, the transferability of the obtained parameters to the condensed phase is verified by comparing experimental refractive indices with the optical dielectric constants that are calculated by applying EPIC at a molecular level on liquid droplets obtained from molecular dynamic simulations. The correlation shows a unitary slope and a coefficient of 0.95.

Free energy of hydration calculations on 485 solutes show that the permanent electrostatic potential, the electronic induction and the electrical response from an implicit solvent model are simultaneously obtained with the same set of parameters. This demonstrates the physical soundness of the EPIC model. The research on implicit solvent models forces us to challenge certain dogmas on the signification of traditionaly used parameters and to propose a 3-zone dilectric function that better reflects the underlying physical principles. The decoupling of each of the fitting steps is a considerable advantage for the generalisation of the EPIC model.

A general polarizable force field has the potential to greatly improve the predictive power of molecular simulations. The accurate electrostatic EPIC model needs few adjustable parameters and, hence, could form the cornerstone for the development of a general and more accurate polarizable force field, which could have important impacts in many areas such as drug design, biological processes understanding, etc.

Keywords : Force field, molecular mechanics, polarisability, electronic polarisation, electrostatic potential, dielectric, Poisson-Boltzmann, parameterization, DRESP,EPIC.

Table des matières

1 INTRODUCTION 1

1.1 La mécanique quantique 4

1.2 La polarisabilité moléculaire 12

1.3 La mécanique moléculaire 17

1.4 Champs de force polarisables : précédents de la littérature 22

1.5 Un solvant implicite à partir du continuum diélectrique 28

1.6 Le modèle polarisable proposé 36

1.7 Problèmes étudiés dans cette thèse 40

1.8 Bibliographie 42

2 ÉLABORATION D'UN MODÈLE PRÉCIS ET RAPIDE FONDÉ SUR L'ÉLECTROSTATIQUE CONTINUE POUR LA PRÉDICTION DE LA POLARISABILITÉ MOLÉCULAIRE 49

2.1 Introduction 53

2.2 Existing empirical polarizable models 55

2.2.1 Point inducible dipole 55

2.2.2 Drude oscillator 55

2.2.3 Fluctuating charges 56

2.2.4 Limitations with the PID related methods 56

2.3 Dielectric polarizability model 57

2.3.1 The model 57

2.3.2 Spherical dielectric 59

2.4 Methods 61

2.4.1 Calculations 61

2.4.2 Fitting procedure 62

2.4.3 Definitions 63

2.4.4 Molecule datasets 64
2.5 Results 66
2.5.1 Diatomics: the Cl_2 polarizability hypersurface 66
2.5.2 Diatomics: polarizability 69
2.5.3 Organic datasets: typical PB parameters 71
2.5.4 Alkanes and aromatics 72
2.5.5 Conformational dependency of polarizability 78
2.6 Discussion 79
2.6.1 Transferability 79
2.6.2 Inner dielectrics 80
2.6.3 Link to the optical dielectric constants 81
2.7 Conclusion 83
2.8 Bibliography 85

3 UTILISER LA POLARISATION PROVENANT DU CONTINUUM INTERNE POUR LES INTERACTIONS INTERMOLÉCULAIRES 94

3.1 Introduction 98
3.2 Methods 99
3.2.1 A Least-Squares Method 100
3.2.2 Computational details 103
3.2.3 Induced polarization 104
3.2.4 Molecule dataset 105
3.3 Results and discussion 105
3.3.1 DRESP vs RESP 107
3.3.2 Use of an existing charge model: the AM1-BCC/DRESP example 108
3.3.3 Induced electrostatic potential 110
3.3.4 Induction by a symmetric field 112

3.3.5 Cation-π interactions 115
3.3.6 H-bond of the pyridine-4(1*H*)-one dimer 118
3.4 Conclusion 124
3.5 Supporting Information 127
3.6 Bibliography 130

4 APPROCHES INTEGREES FONDEES SUR UN CONTINUUM DIELECTRIQUE POUR TRAITER LA POLARISABILITE MOLECULAIRE ET LA PHASE CONDENSEE: INDICE DE REFRACTION ET LA SOLVATATION IMPLICITE 137

4.1 Introduction 141
4.2 Theory and Methods 143
4.2.1 3-Zone dielectric in implicit solvents 143
4.2.2 Molecular polarizability tensor 148
4.2.3 Refractive index calculations 153
4.2.4 Free energy of hydration 156
4.2.5 Quantum calculations 160
4.3 Datasets 160
4.3.1 Polarizability training dataset (PTD) 160
4.3.2 Polarizability validation dataset 161
4.3.3 Polarizability dataset 161
4.3.4 Hydration free Energy Dataset 161
4.3.5 Refractive indices dataset 162
4.4 Results and discussion 162
4.4.1 Polarizability tensor 162
4.4.2 Refractive indices 172
4.4.3 Hydration free energies 175
4.5 Conclusion 181

4.6 Acknowledgment 184
4.7 Supporting Information 184
4.8 Bibliography 185

5 CONCLUSION 195

ANNEXE I DÉCLARATIONS DES COAUTEURS I

ANNEXE II DÉRIVATION DE L'ÉQUATION DE POISSON IV

ANNEXE III OÙ LA DENSITÉ DE CHARGE INDUITE APPARAÎT-ELLE? IX

ANNEXE IV UNITÉS DES CALCULS PAR DIFFÉRENCES FINIES XI

ANNEXE V POLARISABILITÉS MOLÉCULAIRES B3LYP XV

Liste des tableaux

1.1. Permittivité relative (ε_{solv}) expérimentale de quelques liquides (20°C et 1 atm) 33

2.1. Compared polarizabilities (a.u.) of diatomic molecules when the radii and ε_{in} are fit to B3LYP/aug-cc-pVTZ polarizabilities – two fitting methods are involved: 1 radius and 1 dielectric per element, 1 radius per element and a single dielectric for all five..... 70

2.2. Unsigned average errors for all molecules in Figure 2.2, relative to B3LYP/aug-cc-pVTZ, of average polarizability and anisotropy obtained with various parameters typically used in PB applications. 72

2.3. Optimized radii (Å) and inner dielectrics with sensitivity accounting for all molecule sets (Figure 2.2) – parameter sets P2E and P1E..... 73

2.4. Average errors and stdev against experiment[a] for all molecules in Figure 2.2..... 78

4.1. Reported optimal polarization radii (σ_{in}) and atom typing for the four G1 sets defining the internal dielectric (c.f. eq.(4.1))..... 166

4.2. Error obtained with the optimized polarization radii of the G1 sets when EPIC molecular polarizability tensors are compared to B3LYP for different molecule datasets. 170

4.3. Solvent cavity atomic radii (σ_{cavity}) and γ for the 3-zone dielectric model optimized on 485 experimental free energy of hydration with different G1-ε_{in} solute models and ΔG_{np} sources..... 177

4.4. Effects of using different solvent cavity radii set (Table 4.3) with the G1-12 solute model (Table 4.1) on ΔG_{hyd}..... 179

Liste des figures

1.1. Le potentiel électrostatique permanent φ^0 (a) d'une molécule d'eau dans le plan des atomes est perturbée par un monopole localisé à différents endroits. Le potentiel total φ est donné à gauche (b et d) et le potentiel induit φ^{ind} à droite (c et e). Un potentiel très positif est rouge et très négatif est bleu négatif. Le monopole induit un dipôle, en polarisant les électrons de la molécule d'eau, qui diffère selon sa position et déforme le potentiel électrostatique de façon qualitativement et quantitativement appréciable. Obtenu avec EPIC. 11

1.2 Polarisabilité selon les axes principaux en unités atomiques (calculs B3LYP) pour quatre molécules. Le cercle indique l'axe qui sort du plan de la feuille. 14

1.3. Potentiel électrostatique de la molécule 4-pyridone obtenu (a) avec des charges atomiques partielles selon l'éq. (1.31) et (b) par la mécanique quantique où l'éq. (1.14) est calcué avec HF/6-31G*. L'approximation des charges atomiques permet de reproduire précisément le potentiel électrostatique produit par la densité électronique et les noyaux ($\Gamma(\vec{r})$) à une distance où les interactions intermoléculaires prennent place. 22

1.4. Illustration du modèle des dipôles atomiques induits pour une molécule de benzéne avec des atomes de carbone polarisables qui subissent l'influencent d'un champ électrique externe complexe (flèches bleueus). Les charges atomiques ne sont pas montrées, mais contribuent à induire les dipôles atomiques en produisant un champ électrique interne qui s'ajoute à celui créé par les dipôles induits et au champ électrique externe. 25

1.5 Le modèle de solvatation implicite suggère que le champ électrique produit par un soluté polaire organisera les configurations des molécules d'eau (a) de telle sorte que l'effet moyen peut être décrit par un champ de polarisation du solvant (b). L'accumulation des charges partielles positives des H et négatives des O de l'eau sont illustrées par du rouge et du bleu respectivement. 29

1.6 Une cavité dans un diélectrique est formée pour une molécule de benzène. Des sphères sont positionnées sur les atomes de la molécule avec différents rayons. L'enveloppe extérieure défini la cavité. Il s'agit d'une surface dite de van der Waals (vdW). 34

1.7. Cette figure illustre le concept d'un continuum polarisable. Un champ électrique externe est appliqué (flèches bleues) à la molécule de benzène (a). L'intérieur du volume polarisable qui enveloppe la molécule est polarisé selon les lignes du champ électrique externe. À chaque point de l'espace, un dipôle infinitésimal est induit. Si le champ externe n'est plus uniforme (b), les dipôles induits s'alignent toujours selon le champ externe. 37

1.8 Les lignes du champ électrique total sont représentées pour un diélectrique sphérique dans le vide qui subit un champ électrique externe uniforme orienté selon $\hat{x}$ (a) À l'intérieur de la sphère le champ électrique est inférieur au champ externe dû à la polarisation. Le potentiel électrostatique induit en (b), produit par la polarisation de la sphère, est celui d'un dipôle ponctuel placé au centre de la sphère et résulte de charges liées (*bound charges*) accumulées à la surface de la sphère. 38

2.1. The dielectric contribution to the sphere dielectric continuum polarizability goes asymptotically to one and most of the contributions are below εin = 10..... 60

2.2 The molecules used are divided in 12 datasets and six chemical classes: the heteroaromatics training set 'aromatics-t' (a), the heteroaromatics validation set 'aromatics-v' (b), the pyridones training set 'pyridones-t' (c), the pyridones validation set 'pyridones-v' (d), the furans training set 'furans-t', the pyrroles training set 'pyrroles-t', the thiophenes training set 'thiophenes-t' (e), the furans validation set 'furans-v', the pyrroles validation set 'pyrroles-v', the thiophenes validation set 'thiophenes-v' (f), the alkanes training set 'alkanes-t' (g) and the alkanes validation set 'alkanes-v' (h). In the case of n-butane, n-hexane and n-octane, two conformers are considered: all trans (t) and gauche (g). The X atoms in a molecule are either all O, all S, or all NH. 65

2.3. The EPIC model behavior is explored for Cl_2. The average polarizability (a) and the anisotropy (b) isolines (in a.u.) are plotted as a function of the Cl atomic radius, used to define the vdW surface, and the value of the inner dielectric. The target Cl_2 B3LYP values are 31.43 (average) and 18.24 (anisotropy) (c.f.Table 2.1). The polarizability tensor error function isolines in (c) identify the regions where the EPIC model matches the B3LYP polarizability tensor. The external dielectric is set to one and the inter-nuclear distance of Cl_2 is fixed at 2.05Å. These figures show that a high dielectric value is required to match the QM anisotropy, and that a number of minima can be found on the error hypersurface. 68

2.4 Comparison between B3LYP/aug-cc-pVTZ polarizabilities and EPIC models P2E and P1E for all molecules from Figure 2.2. The averaged relative error on average polarizability (eqn (2.10)), anisotropy (eqn (2.9)) and the deviation angle of the eigenvectors (eqn (2.11)) are shown together with the corresponding STDEV reported as error bars. The results for the 2-dielectric fit (P2E) training sets (a) and validation sets (b) show small errors in the average polarizability and relatively small errors in the anisotropy. The results for the 1-dielectric fit (P1E) training sets (c) and the validation sets (d) show larger errors in the alkanes anisotropy and generally larger errors than the P2E parameters (shown under combined P2E). Combined errors of the training and validation sets are similar..... 75

2.5. Correlation between B3LYP/aug-cc-pVTZ polarizability components and the EPIC model P2E. In (a), the polarizability components for all sets of Figure 2.2 are correlated and the ±10% error lines are illustrated. The linear regression shows excellent agreement, especially for polarizabilities smaller than 150 a.u. In (b), 13 stable conformers of n-octane are examined. The all trans conformation polarizabilities are identified with circles. The average polarizability error on the 13 conformers is 1.9% and the anisotropy error is 5.8%. A linear regression gives a R^2 of 0.997, a slope of 1.21 and an ordinate at the origin of -19.5. This means that the EPIC

model P2E overestimates the polarizability of n-octane consistently through all conformers. 76

3.1. Molecule dataset which contains aryls and alkanes chemical classes. These 15 aromatic and 21 alkane molecules are extracted from reference [1]. We use the notation 't' to indicate the all-trans conformation and 'g' when one or more gauche dihedrals are present; these are separate entries. 106

3.2. Benzene, pyridine and cyclopropane optimal charges fitting equally the same electrostatic potential with a dielectric of one (RESP, non-polarizable) and the P2E model (DRESP). The significantly higher charges with the P2E model comes from the internal dielectric screening of the point charges..... 108

3.3. Correlation plot of the RRMSD obtained with AM1-BCC and AM1-BCC/DRESP charging schemes. The RRMSD are calculated against the B3LYP permanent electrostatic potential on the FCC grid..... 109

3.4. Average RRMSD on the induced ESP maps as calculated with EPIC using three parameter sets (P2E, P1E and Bondi, see the text). The ESP maps are generated by a +0.5e located at non-redundant positions and the reference induced ESP is calculated from B3LYP 6-311++G(3df,3dp)..... 111

3.5. The induced electrostatic potential of an argon atom sandwiched in between two +0.5e charges positioned at 3.0 Å of the nucleus. The solid iso-surface corresponds to the B3LYP/6-311++G(3df,3pd) and the mesh iso-surface to the result from EPIC (radius=1.3 Å, dielectric=7.4). The induced moment is an induced quadrupole with the d_z^2 orbital symmetry. The traditional atomic polarizable approaches have a zero induced electrostatic potential..... 114

3.6. The induced electrostatic potential for benzene is shown by iso-contour lines spaced by 5kcal/mol/*e* on one of the six symmetry plan perpendicular to the ring. The external perturbing potential is produced by two +1*e* charges positioned 2Å above and below the benzene ring. The induced potential obtained at the B3LYP/6-311++G(3df,3pd)

level (a) is compared to the EPIC/P2E (b) and AMOEBA, a good quality QM derived point-inducible model (c). 117

3.7. The electrostatic interaction energy between a benzene molecule and H+ displaced along the C_6 symmetry axis corresponds to the electrostatic component of a cation-π system formed with an atomic cation. The non-polarizable model using ESP derived charges is far from the B3LYP calculated energy, the EPIC/P2E model closely follows the B3LYP curve and AMOEBA captures most of the electronic polarization of B3LYP. The EPIC electrostatics fitted on B3LYP monomer reproduces the correct cation-π electrostatic energy without adjustment 119

3.8. The reported electrostatic interaction energies of the H-bonded 4-pyridone dimer (a) show that the EPIC/P2E model produces the appropriate polarization as opposed to the non-polarizable permanent charge model (Coulomb) when compared to B3LYP (3.15). The EPIC/Bondi calculations produce the correct electronic response at long ranges of H-bond distances but saturates as the vdW dielectric surfaces of the monomers start overlapping. The observed deviations are a result of the numerical instability that occur when the dielectric spheres come into contact at 2.6 Å. (b) The BSSE/corrected B3LYP interaction energies of the dimer unveils a very strong H-bond of -10.8 kcal/mol at the minimum located at 1.78 Å. The reported classical approaches combined the electrostatic energies (shown in a) and a fitted repulsive vdW term. EPIC/P2E matches the B3LYP energies over the examined range whereas the Coulomb non-polarizable model deviates at longer distances as a result of the difficulty for such a model to match both regions. 123

4.1. This figure shows the smooth dielectric function used in this work for a single atom with $\sigma = 0.95$Å, $\varepsilon_{in} = 12$, $\varepsilon_{ext} = 1$ and $A = 10.0$ (Cl of the G1-12 set). Starting from the center of the atom (r = 0), the dielectric (blue curve) stays constant until the 'density', expressed with a sum of Gaussians (pink curve), reaches a certain small value that causes the dielectric to smoothly transition to the external dielectric value. The steepness and the position of the switching region depends on the value of the A

parameter. The sum-of-Gaussians density expression is explained in equation 4.2 (see text). 145

4.2. The 3-zone dielectric function allows an accurate description of both the solute polarization and the solvent polarization within the EPIC approach. (a) The radial component of the dielectric for a single atom (G1-12 aromatic carbon) is shown together with the polarization (σ_{in}) and the solvent cavity (σ_{cavity}) atomic radii. Each plateau of the dielectric function defines a zone. The intermediate zone corresponds to the solute/solvent contact distance. (b) The resulting dielectric function is also shown in the ring plane of 4-pyridone (b) when applying the G2-12 parameters..... 147

4.3. The iso-contour plot of the RRMS error between B3LYP/aug-cc-pVTZ and EPIC polarizability tensors are shown as a function of the ε_{in} and A parameters of eq. (4.1). This RRMS surface was generated from a simultaneous fit of the H, alkyl C, aromatic C and aromatic N atomic polarization radii on training set of 11 aromatic and 14 alkane molecules against their B3LYP polarizabilities. It shows that in order for a single dielectric model to fit the polarizabilities of these two chemical classes to within 10% error, the ε_{in} needs to be sufficiently large (>6). Deviations in the anisotropy of the polarizability are the main source of error for lower values of ε_{in}. 164

4.4. Correlation graph between the B3LYP/aug-cc-pVTZ directional polarizabilities (α_1 black circles, α_2 red triangles and α_3 green squares in a.u.) for three G1 dielectric parameter sets (c.f. Table 4.1). Each figure shows the data for 707 molecules for a total of 2121 points. From these figures, it is clear that a small number of parameters (optimized on 265 molecules) can generalize well. A large ε_{in}= 24 (a) produces the best fit, a medium range ε_{in}= 12 produces slightly larger discrepancies and a small ε_{in}= 4 produces significantly larger deviations, in keeping with the results of the range-finding study on the small dataset. 171

4.5. The calculated refractive indices (n) of 23 organic molecules are compared to experiment. Three dielectric parameter sets are used a) G1-24 b) G1-12 and c) G1-4 (Table 4.1). For each set, the pre-optimized radii can be found in Table 4.1. The

reported refractive indices (n) were obtained by polarizing a liquid droplet formed by carving spheres from periodic MD liquid simulation snapshots. The Clausius-Mossoti equation leads to $n^2 = \varepsilon_\infty$ close to experiment, in spite of the large ε_{in}. The predicted values are systematically higher than experiment, which can be explained by potential artifacts or a polarizability shift when passing from vacuum to condensed phase (see text). As with the polarizabilities, the predictions deteriorate with decreasing ε_{in}, in keeping with the results of the range-finding study on the small dataset. 173

4.6. One of the 50 CCl_4 droplets is cut in its center and three dielectric functions (eq. (4.1)) are plotted: a) G1-24 b) G1-12 and c) G1-4. The red color is associated with $\varepsilon(r)=\varepsilon_{in}$ and blue to $\varepsilon(r)=1$. 174

Liste des sigles

AE	Average error
AUE	Averaged unsigned error
B3LYP	Fonctionnelle d'échange-corrélation Becke-Lee/Yang/Parr
D	Debye (dipole moment units)
DFT	Density functional theory
DRESP	Dielectric RESP
EPIC	Electronic Polarization from Internal Continuum
ESP	Electrostatic potential
PB	Poisson-Boltzmann
PE	Poissons's Equation
RESP	Restrained ESP
RMS	Root-mean-square deviation
RRMS	Relative root-mean-square deviation
S.I.	Système international d'unités
TFD	Théorie de la fonctionnelle de la densité
u.a./a.u.	Système des unités atomiques / *atomic units*

Notation

$\vec{v}$	Vecteur tridimensionnel
$\bar{\bar{M}}$	Matrice M
$\hat{O}$	Opérateur différentiel O
$\langle O \rangle$	Valeur moyenne de l'observable O
E_x	La composante x du vecteur $\vec{E}$.
$\vec{\nabla} \cdot \vec{E}(\vec{r})$	La divergence du champ vectoriel $\vec{E}$ au point $\vec{r}$
$\vec{\nabla} f(\vec{r})$	Le gradient de la fonction f au point $\vec{r}$
$\nabla^2 f(\vec{r})$	Le laplacien de la fonction f au point $\vec{r}$
ε_0	Permittivité du vide
ε	Permittivité relative au vide / diélectrique
ε_∞	Constante diélectrique optique
U	Énergie
$\vec{E}$	Champ électrique
φ	Potentiel électrostatique
ρ	Densité de charge
$\delta(\vec{r})$	Fonction delta de Dirac
α	Polarisabilité
h	Constante de Planck

μ	Moment dipolaire
q	Charge
n	Indice de réfraction
χ	Susceptibilité électrique
$\vec{P}$	Polarisation électrique

À Sheila, Julianne et Catherine

Remerciements

Je voudrais d'abord remercier mes trois directeurs et codirecteurs de thèse. Merci à Christopher Bayly pour sa disponibilité, sa grande générosité, sa patience et son support si important tout au long de mes travaux. Son mentorat laissera une empreinte majeure sur ma façon de faire de la science. Je voudrais également remercier Radu Iftimie qui m'a accueilli dans son laboratoire à bras ouverts. Sa disponibilité, sa facilité à transmettre les connaissances, sa gentillesse et la confiance dont il m'a fait preuve furent très appréciées. Merci à Benoît Roux pour son temps, sa contribution scientifique et son enseignement.

Je ne pourrais passer sous silence l'aide qu'Anthony Nicholls m'a apportée tout au long de mes travaux. Son aide, ses conseils et sa connaissance profonde du sujet ont été d'une importance capitale pour l'avancement des travaux de recherche et ma formation. Il a fait preuve d'une rare générosité à mon égard.

J'aimerais remercier mon employeur, Merck Frosst, de m'avoir permis de me concentrer exclusivement sur les travaux de recherche qui font l'objet de cette thèse dans le cadre du Merck Research Laboratories Doctoral Program I. Je me dois de remercier le Conseil de Recherches en Sciences Naturelles et Génie pour une bourse d'études supérieures du Canada. Il faut souligner également les ressources informatiques mises à ma disposition par le Réseau Québécois de Calcul Haute Performance et Merck & Co. Merci à la compagnie OpenEye Inc. pour des licences gratuites pour une panoplie de bibliothèques informatiques qui ont constitué le cœur de beaucoup d'outils calculatoires que j'ai écrits.

Je tiens à remercier mes collègues de l'Université de Montréal: Patrick Maurer, Vibin Thomas et Titus Sandu pour les nombreuses discussions scientifiques. Un merci particulier à François 'Ti Pou' Goyer et Étienne Lanthier pour leur camaraderie et les dîners

parfois colorés. Merci à mes collègues de Merck Frosst: Daniel McKay et Sathesh Bhat pour leur expertise et leur amitié.

J'aimerais remercier mes parents, Lise Truchon et Luc Dubé, pour leur amour, l'énergie qu'ils ont mise à me faire grandir et les outils qu'ils m'ont donnés. Merci à mes grands-parents Sabin Truchon, Andrée Saint-Louis et Gabrielle Jean pour leur affection et leur support spécial pendant mes études.

Merci à mon épouse Sheila Perriard d'avoir révisé ce document. J'aimerais souligner son amour, son dévouement exceptionnel, son aide et ses encouragements dans les moments plus difficiles. Merci pour ces deux belles filles, Julianne et Catherine, qui embellissent ma vie et font vibrer mon âme.

1 Introduction

La chimie est la science qui se penche sur la composition de la matière, de sa transformation par des réactions chimiques et de son comportement. L'élément d'étude est l'atome, une entité d'une dimension si petite qu'elle dépasse l'entendement humain. À travers les siècles, l'existence de l'atome a été le sujet de vifs débats entre philosophes ou scientifiques respectés. Fort étrangement, ce sont les physiciens qui ont mis en évidence le monde atomique de façon définitive. En particulier, la théorie cinétique des gaz, élaborée par James Clerk Maxwell (1831-1879) et Ludwig Boltzmann (1844-1906), a permis de calculer théoriquement la masse des particules gazeuses, d'expliquer l'origine de la température et de la pression par des hypothèses atomistiques et d'obtenir la loi des gaz parfaits, une loi jusque-là uniquement empirique. Ce fut la naissance de la mécanique statistique qui appliquait les lois de la probabilité aux atomes qui devaient être animés par les lois macroscopiques de Newton. Poussant plus loin cette théorie, Einstein publie en 1905 un article dans lequel un modèle basé sur la mécanique statistique permet d'expliquer le mouvement brownien (mouvement perpétuel et aléatoire de particules micrométriques flottant sur un liquide). Plus important encore, l'année suivante, il utilise son modèle théorique pour prédire le comportement de rotation de ces particules micrométriques, prédictions vérifiées trois ans plus tard par Jean Perrin, un physicien français. Cette réussite, impressionnante pour l'époque, fut considérée comme une preuve irréfutable de la nature atomique de la matière[1]. Pour la chimie, ce fut le début de l'ère des modèles physiques théoriques.

Les équations mathématiques fondamentales nécessaires à la compréhension de la chimie moderne ont presque toutes été découvertes pendant la première moitié du XXe siècle; la mécanique quantique et l'équation de Schrödinger étant les piliers théoriques de la chimie moderne. Depuis, le défi majeur de la chimie théorique a été de concevoir des modèles permettant de prédire et de comprendre les phénomènes expérimentaux qui surviennent à une échelle incommensurablement petite, comme l'avait fait Einstein. L'avènement de l'ordinateur et sa démocratisation ont permis des possibilités d'application de ces modèles théoriques jusque-là inimaginables. En effet, la nature statistique de la

chimie exige souvent de considérer un grand nombre de molécules et de résoudre des équations mathématiques de plus en plus difficiles. Donc, aujourd'hui, la machine à calculer qu'est l'ordinateur occupe une place importante dans la découverte scientifique et technologique. Bien que l'ordinateur ait révolutionné la science, sa puissance de calcul n'est pas infinie et nécessite presque toujours de procéder par approximations par rapport aux équations fondamentales, ce qui limite le domaine de validité d'un modèle. Dans cette perspective, la présente thèse a pour objet de développer un nouveau modèle mathématique basé sur des approximations raisonnables.

Souvent, ces approximations comportent des paramètres empiriques et un niveau de complexité variable. Le principe du rasoir d'Occam veut qu'un phénomène soit expliqué et prédit avec le nombre minimum d'hypothèses*. Or, il s'avère souvent en science que les théories les plus générales et prédictives contiennent un nombre minimum d'hypothèses ou de postulats, par exemple la mécanique quantique. Les représentations les plus simples ont aussi souvent l'avantage de présenter un pouvoir d'interprétation plus grand. Du point de vue théorique, il serait souhaitable d'utiliser la mécanique quantique le plus souvent possible pour nos études *in silico*. Il s'agit de la théorie la plus fondamentale et la plus exacte pour décrire et prédire le comportement des électrons et des noyaux, éléments constitutifs des atomes et des molécules. D'un point de vue pragmatique, cependant, la complexité des équations mathématiques qui doivent être résolues en mécanique quantique représente une limite technique importante. Par exemple, le calcul de la constante d'équilibre pour la liaison d'une molécule-médicament à sa cible enzymatique est vraiment hors d'atteinte si l'on utilise seulement les principes fondamentaux de la physique. Or, de façon générale, prédire des propriétés thermodynamiques de nouvelles molécules pas

* Bien que le rasoir d'Occam repose sur la notion arbitraire de simplicité, ce principe assure que la théorie soit générale et prédictive. Il faut cependant ajouter qu'une bonne théorie, comme la mécanique quantique, établit un lien logique avec les théories qui la précèdent. Ceci n'est pas considéré par le principe du rasoir d'Occam, mais a été une source de succès en science.

encore synthétisées est d'une grande importance scientifique et technologique. Depuis longtemps, donc, des modèles plus rapides à calculer et plus exacts sont développés dans ce but.

Il est apparu très tôt qu'il était possible de prendre des fonctions d'énergie potentielle simples, issues de la mécanique classique, et d'en ajuster les paramètres pour en faire des modèles fiables. Dans certains cas, ces modèles empiriques sont plus exacts que les méthodes fondées sur la mécanique quantique qui, elles aussi, sont presque toujours entachées d'approximations par rapport à aux équations fondamentales. La fonction d'énergie potentielle est d'une importance capitale puisqu'elle constitue l'élément clé de la plupart des théories physiques qui font le lien avec l'expérience, entre autres la mécanique statistique.

Cependant, les méthodologies plus rapides commencent à montrer leurs limites. En effet, Mobley *et al*[2] ont établi que la précision théorique pour le calcul d'énergie libre de liaison d'un inhibiteur à une enzyme est d'environ 2 kcal/mol pour une enzyme plutôt simple. Cette erreur est énorme et constitue un obstacle majeur pour le succès des méthodes théoriques dans le monde du développement de médicaments. Pour le comprendre, définissons une constante d'inhibition K_i et l'énergie libre de complexation ΔG_i. La thermodynamique nous apprend que $\Delta G_i = -RT\ln(K_i)$. On peut facilement montrer que $K_i^{calc} / K_i^{\exp} = \exp(-E/RT)$ où E est l'erreur du calcul. Donc, une surestimation de 2 kcal/mol de ΔG_i rend la constante d'équilibre 30 fois trop petite! Le calcul d'énergie libre d'hydratation de molécules bioorganiques est un autre exemple pour lequel les mêmes auteurs[3] ont montré une erreur moyenne de plus de 1 kcal/mol. Comme la fonction d'énergie potentielle est une source connue d'erreurs importantes et qu'elle constitue le pilier des méthodes utilisées pour ces calculs d'énergie libre, elle est une cible de choix pour l'amélioration des modèles et de leur prédictibilité. Nous voudrons donc, dans cette thèse, améliorer la fonction d'énergie qui est le fondement pour de nombreuses applications de la théorie.

Plus précisément, nous étudierons une voie, encore inexplorée, pour introduire la polarisation électronique dans la fonction d'énergie potentielle fondée sur la mécanique classique. Au cours des chapitres suivants, nous caractériserons cette nouvelle méthodologie. Bien que l'idée d'introduire la polarisation électronique ne soit pas nouvelle, la généralisation de ces approches à l'ensemble de la chimie bioorganique n'a pas été faite, malgré les 30 ans de travaux dans ce domaine. L'utilisation efficace d'une méthode qui tient compte de la réponse des électrons dans le contexte de développement de médicaments exige une grande polyvalence et une grande précision. Pour s'en rendre compte, il suffit d'examiner la complexité et le degré de fonctionalisation chimique des médicaments. Ceci nous amène à formuler un autre objectif des travaux présentés qui est de vérifier les avantages de généralisation, de polyvalence et d'exactitude amenés par l'approche que nous proposons.

Pour comprendre davantage la nature des améliorations qui seront apportées et être en mesure de formuler des objectifs clairs, il est approprié de mettre en place le cadre théorique et la méthodologie qui sous-tendent la recherche présentée.

1.1 La mécanique quantique

La mécanique quantique est une théorie physique qui permet, entre autres, d'expliquer le comportement des électrons et des noyaux dans une molécule. Elle prévoit que les électrons et les noyaux d'une molécule, au lieu de suivre une trajectoire comme le prédirait la mécanique classique, sont distribués dans l'espace avec une fonction de densité de probabilité donnée par le carré de la fonction d'onde $\left|\Phi(\bar{R}_n,\bar{r}_e,\bar{\sigma}_e)\right|^2$ où $\bar{R}_n$ est un vecteur contenant la position cartésienne des M noyaux et $\bar{r}_e$ la position des N électrons dont les N états de spin correspondants sont alignés dans le vecteur $\bar{\sigma}_e$. La fonction d'onde est complètement définie comme étant une solution à l'équation de Schrödinger (indépendante du temps)

$$\hat{H}\Phi(\vec{R}_n,\vec{r}_e,\sigma_e)=U\Phi(\vec{R}_n,\vec{r}_e,\sigma_e) \tag{1.1}$$

où $\hat{H}$ est un opérateur hermitique et U l'énergie totale du système, une valeur propre quantifiée. Une simplification importante s'opère lorsque l'on tire avantage de la grande légèreté des électrons comparativement aux noyaux. Les électrons peuvent ainsi être considérés dans leur état d'énergie minimum pour toute position des noyaux dans une fonction d'onde électronique fondamentale. Cette approximation donne lieu à une surface d'énergie potentielle électronique de Born-Oppenheimer sur laquelle les noyaux se déplacent. Ainsi, la fonction d'onde totale se simplifie en un produit de fonction d'onde nucléaire (Ψ_n) et électronique (Ψ) qui dépend paramétriquement de la position des noyaux:

$$\Phi(\vec{R}_n,\vec{r}_e,\sigma_e)=\Psi_n(\vec{R}_n)\times\Psi(\vec{r}_e,\sigma_e;\vec{R}_n) \tag{1.2}$$

Pour simplifier la notation, nous omettrons la dépendance paramétrique de la fonction d'onde électronique sur la position de noyaux. La dynamique des noyaux peut alors être étudiée en résolvant l'équation de Schrödinger qui dépend seulement de la position des noyaux soumis à un potentiel électronique effectif qui est donné par l'énergie de l'état fondamental de la fonction d'onde du problème électronique. L'équation de Schrödinger pour les noyaux devient alors

$$\left[-\frac{h^2}{8\pi^2}\sum_n^M\frac{1}{m_n}\nabla_n^2+U_{elec}(\vec{R}_n)\right]\Psi_n(\vec{R}_n)=U_{noyaux}\Psi_n(\vec{R}_n) \tag{1.3}$$

avec m_n la masse du noyau n et U_{elec} la surface de Born-Oppenheimer. Dans cette thèse, l'approximation de Born-Oppenheimer est sous-entendue.

Se concentrant désormais sur la fonction d'onde électronique, la théorie prévoit qu'en moyenne une quantité observable A prend une valeur donnée par

$$\langle A\rangle=\int\Psi(\vec{r}_e,\vec{\sigma}_e)\hat{A}\Psi^*(\vec{r}_e,\vec{\sigma}_e)d\vec{r}_e\,d\vec{\sigma}_e \tag{1.4}$$

où $\hat{A}$ est l'opérateur associé à A. Par exemple, la densité électronique $\rho(\vec{r})$ qui correspond au nombre d'électrons par unité de volume à chaque point de l'espace est calculée en utilisant

$$\rho(\vec{r}) = \sum_{\sigma_1} \cdots \sum_{\sigma_N} \int d\vec{r}_{e,2} \cdots \int d\vec{r}_{e,N} \Psi(\vec{r}_e, \vec{\sigma}_e) \Psi^*(\vec{r}_e, \vec{\sigma}_e) \tag{1.5}$$

La densité électronique ne dépend que des coordonnées spatiales et peut être observée expérimentalement par des méthodes basées sur la diffraction des rayons X, par exemple.

En mécanique quantique électronique, tout le travail mathématique consiste à trouver la fonction d'onde qui obéit à une équation différentielle aux valeurs propres. Pour un système électronique, l'équation de Schrödinger se décline comme suit*

$$-\frac{1}{2} \sum_{i}^{Electrons} \nabla_i^2 \Psi - \Psi \sum_{A}^{Noyaux} \sum_{i}^{Electrons} \frac{Z_A}{\left|\vec{R}_A - \vec{r}_i\right|} + \Psi \sum_{i<j} \frac{1}{\left|\vec{r}_i - \vec{r}_j\right|} = U\Psi \tag{1.6}$$

Dans l'éq. (1.6), le membre de gauche comporte un terme d'énergie cinétique des électrons, un potentiel de Coulomb pour l'interaction électron-noyau et un autre terme de Coulomb pour l'interaction électron-électron. En fait, l'éq. (1.6) peut se réécrire sous forme d'une équation aux valeurs propres

$$\hat{H}\Psi = U\Psi \tag{1.7}$$

avec un opérateur hamiltonien $\hat{H}$. L'équation de Schrödinger établit la théorie cible pour la chimie. Ce sera, en quelque sorte, un point de repère théorique pour les travaux de cette thèse. L'équation de Schrödinger n'a pas de solution analytique pour les systèmes à plus d'un électron (H, H_2^+, He^+, etc.) de telle sorte que des méthodes numériques sont nécessaires. Pire encore, même une solution numérique est très difficile à obtenir et il faut recourir à des approximations.

---

* Les unités atomiques sont utilisées à partir d'ici dans cette section.

Pour nous aider, le principe variationnel stipule que la fonction d'onde optimale a la propriété suivante

$$U = \min_{\Psi} U[\Psi] = \min_{\Psi} \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{1.8}$$

c'est-à-dire que la fonction propre de l'Hamiltonien minimise l'énergie U. Ceci constitue un guide mathématique pour le développement de méthodes approximatives comme la méthode de Hartree-Fock qui est le point de départ des méthodes dites *ab initio*[4,5]. La méthode de Hartree-Fock fait uniquement une approximation sur la forme de la fonction d'onde en posant qu'elle est le déterminant d'un produit de fonctions mono-électroniques. L'avantage de prendre le déterminant est que la fonction ainsi obtenue obéit à la propriété d'antisymétrie des électrons. Ce déterminant est à l'origine de ce que l'on nomme l'échange Hartree-Fock et a pour effet de créer une répulsion entre les électrons de spins parallèles. C'est le seul terme d'interaction électron-électron non classique de la méthode Hartree-Fock, le terme classique étant l'énergie coulombienne répulsive entre électrons. Mis à part sa calculabilité, la méthode Hartree-Fock offre une façon systématique d'inclure le terme d'énergie manquant appelé l'énergie de corrélation.

Une autre théorie quantique est utilisée en chimie: la théorie de la fonctionnelle de la densité (TFD)[6,7]. La TFD repose sur un formalisme parallèle à celui de la fonction d'onde. L'énergie électronique peut aussi être exactement décrite par une équation complexe de la densité électronique (c.f. éq. (1.5)), une fonction beaucoup plus simple que la fonction d'onde puisqu'elle ne dépend que des trois variables spatiales. En TFD, on écrit l'énergie d'un système électrons-noyaux comme

$$U[\rho(\vec{r})] = T_{ni}[\rho(\vec{r})] + V_{ne}[\rho(\vec{r})] + V_{ee}[\rho(\vec{r})] + U_{xc}[\rho(\vec{r})] \tag{1.9}$$

où le premier terme du membre de droite est une énergie cinétique d'un système d'électrons non interagissant, le deuxième terme correspond à l'énergie coulombienne d'interaction entre noyaux et électrons, le troisième terme est l'énergie d'interaction coulombienne entre électrons et le dernier terme est l'énergie d'échange-corrélation. Tous les termes de l'éq.

(1.9) sont faciles à définir analytiquement à l'exception de U_{xc} qui est un véritable fourre-tout, car il contient: la correction pour l'énergie cinétique des électrons interagissant, la correction pour l'auto-énergie (*self energy*) introduite par V_{ee}, l'énergie d'échange qui provient de la répulsion de Pauli (répulsion des électrons de même spin) et l'énergie de corrélation. La TFD aussi a un théorème variationnel

$$U = \min_{\rho} U[\rho] \tag{1.10}$$

Tout comme la théorie de la fonction d'onde, l'éq. (1.10) mène à des équations aux valeurs propres à Hamiltonien monoélectronique, appelées les équations de Kohn-Sham[6]. Par contre, dans le cas de la TFD, il faut inventer un système conceptuel d'électrons non interagissant. Les équations de Kohn-Sham sont celles utilisées en pratique. L'avantage certain de la TFD est la simplicité des équations et la vitesse avec laquelle les ordinateurs peuvent résoudre les équations de Kohn-Sham, mais la véritable forme mathématique de U_{xc} est inconnue, ce qui constitue une limite quant à la précision que l'on peut obtenir. Fort heureusement, grâce aux travaux pionniers d'Axel Becke[8-10], les fonctionnelles d'échange-corrélation permettent de calculer des propriétés moléculaires de façon très précise. Une panoplie impressionnante de fonctionnelles d'échange-corrélation est maintenant disponible[5,11]. Presque tous nos calculs de référence seront faits avec la TFD.

Pour terminer cette section sur la mécanique quantique, définissons le potentiel électrostatique, une quantité physique qui nous sera très utile. Supposons une molécule A ayant des noyaux atomiques de charges Z_A positionnés en $\vec{R}_A$ et une densité électronique $\rho(\vec{r})_A$. L'énergie d'interaction électrostatique entre la molécule A et une molécule B se calcule de la façon suivante

$$U_{AB} = -\sum_A Z_A \int_V \frac{\rho_B(\vec{r})}{\left|\vec{R}_A - \vec{r}\right|} d\vec{r}^{\,3} - \sum_B Z_B \int_V \frac{\rho_A(\vec{r})}{\left|\vec{R}_B - \vec{r}\right|} d\vec{r}^{\,3} + \frac{1}{2} \iint_{VV} \frac{\rho_A(\vec{r}\,')\rho_B(\vec{r})}{\left|\vec{r}\,' - \vec{r}\right|} d\vec{r}^{\,'3}\, d\vec{r}^{\,3}$$
$$+ \sum_A \sum_B \frac{Z_A Z_B}{\left|\vec{R}_A - \vec{R}_B\right|} \tag{1.11}$$

$$U_{AB} = -\frac{1}{2}\iint_{VV}\frac{\eta_A(\vec{r}\,')\rho_B(\vec{r})}{|\vec{r}\,'-\vec{r}|}d\vec{r}\,'^3\,d\vec{r}^3 - \frac{1}{2}\iint_{VV}\frac{\eta_B(\vec{r}\,')\rho_A(\vec{r})}{|\vec{r}\,'-\vec{r}|}d\vec{r}\,'^3\,d\vec{r}^3$$
$$+\frac{1}{2}\iint_{VV}\frac{\rho_A(\vec{r}\,')\rho_B(\vec{r})}{|\vec{r}\,'-\vec{r}|}d\vec{r}\,'^3\,d\vec{r}^3 + \frac{1}{2}\iint_{VV}\frac{\eta_A(\vec{r}\,')\eta_B(\vec{r})}{|\vec{r}\,'-\vec{r}|}d\vec{r}\,'^3\,d\vec{r}^3 \qquad (1.12)$$
$$\text{avec } \eta_A(\vec{r}) = \sum_A Z_A\delta(\vec{r}-\vec{R}_A)$$

$$U_{AB} = \frac{1}{2}\iint_{VV}\frac{[\eta_A(\vec{r}\,')-\rho_A(\vec{r}\,')][\eta_B(\vec{r})-\rho_B(\vec{r})]}{|\vec{r}\,'-\vec{r}|}d\vec{r}\,'^3\,d\vec{r}^3$$
$$= \frac{1}{2}\iint_{VV}\frac{\Gamma_A(\vec{r}\,')\Gamma_B(\vec{r})}{|\vec{r}\,'-\vec{r}|}d\vec{r}\,'^3\,d\vec{r}^3 \qquad (1.13)$$

Pour passer de l'éq. (1.11) à l'éq. (1.12), nous formulons une densité de charge des noyaux $\eta_A(\vec{r})$ à l'aide de la fonction delta de Dirac, ce qui nous permet de réécrire chacun des termes comme une intégrale. Ce faisant, nous introduisons un double comptage des interactions, c'est-à-dire qu'une même distance *r* entre deux points de l'espace se produit de deux façons dans l'intégrale double, alors que chacune des distances entre les deux mêmes points ne devraient être comptées qu'une seule fois. Cela explique les facteurs d'une demie. Nous pouvons réécrire ces équations en unissant la densité de charge des noyaux et des électrons d'une molécule dans une fonction de densité de charge totale Γ, ce qui mène à l'éq. (1.13). Nous pouvons maintenant définir une fonction fort utile, le potentiel électrostatique d'une molécule

$$\varphi_A(\vec{r}) = \int_V \frac{\Gamma_A(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}d\vec{r}\,'^3 \qquad (1.14)$$

qui permet de calculer l'énergie d'interaction électrostatique entre A et B par une intégrale sur l'espace en 3 dimensions

$$U_{AB} = \frac{1}{2}\int_V \varphi_A(\vec{r})\Gamma_B(\vec{r})d\vec{r}^3 = \frac{1}{2}\int_V \varphi_B(\vec{r})\Gamma_A(\vec{r})d\vec{r}^3 \qquad (1.15)$$

Le potentiel électrostatique peut être compris comme l'énergie électrique qu'il faut fournir pour amener un proton d'une distance infinie, d'énergie totale nulle par définition, à une

distance à laquelle l'interaction prend place (ce qui est évident si $\Gamma_A(\vec{r})$ est remplacé par une fonction delta dans l'éq. (1.15)).

Il est nécessaire de mentionner que les densités électroniques utilisées pour écrire l'éq. (1.11) devraient se modifier quand les molécules A et B se rapprochent ou changent simplement leurs orientations. Nous pouvons alors mettre en évidence le phénomène de polarisation de façon explicite

$$\varphi_A(\vec{r}) = \varphi_A^0(\vec{r}) + \varphi_A^{ind}(\vec{r}) = \int_V \frac{\Gamma_A^0(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} d\vec{r}\,'^3 + \int_V \frac{-\delta\rho_A(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} d\vec{r}\,'^3 \tag{1.16}$$

où φ^0 est le potentiel électrostatique permanent de la molécule (molécule isolée) et, φ^{ind}, le résultat d'un déplacement de la densité électronique $\delta\rho$ causé par la présence de la molécule B (ou d'une perturbation quelconque). La Figure 1.1a illustre φ^0 pour la molécule d'eau à l'aide d'isolignes tracées dans le plan des atomes qui est perturbé par un monopole placé à trois endroits différents. Le potentiel total φ est montré à gauche (b et d) et le potentiel induit φ^{ind} correspondant à droite (c et g). La perturbation du potentiel électrostatique amène des changements substantiels qui ont un impact quantitatif sur l'énergie. L'énergie d'interaction entre deux molécules A et B qui se perturbent coopérativement s'écrit

$$\begin{aligned} U_{AB} &= \frac{1}{2}\int_V \left[\varphi_A^0(\vec{r}) + \varphi_A^{ind}(\vec{r})\right][\Gamma_B^0(\vec{r}) + \delta\rho_B]d\vec{r}^3 \\ &= \frac{1}{2}\int_V \left[\varphi_B^0(\vec{r}) + \varphi_B^{ind}(\vec{r})\right][\Gamma_A^0(\vec{r}) + \delta\rho_A]d\vec{r}^3 \end{aligned} \tag{1.17}$$

L'énergie de polarisation, aussi appelée énergie inductive, est toujours attractive. Il s'agit en fait de degrés de liberté supplémentaires donnés au système pour réduire son énergie totale. Le concept de potentiel électrostatique induit sera abondamment utilisé dans les chapitres suivants, puisqu'ils traiteront principalement de la polarisation électronique. Une des manifestations de la polarisation électronique qui est quantifiable est la polarisabilité moléculaire, traitée dans la prochaine section.

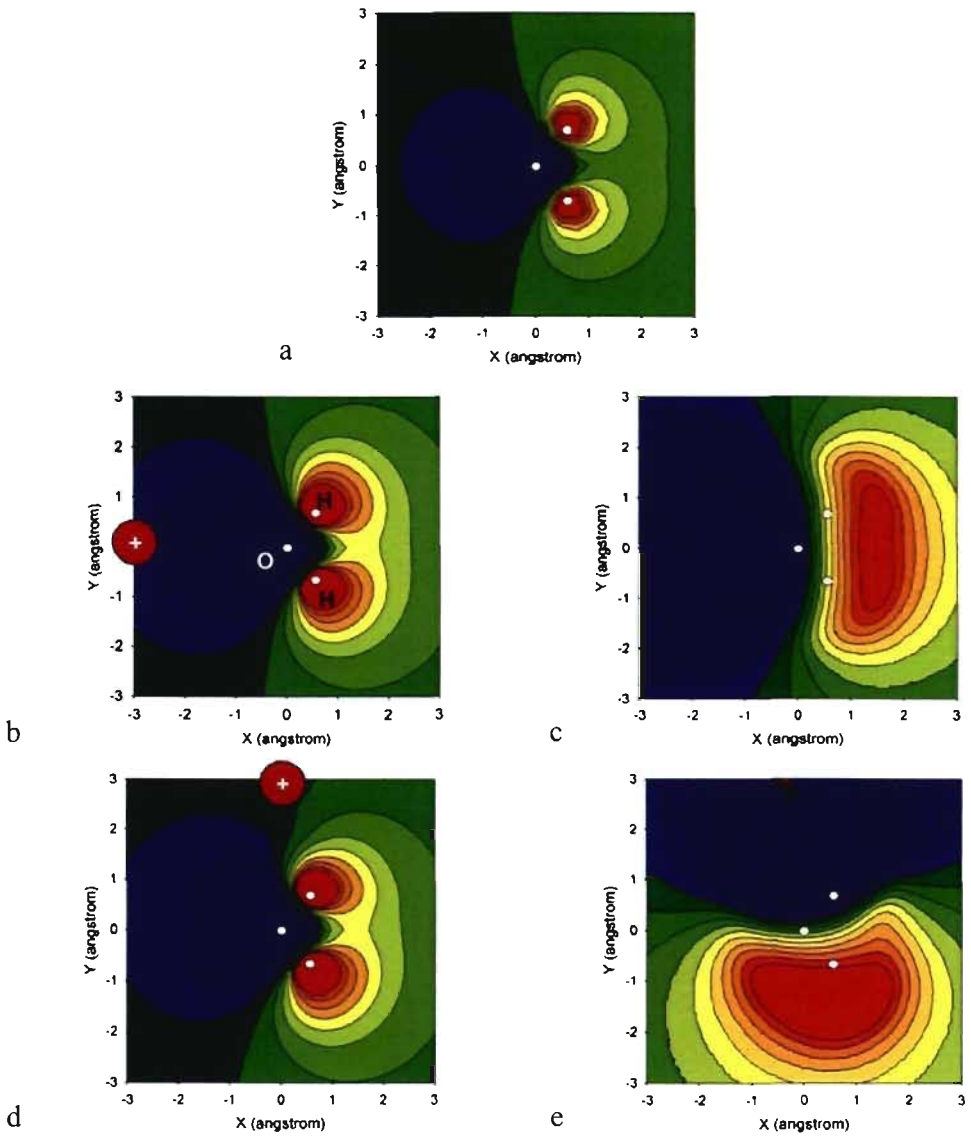


Figure 1.1. Le potentiel électrostatique permanent φ^0 (a) d'une molécule d'eau dans le plan des atomes est perturbé par un monopole localisé à différents endroits. Le potentiel total φ est donné à gauche (b et d) et le potentiel induit φ^{ind} à droite (c et e). Un potentiel très positif est rouge et très négatif est bleu. Le monopole induit un dipôle, en polarisant les électrons de la molécule d'eau, qui diffère selon sa position et déforme le potentiel électrostatique de façon qualitativement et quantitativement appréciable. Obtenu avec EPIC.

1.2 La polarisabilité moléculaire

La polarisabilité moléculaire jouera un rôle central dans cette thèse, il est donc opportun de la définir et de brièvement expliquer comment elle se calcule en mécanique quantique. Nous pouvons dire que la polarisabilité est le résultat de la déformation de la densité électronique lorsqu'un champ électrique externe perturbe celle-ci. De façon générale, le champ électrique perturbant peut prendre n'importe quelle forme et son origine peut être diverse : onde électromagnétique, autre molécule, dispositif électronique, etc. Aussi, le champ électrique peut varier dans le temps, comme dans le cas d'une onde lumineuse. Ici, et dans le reste de cette thèse, nous traiterons principalement du cas où le champ électrique externe (ou perturbateur) est uniforme et varie suffisamment lentement pour que les électrons soient toujours à l'état fondamental. Un champ électrique uniforme a la même valeur et la même direction partout dans l'espace sauf, peut-être, à proximité de la molécule. De façon mathématique, la polarisabilité dipolaire (appelée polarisabilité dans ce travail) est définie par un développement en série de Taylor du moment dipolaire en fonction des trois composantes du champ électrique **uniforme** externe d'une molécule qui s'écrit

$$\begin{aligned}\mu_\alpha(E_x,E_y,E_z) &= \mu_\alpha^0 + \left[\frac{\partial \mu_\alpha}{\partial E_\beta}\right]_{E_\beta=0} \cdot E_\beta + \frac{1}{2}\left[\frac{\partial^2 \mu_\alpha}{\partial E_\beta \partial E_\gamma}\right]_{\substack{E_\beta=0\\E_\alpha=0}} \cdot E_\beta E_\gamma + \ldots \\ &= \mu_\alpha^0 + \alpha_{\alpha\beta}E_\beta + \frac{1}{2}\beta_{\alpha\beta\gamma}E_{\beta\gamma} + \ldots \end{aligned} \tag{1.18}$$

Notons que l'éq. (1.18) utilise la notation tensorielle d'Einstein*. Par définition, donc, le dipôle total est une somme du dipôle permanent et des dérivées du dipôle par rapport au champ électrique quand ce dernier est nul. Il est clair que si le champ électrique est petit,

---

* Un indice en lettre grecque peut être remplacé par x, y ou z et quand deux quantités qui portent le même indice se multiplient, il s'agit d'une sommation sur l'indice. Si les deux mêmes indices se retrouvent sur les variables multipliées, il s'agit d'une double sommation.

les premières dérivées seront suffisantes pour expliquer la déviation du dipôle. Les six premières dérivées indépendantes définissent le tenseur de polarisabilité qui peut être écrit comme une matrice 3x3 symétrique. Quand le champ augmente, la variation n'est plus linéaire et le tenseur d'hyperpolarisabilité ($\beta_{\alpha\beta\gamma}$) devient nécessaire. L'ordre de grandeur du champ électrique qui nous intéresse dans cette thèse justifie l'emploi du seul terme linéaire. Dans certaines situations, par exemple avec un cation divalent (Ca^{2+}), le champ électrique local s'est montré suffisamment fort pour produire une déviation à la linéarité dans certains modèles polarisables[12], mais dans le cas de la polarisabilité produite par un champ électrique uniforme, ce cas d'exception n'est pas examiné. Par conséquent, l'approximation linéaire prédit que le dipôle induit est proportionnel au champ électrique

$$\vec{\mu}^{ind} = \bar{\bar{\alpha}}_{3\times3} \cdot \vec{E} \tag{1.19}$$

avec $\bar{\bar{\alpha}}$ la matrice des dérivées, par exemple: $\bar{\bar{\alpha}}_{0,1} = \alpha_{x,y} = \left(\partial\mu_x / \partial E_y\right)_{\vec{E}=0}$. Le rôle du tenseur de polarisabilité est de moduler la grandeur de l'induction selon l'orientation de la molécule par rapport au champ. Malheureusement, $\bar{\bar{\alpha}}$ varie quand une molécule tourne, mais la trace est invariante, ce qui permet de définir la polarisabilité moyenne

$$\alpha_{moy} = \frac{1}{3}\left(\alpha_{xx} + \alpha_{yy} + \alpha_{zz}\right) \tag{1.20}$$

où α_{moy} est une valeur scalaire qui ne dépend pas de l'orientation de la molécule. Lorsque la symétrie de la molécule la rend isotrope, l'éq. (1.19) s'écrit

$$\vec{\mu}^{ind} = \alpha_{moy}\vec{E} \tag{1.21}$$

Cependant, pour plusieurs molécules, cette approximation est erronée. Pour mieux saisir la signification des éléments de $\bar{\bar{\alpha}}$, il est utile de trouver l'orientation de la molécule qui rend ce tenseur diagonal. Dans ce cas, les axes *x*, *y* et *z* indiquent les directions principales de polarisation et le moment dipolaire induit pour un champ électrique de grandeur *E* orienté en *x* se calcule avec $\mu_x = \alpha_{xx}E$ (même chose pour les autres axes). La

Figure 1.2 montre quatre exemples de molécules orientées dans le système d'axe illustré au-dessus de chacune. La molécule adamantane est complètement isotrope, comme indiqué par la valeur numérique sur les axes, et l'éq. (1.21) s'applique. La molécule de benzène a deux axes principaux de polarisabilité dégénérés dans le plan de la molécule avec une polarisabilité environ deux fois moindre dans l'axe qui sort du plan. Les molécules de quinoxaline et d'isothiazole ont trois polarisabilités différentes pour chacun des axes et l'orientation des moments principaux de la polarisabilité pour l'isothiazole n'est pas aussi évidente que pour les autres cas. L'importance du dipôle induit pour un même champ externe dépend fortement de l'orientation de la molécule.

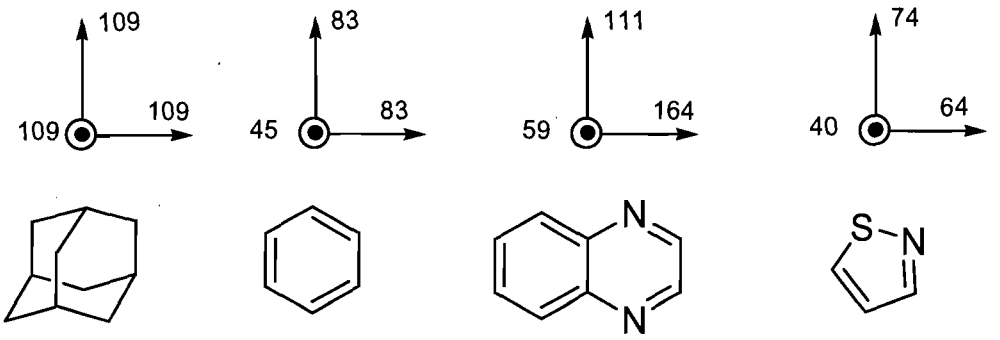


Figure 1.2 Polarisabilité selon les axes principaux en unités atomiques[§] (calculs B3LYP[13]) pour quatre molécules. Le cercle indique l'axe qui sort du plan de la feuille.

Terminons cette section en faisant un survol des méthodes de calcul de la polarisabilité. Prenons d'abord la formule pour calculer l'énergie d'interaction entre un dipôle total et le champ électrique appliqué

[§] Les unités atomiques (u.a.) de la polarisabilité peuvent être obtenues des unités en S.I. (Cm^2/V) en divisant par $4\pi\varepsilon_0$ et en convertissant le m^3 en $bohr^3$. Pour convertir des $Å^3$ en u.a., il faut diviser par 0.148184.

$$\begin{aligned} U(\vec{E}) &= -\mu_\alpha E_\alpha \\ &= -\mu_\alpha^0 E_\alpha - \alpha_{\alpha\beta} E_\beta E_\alpha - \frac{1}{2}\beta_{\alpha\beta\gamma} E_\beta E_\gamma E_\alpha \end{aligned} \tag{1.22}$$

En comparant les éq. (1.18) et (1.22), on retrouve les relations

$$\frac{\partial \mu_\alpha}{\partial E_\beta} = -\frac{\partial^2 U(\vec{E})}{\partial E_\beta \partial E_\alpha} = \alpha_{\alpha\beta} \tag{1.23}$$

qui donnent lieu à deux méthodes de calcul de la polarisabilité. Premièrement, la méthode des différences finies fait un calcul numérique approximatif de la dérivée selon le champ électrique

$$\frac{\partial \mu_\alpha}{\partial E_\beta} = \alpha_{\alpha\beta} \approx \frac{\mu_\alpha(E_\alpha + \Delta E_\alpha) - \mu_\alpha(E_\alpha - \Delta E_\alpha)}{2\Delta E_\alpha} \tag{1.24}$$

où E_α le champ électrique appliqué est généralement petit. Avec les différences finies, le dipôle est calculé avec différentes valeurs du champ électrique séparées par $2\Delta E_\alpha$. Cette méthode est facile à programmer et donne des résultats raisonnables avec des calculs de mécanique quantique[14]. Une deuxième approche utilisée en mécanique quantique fait appel à la théorie des perturbations (formalisme de Rayleigh-Schrödinger)[4,15] qui nécessite de définir un Hamiltonien perturbateur donné, dans le cas du champ électrique uniforme, par $\Delta\hat{H} = \vec{r} \cdot \vec{E}$. Au deuxième ordre, l'énergie d'interaction entre le dipôle induit par la déformation de la densité électronique et le champ électrique est

$$\begin{aligned} \Delta U(\vec{E}_0) &= \sum_{i\neq 0}^{états} \frac{2\langle 0|r_\alpha E_\alpha|i\rangle\langle i|r_\beta E_\beta|0\rangle}{u_0 - u_i} \\ &= E_\alpha E_\beta \sum_{i\neq 0}^{états} \frac{2\langle 0|r_\alpha|i\rangle\langle i|r_\beta|0\rangle}{u_0 - u_i} \\ &= E_\alpha E_\beta \sum_{i\neq 0}^{états} \frac{2\langle 0|\hat{\mu}_\alpha|i\rangle\langle i|\hat{\mu}_\beta|0\rangle}{u_0 - u_i} \end{aligned} \tag{1.25}$$

où $|0\rangle$ est la fonction d'onde de l'état fondamental, $|i\rangle$ la fonction d'onde d'un état excité (fonction propre de l'Hamiltonien non perturbé), u_0 l'énergie de l'état fondamental, u_i l'énergie de $|i\rangle$ et $\hat{\mu}$ l'opérateur pour le moment dipolaire. En appliquant l'éq. (1.23), on obtient une expression pour le tenseur de polarisabilité

$$\alpha_{\alpha\beta} = 2\sum_{i}^{états} \frac{\langle 0|\hat{\mu}_\alpha|i\rangle\langle i|\hat{\mu}_\beta|0\rangle}{u_i - u_0} \tag{1.26}$$

qui nous apprend que la polarisabilité est plus grande pour les molécules ayant une plus grande variation du moment dipolaire lors de transitions à des états excités de plus faibles énergies. Ces états excités sont des déterminants de Slater à excitations simples. Il est intéressant d'observer que le deuxième ordre de perturbation donne accès directement et exactement au tenseur de polarisabilité. Selon l'éq. (1.26), plus une molécule est grande, plus l'étendue spatiale des transitions est grande et plus la polarisabilité est grande, ce qui est en accord avec la notion que la polarisabilité augmente avec la taille des molécules. Par exemple, pour le benzène, il y a plusieurs transitions π–π^* qui ont une composante dans le plan de la molécule alors que les transitions σ–π^* et π–σ^* qui font varier le dipôle hors plan ont un écart énergétique plus élevé. Cela explique que le benzène est plus polarisable dans le plan de la molécule. La théorie des perturbations donne des résultats plus exacts que les différences finies, mais demande des efforts de calcul beaucoup plus grands.

Dans cette thèse, la méthode quantique utilisée pour le calcul des tenseurs de polarisabilités s'appuie sur l'éq. (1.23) et les deuxièmes dérivées de l'énergie par rapport au champ électrique extérieur. La méthode programmée pour ce calcul est celle de *static response function*[16] qui requiert le calcul de dérivées deuxièmes de l'énergie par rapport au champ électrique externe appliqué. Typiquement, le calcul par ordinateur des secondes dérivées demande beaucoup de mémoire et de temps.

1.3 La mécanique moléculaire

La mécanique quantique est inefficace pour simuler de très gros systèmes comme des protéines ou des phénomènes qui surviennent à l'état liquide sur de longues échelles de temps. Or, par des approximations judicieuses, on arrive à simplifier grandement l'équation aux valeurs propres de Schrödinger pour le déplacement des noyaux donnée par l'éq. (1.3).

Premièrement, pour un grand éventail de cas, incluant la plupart des systèmes biomoléculaires simulés sans bris ou formation de liens chimiques, il est acceptable d'ignorer la nature quantique des noyaux et supposer qu'ils se comportent comme des particules classiques mues par la loi de Newton ($\vec{F} = m\vec{a}$). Cela se traduit par l'utilisation de l'énergie cinétique classique et l'éq. (1.3) peut s'écrire

$$\sum_{n}^{M} \frac{|\vec{p}_n|^2}{2m_n} + U_{elec}(\vec{R}_n) = U_{noyaux}(\vec{R}_n) \tag{1.27}$$

avec $\vec{p}_n$ la quantité de mouvement du noyau n. Les noyaux se déplacent dans un potentiel effectif électronique donné par la surface de Born-Oppenheimer $U_{élec}$. Comme il est beaucoup plus facile de résoudre un système d'équations différentielles ordinaires à conditions initiales qu'une équation aux valeurs propres, le problème est grandement simplifié. Il reste toutefois à résoudre l'équation de Schrödinger électronique qui n'est quand même pas une mince tâche.

Une seconde approximation doit intervenir. Nous remplaçons l'équation de Schrödinger électronique par une fonction d'énergie potentielle qui conserve le comportement de $U_{élec}$, mais qui s'écrit par une fonction directe empirique et paramétrique. Le terme mécanique moléculaire signifie que les deux approximations mentionnées ci-dessus sont appliquées.

Dans le jargon de la mécanique moléculaire, nous appelons *champ de force* la fonction d'énergie potentielle empirique des noyaux atomiques. Par exemple, le champ de force utilisé par le logiciel AMBER[17,18] et développé dans le groupe de Peter Kollman (1944-2001) a la forme suivante

$$U(\bar{R}) = U_L(\bar{R}) + U_{NL}(\bar{R}) \tag{1.28}$$

où l'énergie potentielle du système moléculaire, U, est divisée en une énergie intramoléculaire, U_L, qui dépend des coordonnées internes (longueur des liens, angles, etc.) et en un deuxième terme, U_{NL}, qui traite des interactions entre atomes non liés. Plus spécifiquement l'énergie potentielle interne est donnée par

$$\begin{aligned} U_L(\bar{R}_L) = &\sum_{i=1}^{Nb} k_{b,i}(r_i - r_e)^2 + \sum_{i=1}^{Na} k_{a,i}(\theta_i - \theta_e)^2 + \sum_{i=1}^{Np} k_{p,i}(\Phi_i - \Phi_e)^2 \\ &+ \frac{1}{2}\sum_{i=1}^{Nd}\sum_{n} V_n\left[1 + \cos(n\varphi_i - \gamma_i)\right] \end{aligned} \tag{1.29}$$

Dans l'éq. (1.29), l'énergie d'étirement et de compression de chacun des N_b liens chimiques hors de leur longueur d'équilibre (r_e) est traité un comme un oscillateur harmonique dont la constante de force, un paramètre ajustable, est k_b. Une constante de force peut être spécifique à un lien ou bien être utilisée pour plusieurs, en fonction du typage choisi. De la même façon, l'énergie de déformation des N_a angles formés par trois atomes chaînés, comme l'angle H-O-H de la molécule d'eau, est dictée par le paramètre ajustable k_a. Un angle de déformation hors plan (souvent appelé angle impropre), Φ, permet de maintenir un centre trigonal dans le plan. Finalement, un terme d'énergie pour chacun des angles dièdres sert de véritable fourre-tout. Il s'agit d'une expansion en série de cosinus de Fourier (le terme en sinus est inutile étant donné que la parité peut être garantie par un angle de phase γ) qui permet d'exprimer de façon exacte, à la limite d'un grand nombre de termes n, n'importe quel potentiel périodique sur la période $[0, 2\pi]$. Il sert normalement à inclure des effets électroniques reliés à la conjugaison électronique, par exemple, et à corriger les

défauts de tous les autres termes. La fonction U_L définit en grande partie la structure tridimensionnelle des molécules qui peuvent se déformer au gré des forces appliquées.

Le deuxième terme de la fonction d'énergie potentielle totale du champ de force AMBER, comme bien d'autres, se définit comme

$$U_{NL}(\vec{R}) = \sum_{i<j}^{Na} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \sum_{i<j}^{Na} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{1.30}$$

où les sommations sont faites sur toutes les paires distinctes d'atomes non liés. Règle générale, les atomes séparés par moins de trois liens chimiques ne sont pas considérés alors que ceux séparés par quatre liens ou plus peuvent partiellement être considérés. La première sommation est faite sur un terme Lennard-Jones 12-6 qui décrit à la fois les forces répulsives entre les électrons dues à la répulsion de Fermi (ou de l'échange électronique) et les forces attractives de van der Waals. La répulsion de Fermi résulte de l'interpénétration des nuages électroniques qui forcent les électrons de même spin à occuper un espace commun. C'est une force très intense, mais active seulement à courte distance qu'on appelle parfois la distance de contact de Fermi. Des calculs de haut niveau ont montré un comportement exponentiel pour la répulsion de Fermi $\sim\exp(-a\times r)$[15], mais l'avantage calculatoire et la précision de la forme en r^{-12} en ont justifié l'utilisation. Quant aux forces de van der Waals, leurs comportements théoriques suivent une loi en $1/r^6$. Elles sont petites et agissent à courte distance.

Le dernier terme, mais non le moindre, de l'éq. (1.30) est une approximation de l'énergie d'interaction électrostatique avec des charges atomiques partielles et ajustables q_i qui sont positionnées, d'habitude, sur les noyaux. Ces charges n'ont pas de signification physique sinon qu'elles reproduisent le potentiel électrostatique de la molécule qui les porte. Les charges atomiques partielles sont presque toujours ajustées pour reproduire la fonction du potentiel électrostatique $\varphi(r)$ de l'éq. (1.14) qui est calculée à l'aide de la mécanique quantique. Le potentiel électrostatique reproduit se trouve à l'extérieur de la zone de Fermi, c'est-à-dire là où la densité électronique est pratiquement nulle. Cela est

nécessaire puisqu'un modèle qui compte sur des charges atomiques ponctuelles ne peut habituellement pas décrire le potentiel électrostatique complexe près des noyaux qui est d'ailleurs de peu d'intérêt pour traiter les interactions intermoléculaires. Le potentiel électrostatique classique utilisé s'écrit alors

$$\varphi(\vec{r}) = \sum_{i}^{Na} \frac{q_i}{4\pi\varepsilon_0 |\vec{r} - \vec{r}_i|} \tag{1.31}$$

La Figure 1.3 compare le potentiel électrostatique de la molécule 4-pyridone calculé avec la mécanique quantique (éq. (1.14)) en (a) à l'approximation de l'éq. (1.31) en (b). De tous les termes de l'éq. (1.28), l'énergie électrostatique a la plus longue portée et est en grande partie responsable du caractère unique d'une molécule. L'énergie électrostatique peut être répulsive ou attractive mais, en dehors de la zone de Fermi, est presque toujours le terme intermoléculaire dominant.

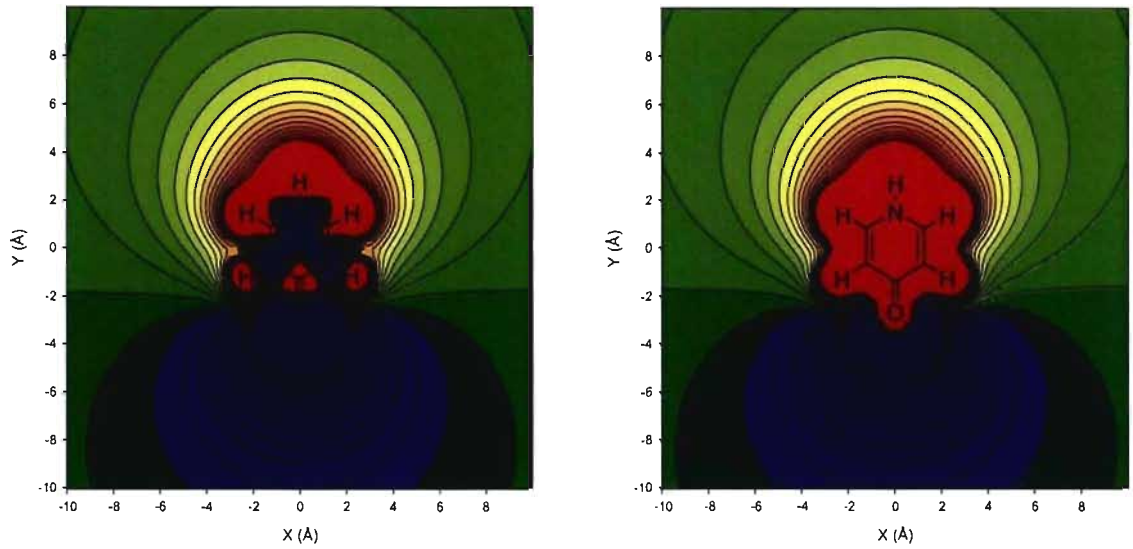
Une ombre se dresse cependant au tableau. La mécanique moléculaire est surtout utilisée pour étudier des systèmes en milieu condensé comme l'eau liquide. Or, il est connu que la polarité des molécules augmente considérablement dans un solvant polaire. Par exemple, la molécule d'eau voit son moment dipolaire passer de 1.8 D à l'état gazeux à 2.8 D à l'état liquide[19,20]. Cette augmentation du moment dipolaire est expliquée par la polarisation électronique qui peut constituer une composante importante de l'énergie d'interaction entre des espèces polaires. La polarisation due au solvant ne peut être négligée. De plus, la polarisation électronique constitue une part importante des interactions dans certains systèmes. Par exemple, l'énergie d'induction dans un système cation-π est aussi importante que l'interaction électrostatique monopole-benzène. Pourtant, les calculs du potentiel électrostatique nécessaires à l'ajustement des charges atomiques, se font dans le vide (état gazeux).

L'inclusion du potentiel électrostatique moléculaire induit φ^{ind} (éq. (1.16)) devient alors une condition *sine qua non* au succès des champs de force. La rédemption vient de deux stratégies plutôt brutales: augmenter par un facteur empirique toutes les charges partielles

atomiques ajustées sur φ^0 [21], utiliser une méthode quantique** qui commet l'erreur de produire systématiquement un moment dipolaire trop élevé pour la phase gazeuse[22]. Dans chacun des cas, toutefois, les charges atomiques partielles demeurent constantes dans les simulations qui utilisent l'éq. (1.30) et l'effet de polarisation obtenu ne s'ajuste pas à l'environnement. Par exemple, un inhibiteur qui se lie au site actif d'une enzyme voit son environnement changer radicalement de polarité. Un soluté qui se trouve à l'interface de deux phases peut être dans une situation analogue. De plus, certaines interactions intermoléculaires ne peuvent être expliquées quantitativement que par la prise en compte de la polarisabilité moléculaire: interactions cation-π[23-27], ponts-H de molécules très polaires[27,28], hydratation d'ions[29-32], affinité enzyme-inhibiteur[33]. Il s'agit d'un problème de longue date qui persiste encore et qui est une source d'erreurs importantes et difficiles à corriger avec les champs de force non polarisables.

Bref, la mécanique moléculaire est un outil de choix qui a été méticuleusement élaboré au cours des 30 dernières années et constitue un fer de lance pour l'étude de biomolécules, entre autres choses. L'omission de la polarisabilité moléculaire pour simplifier la paramétrisation limite encore ce domaine d'étude. Dans cette thèse, nous amènerons une nouvelle méthode qui modélise la polarisation électronique.

** La méthode de Hartree-Fock avec une petite base orbitalaire comme 6-31G*.



a b

Figure 1.3. Potentiel électrostatique de la molécule 4-pyridone obtenu (a) avec des charges atomiques partielles selon l'éq. (1.31) et (b) par la mécanique quantique où l'éq. (1.14) est calculée avec HF/6-31G*. L'approximation des charges atomiques permet de reproduire précisément le potentiel électrostatique produit par la densité électronique et les noyaux ($\Gamma(\vec{r})$) à une distance où les interactions intermoléculaires prennent place.

1.4 Champs de force polarisables : précédents de la littérature

Une des premières simulations qui permettaient à un champ de force d'adapter le potentiel électrostatique des molécules à l'environnement a été réalisée par Warshel et Levitt il y a plus de 30 ans[34]. Depuis ce temps, ce sujet a toujours été d'actualité[††] sans être

[††] ISI Web of Knowledge répertoriait 4154 articles, en date du 20 décembre 2008, dont le sujet est *polariz** en cooccurrence avec *force field** ou *molecular mechanic** ou *AMBER* ou *CHARMM*.

résolu de façon satisfaisante. Deux classes de méthodes, que nous examinerons ici, ont été développées jusqu'à maintenant.

Tout d'abord, le modèle des dipôles atomiques induits[35-39] associe un point polarisable aux atomes d'une molécule et, d'une façon analogue à l'éq. (1.21), un dipôle ponctuel proportionnel au champ électrique total est induit sur chaque noyau polarisable:

$$\vec{\mu}_j^{ind} = \alpha_j \vec{E}(\vec{r}_j) \tag{1.32}$$

Ce modèle est illustré pour le cas d'une molécule de benzène assujettie à un champ électrique complexe à la Figure 1.4. Dans l'éq. (1.32), le champ électrique total $\vec{E}(\vec{r})$ est créé par les charges atomiques partielles responsables du potentiel électrostatique permanent et par les dipôles induits. Si l'on veut utiliser ce modèle dans le cadre d'un champ de force, il faut calculer les dipôles atomiques induits qui produiront un potentiel électrostatique induit en utilisant l'équation[15,35,39]

$$\vec{\mu}_j^{ind} = \alpha_j \times \left[\sum_{i \neq j} \frac{q_i \vec{r}_{ij}}{4\pi\varepsilon_0 r_{ij}^3} - \sum_{i \neq j} \frac{\overline{\overline{T}}_{ij} \bullet \vec{\mu}_i^{ind}}{4\pi\varepsilon_0 r_{ij}^5} + \vec{E}_0(\vec{r}_j) \right] \tag{1.33}$$

avec $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$, le vecteur qui sépare la charge/dipôle source de la position du dipôle induit j; la longueur de cette distance est notée r_{ij}. La matrice $\overline{\overline{T}}_{ij}$ est le tenseur du champ dipolaire donné par

$$\overline{\overline{T}}_{ij} = \frac{3}{r_{ij}^5} \begin{bmatrix} r_{ij,x}^2 - \frac{1}{3} r_{ij}^2 & r_{ij,x} r_{y,ij} & r_{ij,x} r_{ij,z} \\ r_{ij,y} r_{ij,x} & r_{ij,y}^2 - \frac{1}{3} r_{ij}^2 & r_{ij,y} r_{ij,z} \\ r_{ij,z} r_{ij,x} & r_{ij,z} r_{ij,y} & r_{ij,z}^2 - \frac{1}{3} r_{ij}^2 \end{bmatrix} \tag{1.34}$$

avec la composante y du vecteur de distance $\vec{r}_{ij}$ noté $r_{ij,y}$. Il est évident que l'éq. (1.33) nécessite une solution autocohérente. Le terme d'énergie résultant est assez compliqué[15] et nécessite, pour chaque configuration, de résoudre l'éq. (1.33).

Un cas particulier de (1.33) peut être utilisé pour calculer le tenseur de polarisabilité d'une molécule portant des dipôles polarisables. Dans le cas où un champ électrique uniforme externe est appliqué sans les charges partielles atomiques, on trouve

$$\vec{\mu}_j^{ind} = \alpha_j \times \left[\vec{E}_0 - \sum_{i \neq j} \frac{\overline{\overline{T}}_{ij} \bullet \vec{\mu}_i^{ind}}{4\pi\varepsilon_0 r_{ij}^5} \right] \tag{1.35}$$

En réarrangeant (1.35), le tenseur de polarisabilité moléculaire peut être obtenu. Fait à remarquer, la polarisabilité moléculaire est modulée par une interaction intramoléculaire des dipôles induits. À partir de (1.35), on se rend compte que plus les dipôles induits *i* qui entourent l'atome *j* seront grands, moins le champ effectif senti par l'atome *j* sera grand. C'est cet effet intramoléculaire qui est la plus importante source de l'anisotropie du tenseur de polarisabilité.

Le modèle des dipôles ponctuels induits a toutefois une complication importante car, quand deux dipôles sont à proximité, la solution autocohérente de l'éq. (1.33) ou de l'éq. (1.35) peut diverger. Dans le cas de deux points polarisables (ex. molécule diatomique), les composantes de la polarisabilité suivant l'axe de la séparation ($\alpha_\parallel$) et perpendiculaire à l'axe ($\alpha_\perp$) obtenues sont

$$\begin{aligned} \alpha_\parallel &= \frac{\alpha_A + \alpha_A + 4\alpha_A\alpha_A / r^3}{1 - 4\alpha_A\alpha_A / r^6} \\ \alpha_\perp &= \frac{\alpha_A + \alpha_A - 2\alpha_A\alpha_A / r^3}{1 - \alpha_A\alpha_A / r^6} \end{aligned} \tag{1.36}$$

où *r* est la distance qui sépare les deux atomes polarisables[35]. Si le dénominateur est zéro, la solution diverge. Cette divergence est un problème intrinsèque à ce modèle et a été appelée la catastrophe de la polarisabilité (*polarizability catastrophe*[40]). On peut pallier ce problème en modifiant le tenseur du champ dipolaire $\overline{\overline{T}}_{ij}$ de façon à écranter l'interaction entre dipôles en fonction de la distance qui les sépare. La façon la plus fréquente est le terme exponentiel de Thole[38,39] qui utilise une fonction exponentielle de la distance internucléaire. La

méthode des dipôles induits ponctuels s'est avérée très efficace pour modéliser la polarisabilité moyenne des molécules[35,41], mais connait des problèmes considérables à reproduire l'anisotropie du tenseur de polarisabilité, c'est-à-dire la dépendance directionnelle de la grandeur du dipôle induit.

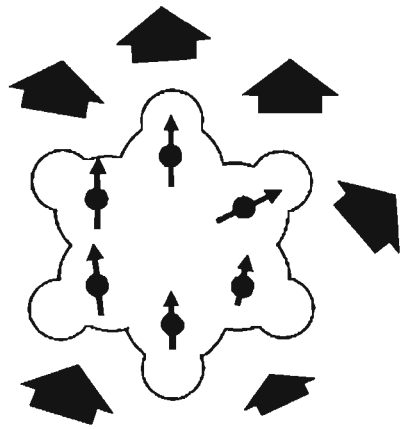


Figure 1.4. Illustration du modèle des dipôles atomiques induits pour une molécule de benzène avec des atomes de carbone polarisables qui subissent l'influence d'un champ électrique externe complexe (flèches bleues). Les charges atomiques ne sont pas montrées, mais contribuent à induire les dipôles atomiques en produisant un champ électrique interne qui s'ajoute à celui créé par les dipôles induits et au champ électrique externe.

Une correction essentielle[37,42] pour le modèle consiste à associer non pas une polarisabilité moyenne à un atome, mais plutôt un tenseur de polarisabilité atomique, ce qui ajoute une grande complexité pour la création de modèles généraux et l'ajustement des paramètres. Avec la méthode qui sera présentée dans cette thèse, nous voudrons à la fois produire un tenseur de polarisabilité adéquat et réduire le nombre de paramètres ajustables ainsi que la complexité.

Pour finir, la méthode des dipôles induits a également des variantes. Les oscillateurs de Drude[12,43] offrent une alternative calculatoire en représentant un dipôle par une charge partielle attachée au noyau par un ressort. La constante de force du ressort est directement reliée à la polarisabilité atomique et les équations ressemblent énormément à celles présentées ici. Cependant, l'utilisation d'un ressort permet d'appliquer des stratégies dans les calculs de dynamique moléculaire qui les rendent plus rapides[43,44]. Un autre modèle analogue utilise une distribution de charges par des fonctions gaussiennes et l'énergie d'interaction interne, responsable pour l'anisotropie, provient de l'intégrale coulombienne du recouvrement des densités de charges. Ce modèle a un comportement équivalent aux dipôles induits ponctuels[40] avec l'avantage que la méthode peut être généralisée aux autres moments atomiques multipolaires induits, avantage qui n'a pas encore trouvé sa niche. Ces méthodes analogues aux dipôles atomiques induits partagent le même problème de la catastrophe de la polarisabilité et de l'anisotropie[12,40,43,44].

La deuxième classe de méthodes utilisées pour introduire la polarisabilité permet aux charges partielles atomiques d'être transférées d'un atome à l'autre dans une molécule[45-50]. Elle est dite méthode des charges fluctuantes (ou *fluctuating charge*). Ce modèle se fonde sur le principe de l'égalisation de l'électronégativité qui peut être comparé au mouvement de chaleur entre bains thermiques qui équilibreront leur température. Ici, la chaleur s'apparente aux charges et la température à l'électronégativité. La justification pour cette théorie relève du champ d'étude de la théorie de la fonctionnelle de la densité[6]. À chaque atome est associée une électronégativité intrinsèque qui dépend de son environnement chimique et cette électronégativité peut être perturbée par le champ électrique qu'il subit. En utilisant le principe de l'égalisation de l'électronégativité, l'énergie de créer une charge partielle sur un atome isolé se développe en série de Taylor

$$U_A(Q_A) = U_A^0 + \chi_A^0 Q_A + \frac{1}{2} J_{AA}^0 Q_A^2 \tag{1.37}$$

avec U_A^0 l'énergie de l'atome A à l'état fondamental, χ_A^0 l'électronégativité intrinsèque de l'atome et J_{AA}^0 est deux fois la dureté de l'électronégativité de l'atome. Ces quantités peuvent se calculer à partir de la mécanique quantique, mais sont toujours, en pratique, considérées comme des paramètres ajustables. La fonction d'énergie électrostatique pour un ensemble moléculaire s'écrit alors

$$U = \sum_{A}^{atomes} U_A(Q_A) + \sum_{A<B} J_{AB} Q_A Q_B \tag{1.38}$$

où J_{AB} prend la forme de l'interaction de Coulomb pour des atomes A et B appartenant à des molécules différentes et devient une intégrale de recouvrement de Coulomb entre orbitales de Slater centrée sur les atomes d'une même molécule. Comme pour les dipôles ponctuels induits, il faut résoudre un système d'équations linéaires. Celles-ci sont obtenues en minimisant la déviation à une électronégativité atomique unique pour tout le système

$$\tilde{\chi} = \frac{\partial U}{\partial Q_A} \tag{1.39}$$

qui doit être maintenue pour chaque atome. Aucune charge n'est transférée entre molécules. Un problème pour ce modèle est que la polarisabilité ne peut se produire que dans la direction des liens chimiques. Par exemple, la molécule de benzène, selon les principes de ce modèle, a une polarisabilité nulle en dehors du plan; pour une molécule diatomique on a $\alpha_\perp = 0$. Une des approches utilisées pour remédier à ce problème est l'ajout des sites auxiliaires qui, dans le cas de la molécule de benzène, se traduit par des charges au-dessus et en dessous du plan pour la molécule de benzène. Toutefois, cette complication n'est pas générale et constitue une barrière à la généralisation du modèle et à la détermination des autres paramètres du champ de force.

Bref, avec les dipôles induits ponctuels et les charges fluctuantes, l'atteinte d'une bonne précision nécessite l'ajout de paramètres et d'atteindre un niveau de complexité qui en réduit la généralité. Une autre problématique qui limite la création de champ de forces

polarisables, généraux et fiables est le couplage des charges atomiques, responsables en partie de la polarisation intramoléculaire, et des paramètres propres à la polarisation. Ce couplage rend difficile l'optimisation des paramètres puisque des charges produisant un dipôle statique plus élevé peuvent compenser pour une polarisabilité moindre et vice-versa. C'est une forme d'indétermination.

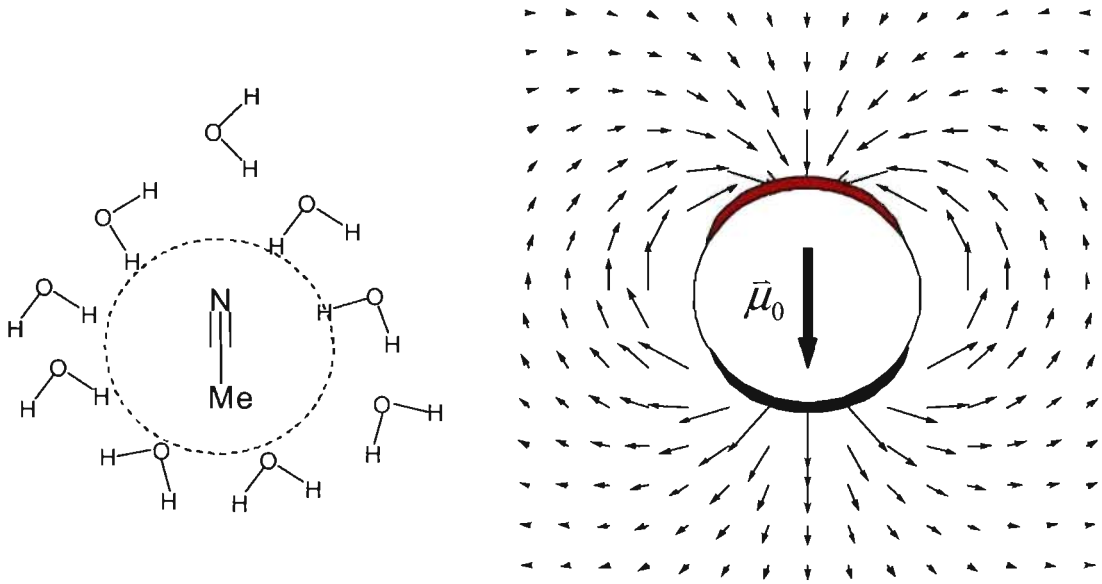
1.5 Un solvant implicite à partir du continuum diélectrique

Le sujet présenté ici fait contraste avec ceux abordés dans les sections précédentes. Il est néanmoins nécessaire, pour bien comprendre l'origine des concepts utilisés dans cette thèse, de survoler comment les effets de solvant peuvent être modélisés par un continuum diélectrique. Il s'agit ici de la polarisation moyenne des molécules de solvant par l'interaction avec un soluté. Ce type de modèle n'est pas seulement utilisé pour calculer des énergies libres d'hydratation, mais aussi pour modéliser les effets électrostatiques du solvant sur des biomolécules comme des protéines complexées à des petites molécules ou de l'ADN. Ces approches servent principalement à réduire l'échantillonnage des configurations du solvant. Elles sont donc très rapides avec une précision souvent satisfaisante. On les utilise dans le contexte de la mécanique moléculaire et de la dynamique moléculaire[51,52].

Un soluté polaire comme l'acétonitrile, placé dans un solvant polaire comme l'eau, formera des interactions polaires par l'intermédiaire des moments électriques et des ponts-H. La Figure 1.5a donne un exemple de configuration des molécules d'eau autour de l'acétonitrile dans l'eau. En fait, cette configuration statique n'est qu'un instantané (*snapshot*) des configurations que la première couche de molécules de solvant peut adopter.

Contrairement aux modèles de solvants explicites, où les coordonnées des molécules d'eau seraient échantillonnées selon une distribution de Boltzmann, un modèle

de **solvant implicite** moyenne les effets en supposant que le soluté est en équilibre avec le mouvement du solvant. Le soluté subit donc un champ électrique moyen.



a b

Figure 1.5 Le modèle de solvatation implicite suggère que le champ électrique produit par un soluté polaire organisera les configurations des molécules d'eau (a) de telle sorte que l'effet moyen peut être décrit par un champ de polarisation du solvant (b). L'accumulation des charges partielles positives des H et négatives des O de l'eau sont illustrées par du rouge et du bleu respectivement.

La Figure 1.5b donne une image du fonctionnement du modèle. Le continuum qui représente l'eau entoure une cavité sphérique qui délimite la distance d'approche des molécules. À la manière des molécules d'eau qui aligneront, en moyenne, leur dipôle selon le champ électrique, le continuum diélectrique est modélisé par une densité de polarisabilité, $\chi(\vec{r})$, appelé susceptibilité électrique, qui est une polarisabilité volumique locale. Ainsi, chaque point de l'espace se voit attribué une densité dipolaire, la

polarisation $\bar{P}(\bar{r})$, qui est un champ vectoriel volumique. Utilisant une réponse linéaire, nous obtenons

$$\bar{P}(\bar{r}) = \chi(\bar{r})\bar{E}(\bar{r}) \tag{1.40}$$

qui est semblable à l'éq. (1.32), à l'exception ici que chaque point de l'espace possède un dipôle induit. Donc, pour calculer $\bar{P}(\bar{r})$, il faut connaître le champ électrique $\bar{E}(\bar{r})$ qui peut être calculé en résolvant l'équation de Poisson

$$\bar{\nabla} \cdot (\varepsilon(\bar{r})\bar{E}(\bar{r})) = \frac{\rho^f(\bar{r})}{\varepsilon_0} \tag{1.41}$$

où

$$\varepsilon(\bar{r}) = \chi(\bar{r}) + 1 \tag{1.42}$$

est appelé fonction de diélectrique et $\rho^f(\bar{r})$ la densité de charge libre (*free charge density*) qui provient de charges statiques, par exemple les charges atomiques partielles qui reproduisent le potentiel électrostatique. Une dérivation de l'équation de Poisson est donnée en Appendice II. Dans l'éq. (1.41), ε_0 est la permittivité du vide et $\varepsilon(\bar{r})\varepsilon_0$ la permittivité locale. Ainsi, la constante diélectrique est sans unité et représente la permittivité relative au vide. Quand ε=1, la polarisabilité est nulle et quand $\varepsilon > 1$ le milieu est polarisé est s'oppose au champ électrique. La Figure 1.5b illustre le champ vectoriel de $\bar{P}(\bar{r})$ pour le problème d'un dipôle à l'intérieur d'une cavité sphérique. Une conséquence de la formation du champ de polarisation est l'apparition d'une densité de charge induite, aussi appelée densité des charges liées (*bound charge density*), qui se calcule

$$\begin{aligned} \rho^b(\bar{r}) &= -\bar{\nabla} \cdot (\chi(\bar{r})\bar{E}(\bar{r})) \\ &= -\bar{\nabla} \cdot \bar{P}(\bar{r}) \end{aligned} \tag{1.43}$$

L'appendice II démontre mathématiquement l'origine de cette équation. Ici cependant, nous pouvons comprendre intuitivement que les dipôles du solvant implicite qui s'arrêtent à la frontière de la cavité auront une extrémité soit chargée négativement (près du méthyle de l'acétonitrile) soit chargée positivement. Cela est représenté par les zones rouges et bleues à

la Figure 1.5b. Par contre, les dipôles qui se suivent à la queue dans le solvant verront leurs charges se neutraliser. Finalement, il ne faut pas croire que le modèle sous-entend que les molécules d'eau s'alignent parfaitement selon le champ vectoriel tracé à la Figure 1.5b. Le modèle prévoit plutôt un alignement moyen du solvant de telle sorte que chacun des dipôles tracés ne représente pas une molécule d'eau, mais plutôt l'effet moyen à cet endroit.

En plus d'avoir une valeur conceptuelle intéressante, la solvatation implicite a surtout été utilisée pour faire des calculs quantitatifs. Historiquement, les méthodes électrostatiques continues ont été parmi les premières approches à considérer les effets de solvant[53-55] et sont encore aujourd'hui très populaires pour estimer l'énergie libre de solvatation[51,56,57]. On peut écrire l'énergie libre pour solvater (ou hydrater dans l'eau) un soluté ΔG_{solv} en une somme de deux processus

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{np} \tag{1.44}$$

où ΔG_{np} est l'énergie libre non polaire nécessaire pour former une cavité dans le solvant et englobe la variation d'énergie et d'entropie du solvant causé par l'insertion du soluté. ΔG_{elec} est l'énergie libre électrostatique qui provient du processus qui consiste à rendre polaire la cavité formée. Ce dernier terme englobe l'entropie et l'énergie reliées à la réorganisation du solvant causé par la polarité du soluté. Dans le cas où la polarisation du soluté est prise en compte, ΔG_{elec} contient également l'énergie libérée par le solvant due à la réorganisation de la distribution de charge à l'intérieur du soluté. Théoriquement, l'éq. (1.44) est exacte puisque ΔG_{solv} est une fonction d'état et que tout chemin impliquant des transformations réversibles donne accès au travail thermodynamique entre deux états. En pratique, calculer ΔG_{solv} exactement est un défi. L'équation de Poisson est uniquement utilisée pour calculer ΔG_{elec}. Comme il en sera question à plusieurs reprises dans cette thèse, la forme de la cavité est très importante pour le calcul de ΔG_{elec}.

L'énergie libérée en mettant un dipôle statique $\bar{\mu}_0$ dans une cavité sphérique a une solution mathématique donnée par l'équation

$$\Delta G_{elec} = -\frac{1}{4\pi\varepsilon_0 a^3}\left(\frac{\varepsilon_{solv}-1}{2\varepsilon_{solv}+1}\right)|\bar{\mu}_0|^2 \qquad (1.45)$$

où a est le rayon de la cavité et ε_{solv} la constante diélectrique du solvant. En fait, ε_{solv} apparaît dans les équations à cause des éq. (1.41) et (1.42). Le continuum a une fonction diélectrique $\varepsilon(\bar{r}) = \varepsilon_{solv}$, une constante, à l'extérieur de la cavité et change abruptement à l'intérieur de celle-ci étant donné qu'aucune molécule de solvant ne peut s'y polariser. La constante diélectrique d'un liquide augmente grossièrement avec la polarité (dont le terme le plus imporant pour une molécule neutre est le moment dipolaire) de celui-ci, comme le rapporte la Table 1.1 qui met quand même en évidence que c'est plutôt la densité dipolaire (la somme des dipôles moléculaires par unité de volume) qui est corrélée avec ε_{solv}. Cela signifie que la capacité d'un liquide à se polariser dépend surtout de la densité dipolaire. L'eau constitue une exception qui est cependant assez peu significative étant donné que la différence entre deux constantes diélectriques élevées est beaucoup moins importante qu'entre deux plus petites comme l'on peut le calculer avec l'éq. (1.45).

Onsager, dans un article marquant[‡‡] publié en 1936[54], a résolu le problème posé à l'époque par les solvants polaries en ajoutant un point polarisable au centre de la sphère et résolu les équations de nouveau.

[‡‡] Cet article est le sixième plus cité de la revue *Journal of the American Chemical Society* de tous les temps.

Table 1.1. Permittivité relative (ε_{solv}) expérimentale[58] de quelques liquides (20 °C et 1 atm)

Substance	ε_{solv}	Moment dipolaire (Debye)	Volume molaire (mL/mol)	Densité dipolaire[a] (Debye×mol/mL)
Formamide	109.0	3.73	39.7	0.009
Eau	78.4	1.85	18.1	0.010
Acétonitrile	37.5	3.92	52.2	0.075
Métanol	32.7	1.70	40.5	0.042
Pyridine	12.3	2.22	80.6	0.027
Benzène	2.3	0	89.1	0.025
Cyclohexane	2.0	0	108.1	0.019

[a]Moment dipolaire / Volume molaire.

Le résultat obtenu par Onsager donne:

$$\Delta G_{elec} = -\frac{f\left|\vec{\mu}_0\right|^2}{(1-f\alpha)} \qquad (1.46)$$

$$\text{où } f = \frac{1}{4\pi\varepsilon_0 a^3}\left(\frac{\varepsilon_{solv}-1}{2\varepsilon_{solv}+1}\right)$$

avec α la polarisabilité moyenne du soluté qu'Onsager a obtenue par la formule suivante

$$\alpha = \left(\frac{n_D^2-1}{n_D^2+2}\right)\left(\frac{3V_m}{4\pi N_A}\right) \qquad (1.47)$$

qui utilise l'indice de réfraction n_d, le volume molaire V_m et la constante d'Avogadro N_A. Nous reviendrons sur l'origine de l'éq. (1.47) à la section suivante. Ici, il suffit de dire que le

facteur le plus à droite représente le cube du rayon d'une sphère qui a le volume qu'occupe en moyenne une seule molécule.

Ces modèles, développés avant l'avènement de super ordinateurs, ont été cruciaux pour la compréhension de la solvatation et du comportement électrique des molécules en interactions. Depuis le milieu des années 1980, des méthodes numériques ont permis de délaisser la cavité sphérique et de résoudre l'équation de Poisson avec des cavités qui épousent davantage la forme des molécules. La Figure 1.6 montre la surface de van der Waals du benzène qui se forme en prenant l'enveloppe extérieure du volume formé par des sphères positionnées sur les atomes. Généralement, les rayons des sphères atomiques proviennent de distances de contact tirées de données cristallographiques[59]. La constante diélectrique à l'extérieur de la cavité provient des valeurs expérimentales alors que le diélectrique interne est souvent 1.

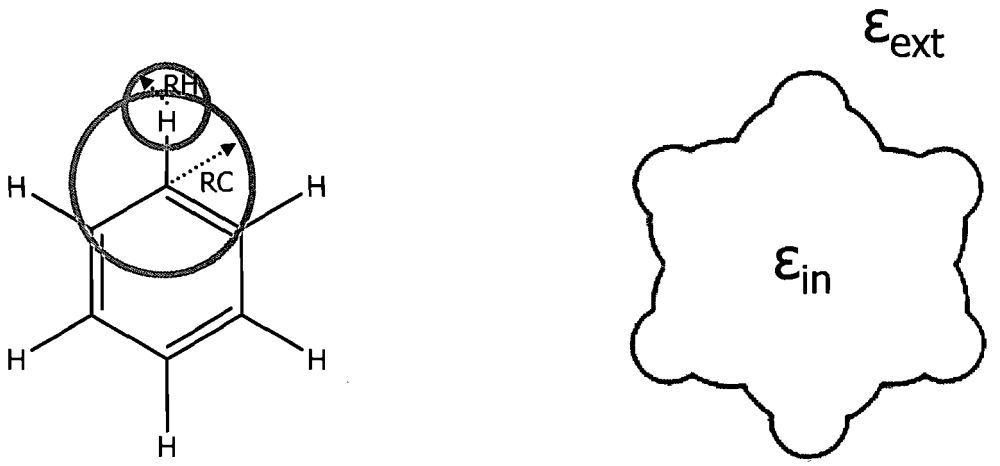


Figure 1.6. Une cavité dans un diélectrique est formée pour une molécule de benzène. Des sphères sont positionnées sur les atomes de la molécule avec différents rayons. L'enveloppe extérieure définit la cavité. Il s'agit d'une surface dite de van der Waals (vdW).

Finalement, il est opportun de faire mention d'une méthode appelée Poisson-Boltzmann (PB) qui ajoute l'effet des sels sur les énergies de solvatation. Ici, on reprend l'équation de Poisson donnée par l'éq. (1.41) et on ajoute au terme de droite une distribution de charges supplémentaires

$$\bar{\nabla}(\varepsilon_0 \varepsilon(\bar{r})\bar{\nabla}\varphi(\bar{r})) = -\rho^f(\bar{r}) - \sum_i^{ions} c_i z_i \lambda(\bar{r}) \exp\left(-\frac{z_i \varphi(\bar{r})}{k_B T}\right) \tag{1.48}$$

avec $\varphi(\bar{r})$ le potentiel électrostatique[§§], c_i la concentration de l'ion i de charge z_i, $\lambda(\bar{r})$ une surface qui délimite la distance d'approche des ions de la molécule, k_B la constante de Boltzmann et T la température. Ce que l'éq. (1.48) ajoute, c'est une distribution uniforme de charges selon une distribution de Boltzmann dont l'énergie est simplement donnée par l'interaction des ions avec le potentiel électrostatique combiné du solvant et du soluté. Des zones avec potentiel négatif seront davantage peuplées par des cations Na^+, ce qui se traduit par une densité de charge positive dans cette région. L'éq. (1.48) peut aussi être rendue linéaire quand la concentration des sels est faible, comme suggéré par la théorie Debye-Hückel. Un terme linéaire remplace alors l'exponentielle par le premier terme dans un développement en série de Taylor et accélère grandement les solutions numériques. L'équation de Poisson utilisée dans le restant de cette thèse est un cas particulier de l'équation de Poisson Boltzmann pour laquelle la concentration des ions est zéro.

Les modèles de solvants implicites ont fait leurs preuves pour le calcul d'énergies libres d'hydratation et se sont montrés utiles pour la simulation de biomolécules. Ils présentent cependant des limites dont il faut être conscient. Premièrement, ils permettent d'étudier des phénomènes en équilibre thermodynamique pour lesquels les degrés de liberté configurationels du solvant peuvent être échantillonnés pour un processus donné. Cela suppose que la relaxation dipolaire est adiabatique au processus étudié. Deuxièment, les molécules de solvant sont considérées isotropes dans le sens où les détails de leur potentiel

---

[§§] La relation entre le potentiel électrostatique et le champ électrique est $-\bar{\nabla}\varphi(\bar{r}) = \bar{E}(\bar{r})$.

électrostatique sont réduits à un simple dipôle[60]. Troisièmement, l'absence de molécules explicites enlève complètement l'effet de viscosité du solvant.

1.6 Le modèle polarisable proposé

L'idée principale de cette thèse est d'utiliser le principe de diélectrique continu non pas pour décrire la solvatation, mais plutôt pour modéliser la polarisation électronique. L'approche mathématique considérée, comme pour la solvatation, est l'équation de Poisson donnée par l'éq. (1.41).

Reprenant l'idée d'un continuum polarisable, l'intérieur de la molécule est dit continument polarisable. Comme schématisé à la Figure 1.7a, l'intérieur de la molécule a une constante diélectrique plus élevée alors que le diélectrique externe est celui du vide. En d'autres mots, nous venons de créer une molécule de benzène dans le vide où ρ^b de l'éq. (1.43) joue le rôle du $\delta\rho$ de l'éq. (1.16). Le principe de la polarisation interne est schématisé à la Figure 1.7b où un champ électrique complexe externe (comme des molécules polaires voisines) induit des dipôles sur l'ensemble du volume de la molécule. Cela fait contraste avec le modèle des dipôles ponctuels induits et des charges fluctuantes étant donné que l'effet est distribué et ne tient pas seulement compte du champ électrique évalué aux positions atomiques, mais plutôt sur l'ensemble du volume moléculaire. Un tel modèle pourrait être utilisé pour tenir compte de la polarisation électronique dans un champ de force polarisable, car l'équation de Poisson nous donnera accès au potentiel électrostatique et donc aux énergies d'interaction intermoléculaires.

Resoulignons l'idée que l'équation de Poisson sous-entend que la polarisation est obtenue de façon adiabatique, c'est-à-dire, que la réponse du milieu polarisé est considérée comme étant instantannée. De ce point de vue, les électrons ont une réponse énormément plus rapide que les noyaux, et l'équation de Poisson embrasse complètement l'idée de la

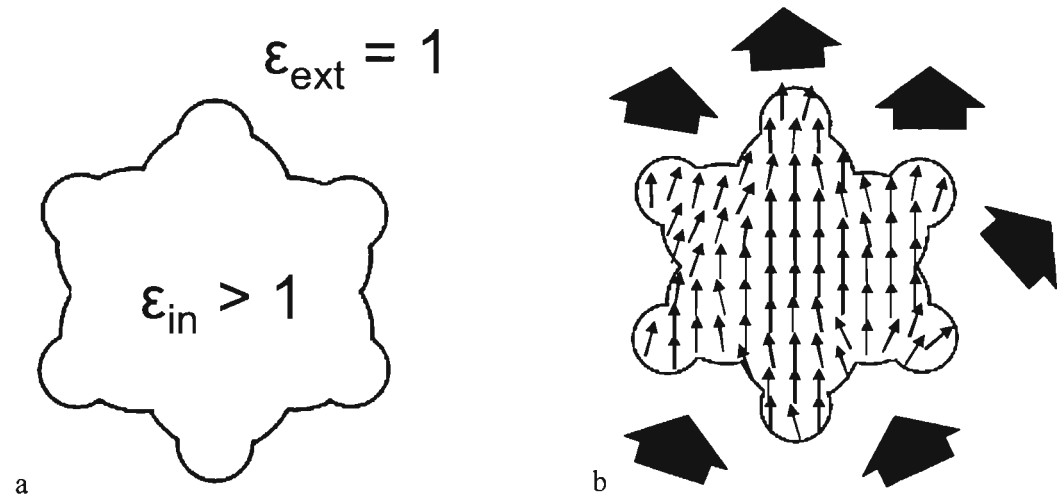


Figure 1.7. Cette figure illustre le concept d'un continuum polarisable. Un champ électrique externe est appliqué (flèches bleues) à la molécule de benzène (a). L'intérieur du volume polarisable qui enveloppe la molécule est polarisé selon les lignes du champ électrique externe. À chaque point de l'espace, un dipôle infinitésimal est induit. Si le champ externe n'est plus uniforme (b), les dipôles induits s'alignent toujours selon le champ externe.

surface de Born-Oppenheimer. Même qu'en ce sens, l'équation de Poisson est plus appropriée pour modéliser la polarisation électronique que la polarisation du solvant.

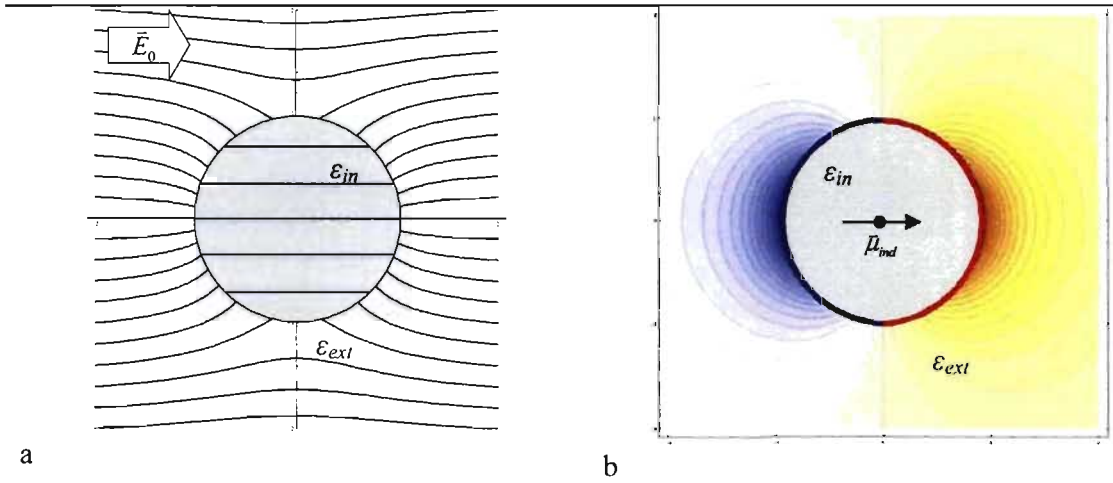
Il est utile à ce point d'étudier un problème pour lequel une solution analytique existe. Prenons une sphère avec une constante diélectrique ε_{in} à l'intérieur et ε_{ext}= 1 à l'extérieur. En appliquant un champ électrique uniforme $\vec{E}_0$, nous pouvons calculer que le potentiel électrostatique induit est celui d'un dipôle ponctuel situé au centre de la sphère[61] et prend la valeur

$$\bar{\mu}_{ind} = a^3 \left(\frac{\varepsilon_{in} - 1}{\varepsilon_{in} + 2} \right) \times \bar{E}_0 \tag{1.49}$$

où a est le rayon de la sphère. Les lignes du champ électrique*** ainsi que le dipôle induit pour ce problème sont illustrés à la Figure 1.8. En comparant les éq. (1.49) et (1.21), on observe que la polarisabilité de la sphère prend la valeur

$$\alpha_{sphère} = a^3 \left(\frac{\varepsilon_{in} - 1}{\varepsilon_{in} + 2} \right). \tag{1.50}$$

Il est intéressant de souligner que pour une sphère, à tout le moins, le modèle des dipôles induits ponctuels et le diélectrique continu sont équivalents. Par contre, il deviendra évident dans les chapitres suivants que cela n'est pas le cas de façon générale.



a b

Figure 1.8 Les lignes du champ électrique total sont représentées pour un diélectrique sphérique dans le vide qui subit un champ électrique externe uniforme orienté selon $\hat{x}$ (a) À l'intérieur de la sphère le champ électrique est inférieur au champ externe dû à la polarisation. Le potentiel électrostatique induit en (b), produit par la polarisation de la sphère, est celui d'un dipôle ponctuel placé au centre de la sphère et résulte de charges liées (*bound charges*) accumulées à la surface de la sphère.

*** Une ligne de champ électrique est la trajectoire que prendrait une particule chargée soumise au champ électrique en question. Plus les lignes sont denses, plus le champ augmente. La direction va du potentiel élevé vers le potentiel plus petit.

Nous avons discuté de la notion de continuum diélectrique dans les contextes très différents de la polarisation du solvant et d'électrons. Pour faire le lien entre les différentes valeurs expérimentales de diélectrique que peut prendre une seule substance avec les processus moléculaires sous-jacents, il faut parler du phénomène de relaxation diélectrique. Une onde électromagnétique est une source fréquemment utilisée de champ électrique qui varie en fonction du temps. La réponse d'une substance au champ électrique dépend de sa capacité à relaxer durant une période d'oscillation. Si l'oscillation est trop rapide, le champ moyen senti sera nul et le processus de relaxation n'aura pas le temps de prendre place[62]. La polarisation électronique survient dans le domaine du visible avec un temps de relaxation sous la femtoseconde. La polarisation atomique, qui résulte d'une réponse vibrationnelle, se produit dans le domaine de l'infrarouge avec un temps de relaxation de quelques de femtosecondes. Finalement, la relaxation dipolaire, qui résulte de la réorientation des dipôles suite à la rotation moléculaire, survient dans le domaine des micro-ondes, et prend place à l'intérieur d'une nanoseconde. Il est à noter que, pour un champ électrique donné, tous les phénomènes plus rapides jouent un rôle. La constante diélectrique du solvant est principalement due à la relaxation dipolaire, mais aussi contient les contributions de toutes les autres relaxations. La constante diélectrique qui correspond à la relaxation électronique a un lien étroit avec l'indice de réfraction. En effet, la relation entre l'indice de réfraction et le diélectrique de polarisation électronique est donnée par[61]

$$\varepsilon_\infty = n^2 \tag{1.51}$$

avec n l'indice de réfraction et ε_∞ la constante diélectrique dans la limite d'une fréquence infinie, c'est-à-dire quand seulement les électrons répondent au champ[†††]. En pratique, la bande jaune du sodium à 589 nm (période de 2 fs) est souvent utilisée pour mesurer ε_∞.

[†††] L'éq. (1.51) est donnée par un tenseur si le matériau est anisotrope, c'est-à-dire que l'arrangement des molécules est différent, en moyenne, selon les axes. Un matériau biréfringent a cette propriété. Les liquides, cependant, en l'absence d'un champ électrique externe macroscopique, sont isotropes.

Comme l'indice de réfraction de substances bioorganiques est compris entre 1.3 et 1.7, ε_∞ adopte des valeurs qui s'étendent entre 1.7 et 2.9.

L'idée que la composante électronique de la polarisation macroscopique peut être représentée par un diélectrique n'est donc pas nouvelle dans cette thèse[63-65]. Par contre, ce concept n'a été appliqué que dans les calculs avec solvant implicite sans même vouloir décrire les détails moléculaires de la polarisation. Dans les calculs avec solvant implicite, il est ainsi parfois reconnu que ces effets électroniques peuvent être bénéfiques et, quand ils sont inclus, la valeur approximative utilisée pour ε_{in} est 2.[54,57,64,66-70] Ce raisonnement se fonde sur les concepts énoncés au paragraphe précédent et donc, par analogie, on admet l'égalité $\varepsilon_{in}= \varepsilon_\infty$. Le lien entre ε_{in} et ε_∞ date de l'article d'Onsager de 1936[54] dans lequel il utilise le résultat de l'éq. (1.50) avec $\varepsilon_{in} = \varepsilon_\infty$ pour obtenir l'éq. (1.47). Nous verrons aux chapitres 2 et 4 que ce choix n'est pas approprié pour l'obtention de polarisabilité moléculaire et nous proposerons une explication.

1.7 Problèmes étudiés dans cette thèse

Tout d'abord, l'objectif de l'utilisation des travaux présentés dans cette thèse est d'inclure la polarisation électronique dans un champ de force polarisable dont le but ultime est de pouvoir étudier, avec plus d'exactitude, une panoplie de problèmes qui vont de l'énergie libre de liaison d'un inhibiteur à une enzyme, à l'étude des propriétés de l'état liquide.

La nouvelle approche méthodologique fondée sur l'électrostatique continue est le point central des travaux présentés dans cette thèse. Sa validation occupera une place prépondérante. Au chapitre 2, nous ferons la démonstration que le concept proposé fonctionne pour modéliser la polarisabilité moléculaire de molécules isolées à caractère différent. Au chapitre 3, nous poserons les bases méthodologiques pour adapter le nouveau modèle polarisable aux calculs d'énergie intermoléculaires, quantités nécessaires dans un

champ de force. Au chapitre 4, nous ferons la démonstration que le modèle développé est généralisable à un très grand nombre de fonctionnalités chimiques. Nous développerons une nouvelle fonction de diélectrique qui permet d'inclure de façon précise la polarisabilité moléculaire dans les calculs d'énergie libre d'hydratation avec un modèle de solvent continu implicite. Ce chapitre démontre également que des propriétés propres à la phase condensée peuvent être obtenues de façon précise. Nous clôturerons cette thèse avec une conclusion.

1.8 Bibliographie

1. Gingras, Y. *Parlons sciences;* Les Éditions du Boréal: Montréal, 2008.

2. Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.*, **2007,** *371*, 1118-1134.

3. Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B*, **2007,** *111*, 2242-2254.

4. Szabo, A. *Modern Quantum Chemistry,* McGraw-Hill: New York, 1989

5. Cramer, C. J. *Essentials of Computational Chemistry Theories and Models;* 2 ed.; John Wiley & Sons: San Francisco, 2004.

6. Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules;* Oxford Science Publications: New York, 2009.

7. Koch, W.; Haulthausen, M. C. *A Chemist's Guide to Density Functional Theory;* 2 ed.; Wiley-VCH: Weinheim, 2001.

8. Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A*, **1988,** *38*, 3098-3100.

9. Becke, A. D. Density-Functional Thermochemistry .3. the Role of Exact Exchange. *J. Chem. Phys.*, **1993,** *98*, 5648-5652.

10. Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.*, **1993,** *98*, 1372-1377.

11. Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, **1996,** *77*, 3865-3868.

12. Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D.; Roux, B. Atomic level anisotropy in the electrostatic modeling of lone pairs for a polarizable force field based on the classical Drude oscillator. *J. Chem. Theory Comput.*, **2006,** *2*, 1587-1597.

13. Truchon, J.-F.; Nicholls, A.; Iftimie, R. I.; Roux, B.; Bayly, C. I. Accurate Molecular Polarizability based on Continuum Electrostatics. *J. Chem. Theory Comput.*, **2008,** *4*, 1480-1493.

14. Calaminici, P.; Jug, K.; Koster, A. M. Density functional calculations of molecular polarizabilities and hyperpolarizabilities. *The Journal of Chemical Physics*, **1998,** *109*, 7756-7763.

15. Stone, A. J. *The Theory of Intermolecular Forces;* Oxford University Press: Oxford, 2002; Vol. 32.

16. Dykstra, C. E.; Jasien, P. G. Derivative Hartree-Fock Theory to All Orders. *Chem. Phys. Lett.*, **1984,** *109*, 388-393.

17. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A 2Nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.*, **1995,** *117*, 5179-5197.

18. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.*, **1993,** *97*, 10269-10280.

19. Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.*, **2003,** *119*, 5185-5197.

20. Silvestrelli, P. L.; Parrinello, M. Water molecule dipole in the gas and in the liquid phase. *Phys. Rev. Lett.*, **1999,** *82*, 3308-3311.

21. Cox, S. R.; Williams, D. E. Representation of the Molecular Electrostatic Potential by A Net Atomic Charge Model. *J. Comput. Chem.*, **1981,** *2*, 304-323.

22. Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.*, **1984,** *5*, 129-145.

23. Dougherty, D. A. Cation-pi interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science*, **1996,** *271*, 163-168.

24. Ma, J. C.; Dougherty, D. A. The cation-pi interaction. *Chem. Rev.*, **1997,** *97*, 1303-1324.

25. Gallivan, J. P.; Dougherty, D. A. Cation-pi interactions in structural biology. *P. Natl. Acad. Sci. U. S. A.*, **1999,** *96*, 9459-9464.

26. Caldwell, J. W.; Kollman, P. A. Cation-Pi Interactions - Nonadditive Effects Are Critical in Their Accurate Representation. *J. Am. Chem. Soc.*, **1995,** *117*, 4177-4178.

27. Truchon, J.-F.; Nicholls, A.; Grant, J. A.; Iftimie, R. I.; Roux, B.; Bayly, C. I. Extending the Polarizable Continuum Dielectric Model to Account for Electronic Polarization in Intermolecular Interactions. *Submitted*, **2008.**

28. Chen, Y. F.; Dannenberg, J. J. Cooperative 4-pyridone H-bonds with extraordinary stability. A DFT molecular orbital study. *J. Am. Chem. Soc.*, **2006,** *128*, 8100-8101.

29. Wick, C.; Dang, L. X. Molecular mechanism of transporting a polarizable iodide anion across the water-CCl4 liquid/liquid interface. *J. Chem. Phys.*, **2007,** *126*, 134702.

30. Jungwirth, P.; Tobias, D. J. Chloride anion on aqueous clusters, at the air-water interface, and in liquid water: Solvent effects on Cl- polarizability. *J. Phys. Chem. A*, **2002,** *106*, 379-383.

31. Lamoureux, G.; Roux, B. Absolute hydration free energy scale for alkali and halide ions established from simulations with a polarizable force field. *J. Phys. Chem. B*, **2006,** *110*, 3308-3322.

32. Dang, L. X.; Rice, J. E.; Caldwell, J.; Kollman, P. A. Ion Solvation in Polarizable Water - Molecular-Dynamics Simulations. *J. Am. Chem. Soc.*, **1991,** *113*, 2481-2486.

33. Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. U. S. A.*, **2008,** *105*, 6290-6295.

34. Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, **1976,** *103*, 227-249.

35. Applequist, J.; Carl, J. R.; Fung, K. K. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.*, **1972,** *94*, 2952-2960.

36. Applequist, J. Atom Dipole Interaction-Model for Molecular Optical-Properties. *Acc. Chem. Res.*, **1977,** *10*, 79-85.

37. Birge, R. R. Calculation of Molecular Polarizabilities Using An Anisotropic Atom Point Dipole Interaction-Model Which Includes the Effect of Electron Repulsion. *J. Chem. Phys.*, **1980,** *72*, 5312-5319.

38. Thole, B. T. Molecular Polarizabilities Calculated with A Modified Dipole Interaction. *Chem. Phys.*, **1981,** *59*, 341-350.

39. van Duijnen, P. T.; Swart, M. Molecular and atomic polarizabilities: Thole's model revisited. *J. Phys. Chem. A*, **1998,** *102*, 2399-2407.

40. Elking, D.; Darden, T.; Woods, R. J. Gaussian induced dipole polarization model. *J. Comput. Chem.*, **2007,** *28*, 1261-1274.

41. Miller, K. J.; Savchik, J. A. New Empirical-Method to Calculate Average Molecular Polarizabilities. *J. Am. Chem. Soc.*, **1979,** *101*, 7206-7213.

42. Harder, E.; Anisimov, V. M.; Whitfield, T.; MacKerell, A. D.; Roux, B. Understanding the Dielectric Properties of Liquid Amides from a Polarizable Force Field. *J. Phys. Chem. B*, **2008,** *112*, 3509-3521.

43. Lamoureux, G.; Roux, B. Polarizable force-field based on drude oscillators: Molecular dynamics simulation and parametrization. *Biophys. J.*, **2001,** *80*, 328A-328A.

44. Lamoureux, G.; Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.*, **2003,** *119*, 3025-3039.

45. Gasteiger, J.; Marsili, M. New Model for Calculating Atomic Charges in Molecules. *Tetrahedron Lett.*, **1978,** 3181-3184.

46. Rappe, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.*, **1991,** *95*, 3358-3363.

47. Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.*, **1994,** *101*, 6141-6156.

48. Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R. H.; Berne, B. J.; Friesner, R. A. Fluctuating charge, polarizable dipole, and combined models: Parameterization from ab initio quantum chemistry. *J. Phys. Chem. B*, **1999,** *103*, 4730-4737.

49. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X. X.; Murphy, R. B.; Zhou, R. H.; Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *J. Comput. Chem.*, **2002,** *23*, 1515-1531.

50. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. *J. Phys. Chem. A*, **2004,** *108*, 621-627.

51. Prabhu, N. V.; Zhu, P. J.; Sharp, K. A. Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method. *J. Comput. Chem.*, **2004,** *25*, 2049-2064.

52. Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.*, **2002,** *23*, 1244-1253.

53. Born, M. Volumen und Hydratationswärme der Ionen. *Z. Phys. A: Hadrons Nucl.*, **1920,** *1*, 45-48.

54. Onsager, L. Electric Moments of Molecules in Liquids. *J. Am. Chem. Soc.*, **1936,** *58*, 1486-1493.

55. McQuarrie, D. A. *Statistical Mechanics;* University Science Books: Sausalito, 2000.

56. Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspectives in Drug Discovery and Design*, **2000,** *18*, 113-135.

57. Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. *J. Med. Chem.*, **2008,** *51*, 769-779.

58. *CRC Handbook of Chemistry and Physics;* 74 ed.; CRC Press, Inc.: Boca Raton, 1993.

59. Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.*, **1964,** *68*, 441-451.

60. Mobley, D. L.; Barber, A. E.; Fennell, C. J.; Dill, K. A. Charge asymmetries in hydration of polar solutes. *J. Phys. Chem. B*, **2008,** *112*, 2405-2414.

61. Bottcher, C. J. F. *Theory of Electric Polarization;* 2 ed.; Elsevier Scientific Publishing Company: New York, 1973; Vol. 1.

62. *Broadband Dielectric Spectroscopy;* Springer-Verlag: Berlin, 2003.

63. Sharp, K.; Jean-Charles, A.; Honig, B. A Local Dielectric-Constant Model for Solvation Free-Energies Which Accounts for Solute Polarizability. *J. Phys. Chem.*, **1992,** *96*, 3822-3828.

64. Tan, Y. H.; Luo, R. Continuum treatment of electronic polarization effect. *J. Chem. Phys.*, **2007,** *126*, 094103.

65. Tan, Y. H.; Tan, C. H.; Wang, J.; Luo, R. Continuum polarizable force field within the Poisson-Boltzmann framework. *J. Phys. Chem. B*, **2008,** *112*, 7675-7688.

66. Warwicker, J.; Watson, H. C. Calculation of the Electric-Potential in the Active-Site Cleft Due to Alpha-Helix Dipoles. *J. Mol. Biol.*, **1982,** *157*, 671-679.

67. Jayaram, B.; Liu, Y.; Beveridge, D. L. A modification of the generalized Born theory for improved estimates of solvation energies and pK shifts. *J. Chem. Phys.*, **1998,** *109*, 1465-1471.

68. Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Modeling electrostatic effects in proteins. *Biochimica et Biophysica Acta-Proteins and Proteomics*, **2006,** *1764*, 1647-1676.

69. Green, D. F.; Tidor, B. Evaluation of a initio charge determination methods for use in continuum solvation calculations. *J. Phys. Chem. B*, **2003,** *107*, 10261-10273.

70. Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J. Phys. Chem.*, **1994,** *98*, 1978-1988.

2 Élaboration d'un modèle précis et rapide fondé sur l'électrostatique continue pour la prédiction de la polarisabilité moléculaire

Résumé. Une nouvelle approche pour modéliser la polarisation électronique par un continuum diélectrique est introduite. Il est démontré, à l'aide de solutions à l'équation de Poisson par différences finies, que le modèle de la polarisation électronique par le continuum interne produit des tenseurs de polarisabilité moléculaire en phase gazeuse précis pour un ensemble de 98 molécules reconnues difficiles. Cet ensemble est composé d'hétéroaromatiques, d'alcanes et de diatomiques. La source de la polarisation électronique est un diélectrique interne élevé capable de calculer des polarisabilités en accord avec les valeurs produites par B3LYP/aug-cc-pVTZ et l'expérience. À la différence d'autres approches qui modélisent la polarisation électronique, ce modèle simple évite la catastrophe de la polarisabilité et permet de calculer des anisotropies moléculaires précises à l'aide de très peu de paramètres ajustables, et ce, sans recourir à des sites auxiliaires ou à des centres atomiques anisotropes. En moyenne, l'erreur positive de la polarisabilité moyenne et l'erreur produite sur l'anisotropie, relativement à B3LYP, sont de 2 % et de 5 %, respectivement. La corrélation entre les moments principaux de la polarisabilité de B3LYP et ceux obtenus avec cette approche donne un R^2 de 0.990 et une pente de 0.999. Même l'anisotropie de F_2, une véritable gageure pour les modèles existants, peut être reproduite avec une erreur de 2 %. En plus de rendre disponibles de nouveaux paramètres pour une méthode rapide et applicable telle quelle au calcul des polarisabilités, ce travail élargit l'usage de l'équation de Poisson à des domaines qui requièrent des polarisabilités moléculaires précises.

Mise en contexte. Le premier pas à faire dans l'élaboration d'une nouvelle méthodologie est de la valider. Dans ce chapitre, nous vérifions que l'idée d'utiliser un continuum diélectrique pour modéliser la polarisabilité moléculaire fonctionne. Nous vérifions également si elle est facilement généralisable, c'est-à-dire si l'on peut optimiser un petit

nombre de paramètres tout en maintenant un bon niveau de précision. Cette caractéristique est importante si l'on veut être en mesure de développer des paramètres pour un grand nombre de fonctionnalités chimiques, comme celles que l'on retrouve dans les molécules biologiquement actives. Un des éléments clés pour le succès de notre méthode est l'idée d'optimiser les paramètres sur les tenseurs de polarisabilités quantiques (ou expérimentaux). Or, pour en être capable, il a fallu développer une méthode pour calculer la polarisabilité d'un diélectrique de volume quelconque dans le vide. Cette technique n'avait pas été rapportée dans la littérature et constitue une nouveauté qui paraît pratiquement banale dans l'article, mais qui est au cœur même de travail présenté dans cette thèse.

Dans ce chapitre comme dans les suivants, les calculs des tenseurs de polarisabilités qui servent de référence proviennent de la théorie de la fonctionelle de la densité et de la fonctionnelle d'échange et corrélation B3LYP. Ce choix permet d'obtenir des polarisabilités moyennes avec une erreur de 5 % en moyenne par rapport à l'expérience quand une base orbitalaire avec suffisamment de fonctions diffuses est utilisée. Des erreurs systématiques plus élevées ont été rapportées et caractérisées par d'autres chercheurs et nous en avons tenu compte. Nous avons estimé que la méthode B3LYP est la plus exacte étant donné les ressources de calcul dont nous disposions. Nous estimons que si nous pouvons reproduire le niveau de théorie B3LYP, notre méthode empirique sera utile et pourra éventuellement être améliorée avec des ressources calculatoires supplémentaires. Cependant, il est probable que B3LYP soit une approche suffisamment exacte pour l'élaboration de modèles empiriques.

Reproduit avec la permission de : Jean-François Truchon, Anthony Nicholls, Radu I. Iftimie, Benoît Roux et Christopher I. Bayly, Accurate Molecular Polarizabilities Based Continuum Electrostatics, *Journal of Chemical Theory and Computation* **4**, 1480 (2008). Copyright 2008 American Chemical Society.

Accurate molecular polarizabilities based on continuum electrostatics

Jean-François Truchon[a,b], Anthony Nicholls[c], Radu I. Iftimie[a], Benoît Roux[d] and Christopher I. Bayly[b]

[a]Département de chimie, Université de Montréal, C.P. 6128 Succursale centre-ville, Montréal, Québec, Canada H3C 3J7 [b]Merck Frosst Canada Ltd., 16711 TransCanada Highway, Kirkland, Québec, Canada H9H 3L1 [c]OpenEye Scientific Software, Inc., Santa Fe, New Mexico 87508 [d]Institute of Molecular Pediatric Sciences, Gordon Center for Integrative Science, University of Chicago, Illinois 929 East 57th Street, Chicago, Illinois 60637

Abstract

A novel approach for representing the intramolecular polarizability as a continuum dielectric is introduced to account for molecular electronic polarization. It is shown using a finite-difference solution to the Poisson equation, that the Electronic Polarization from Internal Continuum (EPIC) model yields accurate gas-phase molecular polarizability tensors for a test set of 98 challenging molecules composed of heteroaromatics, alkanes and diatomics. The electronic polarization originates from a high intramolecular dielectric that produces polarizabilities consistent with B3LYP/aug-cc-pVTZ and experimental values when surrounded by vacuum dielectric. In contrast to other approaches to model electronic polarization, this simple model avoids the polarizability catastrophe and accurately calculates molecular anisotropy with the use of very few fitted parameters and without resorting to auxiliary sites or anisotropic atomic centers. On average, the unsigned error in the average polarizability and anisotropy compared to B3LYP are 2% and 5%, respectively. The correlation between the polarizability components from B3LYP and this approach lead to a R^2 of 0.990 and a slope of 0.999. Even the F_2 anisotropy, shown to be a difficult case

for existing polarizability models, can be reproduced within 2% error. In addition to providing new parameters for a rapid method directly applicable to the calculation of polarizabilities, this work extends the widely used Poisson equation to areas where accurate molecular polarizabilities matter.

2.1 Introduction

The linear response of the electronic charge distribution of a molecule to an external electric field, the polarizability, is at the origin of many chemical phenomena such as electron scattering[1], circular dichroism[2], optics[3], Raman scattering[4], softness and hardness[5], electronegativity[6], etc. In atomistic simulations, polarizability is believed to play an important and unique role in intermolecular interactions of heterogeneous media such as ions passing through ion channel in cell membranes[7], in the study of interfaces[8] and in protein-ligand binding[9].

Polarizability is considered to be a difficult and important problem from a theoretical point of view. Much effort has been invested in the calculation of molecular polarizability at different levels of approximation. At the most fundamental level, electronic polarization is described by quantum mechanics (QM) electronic structure theory such as extended basis set density functional theory (DFT) and ab initio molecular orbital theory. However, the extent of the computational resources required is an impediment to the wide application of these methods on large molecular sets or on large molecular systems such as drug-like molecules[10]. In order to circumvent these limitations, empirical physical models, based on classical mechanics, have been parameterized to fit experimental or quantum mechanical polarizabilities.

In this article, we explore a new empirical physical model to account for electronic polarizability in molecules. The Electronic Polarization from Internal Continuum (EPIC) model uses a dielectric constant and atomic radii to define the electronic volume of a molecule. The molecular polarizability tensor is calculated by solving the Poisson equation (PE) with a finite difference algorithm. The concept that a dielectric continuum can account for solute polarizability has been examined previously. For example, Sharp *et al.*[11] showed that condensed phase induced molecular dipole moments are accounted for with the continuum solvent approach and that it leads to accurate electrostatic free energy of

solvation. More recently Tan and Luo[12] have attempted to find an optimal inner dielectric value that reproduces condensed phase dipole moments in different continuum solvents. In spite of these efforts, we found that none of these models can account correctly for molecular polarizability. Here, the concept is explored with the objective of producing a high accuracy polarizable electrostatic model. Therefore, we focus on the optimization of atomic radii and inner dielectrics to reproduce the B3LYP/aug-cc-pVTZ polarizability tensor.

In this preliminary work, we seek to establish the soundness and accuracy of the EPIC model in the calculation of the molecular polarizability tensor on three classes of molecules: homonuclear diatomics, heteroaromatics and alkanes. These molecular classes required a special attention with previous polarizable models due to their high anisotropy[13-15]. Overall, 53 different molecules are used to fit our model and 45 molecules to validate the results. Five specific questions are addressed: Can this model accurately calculate the average polarizability? If so, can it further account for the anisotropy and the orientation of the polarizability components? How few parameters are needed to account for highly anisotropic molecules and how does this compare to other polarizable models? How transferable are the parameters obtained with this model? Is the model able to account for conformational dependency? In answering these questions, we obtained a fast and validated method with optimized parameters to accurately calculate the molecular polarizability tensor for a large variety of heteroaromatics not previously considered.

The remainder of the article is organized as follows. In the next section, we briefly review the most successful existing polarizable approaches. Then we introduce the dielectric polarizable method with a polarizable sphere analytical model. A methodology section in which we outline the computational details follows. The molecular polarizability results are then reported. This is followed by a discussion and conclusion.

2.2 Existing empirical polarizable models

In this section, we briefly review existing models, focusing on aspects relevant to this study.

2.2.1 Point inducible dipole

The point inducible dipole model (PID) was first outlined by Silberstein in 1902[16]. This model has been extensively used to calculate molecular polarizability[14,15,17-22] and to account for many-body effects in condensed phase simulations[23-25]. Typically, in the PID, an atom is a polarizable site where the electric field direction and strength together with the atomic polarizability define the induced atomic dipole moment. Since the electric field at an atomic position is in part due to other atoms' induced dipoles, the set of equations must be solved iteratively (or through a matrix inversion). In 1972, Applequist[19] showed that the PID can accurately reproduce average molecular polarizability of a diverse set of molecules, but also that the mathematical formulation of the PID can lead to a polarizability catastrophe. Briefly, when two polarizable atoms are close to each other, the solution to the mathematical equations involved is either undetermined (with the matrix inversion technique) or the neighbor dipole moments cooperatively increase to infinity. To circumvent this problem, Thole[14,22] modified the dipole field tensor with a damping function, which depends on a lengthscale parameter meant to represent the spatial extent of the polarized electronic clouds; his proposed exponential modification is still important and remains in use[13,14,26].

2.2.2 Drude oscillator

The Drude oscillator (DO) represents electronic polarization by introducing a massless charged particle attached to each polarizable atom by a harmonic spring[27]. When the Drude charge is large and tightly bound to its atom, the induced dipole essentially

behaves like a PID. The DO model is attractive because it preserves the simple charge-charge radial Coulomb electrostatic term already present and it can be used in molecular dynamics simulation packages without extensive modifications. The DO model has not yet been extensively parameterized to reproduce molecular polarizability tensors, but recent results suggest that it could perform as well as PID methods. Finally, the DO model also requires a damping function to avoid the polarizability catastrophe[26].

2.2.3 Fluctuating charges

A third class of empirical model, called fluctuating charge (FQ), was first published in a study by Gasteiger and Marsili[28] in 1978 to rapidly estimate atomic charges. Subsequently, FQ was adapted to reproduce molecular polarizability and applied in molecular dynamic simulations[29,30]. It is based on the concept that partial atomic charges can flow from one atomic center to another based on the local electrostatic environment surrounding each atom. The equilibrium point is reached when the defined atomic electronegativities are equal. The FQ model, like the DO, has mainly been used in condensed phase simulations and not specifically parameterized to reproduce molecular polarizabilities. A major problem with FQ is the calculation of directional polarizabilities (eigenvalues of the polarizability tensor). For planar or linear chemical moieties (ketone, aromatics, alkane chains, etc.) the induced dipole can only have a component in the plane of the ring or in line with the chain. For instance, out-of-plan polarizability of benzene can only be correctly calculated if out-of-plane auxiliary sites are built. For alkane chains, though, there is no simple solution[31]. For this reason, the ability of the FQ model to accurately represent complex molecular polarizabilities is clearly limited.

2.2.4 Limitations with the PID related methods

The PID and the related models have been parameterized and show an average error on the average polarizability around 5%. However, errors in the anisotropy are often

around 20% or higher[15,20]. Diatomic molecules are not handled correctly by any of these methods leading to errors of 82% in the anisotropy for F2 for example[13,14]. Heteroaromatics, which are abundant moities in drugs, are often poorly described by PID methods. This limitation is due to the source of anisotropy in the PID model i.e. the interatomic dipole interaction located at static atom positions. It is nevertheless possible to improve these models. For example, using full atomic polarizability tensors instead of isotropic polarizabilities have reduced the errors in polarizability components from 20% to 7%[20,21]. In the case of the DO model, acetamide polarizabilities have been corrected by the addition of atom-type-dependent damping parameters and anisotropic harmonic springs[32]. In these cases, the improvement required a significant amount of additional parameters which brings an additional level of difficulty in their generalization. As illustrated below, our model seems to address most of these complications without additional parameters and complexity.

2.3 Dielectric polarizability model

The mathematical model that we explore in this article is based on simple concepts that have proved extremely useful in chemistry[33-38]. We propose a specific usage that we clarify and describe in this section.

2.3.1 The model

Traditionally in Poisson-Boltzmann (PB) continuum solvent calculations, the solute is described as a region of low dielectric containing a set of distributed point charges; the polar continuum solvent (usually water) is described by a region of high dielectric. This theoretical approach gives the choice to either include average solution salt effects (PB) or to use the pure solvent (PE). Solving PE for such a system is equivalent to calculating a charge density around the solute at the boundary (where the dielectric changes)[39]. This, among other things, allows the calculation of the free energy of charging of a cavity in a

continuum solvent where, at least in the case of water, polarization comes mostly from solvent nuclear motion averaging. While the dielectric boundary is de facto representing the molecular polarization, the dielectric constants and radii employed traditionally are parameterized by fitting to energies (such as solvation or binding free energies) without regard for the molecular polarizabilities themselves. These energies are also dependent on details of the molecular electronic charge distribution, the solvent/solute boundary, and sometimes the nonpolar energy terms, all of which obfuscate the parameterization with respect to the key property of molecular polarizability.

Our approach is to use an intramolecular effective dielectric constant, together with associated atomic radii, to accurately represent the detailed molecular polarizability. For this to be a widely applicable model of polarizability, the generality between related chemical species of a given set of intramolecular effective dielectric constants and associated atomic radii would have to be demonstrated. Such a polarizability model, independent per se of the molecule's charge distribution, could then subsequently be combined with a suitable static charge model to produce a polarizable electrostatic term applicable to force fields.

To evaluate the model, the simplest starting point is gas-phase polarizabilities, using a higher dielectric value inside the molecule and vacuum dielectric outside[40]. This way, the charge density formed at the exterior/interior boundary comes from the polarization of the molecule alone. Comparison of the polarizability tensors from such calculations directly to those from B3LYP/aug-cc-pVTZ calculations allows proof-of-concept of the model. The resulting parameters can be used to rapidly calculate molecular polarizabilities on large molecules.

To calculate the molecular polarizability, we first solve EPIC for a system in which the interior/exterior boundary is described by a van der Waals (vdW) surface, an inner dielectric and a uniform electric field. The electric field is simply produced from the boundary conditions when solving on a grid (electric clamp). From the obtained solution, it

is possible to calculate the charge density from Gauss' law (i.e. from the numerical divergence of the electric field) and the induced dipole moment is simply the sum of the grid charge times its position as shown by eqn (2.1) below

$$\bar{\mu}^{ind} = \sum_{i=1}^{grid} q_i \cdot \bar{r}_i \quad . \tag{2.1}$$

Knowing the applied electric field, it is then possible, as shown in eqn (2.2), to compute the polarizability tensor given that three calculations are done with the electric field applied in orthogonal directions; in eqn (2.2), i and j can be x, y or z.

$$\alpha_{ij} = \mu_i^{ind} / E_j \tag{2.2}$$

2.3.2 Spherical dielectric

For the sake of clarifying the internal structure of the model, let us first consider the induced polarization of a single atom in vacuum under the influence of a uniform external electric field – the EPIC model for an atom. Given a sphere of radius R, a unitless inner dielectric ε_{in} and the uniform electric field E, we can exactly calculate the induced dipole moment with eqn (2.3).

$$\bar{\mu}^{ind} = 4\pi\varepsilon_0 \left(\frac{\varepsilon_{in} - 1}{\varepsilon_{in} + 2} \right) R^3 \cdot \bar{E} \tag{2.3}$$

Clearly, in this problem, the atomic polarizability is given by the electric field E pre-factor, which is a scalar given the symmetry of the problem. The induced dipole moment originates from the accumulation of charge density at the boundary of the sphere opposing the uniform electric field[39]. From eqn (2.3), we see that the polarizability has a cubic dependency on the sphere radius and that the inner dielectric can reduce the polarizability to zero (ε_{in}=1), while the upper limit of its contribution is a factor of 1 (ε_{in} >> 1). The

contribution of ε_{in} to the atomic polarizability asymptotically reaches a plateau as shown in Figure 2.1.

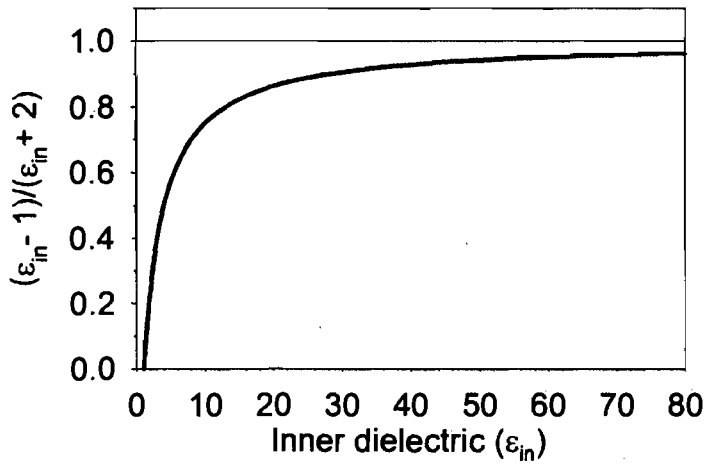


Figure 2.1. The dielectric contribution to the sphere dielectric continuum polarizability goes asymptotically to one and most of the contributions are below $\varepsilon_{in} = 10$.

It is interesting to make a parallel between eqn (2.3) and the PID model, where the polarizable point would be located exactly at the nucleus. In this particular case, it is possible to equate the polarizability from PE, induced by the radius and the dielectric, to any point polarizability[11]. However, when the electric field is not uniform, the PID induced atomic dipole originating from the evaluation of the electric field at a single point may not be representative, leading to inaccuracies[41]. This is in contrast with the EPIC model that builds the response based on the electric field lines passing through the entire surface defined by the sphere, allowing a response more complex than that of a point dipole. In molecules, the atomic polarizabilities of the PID model do not find their counterparts in the EPIC model since it is difficult to assign non-overlapping dielectric spheres to atoms and obtain the correct molecular behavior. The Cl_2 molecule studied in this work is an example.

2.4 Methods

2.4.1 Calculations

Prior to the DFT calculation, SMILES[42-44] strings of the desired structures were transformed into hydrogen-capped three-dimensional structures with the program OMEGA[45]. The n-octane conformer set was also obtained from OMEGA. The resulting geometries were optimized with the Gaussian'03[46] program using B3LYP[47-49] with a 6-31++G(d,p) basis set[50,51] without symmetry. The atomic radii and molecular inner dielectrics were fit based on molecular polarizability tensors calculated at the B3LYP level of theory[52] with the Gaussian'03 program. The extended Dunning's aug-cc-pVTZ basis set[53,54], known to lead to accurate gas phase polarizabilities, was used[55]. An extended basis set is required to obtain accurate gas phase polarizabilities that would otherwise be underestimated.

The solutions to the PE were obtained with the finite difference PB solver Zap[56] from OpenEye Inc. modified to allow voltage clamping of box boundaries to create a uniform electric field. The electric field is applied perpendicularly to two facing box sides (along the z axis). The difference between the fixed potential values on the boundaries is set to meet: $\Delta\varphi = E_z \times \Delta Z$, where $\Delta\varphi$ is the difference in potential, E_z is the magnitude of the uniform electric field and ΔZ is the grid length in the z direction. The salt concentration was set to zero and the dielectric boundary was defined by the vdW surfaces. The grid spacing was set to 0.3 Å and the extent of the grid was set such that at least 5 Å separated the box wall from any point on the vdW surface. As detailed in the Supporting Information, grid spacing below 0.6 Å did not show significant deterioration of the results. Small charges of ±0.001e were randomly assigned to the atoms to ensure ZAP would run, typically converging to 0.000001 kT.

In tables where optimized parameters are reported, a sensitivity value associated with each fitted parameter is also reported. The sensitivity of a parameter corresponds to its smallest variation producing an additional 1% error in the fitness function considering only molecules using this parameter. The sensitivity is calculated with a three-point parabolic fit around the optimal parameter value and the change required obtaining the 1% extra error is extrapolated. Therefore, the reported sensitivity indicates the level of precision for a given parameter and whether or not some parameters could be eventually merged.

2.4.2 Fitting procedure

Eqn (2.4) shows the fitness function F utilized in the fitting of the atomic radii, and the inner dielectrics.

$$F(\{R\},\{\varepsilon\}) = \frac{1}{3N}\sum_{i=1}^{N}\sum_{j=xx,yy,zz}\frac{|\alpha_{ij}^{QM}-\alpha_{ij}|}{\alpha_{ij}^{QM}} + \frac{1}{N_\theta}\sum_{i=1}^{N_\theta}\frac{1-\left|\vec{v}_{ij}^{QM}\cdot\vec{v}_{ij}\right|}{1-\cos 45^\circ} \qquad (2.4)$$

In eqn (2.4), N corresponds to the number of molecules used in the fit, α_{ij} to the polarizability component j of the molecule i and v_{ij} to the eigenvector of the polarizability component j of molecule i. N_θ is the number of non-degenerate eigenvectors found in all the molecules. This fitness function is minimal when the three calculated polarizability components are identical to the QM values and when the corresponding component directions are aligned with the QM eigenvectors of the polarizability tensor.

As shown in the Cl_2 example (c.f. section 2.5.1), the hypersurface of eqn (2.4) has a number of local minima; it is important that our fitting procedure allows these to be examined. Because the calculations were fast, we decided to proceed in two steps: First, a systematic search was carried out varying each fitted parameter over a range and testing all combinations. The 30 best sets of parameters were then relaxed using a Powell minimization algorithm and the set of optimized parameters leading to the smallest error was kept.

2.4.3 Definitions

The polarizability tensor is a symmetric 3x3 matrix derived from six unique values. It can be used to calculate the induced dipole moment μ_i (*i* takes the value *x*, *y* and *z*) given a field vector E

$$\mu_i^{ind} = \alpha_{ix}E_x + \alpha_{iy}E_y + \alpha_{iz}E_z \tag{2.5}$$

In this work, we use the eigenvalues and eigenvectors of the polarizability tensor. The eigenvalues are rotationally invariant and their corresponding eigenvectors indicate the direction of the principal polarizability components. The three molecular eigenvalues are named α_{xx}, α_{yy}, α_{zz} and by convention $\alpha_{xx} \leq \alpha_{yy} \leq \alpha_{zz}$. The average polarizability (or isotropic polarizability) is calculated with eqn (2.6) below. We also define the polarizability anisotropy in eqn (2.7). This particular definition of anisotropy is an invariant in the Kerr effect and has been often used in the literature[57].

$$\alpha_{avg} = \frac{\alpha_{xx} + \alpha_{yy} + \alpha_{zz}}{3} \tag{2.6}$$

$$\Delta\alpha = \sqrt{\frac{(\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{xx} - \alpha_{zz})^2 + (\alpha_{yy} - \alpha_{zz})^2}{2}} \tag{2.7}$$

Eqn (2.7) can be rewritten in terms of only two independent differences in the polarizabilities as shown in eqn (2.8),

$$\Delta\alpha = \sqrt{a^2 + b^2 + ab} \tag{2.8}$$

where $a = \alpha_{zz} - \alpha_{yy}$ and $b = \alpha_{yy} - \alpha_{xx}$. In the case of degenerate molecules as in diatomics, eqn (2.8) reduces to the unsigned difference between two different polarizability eigenvectors.

We now define errors as used in the rest of this article. Eqn (2.9) gives the average unsigned error of the approximated anisotropy ($\Delta\alpha$) where N corresponds to the number of molecules, $\alpha_{i,avg}$ to the average polarizability (eqn (2.6)) of molecule i and QM corresponds to the DFT values.

$$\delta_{aniso} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\Delta\alpha_i^{QM} - \Delta\alpha_i\right|}{\alpha_{i,avg}^{QM}} \tag{2.9}$$

Similarly, the average unsigned error of the average polarizability is defined by

$$\delta_{avg} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\alpha_{i,avg}^{QM} - \alpha_{i,avg}\right|}{\alpha_{i,avg}^{QM}} \tag{2.10}$$

Finally we define an average angle error between the eigenvectors v from QM and our parameterized model as

$$\theta = \frac{1}{N_\theta}\sum_{i=1}^{N_\theta}\left|\cos^{-1}(\vec{v}_i \cdot \vec{v}_i^{QM})\right| \tag{2.11}$$

We prefer the use of the error in the average polarizabiliy, the anisotropy and the deviation angle over the error in the polarizability components or the tensor elements. This allows us to analyze the physical origin of the errors, and in particular how much comes from anisotropy, normally a more stringent property to fit.

2.4.4 Molecule datasets

Our dataset is made to challenge the EPIC model with anisotropic cases known to be difficult. It is formed from three chemical classes: diatomics, heteroaromatics, and the alkanes. While not comprehensive, these datasets were deemed sufficient for proof of concept. Except for the diatomics, all the molecules examined are subdivided into 12

datasets and 6 chemical classes as in Figure 2.2. For each class there is a training set ('-t'

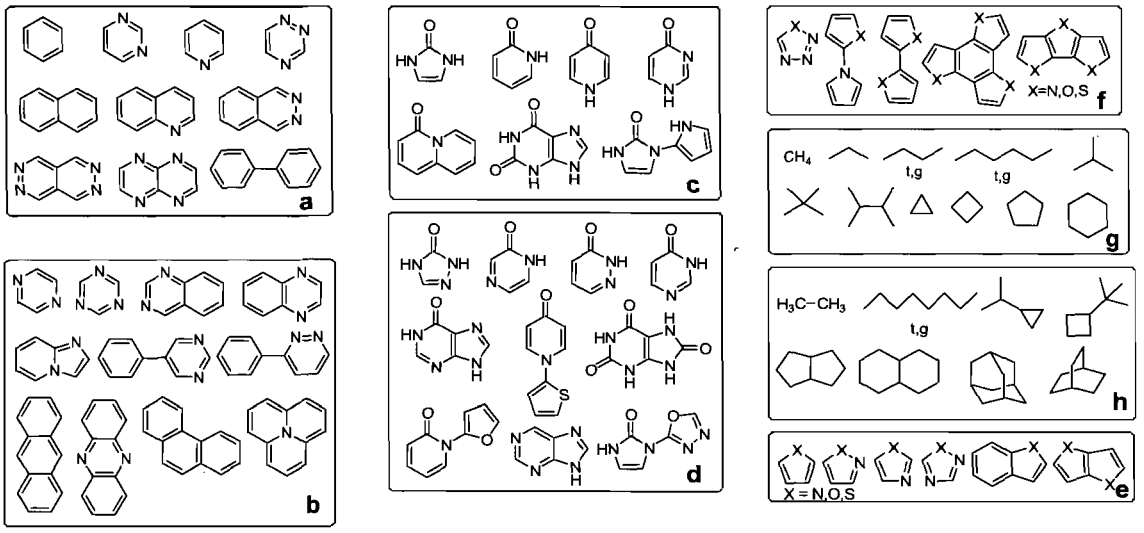


Figure 2.2 The molecules used are divided in 12 datasets and six chemical classes: the heteroaromatics training set 'aromatics-t' (a), the heteroaromatics validation set 'aromatics-v' (b), the pyridones training set 'pyridones-t' (c), the pyridones validation set 'pyridones-v' (d), the furans training set 'furans-t', the pyrroles training set 'pyrroles-t', the thiophenes training set 'thiophenes-t' (e), the furans validation set 'furans-v', the pyrroles validation set 'pyrroles-v', the thiophenes validation set 'thiophenes-v' (f), the alkanes training set 'alkanes-t' (g) and the alkanes validation set 'alkanes-v' (h). In the case of n-butane, n-hexane and n-octane, two conformers are considered: all trans (t) and gauche (g). The X atoms in a molecule are either all O, all S, or all NH.

postfix), used in the parameterization, and a validation set ('-v' postfix) to verify the transferability of the obtained parameters.

Trying to cover most of the unsubstituted aromatic molecules, we selected 5 classes of aromatics: aromatics, pyridones, pyrroles, furans and thiophenes. The aromatics are limited to C, H and divalent N atoms. The pyridones contain aromatic amides; while these also exist as their hydroxypyridine tautomers, in water the equilibrium is strongly driven

toward the pyridone form, which we exclusively study. The pyrroles, furans and thiophenes classes are made from the same scaffolds except differing by one atomic element for each class. In the training sets, balancing the number of molecules is important to avoid overfitting. Each non-degenerate molecular polarizability tensor contributes six datapoints (i.e. from six independent tensor elements). Degenerate molecules contribute either four or one independent data points, depending on the degree of symmetry. The pyridones-v, the pyrroles-v, the thiophenes-v and the furans-v sets all contain multiple functional groups.

The alkanes-t set contains both small and large isotropic molecules (methane and neopentane). It also contains anisotropic molecules like trans-hexane. We included two conformers of butane and hexane because their isotropic polarizability is similar but their anisotropy differs. Cyclic species are also included due to their special nature. The alkanes-v set contains fused cyclic alkanes and an octane in two different conformations of which the trans form is highly anisotropic. We also mixed cyclic alkanes with chain alkanes in the validation set; all this with the desire of having a validation set significantly different from the training set to really assess the transferability of the fitted parameters. For this reason, none of the molecules from the validation sets are used in the parameterization.

2.5 Results

2.5.1 Diatomics: the Cl_2 polarizability hypersurface

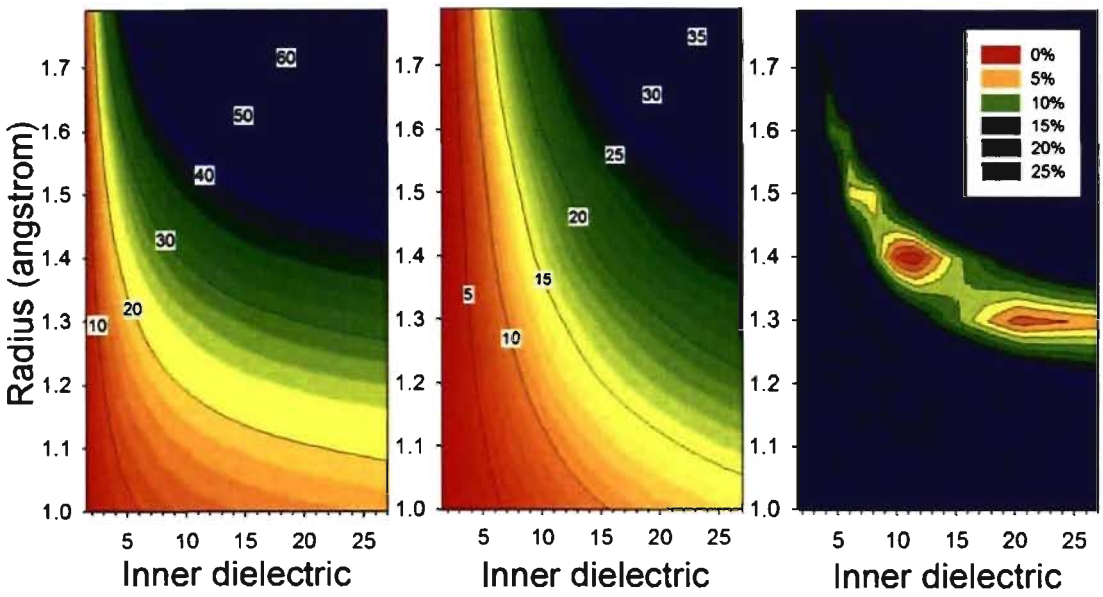
The Cl_2 homonuclear diatomic is the simplest molecule that unveils the dependency of the polarizabilities on the radius and the inner dielectric. In Figure 2.3, parameter hypersurfaces are illustrated for Cl_2 made of two spheres of radius R separated by 2.05 Å (DFT equilibrium distance) within which the inner dielectric is higher than one and the outer dielectric set to the vacuum value of one. When the two spheres overlap (R >1 Å), the molecular volume is described by a vdW surface. Figure 2.3a shows the contour plot of the average polarizability of the molecule as a function of the Cl radius and inner dielectric. As

with the sphere polarizability, the radius has a strong impact on the average polarizability and the influence of the inner dielectric is significantly reduced beyond a value of 10. The anisotropy, however, is more affected by the dielectric constant and varies less rapidly and over a larger range of radius and dielectric than the average polarizability. The Cl_2 example illustrates the need for high dielectric compared to experimental values and this is especially true when a molecule is highly anisotropic. Figure 2.3b shows that for low values of the inner dielectric, the dependence of the anisotropy on the radius diminishes.

Importantly, it is clear that the EPIC model does not have the polarizability catastrophe problem associated with the PID family of polarizable models. When two polarized spheres start to overlap, the interaction between the induced dipoles does not diverge. One reason for this is that the induced polarization is spread over space, rather than being concentrated at a point. Also, when two atoms approach each other, their volumes and hence the total polarizability is decreased. Hence, the atomic radii in the EPIC model play a role somewhat similar to the Thole shielding factor used in PID and DO models.

The Cl_2 bond-parallel and -perpendicular polarizabilities obtained by DFT are 25.4 and 43.6 a.u. respectively, leading to an average polarizability of 31.4 a.u. and an anisotropy of 18.2 a.u. Pairs of radius and dielectric that can reproduce the DFT values can be visually identified by plotting the isolines of the fitness function (2.12) as shown in Figure 2.3c.

$$F(R,\varepsilon) = \frac{2\left|\alpha_{\perp}(R,\varepsilon) - \alpha_{\perp}^{QM}\right| + \left|\alpha_{\parallel}(R,\varepsilon) - \alpha_{\parallel}^{QM}\right|}{3\alpha_{avg}^{QM}} \tag{2.12}$$



a b c

Figure 2.3. The EPIC model behavior is explored for Cl_2. The average polarizability (a) and the anisotropy (b) isolines (in a.u.) are plotted as a function of the Cl atomic radius, used to define the vdW surface, and the value of the inner dielectric. The target Cl2 B3LYP values are 31.43 (average) and 18.24 (anisotropy) (c.f.Table 2.1). The polarizability tensor error function isolines in (c) identify the regions where the EPIC model matches the B3LYP polarizability tensor. The external dielectric is set to one and the inter-nuclear distance of Cl_2 is fixed at 2.05Å. These figures show that a high dielectric value is required to match the QM anisotropy, and that a number of minima can be found on the error hypersurface.

Four local minima are identified (three are obvious from the figure) from which two, located at (R=1.4,ε=11.5) and (R=1.3,ε=20.0) produce an overall error less than 5%. The existence of the multiple minima is due to the multi-objective nature of the fitness function. For instance, at (R=1.5, ε=7.0) the average value is matched but not the anisotropy. Similar hypersurfaces have been found with PE in a different context[37,58]. The error surface has minima where the isolines of ~30 a.u. in Figure 2.3b and the isoline of ~20 a.u. in Figure

2.3c are close to each other, simultaneously matching the DFT values. Higher minima are found when only one of the anisotropy or the average polarizability match the DFT values.

Finally, it is interesting to note, as alluded to in the previous section, that for Cl_2 it is not possible to assign a small sphere (< 1 Å) to each atom, no matter how large is the dielectric, and reproduce the correct polarizability. This clarifies the difference between the EPIC and PID models. Although they both serve the same purpose, the two models do not present identical physical pictures. For instance, shielding must be introduced explicitly in PID whereas it is an intrinsic feature of the EPIC model.

2.5.2 Diatomics: polarizability

Homonuclear diatomic molecules constitute a difficult test for a polarizable model. For example, the FQ model does not allow for bond-perpendicular polarizability, which is typically half of the bond-parallel polarizability. van Duijnen *et al.*[14] have re-parameterized the PID-Thole model and they obtained 22% error on the average polarizabilities of H_2, N_2 and Cl_2. Their error in the anisotropy is significantly larger. More recently, a special parameterization for homo-halides with the PID-Thole model gave an error of 9% and 82% on the average polarizability and anisotropy of F_2 respectively[13]. In the case of Cl_2, the error on the average polarizability and anisotropy are 2% and 20%; finally for Br_2 the same authors found 0.8% and 13%. However, Birge[20] assigned anisotropic atomic polarizabilities and obtained the experimental values for H_2 and N_2. These large errors of the models without atomic anisotropy have been attributed to the difficulty of increasing the atomic induced dipoles interaction. Fitting our model to match B3LYP/aug-cc-pVTZ molecular polarizabilities lead to significantly smaller errors as shown in Table 2.1. In the best case, we fit different inner dielectric and radius for each element. This is a good example of overfitting since two parameters are used to reproduce two polarizabilities. However, it is a way to verify that the dielectric model is flexible enough to deal with the diatomics without using atomic anisotropy parameters. Table 2.1 shows the results for five diatomic

Table 2.1. Compared polarizabilities (a.u.) of diatomic molecules when the radii and ε_{in} are fit to B3LYP/aug-cc-pVTZ polarizabilities – two fitting methods are involved: 1 radius and 1 dielectric per element, 1 radius per element and a single dielectric for all five.

		$\alpha_{\perp}$	$\alpha_{\parallel}$	α_{avg}	$\Delta\alpha$	δ_{avg} (%)[b]	δ_{aniso} (%)[b]
H_2	EPIC (0.88, 7.8)[c]	4.92	6.83	5.55	1.91	0.1	0.3
	(0.83)[d]	4.47	6.60	5.18	2.12	6.7	4.1
	B3LYP	4.92	6.81	5.55	1.89		
	exp[a]	4.86	6.28	5.33	1.42		
N_2	EPIC (1.02,19.5)[c]	10.49	15.89	12.29	5.40	1.8	3.7
	(1.03)[d]	10.35	15.58	12.09	5.23	0.2	2.3
	B3LYP	10.42	15.38	12.07	4.96		
	exp[a]	9.8	16.1	11.90	6.3		
F_2	EPIC (0.86,20.5)[c]	6.26	12.64	8.39	6.37	0.5	1.5
	(0.84)[d]	6.06	11.20	7.77	5.14	6.9	16.3
	B3LYP	6.18	12.68	8.35	6.50		
Cl_2	EPIC (1.34,19.3)[c]	25.64	43.90	31.73	18.26	0.9	0.1
	(1.34)[d]	25.38	43.03	31.26	17.65	0.7	1.9
	B3LYP	25.35	43.59	31.43	18.24		
	exp[a]	24.5	44.6	31.15	20.1		
Br_2	EPIC (1.53,17.5)[c]	36.84	62.42	45.37	25.57	1.0	2.2
	(1.52)[d]	36.19	62.73	45.04	26.54	1.7	0.1
	B3LYP	36.96	63.53	45.82	26.57		

[a]Experimental values are from ref [19].
[b]Error relative to B3LYP values.
[c]The number in the parentheses are the optimal (radius Å, dielectric) individually fit for each molecule.
[d]The optimal radius (in Å) fit for each individual diatomic is reported in parentheses given a globally fit dielectric of 18.0.

molecules and the reported errors for the average polarizability and anisotropy are: 0.1% and 0.3% for H_2, 1.8% and 3.7% for N_2, 0.5% and 1.5% for F_2, 0.9% and 0.1% for Cl_2, 1.0% and 2.2% for Br_2. These results clearly show enough flexibility to account for both

average polarizability and anisotropy. The second fitting scenario involved a single dielectric for all five molecules and five atomic radii, fitting 6 parameters to 10 data points. The optimal parameters give results still in relatively good agreement with DFT with a maximum of 16% error made in the case of F_2 anisotropy. For both optimal parameter sets, the radii and dielectrics are reported in Table 2.1 in parenthesis. These encouraging results on diatomics show that the EPIC model can correctly account for polarizability on a minimal group of two atoms. Therefore, we expect that the local polarizability may be well represented in larger molecules.

2.5.3 Organic datasets: typical PB parameters

As an initial check on how well typical radii and inner dielectric used in PB applications could reproduce the molecular polarizabilities, we first examined the set of parameters obtained by Tan and Luo[12] that lead to reasonable dipole moments in different continuum external dielectrics. In their work, they not only fit the inner dielectric but also the atomic charges. They use the PCM radii and obtained a best inner dielectric of 4. This combination of parameters produces an error of 52% in the average polarizability (eqn (2.10)) compared to B3LYP (all molecules from Figure 2.2 and an error of 18% (eqn (2.9)) in the anisotropy as outlined in Table 2.2. In both cases, the standard deviations (STDEV) of the errors are large. The other two sets of radii examined are those from CHARM22[59] and Bondi[60]. We applied four representative inner dielectrics: 2, 4, 8 and 16 spanning the range of dielectrics often reported to be optimal. Table 2.2 shows very high errors for all the combinations, the best being Bondi radii with an inner dielectric of 4 which lead to an average polarizability error of 9% with a STDEV of 6% and an anisotropy error of 26% with a STDEV of 15%. These particular parameters have a bimodal error distribution producing smaller errors for alkanes than for aromatics, which is consistent with other findings (*vide infra*). Clearly, the parameters from previous studies are not appropriate for the calculation of vacuum molecular polarizabilities and they do not accurately account for the electronic polarization. When attempting to only optimize the inner dielectric, while

keeping the atomic radii to their Bondi values, it was not possible to obtain small errors on the anisotropy. In the next sections, we present details about new parameterizations that are in much better agreements with DFT values. As outlined in Table 2.2, we reduced the error produced by the best Bondi combination by a factor of 4 for both the average polarizability and anisotropy. The STDEV is also greatly reduced allowing for more confidence and robustness in the polarizability predictions.

Table 2.2. Unsigned average errors for all molecules in Figure 2.2, relative to B3LYP/aug-cc-pVTZ, of average polarizability and anisotropy obtained with various parameters typically used in PB applications.

Radii	ε_{in}	δ_{avg} (%)	STDEV (%)	δ_{aniso} (%)	STDEV (%)
Tan *et al.*[a]	4	52	20	18	10
CHARM22[b]	2	40	13	47	23
	4	26	26	28	13
	8	84	40	17	26
	16	129	50	54	44
Bondi [c]	2	51	6	47	23
	4	9	6	26	15
	8	51	15	14	16
	16	91	17	52	29
EPIC/P2E[d]	4.98 14.55	2	2	5	4
EPIC/P1E[d]	11.7	2	2	6	6

[a]Reference 12.
[b]Reference 59.
[c]Bondi radii from reference 60. The Hydrogen radius is set to 1.1 Å following Rowland and Taylor's recommendations[61].
[d]EPIC used with parameters fit in this work reported in ref. 1.

2.5.4 Alkanes and aromatics

Figure 2.4a and b summarize the results obtained with the best parameter set, fitted with two inner dielectrics (P2E), for the 12 sets formed by the 6 classes: alkanes, aromatics, pyridones, pyrroles, furans and thiophenes. The optimal parameters with the atom-typing

scheme used to generate the molecular polarizabilities are given in Table 2.3, along with Bondi radii[60]. In Figure 2.4, the comparisons are between the DFT polarizabilities and the EPIC model. The errors are reported with histograms and error bars corresponding to the average unsigned errors (eqns (2.9), (2.10) and (2.11)) and the corresponding STDEV indicating the range of variation of the errors.

Table 2.3. Optimized radii (Å) and inner dielectrics with sensitivity[a] accounting for all molecule sets (Figure 2.2) – parameter sets P2E and P1E.

Atom type description	Optimal (P2E)	Sensitivity (P2E)	Optimal (P1E)	Sensitivity (P1E)	Bondi Radii[b]
alkanes					
C alkyl	1.39	0.04	1.13	0.03	1.70
H bond on an alkyl C	0.99	0.02	0.78	0.05	1.20
Dielectric alkanes	4.98	0.27	11.70	1.18	
aromatics					
C aromatic	1.32	0.05	1.30	0.04	1.70
H bonded to aromatic C or N	0.64	0.09	0.78	0.05	1.20
N aromatic	1.06	0.16	1.10	0.14	1.55
O furan-like aromatic	0.74	0.23	0.75	0.27	1.52
O in pyridone carbonyl	0.95	0.25	1.03	0.16	1.52
S thiophene-like	1.50	0.06	1.58	0.05	1.80
Dielectric aromatics	14.56	1.50	11.70	1.18	

[a]Smallest parameter variation required to produce a 1% additional error in fitting function (see Method section for details).
[b]Reference 60.

In Figure 2.4a, the error on the average polarizabilities is less than 3% for all classes of the training sets, less than 1% for the thiophenes-t set and the combined average error is less than 2%. The corresponding error on the average polarizabilities for the validation sets in Figure 2.4b is slightly higher with a maximum of 3.2% for the pyrrole-v set; the combined error is 2.4%.

While this low level of error obtained in the average polarizability has also been observed with other polarizable methods, the anisotropy of the polarizability is less

tractable. Previous models normally require the use of directional atomic polarizabilities[15,20,21] especially for aromatics. In our training sets, as shown in Figure 2.4a, we obtain a combined error for the anisotropy of 4%. The worst set, pyridones-t, has an average error of only 7.1%. Although this class is found in biologically active molecules, we could not find published results from other empirical polarizable models for molecular polarizability tensors. We believe that this class might be particularly difficult due to variable aromaticity and accounting for a range of chemical functionalities with the same parameters (imidazolones, 2-pyridones, 4-pyridones, etc.).

The anisotropy average error on the validation set in Figure 2.4b ranges from 2.5% for the alkanes-v up to 7.4% for the aromatics-v. It is not surprising that the error is larger for the validation sets than for the training sets. Overall, however, when comparing the anisotropy error made on the combined sets, it is not significantly higher: 5.3% for the validation sets versus 4% for the training sets. On the other hand, the STDEV is significantly higher in the validation set.

The aromatics class shows the highest anisotropy shift from the training set to the validation set. Phenazine and phenanthrene are responsible for two out of three large discrepancies between B3LYP and EPIC. It is interesting to note that when comparing B3LYP average polarizability and anisotropy to experiment, the errors are 11% and 30% for phenazine, 17% and 20% for anthracene. The same errors, when comparing our model and experiment, are 5% and 15% for phenazine, 1.7% and 1.4% for anthracene. The EPIC model is thus more accurate for these molecules, which can be partly explained by the known size-consistency defect of DFT for oligocenes (benzene, naphthalene, anthracene, tetracene, etc.) that are usually too anisotropic[55]. In general, DFT methods have problems reproducing the polarizability of long delocalized molecules and this has been attributed to deficiency of the currently used functionals to account for a self-interaction correction[62]. It

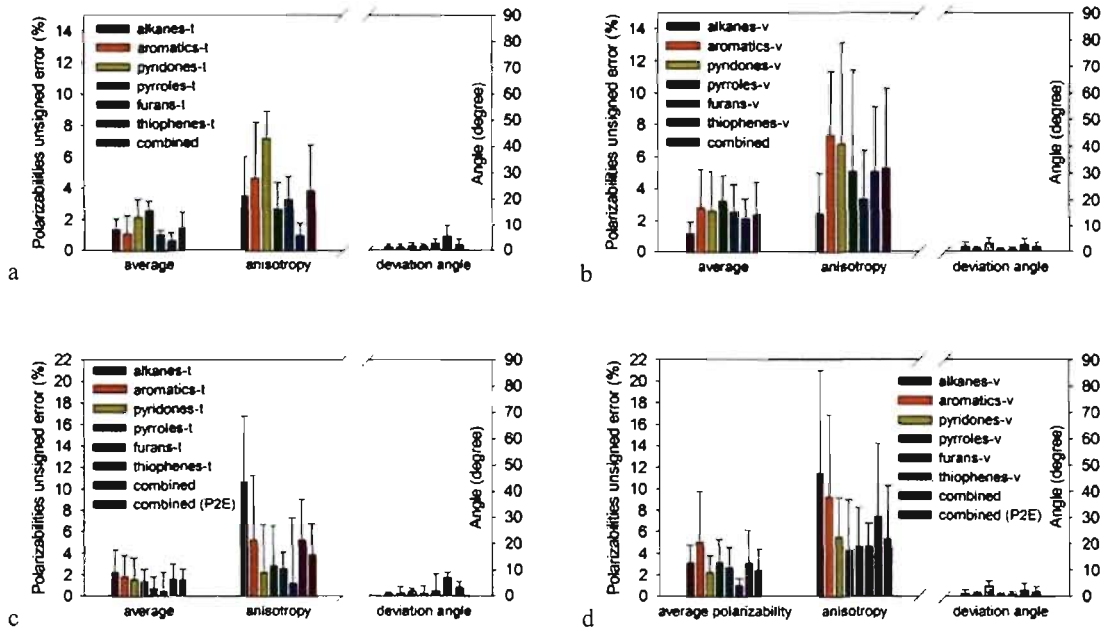


Figure 2.4 Comparison between B3LYP/aug-cc-pVTZ polarizabilities and EPIC models P2E and P1E for all molecules from Figure 2.2. The averaged relative error on average polarizability (eqn (2.10)), anisotropy (eqn (2.9)) and the deviation angle of the eigenvectors (eqn (2.11)) are shown together with the corresponding STDEV reported as error bars. The results for the 2-dielectric fit (P2E) training sets (a) and validation sets (b) show small errors in the average polarizability and relatively small errors in the anisotropy. The results for the 1-dielectric fit (P1E) training sets (c) and the validation sets (d) show larger errors in the alkanes anisotropy and generally larger errors than the P2E parameters (shown under combined P2E). Combined errors of the training and validation sets are similar.

is therefore possible that our model, fit on smaller molecules, tend to produce better behavior on these delocalized molecules. Another implication is that large molecules should not be used for the training of a polarizable model with DFT methods. Figure 2.5a shows that in fact the correlation between the polarizability components of the entire set of molecules of Figure 2.2 is excellent up to 150 a.u. Part of the discrepancy might be attributable to a different behavior of DFT methods in that range of polarizabilities. In this respect, optimized effective potential (OEP) and time-dependent DFT methods have shown

significant improvement[63-65], but these are still considerably more resources-intensive. The third worst anisotropy discrepancy between B3LYP and EPIC of this aromatics-v set comes from the cycl[3.3.3]azine molecule which has already shown differences with regular polyacenes in terms of excited states[66]. The transferability for that particular molecule is good, all things considered, with an average polarizability error of 8.6% and anisotropy error of 12.8%.

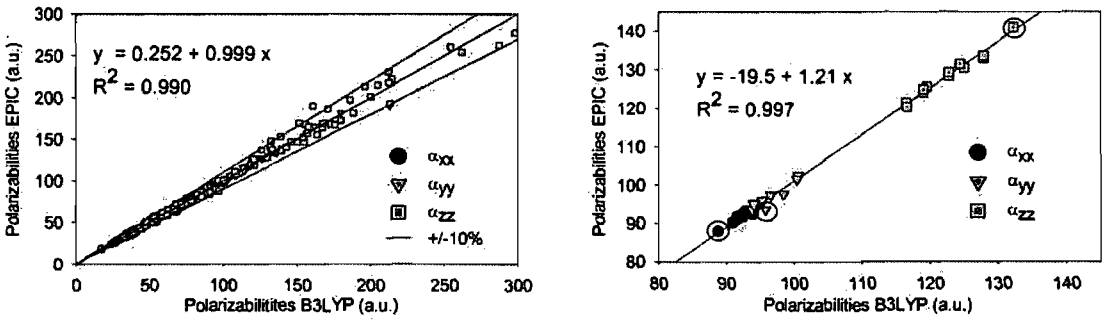


Figure 2.5. Correlation between B3LYP/aug-cc-pVTZ polarizability components and the EPIC model P2E. In (a), the polarizability components for all sets of Figure 2.2 are correlated and the ±10% error lines are illustrated. The linear regression shows excellent agreement, especially for polarizabilities smaller than 150 a.u. In (b), 13 stable conformers of n-octane are examined. The all trans conformation polarizabilities are identified with circles. The average polarizability error on the 13 conformers is 1.9% and the anisotropy error is 5.8%. A linear regression gives a R^2 of 0.997, a slope of 1.21 and an ordinate at the origin of -19.5. This means that the EPIC model P2E overestimates the polarizability of n-octane consistently through all conformers.

The pyridones-v set is the most challenging with the highly functionalized purine derivates (purine, hypoxanthine and uric acid) and the substituted pyridones with five member heteroaromatic rings. For example, the geometry optimized 1-(2-thienyl)-pyridin-4-one shows an angle of 58 degrees between the two aromatic rings as opposed to the 1-

(oxadiazol)-imidazolone that has the two connected rings co-planar and a fully delocalized electron π system. This dataset is not dissimilar from the chemical functionalization of drug-like molecules.

The average angles between the eigenvectors of the polarizability components of the B3LYP and EPIC are less than 5.5 degrees in all sets, although in some molecules the angles can be as large as 23 degrees, i.e. for thiazole. For the pyridones-t and pyridones-v sets, the angular diffences remain surprisingly small.

Finally, Table 2.4 shows that compared to experimental values, the parameterized EPIC method performs comparably to B3LYP against the subset of 25 molecules for which experimental data is available. Indeed, EPIC produces a δ_{avg} of 3.9% with experiment compared to 4.1% for B3LYP. It also gives a δ_{aniso} of 9.0% with experiment compared to 10.5% in the case of B3LYP. The STDEV of the errors from B3LYP match EPIC values. The discrepancy between B3LYP and EPIC calculated for the molecules of Figure 2.2 is smaller leading to a δ_{avg} of 1.9% and a δ_{aniso} of 4.6%. The level of error compared to experiment obtained with both B3LYP and EPIC is not necessarily beyond experimental uncertainty.

Table 2.4 Average errors and stdev against experiment[a] for all molecules in Figure 2.2.

Method	δ_{avg} (%)	stdev (%)	δ_{aniso} (%)	stdev (%)
Tan *et al.*[b]	58.4	19.8	13.6	9.4
Bondi[c]	8.3	6.2	22.4	13.5
EPIC/P2E[d]	3.9	4.1	9.0	9.5
EPIC/P1E[d]	3.8	3.1	7.3	6.4
B3LYP	4.1	4.1	10.5	9.9

[a]25 experimental average polarizabilities and 18 anisotropy data. Details given in Supporting Information.
[b]Reference 12.
[c]Bondi radii and $\varepsilon_{in} = 4$.
[d]EPIC used with parameters fit in this work reported in Table 2.3.

2.5.5 Conformational dependency of polarizability

Although we avoided comparing the polarizability of flexible molecules to experimental data, it is obvious that a good empirical method should account for the conformational dependency of the polarizability, the anisotropy and the orientation of the polarizability tensor eigenvectors. In addition to the deliberate choice of a wide range of 3D diversity in our molecular sets, we examined the case of n-octane, the most flexible molecule of the sets. Taking 13 diverse B3LYP geometry optimized conformers of n-octane, we computed the polarizability, anisotropy and the eigenvectors using the P2E parameters. The EPIC method gives average polarizability error and anisotropy error of 1.9% and 5.8% respectively. Figure 2.5b shows a correlation graph between B3LYP polarizability components and our model (α_{xx}, α_{yy}, α_{zz}). The correlation is perfectly linear as shown by a linear regression leading to a R^2 of 0.997 although the slope of the regression is 1.21, consistent with the average errors outlined above. Moreover, in Figure 2.5a, we clearly see that correlation of the polarizability components for all the molecules of Figure 2.2 is excellent with a slope of 1 and a R^2 of 0.990. This result leads to the conclusion that

our model is at least consistently making the same errors for n-octane conformers compared to B3LYP. Finally, the orientations of the polarizability components differ by 0.97 degrees with a maximum value of 3.7 degrees; this is in spite of the broken symmetry in the gauche octane conformers.

2.6 Discussion

2.6.1 Transferability

Shanker and Applequist[15], with a variation of the PID model, studied seven nitrogen heterocyclic molecules that we also included in our sets: pyridine, pyrimidine, pyrazine, 9H-purine, quinoxaline, quinoline and phenazine. Using 12 parameters including directional atomic polarizabilities, they show an average polarizability (eqn (2.10)) and anisotropy errors (eqn (2.9)) of 10% and 12% respectively[67]; the parameterized EPIC (Table 2.2) produces correspondingly 3% and 5% of error with only 4 parameters; we feel that the reduced requirement for fitted parameters is due to a better physical model. Similar comparisons can be made to the work of Miller[21] where it is reported that 6 parameters for benzene, 9 parameters for pyridine, 9 parameters for naphthalene and 12 parameters for quinoline are needed to obtain both the average polarizability and anisotropy. With the EPIC method, again the same 4 parameters do for all.

Recently, Williams and Stone[68] have parameterized a polarizable model on n-propane, n-butane, n-pentane and n-hexane in both their trans and gauche conformations. With their simplest Ctg model, they use 10 atomic polarizability parameters to fit the polarizability tensors to B3LYP values. They obtain a very small error on both the average polarizability and the anisotropy of 1.16% and 2.37% respectively. Making the same comparison with our model, we obtain 1.7% of average polarizability error and 3.99% of anisotropy error. Although the error is slightly larger with our EPIC model, this is obtained with only three parameters also producing similar levels of errors in our extended set of

alkanes. Furthermore, the level of errors reported by Williams *et al.* and our studies are all within the accuracy of B3LYP method.

The small number of parameters (c.f. Table 2.3) needed to fit all the aromatic compounds of Figure 2.2 is a good indication of the transferability and the generality of the method for heteroaromatic compounds. For example the same nitrogen radius could simultaneously fit pyridine, pyridone, pyrrole, and even branched nitrogen. In the case of alkanes, we have examined most characteristic shapes. Moreover, the training and validation sets produce similar errors, thus the expected performance of our method in the general case can be approximated by the errors on the validation sets.

Overall, we obtain the same level of error as the best PID methods parameterized with anisotropic atomic polarizabilities and about threefold more parameters. Although the number of parameters is not an issue for a small and homogenous set of molecules, it would become a serious barrier for further development of a model applicable to the immense functional group complexity of drug-like molecules, one of the main goals of this ongoing effort.

2.6.2 Inner dielectrics

The choice of fitting two inner dielectrics, one for the alkanes and one for the heteroaromatics, makes the calculation of new mixed molecules such as t-butylbenzene not possible unless we have a way to switch from a high dielectric (benzene) to a lower dielectric (t-butyl) intramolecularly. Overall, the value of multiple dielectrics, based on chemical constituency, seems proven as well as being physically reasonable. This is a potentially useful strategy in the development of a future general polarizability model. However, simultaneously fitting the polarizabilities of all the compounds from Figure 2.2 with a single dielectric still gives reasonable results. Table 2.3 reports the values of the optimal parameters used to produce the data of Figure 2.4c and d. We fit one radius per element except for oxygen, which is split into furan-like and pyridone-like, and for carbon

which is split into alkane and aromatic. We first had two hydrogen radii, but there was no significant cost to merge them into one single radius. The results, shown in Figure 2.4c and d, when compared with those of Figure 2.4a and b, show a significant increase in the errors on the alkanes-t and alkanes-v sets although the errors on the heteroaromatics classes remain similarly small. It is nevertheless surprising that the level of error remains low when describing the electronic dielectric with a single constant when, in principle, the electronic local polarization should vary intramolecularly as suggested by Oxtoby[69].

Finally, it is reassuring that the best radii for both reported parameterizations follow the chemical sense of atomic size. The remarkably reduced size of the optimal radii compared to usual solvent vdW radii (like Bondi) is worth few comments. First, our reduced radii explain a different physical reality than the vdW radii: the former relates to the electronic response and the latter to the repulsive vdW forces that keep solvent away, for example. Furthermore, the high dielectric and the small radii were needed to fit the correct polarizability anisotropy. This modulates the shape of the molecules. A benzene molecule is flattened when the carbon radii are reduced and thus the out-of-plane polarizable volume is reduced while the in-the-plane length is more or less conserved. The polarizability density (also called the susceptibility given by $\varepsilon_0 \times [\varepsilon_{in}(r)-1]$) being constant inside the molecular volume, a higher dielectric value is then needed to conserve the molecular polarizability.

2.6.3 Link to the optical dielectric constants

Intramolecular dielectric constants in the context of PE or PB can adopt many values depending on the system and the phenomena involved[35,37,58,70] and have been attributed values from 1 to 20. The optimal inner dielectric of solutes in continuum solvent free energy and in ligand-protein binding calculations do not agree[37]. Here, we attempt to position our work in this jungle of dielectrics.

We are concerned uniquely with the electronic polarization component. None of the optimal dielectric constants fitted in this work match the experimental optical dielectric constants calculated as the square of the refractive index, which normally have values between 1.2 and 4.0. We partly justify the need for larger dielectrics in section 6.2, but there are other factors that should also be considered. It is important to realize that the link between the molecular polarizability and the macroscopic optical dielectric constant is given by the Lorentz-Lorenz relation shown in eqn (2.12) where N is the number of molecule in the volume V and ε is the macroscopic dielectric when the light frequency is high compare to the dipolar or ionic relaxation time (ε_0 is the vacuum permittivity constant).

$$\alpha_{avg} = \frac{3\varepsilon_0 V}{N}\left[\frac{\varepsilon - 1}{\varepsilon + 2}\right] \tag{2.13}$$

In the Lorentz-Lorenz equation a molecule is approximated as a spherical dielectric with an effective molecular volume given by the ratio of the macroscopic space occupied by one molecule. However, from our atomistic perspective the effective volume of a molecule is defined by the electronic density and does not include the empty space between molecules effectively included in eqn (2.12). Hence, in the EPIC model that we parameterize, the average polarizability is the link to refractive index and not the inner dielectric. The main reason for this is the inconsistency between the atomistic and macroscopic definitions of the molecular volume. This raises the point that using experimental optical dielectrics assigned to the solute interior in continuum solvent approaches should be further questioned.

A misleading impression can also come from the non-linear behavior of the dielectric on the polarizability. This brings us back to Figure 2.1 where we can appreciate that if ε_{in} is increased from 4 to 8, the polarizability pre-factor is increased from 0.5 to 0.7 while when ε_{in}=14.6, our optimized aromatic dielectric, the pre-factor reaches 0.82 for an

increase of 64% over the ε_{in}=4 case. Similar effects are also observed in non-spherical shapes and this gives some perspective on the high inner dielectric constants that we obtain in this work.

Finally, we believe that the use of accurate parameters to describe solute polarizability in the context of continuum solvent could improve the quality of these methods. Unfortunately, the radii and dielectrics obtained in the present work cannot be used in continuum solvent models directly. Obviously, the solvent charge density should be found at the regular vdW distances. Therefore, to simultaneously include the solute electronic response and the correct solvent response, there is a dielectric region between our optimal radii and the vdW radii that still needs to be elucidated. We believe that this interesting question needs to be addressed to extend the use of our findings to implicit solvent models. Once done, one could think of obtaining a polarizable model close to the 'polarizable continuum model' (PCM) of Tomasi[71] in which the electronic density would be simply replaced by an 'electronic volume' defined with radii and a dielectric constant.

2.7 Conclusion

In this work, the simple physical picture afforded by a continuum dielectric representation has been used to accurately model molecular dipole polarizability tensors. The molecular inner dielectric in the EPIC model accounts for the electronic polarization. To tackle gas-phase polarizabilities, we capitalized on existing finite difference Poisson-Boltzmann code to calculate the induced dipole moment of a molecule in vacuum in the presence of a uniform electric field. As opposed to the usual use of PE or PB in continuum models, the molecule is a region of higher dielectric and the external dielectric is set to the vacuum value. The calculations are fast and resource-sparing, with equivalently good results up to a grid spacing of 0.5 Å, even though a discrete vdW dielectric boundary is used.

This EPIC model of molecular polarizability possesses some important differences with other approximations such as the point inducible dipole and the fluctuating charge models. It is based on a local differential equation solved on a grid, which brings to the same level of complexity the polarizability and coulombic electrostatic components. In particular, EPIC avoids the polarizability catastrophe found in the other PID based models. Furthermore, it allows, in principle, for a more thorough response to the electric field than the PID or the FQ models based on the fact that the response emerges from the electric field lines that cross the molecule surface instead of evaluations only at atomic nuclear positions.

This study involved the parameterization of atomic radii, used in the definition of the vdW dielectric boundary, and the molecular inner dielectric. Previous values of these variables found in the literature are unacceptably poor at approximating molecular polarizability. We attribute this discrepancy to the fact that previous models simultaneously optimize different kinds of interdependent parameters fitting to a complex energy property instead of focusing on solute polarization. Indeed, the previous purpose of using dielectric continuum was in the context of continuum solvent, often completely neglecting the solute response *per se*. To test the newly proposed method, we selected difficult chemical classes: the homonuclear diatomics, a wide variety of heteroaromatics and a diverse set of alkanes. A total of 5 diatomics plus 48 molecules are part of the training sets, subdivided into 6 chemical classes to which we add 45 molecules for validation purposes.

In previous models, the polarizabilities of these classes of compounds were correctly calculated only when anisotropic atomic polarizabilities were employed (or auxiliary sites in the case of FQ). Already, with about threefold less parameter than other studies with different models, we have obtained averaged polarizability errors smaller than 5% and averaged anisotropy errors less than 8% considering all sets. The polarizability components calculated with the EPIC/P2E model correlates very well with B3LYP/aug-cc-pVTZ with a R^2 of 0.990 and a slope of 0.999. The orientations of the polarizability eigenvectors are also well reproduced. The flexibility of the model even allowed the

calculation of an accurate anisotropy for F_2 without resorting to auxiliary sites or anisotropic parameters. We also found that the EPIC model was able to consistently calculate the molecular polarizabilities on 13 different conformers of n-octane. Because of the success of parsimonious parameterization of the EPIC model on difficult chemical classes, we believe that the parameterization can be generalized for all organic chemistry with adequate accuracy. In doing this, we found that intra-molecularly varying dielectric constant might be needed to account for the molecular anisotropy.

Overall, this study exemplified that a phenomena as complex as electronic polarization can be accurately modeled with a simple dielectric continuum model. The principal implications of these findings are in the areas of Poisson-Boltzmann methods and in polarizable force field development. However, the level of accuracy obtained might also have impact beyond our initial consideration, for example in the field of spectroscopy.

2.8 Bibliography

1. Lane, N. F. Theory of Electron-Molecule Collisions. *Rev. Mod. Phys.*, **1980,** *52*, 29-119.
2. Kirkwood, J. G. On the Theory of Optical Rotatory Power. *J. Chem. Phys.*, **1937,** *5*, 479-491.
3. Wagnière, G. H. *Linear and Nonlinear Optical Properties of Molecules;* VCH ed.; VCH; Helvetica Chimica Acta Publishers: Weinheim, 1993.
4. Maroulis, G.; Hohm, U. Electric polarizabilities of Ge(CH3)(4) from collision-induced light-scattering experiments and ab initio calculations. *Phys. Rev. A*, **2007,** *76*, 032504.

5. Vela, A.; Gazquez, J. L. A Relationship Between the Static Dipole Polarizability, the Global Softness, and the Fukui Function. *J. Am. Chem. Soc.*, **1990,** *112*, 1490-1492.

6. Nagle, J. K. Atomic Polarizability and Electronegativity. *J. Am. Chem. Soc.*, **1990,** *112*, 4741-4747.

7. Allen, T. W.; Andersen, O. S.; Roux, B. Energetics of ion conduction through the gramicidin channel. *Proc. Natl. Acad. Sci. U. S. A.*, **2004,** *101*, 117-122.

8. Wick, C. D.; Kuo, I. F. W.; Mundy, C. J.; Dang, L. X. The Effect of Polarizability for Understanding the Molecular Structure of Aqueous Interfaces. *J. Chem. Theory Comput.*, **2007,** *3*, 2002-2010.

9. Guo, H.; Gresh, N.; Roques, B. P.; Salahub, D. R. Many-body effects in systems of peptide hydrogen-bonded networks and their contributions to ligand binding: A comparison of the performances of DFT and polarizable molecular mechanics. *J. Phys. Chem. B*, **2000,** *104*, 9746-9754.

10. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, **1997,** *23*, 3-25.

11. Sharp, K.; Jean-Charles, A.; Honig, B. A Local Dielectric-Constant Model for Solvation Free-Energies Which Accounts for Solute Polarizability. *J. Phys. Chem.*, **1992,** *96*, 3822-3828.

12. Tan, Y. H.; Luo, R. Continuum treatment of electronic polarization effect. *J. Chem. Phys.*, **2007,** *126*, 094103.

13. Elking, D.; Darden, T.; Woods, R. J. Gaussian induced dipole polarization model. *J. Comput. Chem.*, **2007,** *28*, 1261-1274.

14. van Duijnen, P. T.; Swart, M. Molecular and atomic polarizabilities: Thole's model revisited. *J. Phys. Chem. A*, **1998,** *102*, 2399-2407.

15. Shanker, B.; Applequist, J. Polarizabilities of nitrogen heterocyclic molecules from atom monopole dipole interaction theory. *J. Phys. Chem.*, **1996,** *100*, 3879-3881.

16. Silberstein, L. *Philos. Mag.*, **1917,** *33*, 521-533.

17. Bode, K. A.; Applequist, J. A new optimization of atom polarizabilities in halomethanes, aldehydes, ketones, and amides by way of the atom dipole interaction model. *J. Phys. Chem.*, **1996,** *100*, 17820-17824.

18. Applequist, J. Atom Charge-Transfer in Molecular Polarizabilities - Application of the Olson-Sundberg Model to Aliphatic and Aromatic-Hydrocarbons. *J. Phys. Chem.*, **1993,** *97*, 6016-6023.

19. Applequist, J.; Carl, J. R.; Fung, K. K. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.*, **1972,** *94*, 2952-2960.

20. Birge, R. R. Calculation of Molecular Polarizabilities Using An Anisotropic Atom Point Dipole Interaction-Model Which Includes the Effect of Electron Repulsion. *J. Chem. Phys.*, **1980,** *72*, 5312-5319.

21. Miller, K. J. Calculation of the Molecular Polarizability Tensor. *J. Am. Chem. Soc.*, **1990,** *112*, 8543-8551.

22. Thole, B. T. Molecular Polarizabilities Calculated with A Modified Dipole Interaction. *Chem. Phys.*, **1981,** *59*, 341-350.

23. Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, **1976,** *103*, 227-249.

24. Cieplak, P.; Kollman, P. A.; Lybrand, T. A new water potential including polarization: Appilcation to gas-phase, liquid, and crystal properties of water. *J. Chem. Phys.*, **1990,** *92*, 6755-6760.

25. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. *J. Phys. Chem. A*, **2004,** *108*, 621-627.

26. Noskov, S. Y.; Lamoureux, G.; Roux, B. Molecular dynamics study of hydration in ethanol-water mixtures using a polarizable force field. *J. Phys. Chem. B*, **2005,** *109*, 6705-6713.

27. Lamoureux, G.; Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.*, **2003,** *119*, 3025-3039.

28. Gasteiger, J.; Marsili, M. New Model for Calculating Atomic Charges in Molecules. *Tetrahedron Lett.*, **1978,** 3181-3184.

29. Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.*, **1994,** *101*, 6141-6156.

30. Rappe, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.*, **1991,** *95*, 3358-3363.

31. Chelli, R.; Procacci, P.; Righini, R.; Califano, S. Electrical response in chemical potential equalization schemes. *J. Chem. Phys.*, **1999,** *111*, 8569-8575.

32. Harder, E.; Anisimov, V. M.; Whitfield, T.; MacKerell, A. D.; Roux, B. Understanding the Dielectric Properties of Liquid Amides from a Polarizable Force Field. *J. Phys. Chem. B*, **2008,** *112*, 3509-3521.

33. Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science*, **1995,** *268*, 1144-1149.

34. Roux, B.; MacKinnon, R. The cavity and pore helices the KcsA K+ channel: Electrostatic stabilization of monovalent cations. *Science*, **1999,** *285*, 100-102.

35. Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediction of Ph-Dependent Properties of Proteins. *J. Mol. Biol.*, **1994,** *238*, 415-436.

36. Simonson, T.; Archontis, G.; Karplus, M. A Poisson-Boltzmann study of charge insertion in an enzyme active site: The effect of dielectric relaxation. *J. Phys. Chem. B*, **1999,** *103*, 6142-6156.

37. Naim, M.; Bhat, S.; Rankin, K. N.; Dennis, S.; Chowdhury, S. F.; Siddiqi, I.; Drabik, P.; Sulea, T.; Bayly, C. I.; Jakalian, A.; Purisima, E. O. Solvated Interaction Energy (SIE) for Scoring Protein-Ligand Binding Affinities. 1. Exploring the Parameter Space. *J. Chem. Inf. Model.*, **2007,** *47*, 122-133.

38. Fogolari, F.; Brigo, A.; Molinari, H. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.*, **2002,** *15*, 377-392.

39. Griffiths, D. J. *Introduction to Electrodynamics;* 3rd ed.; Prentice-Hall Inc.: Upper Saddle River, NJ, 1999.

40. Nicholls, A. *The 233rd ACS National Meeting, Chicago, IL, March 25-29*; 2007.

41. Schropp, B.; Tavan, P. The Polarizability of Point-Polarizable Water Models: Density Functional Theory/Molecular Mechanics Results. *J. Phys. Chem. B*, **2008,** *112*, 6233-6240.

42. Weininger, D. Smiles .3. Depict - Graphical Depiction of Chemical Structures. *J. Chem. Inf. Model.*, **1990,** *30*, 237-243.

43. Weininger, D.; Weininger, A.; Weininger, J. L. Smiles .2. Algorithm for Generation of Unique Smiles Notation. *J. Chem. Inf. Model.*, **1989,** *29*, 97-101.

44. Weininger, D. Smiles, A Chemical Language and Information-System .1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.*, **1988,** *28*, 31-36.

45. *OMEGA,* version 2.2.1; Santa Fe, NM, USA, 2007.

46. *Gaussian 03, Revision* C.02; Wallingford CT, USA, 2004.

47. Becke, A. D. Density-Functional Thermochemistry .3. the Role of Exact Exchange. *J. Chem. Phys.*, **1993,** *98*, 5648-5652.

48. Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.*, **1993,** *98*, 1372-1377.

49. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab-Initio Calculation of Vibrational Absorption and Circular-Dichroism Spectra Using Density-Functional Force-Fields. *J. Phys. Chem.*, **1994,** *98*, 11623-11627.

50. Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.*, **1984,** *80*, 3265-3269.

51. Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. Efficient Diffuse Function-Augmented Basis-Sets for Anion Calculations .3. the 3-21+G Basis Set for 1St-Row Elements, Li-F. *J. Comput. Chem.*, **1983,** *4*, 294-301.

52. Rice, J. E.; Handy, N. C. The calculation of frequency-dependent polarizabilities as pseudo-energy derivatives. *J. Chem. Phys.*, **1991,** *94*, 4959-4971.

53. Woon, D. E.; Dunning, J. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.*, **1993,** *98*, 1358-1371.

54. Kendall, R. A.; Dunning, J.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.*, **1992,** *96*, 6796-6806.

55. Hammond, J. R.; Kowalski, K.; deJong, W. A. Dynamic polarizabilities of polyaromatic hydrocarbons using coupled-cluster linear response theory. *J. Chem. Phys.*, **2007,** *127*, 144105.

56. Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Comput. Chem.*, **2001,** *22*, 608-640.

57. Kassimi, N. E. B.; Lin, Z. J. Aza-substituted thiophene derivatives: Structures, dipole moments, and polarizabilities. *J. Phys. Chem. A*, **1998,** *102*, 9906-9911.

58. Rankin, K. N.; Sulea, T.; Purisima, E. O. On the transferability of hydration-parametrized continuum electrostatics models to solvated binding calculations. *J. Comput. Chem.*, **2003,** *24*, 954-962.

59. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, **1998,** *102*, 3586-3616.

60. Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.*, **1964,** *68*, 441-451.

61. Rowland, R. S.; Taylor, R. Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der Waals radii. *J. Phys. Chem.*, **1996,** *100*, 7384-7391.

62. Sekino, H.; Maeda, Y.; Kamiya, M.; Hirao, K. Polarizability and second hyperpolarizability evaluation of long molecules by the density functional theory with long-range correction. *J. Chem. Phys.*, **2007,** *126*, 014107.

63. van Faassen, M.; de Boeij, P. L. Excitation energies,for a benchmark set of molecules obtained within time-dependent current-density functional theory using the Vignale-Kohn functional. *J. Chem. Phys.*, **2004,** *120*, 8353-8363.

64. van Faassen, M.; Jensen, L.; Berger, J. A.; de Boeij, P. L. Size-scaling of the polarizability of tubular fullerenes investigated with time-dependent (current)-density-functional theory. *Chem. Phys. Lett.*, **2004,** *395*, 274-278.

65. van Faassen, M. Time-dependent current-density-functional theory applied to atoms and molecules. *Int. J. Mod. Phys. B*, **2006,** *20*, 3419-3463.

66. Leupin, W.; Berens, S. J.; Magde, D.; Wirz, J. Picosecond Fluorescence from the 2Nd Excited Singlet-State of Cycl[3.3.3]Azine, A Bridged 12-Pi-Perimeter Annulene - Spectra and Kinetics for S2 -] S0 and S2 -] S1 Transitions. *J. Phys. Chem.*, **1984,** *88*, 1376-1379.

67. For purine and quinoxaline, the B3LYP/aug-cc-pVTZ components from this work are used for the comparison since they match the experimental average polarizability reported by Shanker et al. Averaged experimental components reported by Shanker et al. are used for pyrimidine and pyrazine.

68. Williams, G. J.; Stone, A. J. Transferable polarizabilities for the alkanes. *Mol. Phys.*, **2004,** *102*, 985-991.

69. Oxtoby, D. W. Local Polarization Theory for Field-Induced Molecular Multipoles. *J. Chem. Phys.*, **1980,** *72*, 5171-5176.

70. Elcock, A. H.; Sept, D.; McCammon, J. A. Computer simulation of protein-protein interactions. *J. Phys. Chem. B*, **2001,** *105*, 1504-1518.

71. Miertus, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of A Solute with A Continuum - A Direct Utilization of Abinitio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.*, **1981,** *55*, 117-129.

3 Utiliser la polarisation provenant du continuum interne pour les interactions intermoléculaires

Résumé. Récemment, le tenseur de polarisabilité moléculaire dans le vide de molécules variées a été modélisé avec exactitude (Truchon, J.-F. *et al*, *J Chem Theory and Comput* 2008, *4*, 1480) à l'aide d'un diélectrique intramoléculaire continu. Cette étude préliminaire a montré que la polarisation électronique peut être bien modélisée lorsque des rayons atomiques et des diélectriques internes appropriés sont employés. Dans cet article, avec les paramètres ajustés pour reproduire des polarisabilités moléculaires *ab initio* quantiques, nous élargissons l'application de l'approche *polarisation électronique provenant d'un continuum interne* (*electronic polarisation from internal continuum*: EPIC) aux interactions intermoléculaires. Nous dérivons d'abord DRESP (diélectrique RESP) qui est une nouvelle procédure fondée sur la méthode des moindres carrés pour ajuster les charges atomiques partielles sur le potentiel électrostatique quantique en présence d'un diélectrique interne > 1. Nous esquissons également une procédure pour adapter tout modèle de charges atomiques existant à l'approche EPIC. L'étude de la capacité du modèle EPIC à reproduire une polarisation locale, par opposition à une polarisation uniforme, résulte en une racine du carré de la déviation égale à 1 % relativement aux calculs quantiques lorsque moyennée sur 37 molécules dont des hétéroaromatiques et des alcanes. L'avantage du modèle polarisable continu sur les modèles avec polarisabilité centrée sur les atomes est illustré avec un atome et une molécule de benzène qui sont symétriquement perturbés. Nous appliquons EPIC à un système composé d'un cation atomique et d'une molécule de benzène formant une interaction cation-π, ce qui montre que l'approche étudiée donne une bonne énergie d'induction. Finalement, cet article montre que la composante quantique de l'énergie électrostatique dans un pont-H très polarisé, liant le dimère de 4-pyridone (molécule très polaire et très polarisable), est bien reproduite sans ajustement de paramètres supplémentaires.

Mise en contexte. Nous avons démontré au chapitre précédent que l'idée d'utiliser un continuum diélectrique pour reproduire le tenseur de polarisabilité moléculaire fonctionne mieux que les approches précédentes. Or, notre objectif ultime n'est pas seulement de reproduire la polarisabilité moléculaire, mais aussi d'utiliser EPIC dans un champ de force dans le contexte de simulations de biomolécules en solution.

Un premier pas de plus que nous ferons dans ce chapitre sera d'évaluer comment le potentiel électrostatique induit par un champ qui n'est pas uniforme, mais plutôt local, se compare avec les résultats de la mécanique quantique. En effet, en solution, une molécule subit un champ électrique qui est loin d'être uniforme. Il n'est donc pas évident que l'ajustement des rayons atomiques et de la constante diélectrique interne du chapitre précédent puisse être adapté au cas d'intérêt.

Deuxièmement, le couplage entre la polarisation intramoléculaire et les charges atomiques partielles pose un défi considérable à la communauté scientifique qui voudrait généraliser les paramètres. Dans ce chapitre, nous dérivons les équations et la méthodologie nécessaires à l'obtention de charges atomiques partielles qui reproduisent le potentiel électrostatique permanent prédit par la mécanique quantique dans le contexte où ces charges baignent dans un milieu diélectrique. Ceci est essentiel à l'obtention du terme d'interaction électrostatique dans un champ de force. Conservant l'idée d'avoir une méthode polyvalente et des paramètres complètement généraux, nous montrerons que, contrairement aux approches prises jusqu'à maintenant, nous sommes en mesure de découpler l'optimisation de la polarisabilité et des charges atomiques partielles.

Finalement, nous construirons un champ de force pour calculer le terme d'énergie d'interaction de deux dimères dans le vide pour lesquels la polarisabilité joue un rôle crucial.

Using electronic polarization from the internal continuum (EPIC) for intermolecular interactions

Jean-François Truchon[a,b], Anthony Nicholl's[c], J. Andrew Grant[d], Radu I. Iftimie[a], Benoît Roux[e] and Christopher I. Bayly[b]

[a]Département de chimie, Université de Montréal, C.P. 6128 Succursale centre-ville, Montréal, Québec, Canada H3C 3J7 [b]Merck Frosst Canada Ltd., 16711 TransCanada Highway, Kirkland, Québec, Canada H9H 3L1 [c]OpenEye Scientific Software, Inc., Santa Fe, New Mexico 87508 [d]AstraZeneca Pharmaceuticals Mereside, Macclesfield, Cheshire SK10 4TF, England [e]Institute of Molecular Pediatric Sciences, Gordon Center for Integrative Science, University of Chicago, Illinois 929 East 57th Street, Chicago, Illinois 60637

Abstract

Recently, the vacuum-phase molecular polarizability tensor of various molecules has been accurately modeled (Truchon, J.-F. *et al*, *J Chem Theory and Comput* 2008, *4*, 1480) with an intra-molecular continuum dielectric model. This preliminary study showed that electronic polarization can be accurately modeled when combined with appropriate dielectric constants and atomic radii. In this article, using the parameters developed to reproduce *ab initio* Quantum molecular polarizabilities, we extend the application of the 'electronic polarization from internal continuum' (EPIC) approach to intermolecular interactions. We first derive a dielectric-adapted least-square-fit procedure similar to RESP, called DRESP, to generate atomic partial charges based on a fit to a Quantum *ab initio* electrostatic potential. We also outline a procedure to adapt any existing charge model to

EPIC. The ability of this to reproduce local polarization, as opposed to uniform polarization, is also examined leading to an induced electrostatic potential relative root mean square deviation of 1%, relative to *ab initio*, when averaged over 37 molecules including aromatics and alkanes. The advantage of using a continuum model as opposed to an atom-centered polarizable potential is illustrated with a symmetrically perturbed atom and benzene. We apply EPIC to a cation-π binding system formed by an atomic cation and benzene and show that the EPIC approach can accurately account for the induction energy. Finally, this article shows that the *ab initio* electrostatic component in the difficult case of the H-bonded 4-pyridone dimer, a highly polar and polarized interaction, is well reproduced without parameter adjustment.

3.1 Introduction

An intramolecular continuum dielectric model has been recently applied to the calculation of molecular polarizabilities and shown to accurately reproduce those computed using high-level *ab intio* Quantum Mechanical (QM) calculations[1]. The electronic polarization from internal continuum (EPIC) approach showed that, relative to other methods, significantly less parameters were required to describe the anisotropy in molecular polarizability which was illustrated by calculations on a set of aromatic, diatomic and alkane molecules. This work focuses on the ability of the parameterized EPIC model to reproduce electrostatic potentials and in particular the response of this potential to external electric fields typical of those responsible for electronic polarization in intermolecular interactions. This is considered an essential feature for force-field based methodologies.

Many researchers have published work on combining a polarizable force-field with Poisson-Boltzmann (PB) formalism mainly to take advantage of the implicit solvent averaging modeled by a solvent dielectric constant. For example, the 'Polarizable Force Field' (PFF)[2] combines a point induced dipole (PID) model with a continuum solvent. Similarly the 'Atomic Multipole Optimized Energetics for Biomolecular Applications' (AMOEBA) force field couples a static multipolar expansion with atomic polarizabilities[3] in the context of a PB description of solvent. However, such models are computationally complex requiring atomic tensors to model response effects. In contrast, this work demonstrates that solute polarizability can be modeled using simple dielectric response theory, and requires only a small number of fitted atomic radii and isotropic relative permittivities (dielectrics). Here, however, we focus on explicit interactions, not on solvent reaction fields. Complex electron distributions and response are modeled using atom-centered point charges and a continuum dielectric. A crucial feature of the model is that screening effects produced by intramolecular polarization of the dielectric are explicitly

accounted for in calculating atomic charges by least-squares fitting to the electrostatic potential (ESP) computed using *ab initio* Quantum techniques.

This article describes such a modified RESP[4] approach to calculating atomic charges from the ESP. To indicate that dielectric screening is accounted for in the model, the approach is referred to as DRESP. In the rest of this article, we derive the equations relative to DRESP and show the resulting charges on few examples. Additionally, we propose a general way of adapting an existing charge model to behave properly when used with EPIC and we show its performances in reproducing the AM1-BCC[5,6] permanent ESP on selected molecules. The ability of the proposed polarization model to reproduce the response to non-uniform perturbations is also examined. In doing this, we have found that it is not necessary to refit radii and dielectric values optimized based on the vacuum-phase molecular polarizabilities only, thus our previously published parameter sets[1] are directly applied. These findings suggest that a general small molecule polarizable model based on continuum electrostatics could be developed. Finally, we apply the EPIC polarizable model to two problems of fundamental importance in biological applications (namely cation-π binding and H-bonding) to demonstrate its performance.

3.2 Methods

Below, a least-squares method, hereafter named DRESP, is derived for fitting atomic point charges to a QM electrostatic potential in the presence of an internal dielectric. The computational details for both the finite difference Poisson's equation (PE) solver and B3LYP calculations are then presented. Details on the calculation of the induced ESP are also given. Finally, the molecule dataset used to validate the current approaches is described.

3.2.1 A Least-Squares Method

The derivation presented in this section is a generalization of the least-squares method already in use to fit atomic charges to an ESP computed on a grid. Finding the optimal set of atomic charges in the presence of a locally changing dielectric requires a special treatment. For example, Tan and Luo[7] designed an iterative optimization scheme. The objective here is to obtain a set of equations linear in the atomic charges that can be easily and quickly solved computationally. The following derivation uses the Poisson's equation linearity in both the charge density and the potential such that the solution for each individual charge can be superimposed to produce the correct solution for the entire system.

The ESP at a point r of space, $\varphi(r)$, can be written in an integral formulation with a kernel function $G(r,r')$ (a Green's function) that defines the contribution of the local charge density $\rho(r')$ at a point r of space as shown in (3.1). The discrete nature of point charges allows us to write the charge density as a sum of N Dirac delta functions δ centered at each of the N charge positions r_i in (3.2) leading to (3.3):

$$\varphi(\vec{r}) = \int G(\vec{r},\vec{r}')\rho(\vec{r}')d\vec{r}' \tag{3.1}$$

$$= \int G(\vec{r},\vec{r}')(\sum_{i=1}^{n} \delta(\vec{r}'-\vec{r}_i)Q_i)d\vec{r}' \tag{3.2}$$

$$= \sum_{i=1}^{N} G(\vec{r},\vec{r}_i)Q_i \tag{3.3}$$

Now, we define a basis of N atom-source potential $\Phi^i(\vec{r})$ functions which describe the potential produced by a unit charge placed at atom i, setting all the other atomic charges to zero. Thus from (3.3):

$$\Phi^i(\vec{r}) = G(\vec{r},\vec{r}_i) \tag{3.4}$$

where $\Phi^i(\vec{r})$ is a Green's function with a single charge that can be obtained, for example, with a finite difference Poisson's equation solver. The linearity of electrostatic potential (with respect to charge) means (3.1) can be written in terms of the atom source potentials:

$$\varphi(\vec{r}) = \sum_{i=1}^{N} \Phi^i(\vec{r}) Q_i \tag{3.5}$$

The goal is to obtain a set of atomic charges that produces an electrostatic potential φ that best approximates that computed from an accurate Quantum calculation ψ. This can be achieved by minimizing the sum of the squares of the residuals between φ and ψ, evaluated at each grid point m. Defining the residual as χ^2 we have:

$$\chi^2 = \sum_{m=1}^{M} (\varphi_m - \psi_m)^2 \tag{3.6}$$

Substituting (3.5) into (3.6) gives

$$\chi^2 = \sum_{m=1}^{M} \left(\sum_{i=1}^{N} Q_i \Phi^i_m - \psi_m \right)^2 \tag{3.7}$$

Using a finite-difference Poisson solver for (3.7), Φ^i_m is the atom-source potential evaluated at the fitting grid point m by some interpolation scheme. The values of Q_i that minimize the residual χ^2 are obtained by setting to zero the N first derivatives of eq 7 against the atomic charges. This leads to:

$$\frac{\partial \chi^2}{\partial Q_k} = \sum_{m=1}^{M} 2 \left(\sum_{i=1}^{N} Q_i \Phi^i_m - \psi_m \right) \Phi^k_m = 0 \tag{3.8}$$

$$\sum_{i=1}^{N} Q_i \sum_{m=1}^{M} \Phi_m^i \Phi_m^k = \sum_{m=1}^{M} \psi_m \Phi_m^k \tag{3.9}$$

To simplify the notation, we define a matrix $\bar{\bar{A}}$ and a vector $\bar{b}$:

$$A_{ik} = \sum_{m=1}^{M} \Phi_m^i \Phi_m^k \tag{3.10}$$

$$b_i = \sum_{m=1}^{M} \psi_m \Phi_m^i \tag{3.11}$$

The linear system of equations to solve becomes

$$b_i = \sum_{k=1}^{N} A_{ik} Q_k \tag{3.12}$$

If the internal dielectric is set to 1 (vacuum and no polarization) in the computation of atom-source potentials, then the linear equations in (3.12) are those used by RESP[4]. However by introducing an arbitrary dielectric to model solute polarizability in the calculation of the atom-source potentials, it is possible to account for the contribution of polarization effects in obtaining charges that fit to the Quantum electrostatic potential. In practice, one can solve Poisson's equation and interpolate the electrostatic potential on the fitting grid for each Green's function related systems. In general, this procedure requires a solution per point charge. It is also possible to incorporate Lagrangian constraints and regularizing restraints as with RESP[4]. The introduction of the dielectric effects into the overall procedure is indicated by calling the modified procedure Dielectric RESP (DRESP). In this work, while we choose to not use restraints, we constrained topologically equivalent atoms to have identical charges and applied a constraint to the molecular formal charge. For example, the high symmetry of benzene leads to only one charge degree of freedom whereas in 4-pyridone there are 7 such degrees of freedom. In the remainder of this work,

we set the hyperbolic restraints to zero for the all the fit using a dielectric of one (RESP) or higher (DRESP). Normally the RESP hyperbolic restraints reduce the charges to make them more transferable at the expense of reducing the fit quality; for this work restraints were not used. Finally, the general formulation presented here allows to use any kind of dielectric boundary functions (smoothed or not) and multiple dielectric values.

3.2.2 Computational details

Ab initio Quantum electronic calculations were all based on density functional theory (DFT) using the B3LYP[8-10] functional as implemented in the Gaussian 03[11] software. The molecular geometries were optimized as reported in a previous study[1] with a 6-31++G(d,p) basis[12]. Subsequent ESP and energy calculations were performed using these optimal geometries but with an extended 6-311++G(3df,3pd) basis set. Molecular symmetry was not used as part of the calculations and an accurate convergence criteria was specified for the iterative calculation of the electron density matrix (by using the keyword Scf=Tight, which sets the average density change to be smaller than 10^{-8} a.u.). Interaction energies reported in this work were all corrected for basis set superposition error (BSSE) using the Boys and Bernardi counterpoise method[13,14].

Finite difference Poisson calculations were done using the program Zap[15]. A grid spacing of 0.5 Å was used to calculate potentials for both the DRESP charge fitting procedure and in the analysis of induced polarization. A smaller grid spacing of 0.3 Å was needed when solving PE to obtain smooth intermolecular energy curves. The grid boundary was positioned 10 Å away from the closest point on the vdW surface. The convergence grid energy was set to 0.00006 kcal/mol and a Richards vdW surface[16] was used to define the molecular dielectric boundary. The molecule inner dielectric and atomic radii were based on the parameterization that reproduced vacuum QM polarizabilities[1]. The three parameter sets reported previously[1] are examined in this work, namely: "P2E" a parameter set which adopts different dielectric values for alkane and aromatic molecules, "P1E" parameter set in

which a single dielectric is used to describe the solute response of all molecules, and finally "Bondi" which combines a dielectric of 4 with Bondi radii[17] (H radius set to 1.1 Å). The P2E and P1E parameter sets use optimal radii which are systematically smaller than Bondi radii, necessary to accurately reproduce the anisotropy of the molecular polarizability.

In addition to the grid specific to the Zap calculations, a grid was needed to compute the vacuum and the induced ESP. For this purpose, a face-centered cubic (FCC) grid ranging from 1.4 to 2.0 times the atomic Bondi vdW radii was employed[5,6]. These distances were determined to lead to an adequate sampling[18] of the ESP for atomic charge determination. The FCC grid spacing was set to 0.5 Å.

3.2.3 Induced polarization

To examine the accuracy of the dielectric polarization model, molecules were probed with a +0.5e point charge positioned exterior to the vdW surface of molecules. This probe charge has been used by others[19,20] and was recently shown to be well-adapted to examine the polarizability of aromatics[20]. In this work, the probe charge positions were determined using a single Conolly surface[21] with a vdW distance scaling factor of 2.0 and a density of 0.4 point per $Å^2$. Redundant probe charge positions, by symmetry, were partially eliminated. For example, this resulted in benzene having 6 positions for the perturbing charge, whereas 1,2,4-triazine had 47.

The induced potential is obtained by difference relative to the vacuum potential. Care must be taken to eliminate the contribution of the perturbing point charge to the calculation of the induction potential. This is handled automatically in the Quantum case as a consequence of the coding of Gaussian03. However, when using Zap an extra calculation is required to determine the explicit Coulombic potential arising from the perturbing charge. In total, the induced potential required three PE solutions per molecule on a constant grid. The numerical accuracy was ensured with the low energy convergence criterion.

The difference between QM induction potential (ψ) and that from EPIC model (φ) is characterized as a relative root mean square deviation (RRMSD) as follows:

$$\mathrm{RRMSD} = \frac{\sum_{m=1}^{M}\left(\psi_m - \varphi_m\right)^2}{\sum_{m=1}^{M}\left(\psi_m\right)^2} \tag{3.13}$$

3.2.4 Molecule dataset

The molecule dataset used in this work contains 15 polar and non-polar aromatics and 22 alkanes as shown in Figure 3.1. In addition, 4-pyridone is used in the study of a H-bond potential. The aromatic molecules exhibit quite a large spectrum of values for the ESP at the surface studied and possess a wide range of polarity. Due to their anisotropy, they constitute a good challenge to polarizable methods. Because of their non-planarity, biphenyl and its analogs have a different shape of the potential and a different conjugation of the π electron system compared to the other aryl molecules. Finally, the alkanes are quite polarizable in spite of their low polarity and the molecules used exercise many kind of shapes. All the molecule structures in this dataset are from the previous study[1].

3.3 Results and discussion

In this section, the ability of the proposed least-squares fitting method, DRESP, to produce an accurate permanent electrostatic potential is first examined and compared to the usual least-squares fitting RESP approach. We then continue by proposing a general way of coupling an existing charge model to EPIC, illustrated with the AM1-BCC charge model, which is general and shows many advantages in condensed phase simulations[5,6,22,23]. The ability of the EPIC approach to produce an accurate induced ESP in the presence of a locally varying electric field is also examined. Finally, the polarizable and permanent

features of EPIC are used to study the electrostatic energy profiles in two challenging cases: cation-π and 4-pyridone H-bonded dimers.

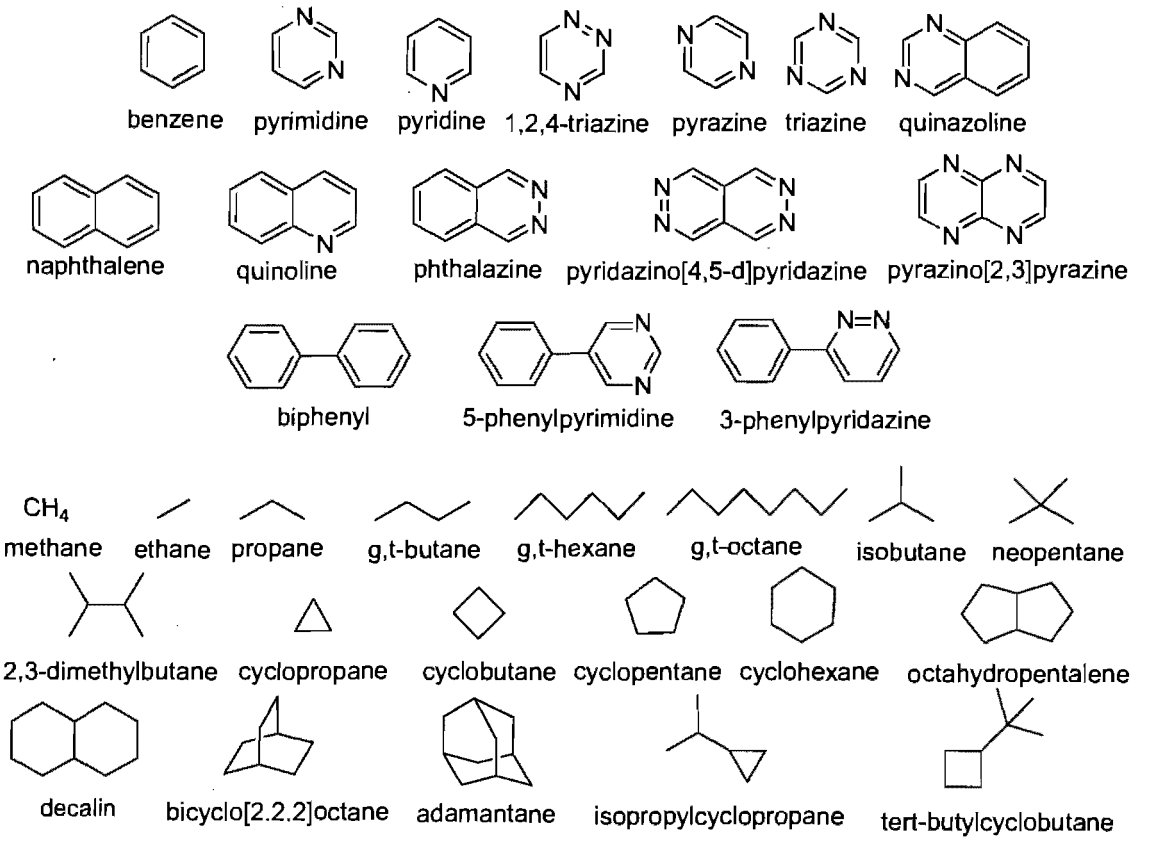
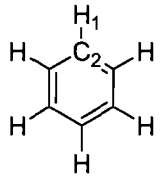


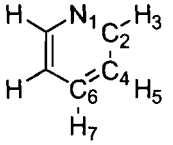
Figure 3.1. Molecule dataset which contains aryls and alkanes chemical classes. These 15 aromatic and 21 alkane molecules are extracted from reference 1. We use the notation 't' to indicate the all-trans conformation and 'g' when one or more gauche dihedrals are present; these are separate entries.

3.3.1 DRESP vs RESP

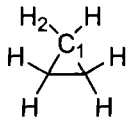
Given the way EPIC incorporates electronic polarization, the usual strategy to fit atomic partial charges to reproduce the QM *ab initio* ESP must be modified. This is due to the presence of a dielectric continuum inside the molecular electronic volume that screens the effects of a charge even at long distances. The DRESP approach, outlined in the Method section, solves this issue and provides a set of linear equations that give the optimal set of charges, in the least-squares sense, that reproduce the QM *ab initio* ESP. We first apply DRESP with the P2E parameters and compare the RRMSD made on the permanent ESP of B3LYP for the molecules of Figure 3.1. The atomic partial charges are also fitted to the same vacuum QM ESP using RESP. The DRESP charges are significantly larger than those obtained with RESP. This is illustrated in Figure 3.2 for benzene, pyridine and cyclopropane: the C/H charges of benzene are found to be ±0.12e as calculated with RESP and ±0.61 with the dielectric P2E model calculated with DRESP, the charges on pyrimidine are also significantly smaller with RESP than with DRESP and similarly for the cyclopropane. Although large and chemically counter-intuitive charges are needed when Poisson's equation is used, they produce the same ESP as the regular coulombic approach. Figure 3.2 also reports comparable RRMSD deviations between DRESP and RESP for benzene, pyridine and cyclopropane. The RRMSD reported on cyclopropane is high, but it is known that for simple alkanes the atomic partial charge approximation is poor[24,25]. However, since the actual ESP for the alkanes is very small, this defect might be negligible in intermolecular interactions although the alkane polarizability is not. Finally, it is important to keep in mind that whenever a radius or an inner dielectric value change, the DRESP charges cannot be transferred but must be re-optimized against the same QM *ab initio* ESP grid.



	Q_{RESP}	Q_{DRESP}
H_1	0.12	0.61
C_2	-0.12	-0.61
RRMS	14%	11%



	Q_{RESP}	Q_{DRESP}
N_1	-0.72	-2.56
C_2	0.57	1.49
H_3	0.01	0.24
C_4	-0.64	-1.26
H_5	0.21	0.60
C_6	0.39	-0.20
H_7	0.05	0.62
RRMS	13%	13%



	Q_{RESP}	Q_{DRESP}
C_1	-0.28	-0.72
H_2	0.14	0.36
RRMS	66%	54%

Figure 3.2. Benzene, pyridine and cyclopropane optimal charges fitting equally the same electrostatic potential with a dielectric of one (RESP, non-polarizable) and the P2E model (DRESP). The significantly higher charges with the P2E model comes from the internal dielectric screening of the point charges.

3.3.2 Use of an existing charge model: the AM1-BCC/DRESP example

AM1-BCC from Jakalian *et al.*[5,6] is an accurate charge model applicable to all small organic molecules and based on the over-polarized HF/6-31* ESP known to produce good charges adapted to polar media such as water. AM1-BCC was shown to be better or equal to much more computationally expensive methods in matching experimental free energies of solvation of small molecules[22,26]. A significant advantage of AM1-BCC is that the atomic point charges are obtained by adjusting with pre-fitted bond-charge corrections the AM1 electronic population analysis charges. Because the bond-charge corrections published by Jakalian *et al.* are adapted to produce polar media charges, a gas-phase AM1-BCC model is needed for use with a polarizable electrostatic model. In this work, we demonstrate proof-of-concept for the coupling of a general AM1-BCC/vacuum model, fitted to vacuum B3LYP/cc-pVTZ ESP[27], to the dielectric polarizable model. The approach

taken here uses the AM1-BCC/vacuum charges to calculate the ESP on a fitting FCC grid, detailed in the Methods section, to which dielectric adapted charges are fitted using the DRESP method. We call this fitting strategy AM1-BCC/DRESP. Although we apply this strategy to the AM1-BCC/vacuum charging scheme, it is general in the sense that any other charge model could be adapted the same way.

In Figure 3.3, we compare the B3LYP/6-311++G(3df,3pd) ESP to both the AM1-BCC/vacuum and the AM1-BCC/DRESP ESP. The correlation between the two RRMSD is excellent for aromatics and the AM1-BCC/DRESP charges produce slightly smaller RRMSD for the alkanes, which we think is not significant given the low level of accuracy for this chemical class. This demonstrates that extending an existing charge model, developed in a non-polarizable context, is easy and accurate. More importantly, the parameterization of the polarizable parameters and of the permanent charges can be fully

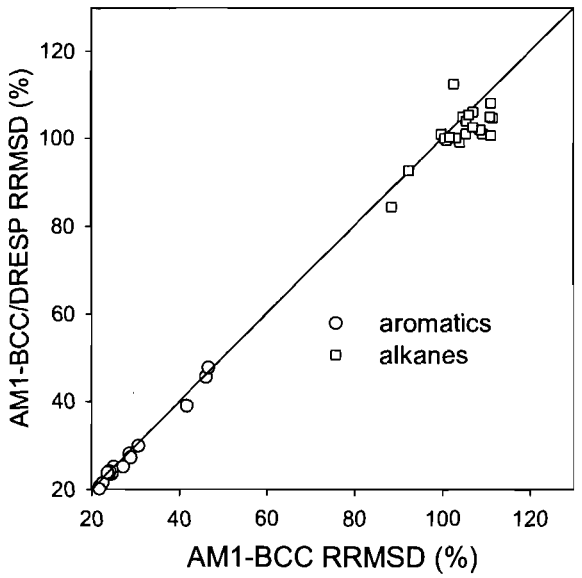


Figure 3.3. Correlation plot of the RRMSD obtained with AM1-BCC and AM1-BCC/DRESP charging schemes. The RRMSD are calculated against the B3LYP permanent electrostatic potential on the FCC grid.

decoupled, hence greatly reducing the fitting complexity.

3.3.3 Induced electrostatic potential

Previously, EPIC was shown to reproduce accurate induced dipole moments on molecules submitted to a uniform electric field[1]. Here we assess the ability of the approach to account for more local perturbations. This is accomplished by examining each of the 37 molecules from Figure 1 for which the electrostatic potential induced by a single probe charge of +0.5e is calculated with both B3LYP and EPIC according to the prescription detailed in the Methods section. Given the numerous placements of the probe charge for each molecule, the RRMSD deviations between B3LYP and EPIC are averaged and reported in Figure 3.4. This includes around 1700 B3LYP single point calculations in total. The RRMSD standard deviation (STDEV) indicates how much the error varies as a function of the position of the probe charge. The three bars per molecule correspond to the results obtained using three different parameter sets from the previous study[1]. The average RRMSD obtained with P2E (c.f. Method section) were used to sort the molecules in Figure 3.1. This method led to an average RRMSD and STDEV across the molecules of 1.06% and 0.3% respectively; the maximum average RRMSD is attributed to methane with 2.1%. The results obtained with the P1E parameter set (see Method) are slightly worse with an average RRMSD across the molecules of 1.70% and a STDEV of 0.8%. In the case of the Bondi parameter set (see Method), the average RRMSD and STDEV are 3.74% and 2.0%, almost a factor of four higher than P2E. The errors reported with Bondi parameters show a bias toward a more accurate description of the alkane polarizabilities.

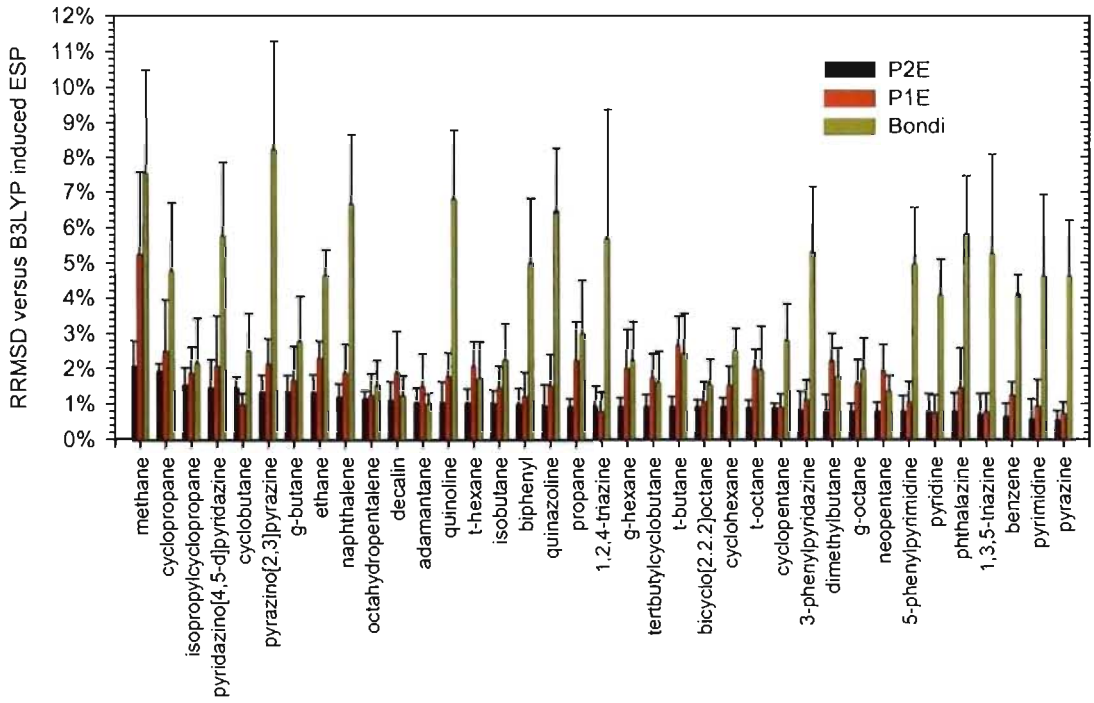


Figure 3.4. Average RRMSD on the induced ESP maps as calculated with EPIC using three parameter sets (P2E, P1E and Bondi, see the text). The ESP maps are generated by a +0.5e located at non-redundant positions and the reference induced ESP is calculated from B3LYP 6-311++G(3df,3dp).

To examine the effect of induction by the +0.5e charge on the alkanes, non-polar aromatics, and polar aromatics, it is of interest to compare the grid unsigned average of the induced and vacuum ESPs for each class. In the case of the alkanes, the unsigned average of the induced ESP is about 3 times higher than the vacuum ESP evaluated on the fitting grid. Benzene and non-polar aromatics have comparable induced and static ESP positive magnitudes and, in the case of polar aromatics such as pyrimidine, the induced ESP is between twofold and threefold smaller than the vacuum ESP. This indicates that the level

of electronic polarization is significant and appropriately challenging to the polarizable model. The RRMSD obtained with the induced ESP is much lower than the RRMSD of the static ESP fit by atomic charges. This shows that the model can accurately account for locally induced polarization although the induced ESP is certainly simpler than the ESP originating from the unperturbed molecule.

Here, we use the induced ESP to test the ability of the EPIC model, *parameterized solely on the QM gas phase polarizability tensor*, in reproducing a molecule's induced QM electronic response to a polarizing charge. The very small RRMSD obtained with the P2E and the P1E parameter sets indicate that the local polarization with a non-uniform electric field is as accurately modeled as the induced dipole moment due to a uniform electric field[1] (molecular polarizability), the basis of the P2E and P1E parameterization. In contrast, most previous polarizable models[28-30] obtain their polarizability parameters (and often simultaneously the charges) by fitting to the induced ESP on polarized molecules.

3.3.4 Induction by a symmetric field

The parameterization of polarizable models usually involves simple fields such as uniform external electric fields or fields produced by a probe point charge or a probe dipole. These external fields induce mainly a molecular dipole moment, which should be reasonably well accommodated given a simple functional form for a polarizable model. However, in condensed media, the electric field is rarely simple and is often transiently symmetric around a molecule. In this section we compare the induced electrostatic potential in a non-trivial electric field applied on an argon atom and benzene; three methods are examined: point inducible dipoles, EPIC and B3LYP.

In the first system, the argon atom is sandwiched between two positive point charges positioned at 3.0 Å above and below the atom. The EPIC argon radius used is set to

1.31 Å and the dielectric=7.36 as fit to reproduce the (gas-phase) B3LYP atomic polarizability of argon (11.1 a.u.). Given the symmetry of this system, the net electric field at the argon nucleus is zero. Since the point inducible dipole[1] and the fluctuating charges[31] polarizable models respond only to the net electric field at the nucleus, neither of these models would show induction in this case. As illustrated on Figure 3.5, the B3LYP induced electrostatic potential is a quadrupole with the d_z^2 hydrogenoid-like orbital symmetry. Remarkably, the EPIC induced electrostatic potential has the same symmetry and is of similar magnitude.

The second system examined consists of benzene sandwiched by two $+1e$ point charges located at 2Å above and below the ring. This is a fairly large perturbation where the point charges are positioned approximately at the Li^+/benzene equilibrium distance. Although in nature it is unlikely that two Li^+ atoms would be stable in such a sandwich system, the perturbing electric field varies quickly from zero at the center of the benzene ring to 2/r at infinity. So within the volume of the benzene, the field varies enough to significantly test the model. The symmetry of this arrangement is such that the external electric field has only a component in the plane of the ring at the atomic positions, although the out-of-plane polarization should be predominant. A similar system was used to show the failure of the fluctuating point charges model by Stern *et al.*[31]. In principle the point inducible and related polarizability models should also have difficulty since the magnitude of the induced potential would then be fully dictated by the in-plane polarizability component. For our comparison, we used the AMOEBA polarizability model for benzene which includes a Thole exponential damping parameter of 0.39 coupled with carbon and hydrogen isotropic polarizabilities of 1.334 $Å^3$ and 0.8 $Å^3$ respectively (as provided in TINKER 4.2 distribution parameter file). The Thole parameter has the role of adjusting the molecular polarizability anisotropy by reducing the atomic induced-dipole/induced-dipole interactions thereby avoiding the polarizability catastrophe, inherent to the PID models.

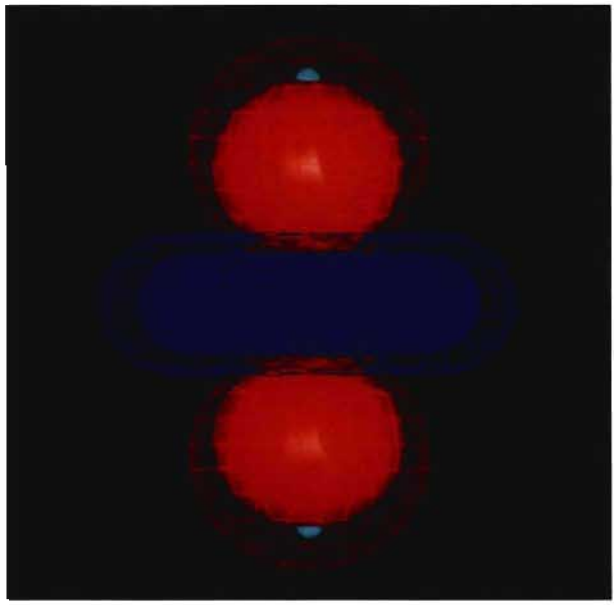


Figure 3.5. The induced electrostatic potential of an argon atom sandwiched in between two +0.5e charges positioned at 3.0 Å of the nucleus. The solid iso-surface corresponds to the B3LYP/6-311++G(3df,3pd) and the mesh iso-surface to the result from EPIC (radius=1.3 Å, dielectric=7.4). The induced moment is an induced quadrupole with the d_z^2 orbital symmetry. The traditional atomic polarizable approaches have a zero induced electrostatic potential.

Calculated with these parameters, the benzene vacuum molecular polarizabilities are 11.4 $Å^3$ in the plane and 6.2 $Å^3$ perpendicular to the plane, which are in close agreement with the values of 12.2 $Å^3$ and 6.7 $Å^3$ obtained with B3LYP/aug-cc-pVTZ. For this comparison, we use EPIC with the P2E parameters which also produces accurate polarizabilities of 12.2 $Å^3$ and 6.6 $Å^3$.

For the charge-sandwiched system, the iso-contour lines of the induced electrostatic potential on one of the six planes of symmetry perpendicular to the ring are plotted in Figure 3.6. The iso-lines are spaced by 5 kcal/mol/*e* and the ESP values are given in kcal/mol/*e*. For the three methods, the induced ESP has the shape and symmetry of a d_z^2 orbital, a quadrupole moment, with the negative lobes oriented along the axis joining the

two probe charges and the positive torus located close to the hydrogen atoms. Good agreement is obtained between the induced ESP of B3LYP and EPIC even intra-molecularly. At the probe charge positions, the B3LYP, EPIC and AMOEBA induced potentials are -52, -60 and -29 kcal/mol/*e* respectively. In general the point inducible dipole potential is too positive moving away from the benzene along the probe charge axis. The relatively large difference between B3LYP and AMOEBA, compared with the smaller error made with EPIC, cannot solely be explained by the smaller in-plane benzene polarizability of AMOEBA because the AMOEBA positive ESP regions, particularly at vdW distances from the H atoms (2 Å away), match the B3LYP values. We attribute the better success of the dielectric-based EPIC model by its departure from the atom centric-polarization to an electron-centric model where the entire molecular boundary responds to the electric field. In order to improve the PID model while retaining the correct average molecular polarizability and anisotropy of benzene, auxiliary polarizable points above and below the ring would have to be added, at the expense of additional complexity and supplemental parameterization. This being said, condensed phase simulations have lead to parameterizations of polarizable PID-based force fields that accounted for important liquid properties of benzene[32] although a polarizable electrostatic term is often not necessary to fit the liquid properties. We believe that the accurate polarizable electrostatic term should make a more important difference in heterogeneous and anisotropic environment such as the active site of an enzyme or a trans-membrane ionic channel. It is encouraging that the EPIC model with default P2E parameters exhibits good physical behaviors at the purely electrostatic level even in the context of strong and complex electric fields.

3.3.5 Cation-π interactions

In the previous sections, we have separately demonstrated that EPIC can handle both the permanent and induced electrostatic potential strictly by comparison to the induced B3LYP electrostatic potentials. Now, we combine the DRESP fitting procedure and the EPIC model to assess the electrostatic interaction energy in cation-π systems. The cation-π

attractive interaction energy between benzene and Li+, Na+ or K+, displaced along the benzene six-fold symmetry axis, was shown to be impossible to describe with nonadditive models[33]. However, the contribution of the induction energy (from electrostatic polarization) is crucial for an accurate description of cation-π binding[33,34]. Furthermore, the presence of cation-π interactions in many biological systems[35-37] makes this application a critical validation for a new polarizable electrostatic term. Therefore, in this section we decompose the total interaction energies and check if the induction found with the P2E parameter set is quantitatively correct.

At the QM level, the total interaction energy between an atomic cation and benzene can be conceptually split into electrostatic, vdW repulsive, and vdW attractive components. Each term is normally represented separately in a force field, although one term often compensates for another containing deficiencies. We are now only interested in examining the electrostatic component and we focus on the proton since its vdW interaction with the benzene can be neglected. Other cations have similar electrostatic profiles outside their vdW range (data not shown). At the force field level, if the proton/benzene interaction energy is described correctly, all that remains to be added is the repulsive vdW term in order to model other, more physiologically relevant, atomic cation-π interactions. This will be shown in the next section in a different application.

The energy of interaction between a benzene molecule and H^+ is calculated at the B3LYP/6-311++G(3df,3pd) level with no basis function positioned on the proton to avoid unphysical stabilization. Figure 3.7 reports the energies as a function of the distance from the center of the benzene ring using B3LYP, a non-polarizable Coulomb potential calculated with RESP-fitted atomic charges, the electrostatic component of the polarizable AMOEBA force field, and our polarizable EPIC/P2E model. For the non-polarizable and AMOEBA models, the parameters mentioned in section 3.4 are used. Figure 3.7 shows that EPIC matches quantitatively the B3LYP energy profile. As found previously[33], the non-

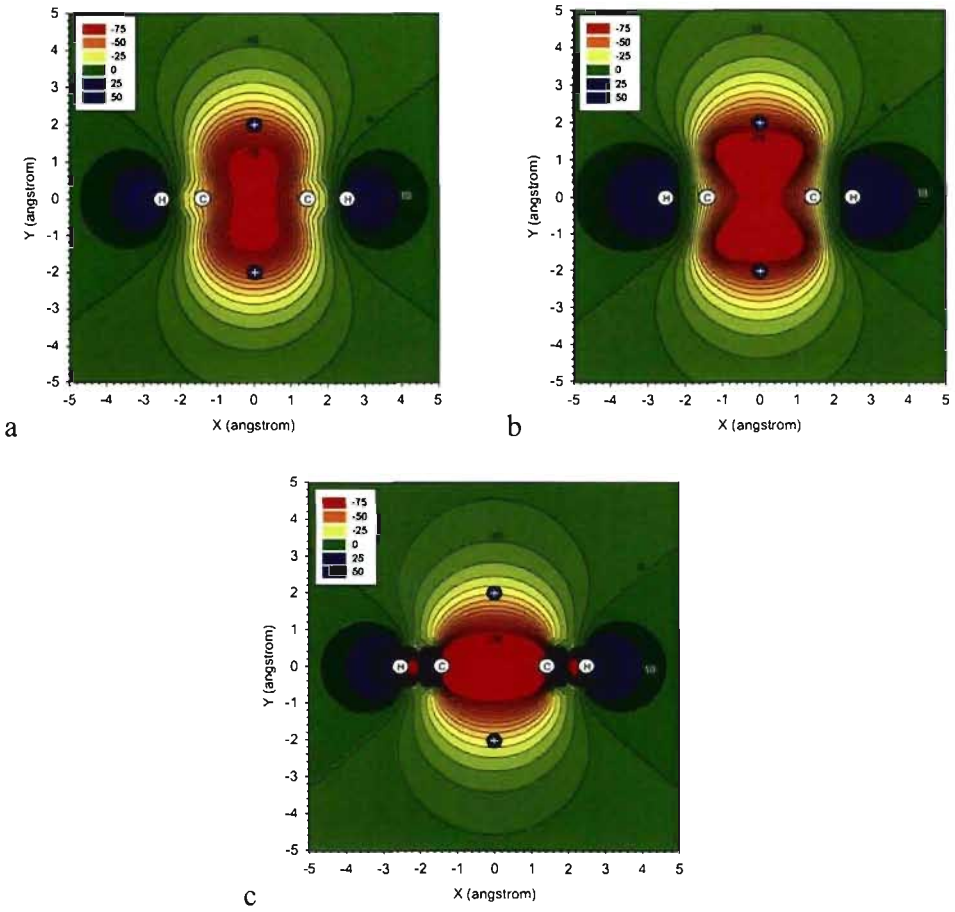


Figure 3.6. The induced electrostatic potential for benzene is shown by iso-contour lines spaced by 5 kcal/mol/*e* on one of the six symmetry plan perpendicular to the ring. The external perturbing potential is produced by two +1*e* charges positioned 2 Å above and below the benzene ring. The induced potential obtained at the B3LYP/6-311++G(3df,3pd) level (a) is compared to the EPIC/P2E (b) and AMOEBA, a good quality QM derived point-inducible model (c).

polarizable potential is inappropriate for describing cation-π interactions, despite the fact that the atomic partial charges of benzene were fitted on the same ESP grid as were the EPIC/DRESP charges. The energy resulting from electronic polarization dominates the electrostatic interaction energy, being twice more stabilizing than the static contribution at typical intermolecular separations – this induction energy remains substantial for intermolecular separations as large as 4 Å. The AMOEBA description of the electrostatic

energy captures most of the induction energy. This model, like EPIC, is derived from *ab initio* calculations, but includes atom-based multipole expansion terms up to the quadrupole for the permanent potential. In addition, AMOEBA adds an atomic polarizability on the hydrogen atoms. The difference between B3LYP and AMOEBA energies can be explained, in part, by AMOEBA's slightly smaller out-of-the-plane polarizability (c.f. section 3.4). Tsuzuki *et al.*[34] have shown that it is possible to get the cation-π system correctly modeled with a PID model if it is parameterized in a specialized manner. In their work they not only use atomic multipoles, but also anisotropic atomic polarizability located on the carbon atoms. Once more, we see that the EPIC model with P2E parameters is robust and general, needing only a small number of default parameters to account for the different aspects of electronic polarization. The accuracy of this model implies that the vdW term will not need to compensate for errors in the close-range electrostatics thus easing its parameterization.

3.3.6 H-bond of the pyridine-4(1*H*)-one dimer

The induction energy in H-bonds is implicitly included in existing additive force fields. The single-minimum interaction energy profile representing an H-bond becomes a trade-off between the attractive electrostatic term and the short-range repulsive vdW term of a force field. For example, in the original AMBER force field a special 12-10 Lennard-Jones potential was initially needed to describe short range H-bonding potentials[38]. In more recently developed non-polarizable force fields[39], both the condensed phase overpolarized HF/6-31G(d) charges and the addition of hydrogen atom types using small Lennard-Jones radius (σ) parameters corrected for the lack of induction energy. Including explicitly induction energies should, in principle, simplify the parameterization of vdW potential and incorporate non-additive condensed phase effects such as H-bond cooperativity[40,41]. In this section, we examine the challenging case of pyridin-4(1*H*)-one, hereafter 4-pyridone, which was shown to form strong intermolecular H-bonds[42] when monomers are co-planar and aligned along the two-fold axis passing through the NH and the CO bonds. This is in part due to the large dipole moment of 7 Debye and the high polarizability component of 97

a.u. both oriented along the H-bond axis that should polarize the monomers and thereby constitute an interesting second application for the validation of polarization using EPIC. This is a case where polarizability needs to be taken into account even at the dimer level. In what follows, we conduct two different studies which objectives are a) to assess if the EPIC model with DRESP charges gives the correct electrostatic interaction energy profile compared to B3LYP and b) to reproduce the dimer H-bond dissociation energy curve obtained by B3LYP.

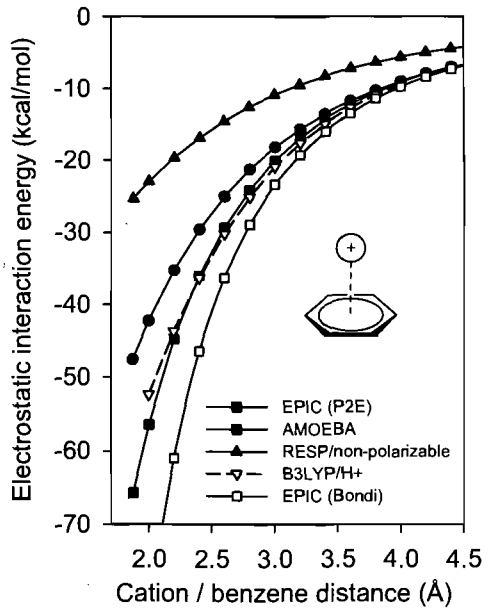


Figure 3.7. The electrostatic interaction energy between a benzene molecule and H+ displaced along the C_6 symmetry axis corresponds to the electrostatic component of a cation-π system formed with an atomic cation. The non-polarizable model using ESP derived charges is far from the B3LYP calculated energy, the EPIC/P2E model closely follows the B3LYP curve and AMOEBA captures most of the electronic polarization of B3LYP. The EPIC electrostatics fitted on B3LYP monomer reproduces the correct cation-π electrostatic energy without adjustment.

3.3.6.1 4-pyridone dimer electrostatic interaction energy

Unlike the previous cation-π application, evaluating the performance of our electrostatic approach here involves partitioning the total electronic structure energy into its components because the vdW term cannot be neglected. We will divide the total energy into an electrostatic (ΔEelec) term and an exchange-repulsion (hereafter called the vdW repulsion and noted ΔEvdw) term. The short range attractive dispersion energy is omitted because B3LYP does not capture it and it remains very small compared to the range of energies involved.

Given this partitioning, we calculate ΔEelec by subtracting ΔEvdw from the total interaction energy (ΔEtot). The delta symbol in front of the energy signifies that this is an interaction energy, meaning the difference in energy between the interacting system and the energies of the monomers. To approximate the $\Delta Evdw^{DFT}$ term, we used benzo-1,4-quinone (quinone) and benzene which both have a relatively small ESP. The benzene hydrogen replaces the 4-pyridone H donor and the quinone oxygen the acceptor, forming a quasi H-bond. The assumption here is that the vdW repulsive term is similar between the two systems. The monomer geometries are held fixed and the carbonyl oxygen to benzene H distances are the same as the 4-pyridone H-bonded dimer distances. The ESP of both molecules is relatively small and can be approximated using RESP-derived atomic point charges ($\Delta Eelec^{RESP}$); thus ΔEelec is calculated using equations (3.14) and (3.15):

$$\Delta Evdw^{DFT} \approx \Delta Etot^{DFT}(\text{benzene, quinone}) - \Delta Eelec^{RESP}(\text{benzene, quinone}) \quad (3.14)$$

$$\Delta Eelec^{DFT} \approx \Delta Etot^{DFT}(\text{4-pyridone,4-pyridone}) - \Delta Evdw^{DFT} \quad (3.15)$$

Figure 3.8a shows $\Delta Eelec^{DFT}$, $\Delta Eelec^{RESP}$ and $\Delta Eelec^{DRESP}$ associated with the 4-pyridone dimer for distances going from 1.6 Å to 4 Å. The equilibrium distance is found at 1.78 Å[42]; details of the energies are reported in the Supplemental Material. As reported in

Figure 3.8a, the induction energy stabilization obtained with EPIC/P2E matches B3LYP over most of the distances. The deviations at close contact (<2 Å) may be attributable to defects in the approximation given by eq. (3.14). Comparing the Coulomb and B3LYP curves at the equilibrium distance, the induction energy is almost -5 kcal/mol, a significant increase over the Coulomb value of -11.6 kcal/mol. The quantitative match between the approximated B3LYP electrostatic H-bond energy shows that the induction interaction is appropriately described by our polarizable EPIC/P2E. It is important to emphasize that the radii and dielectric parameters (P2E) used were not fitted to any energy but to the QM gas-phase molecular polarizability tensors for many molecules simultaneously[1]. In Figure 3.8a, we can also see the EPIC/Bondi interaction curve that uses DRESP derived charges and the Bondi parameter set (c.f. Method). The long range interaction energies are appropriate but when the electronic volumes, defined by the Bondi radii (1.52 Å for oxygen and 1.1 Å for the hydrogen), start to interpenetrate, the induction energy becomes insufficient and undergoes numerical instability that we attribute to the vdW surface used in the finite difference PB solver, known to form cusps.

3.3.6.2 4-pyridone dimer dissociation energy

To get an idea of how well EPIC polarization could be incorporated into an atomic force field, a vdW term was fitted to see how well the QM energy profile could be reproduced; the details of the fitted vdW terms are given in the Supplemental Material. The resulting complete interaction energy profiles for the H-bond formation of the 4-pyridone dimer, ΔEtot(4-pyridone,4-pyridone), are presented in Figure 3.8b. For the DFT profile ($\Delta Etot^{DFT}$), the approximation of (3.15) is not needed because only the total energy is examined. The B3LYP/6-311++G(3df,3pd) BSSE corrected energies show a very stable H-bond with a dissociation energy of -10.8 kcal/mol..

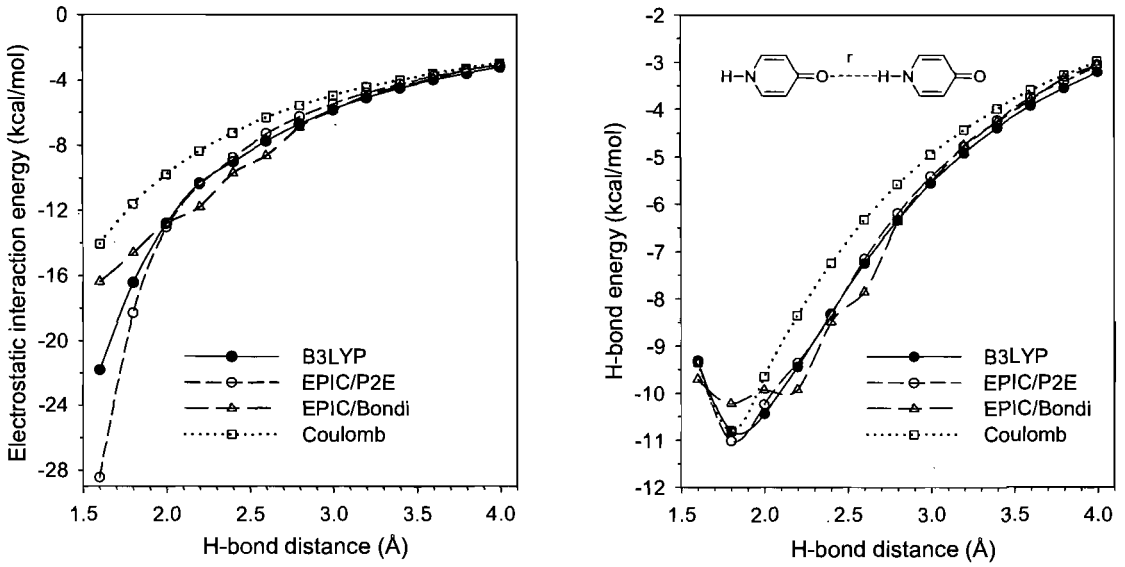
A few comments on the vdW fitting process are in order. In keeping with high-level QM calculation of exchange-repulsion energies,[43] we have used a two-parameter

exponential vdW energy function $\Delta Evdw^{EPIC}(r)$ fitted to the residuals $\Delta Etot^{DFT}(r)$ - $\Delta Eelec^{EPIC}(r)$ calculated along the intermolecular H-bond axis. Fortunately, the vdW dispersion energy, an attractive force, is absent in DFT methods and should not be needed here. The fit of the EPIC/P2E residuals resulted in a dissociation energy curve quantitatively reproducing the B3LYP energies in all ranges examined, giving a correlation coefficient (R^2) of 0.999.

Note that the atomic partial charges derive from a DRESP fit to the monomer only. In contrast, the Coulomb model residuals (using atomic partial charges from the gas-phase monomer without polarizability) could not be fit to an exponential as successfully, exhibiting an attractive potential well of about -1 kcal/mol located at 2.3 Å. This example shows that, at the level of the dimer, the lack of polarizability introduces a requirement for a more complex functional form for the vdW term in order to compensate. Finally, the EPIC/Bondi potential has difficulty capturing the energy minimum and exhibits numerical instability. At very short H-bond distances (<1.5 Å), EPIC/P2E also exhibits similar instability behavior, but fortunately this is less relevant given that the high repulsive vdW energy is dominant. This is due to the smaller dielectric radii assigned to the oxygen and hydrogen atoms in the fit. The use of a smooth dielectric boundary could significantly reduce these effects[15].

While these applications demonstrate the use of EPIC to incorporate electronic polarization into short-range intermolecular interactions, the generalization of this model for use in the condensed phase will require special attention in the parameterization. Indeed, it has been proposed based on various evidences that the condensed phase molecular polarizability per monomer is smaller than its gas phase polarizability. In practice, the correction to a PID model was made by fitting the polarizabilities on *ab initio* values obtained with a relatively small basis set, which systematically produces smaller molecular polarizabilities[43]. Hence for the PID model, the parameterized polarizability cannot accurately account for both the gas- and condensed-phase polarization, one of the

main objectives for a polarizable force field. We are currently examining this problem with EPIC and this will be addressed in a subsequent publication.



a b

Figure 3.8. The reported electrostatic interaction energies of the H-bonded 4-pyridone dimer (a) show that the EPIC/P2E model produces the appropriate polarization as opposed to the non-polarizable permanent charge model (Coulomb) when compared to B3LYP (3.15). The EPIC/Bondi calculations produce the correct electronic response at long ranges of H-bond distances but saturates as the vdW dielectric surfaces of the monomers start overlapping. The observed deviations are a result of the numerical instability that occur when the dielectric spheres come into contact at 2.6 Å. (b) The BSSE/corrected B3LYP interaction energies of the dimer unveils a very strong H-bond of -10.8 kcal/mol at the minimum located at 1.78 Å. The reported classical approaches combined the electrostatic energies (shown in a) and a fitted repulsive vdW term. EPIC/P2E matches the B3LYP energies over the examined range whereas the Coulomb non-polarizable model deviates at longer distances as a result of the difficulty for such a model to match both regions.

3.4 Conclusion

This work extends the proof-of-concept of the EPIC (electronic polarization from intramolecular continuum) approach to intermolecular interactions[1]. We show that in general the presence of an intramolecular continuum dielectric, for example in PB or PE formalisms, requires a modification of the conventional method for obtaining accurate atomic partial charges. We derive a least-squares approach called DRESP, by analogy to RESP, which not only reproduces the *ab initio* electrostatic potential but also offers a way to transfer the atomic partial charges from an existing vacuum charge model. We demonstrate this by successfully transferring the charges from an AM1-BCC/vacuum charging scheme of a set of 37 molecules including polar and non-polar aromatics and alkanes. In general, the charges obtained are significantly larger whenever the dielectric inside the molecule is larger than one. Even more importantly, our results show that the polarizable parameters of EPIC can be derived independently from the permanent electrostatic terms. In fact, the atomic radii and dielectrics fitted to vacuum *ab initio* molecular polarizabilities were derived in the absence of atomic partial charges. We believe that this parameter decoupling will be an important asset to broaden the approach to encompass bio-organic chemistry.

Although EPIC was successfully fitted previously to reproduce *ab initio* dipole moments induced by a uniform electric field[1], it was important to demonstrate that the full electrostatic potential induced by a more local or complex perturbation would be as accurate. To this end, we tested the validity of our parameters with a perturbing +0.5e probe charge moved on a Connolly surface. This is frequently employed to fit the polarizable parameters in force fields. The results were very encouraging leading to an average of 1% RRMSD deviation, relative to B3LYP, obtained from about 1700 calculations done on our set of 37 molecules. This shows that we can fit the EPIC radii and

dielectrics on gas-phase molecular polarizabilities and expect an accurate local electronic polarization response as well.

A potential advantage of EPIC over other polarizable models is its use of 'electronic polarizability density' through the intramolecular continuum dielectric. As opposed to PID models, the electric field induction is effective through the entire molecular volume. A recent study by Schropp and Tavan[44] suggests that the usual approximation that the polarizability is centered on atomic nuclei artificially invalidates the use of gas-phase derived atomic polarizabilities into condensed phase simulations unless non-obvious and non-general corrections are applied. To further examine the ability of EPIC to deal with inhomogeneous environments, an argon atom and a benzene molecule were sandwiched by positive point charges. In both cases EPIC led to an accurate induced electrostatic potential relative to B3LYP. The *ab initio* derived AMOEBA polarizable force field parameters were used for comparison and show a significant quantitative discrepancy of the out-of-plane potential. In agreement with Schropp and Tavan, we attribute this deficiency to the difficulty of atom centered polarizabilities to account for locally varying electric field.

We also applied our independently generated charges, radii and dielectrics to calculate the cation-π electrostatic interaction energy between a benzene molecule and a proton. This energy corresponds, more generally, to the electrostatic component for the binding of Li^+, Na^+ or K^+ to benzene where the induction energy is predominant. The results show that EPIC with parameters derived uniquely on the monomer lead to B3LYP quality binding energies. In addition, the H-bonding electrostatic energy of the 4-pyridone dimer has been examined and we found that EPIC/P2E quantitatively matches the approximated B3LYP electrostatics and total interaction energies whereas the non-polarizable term obtained with fixed charges derived from the ESP were not sufficient. Although we have not covered an exhaustive intermolecular list, these two applications are challenging cases that clearly show that the approach can work. Our fitting strategy of the electrostatic on the monomers can be easily generalized. In the 4-pyridone dimer example,

we also show that the vdW term, needed to obtain the full energy, would not have to compensate for the poor electrostatic at short distances. This is one important and expected advantage from an accurate polarizable electrostatic model.

This work shows a new and potentially advantageous electrostatic term which could be applied to molecular dynamics or Monte Carlo simulations. Recently, molecular dynamics simulations using a PB solver to include implicitly the effects of solvent polarization were successfully carried out [2,3,15,45-48]. In order to fully use the concepts discussed herein as the electrostatic foundation of a force field, there are remaining scientific points to be addressed. The dielectric boundary will require a special attention if stable forces are to be calculated. Also, transferability of the parameters obtained from *ab initio* gas-phase calculations to the condensed phase need to be addressed. The extension of the parameterization will command a major effort not only for the electrostatic term, but also for all the other force field terms which need to balance the electrostatics.

Acknowledgment. The authors thank Dr. Chen from Lamar University for providing to us the coordinates of the optimized 4-pyridone dimer. We thank Daniel J. McKay from Merck Frosst and Georgia McGaughey from Merck & Co. for useful comments on the manuscript. This work was made possible by the computational resources of Réseau Québécois de Calcul Haute Performance (RQCHP). The authors are grateful to OpenEye Inc. for free academic licenses. J.-F.T. thanks the Natural Sciences and Engineering Research Council of Canada (NSERC) for a Canada graduate scholarship (CGS D) and Merck & Co. for support through the MRL Doctoral Program. R. I. I. acknowledges financial support from the NSERC. B. R. is supported by NIH grant GM072558

3.5 Supporting Information

Data related to Figure 3.7 for the interaction energy of a positive point charge (H+) move along the benzene C_6 symetry axis forming a cation-pi interaction.

R[a]	EPIC/P2E[b]	AMOEBA[b]	RESP[b]	B3LYP/H+[b]	EPIC/Bondi[b]
1,87	-65,76	-47,60	-25,37		-102,60
2,00	-56,43	-42,30	-22,98	-52,37	-82,75
2,20	-44,79	-35,26	-19,66	-43,74	-61,00
2,40	-36,02	-29,59	-16,87	-36,36	-46,48
2,60	-29,33	-25,00	-14,52	-30,18	-36,30
2,80	-24,14	-21,25	-12,54	-25,09	-28,90
3,00	-20,08	-18,18	-10,87	-20,95	-23,40
3,20	-16,86	-15,64	-9,47	-17,60	-19,30
3,40	-14,29	-13,53	-8,27	-14,89	-16,00
3,60	-12,22	-11,76	-7,26	-12,69	-13,50
3,80	-10,53	-10,28	-6,40	-10,90	-11,40
4,00	-9,15	-9,02	-5,66	-9,42	-9,78
4,20	-8,02	-7,95	-5,03		-8,43
4,40	-7,09	-7,04	-4,48		-7,32

[a]Distance between the point charge and the center of the benzene ring in Angstrom.
[b]Electrostatic interaction energies in kcal/mol.

Data related to Figure 3.8 for the interaction energies between two H-bonded 4-pyridone molecules. The data is spread over few tables. The notation for the column labels follows the tables.

	Eq. (3.14) rhs term 1	Eq. (3.14) rhs term 2	Eq. (3.14) lhs	Eq. (3.15) rhs term 1 Figure 8b (B3LYP)	Eq. (3.15) lhs
Distance	$\Delta E(B,Q)_{tot}$ DFT	$\Delta E(B,Q)_{elec}$ RESP	ΔE_{vdw} DFT	$\Delta E(P,P)_{tot}$ DFT	ΔE_{elec} DFT
1,600	10,648	-2,018	12,666	-9,340	-22,006
1,800	4,200	-1,527	5,727	-10,800	-16,527
2,000	1,219	-1,181	2,400	-10,440	-12,840
2,200	-0,039	-0,930	0,891	-9,440	-10,331
2,400	-0,484	-0,744	0,260	-8,310	-8,570
2,600	-0,575	-0,603	0,028	-7,250	-7,278
2,800	-0,538	-0,495	-0,043	-6,330	-6,287
3,000	-0,465	-0,411	-0,054	-5,560	-5,506
3,200	-0,394	-0,344	-0,050	-4,920	-4,870
3,400	-0,332	-0,291	-0,041	-4,380	-4,339
3,600	-0,282	-0,247	-0,035	-3,930	-3,895
3,800	-0,240	-0,212	-0,028	-3,550	-3,522
4,000	-0,205	-0,183	-0,022	-3,210	-3,188

	Figure 8a (EPIC/P2E)	Figure8a (EPIC/Bondi)	Figure 8a (Coulomb)	vdW residuals	vdW residuals
Distance	$\Delta E(P,P)_{elec}$ EPIC/P2E	$\Delta E(P,P)_{elec}$ EPIC/BONDI	$\Delta E(P,P)_{elec}$ Coulomb	E_{vdw} EPIC/P2E	E_{vdw} EPIC/Bondi
1,600	-28,47	-16,382	-14,064	19,130	7,042
1,800	-18,31	-14,592	-11,629	7,510	3,792
2,000	-13,02	-12,799	-9,792	2,584	2,359
2,200	-10,41	-11,798	-8,368	0,970	2,358
2,400	-8,74	-9,705	-7,238	0,430	1,395
2,600	-7,31	-8,661	-6,326	0,060	1,411
2,800	-6,25	-6,889	-5,577	-0,080	0,559
3,000	-5,44	-5,884	-4,953	-0,120	0,324
3,200	-4,79	-4,976	-4,428	-0,130	0,056
3,400	-4,23	-4,376	-3,982	-0,150	-0,004
3,600	-3,75	-3,901	-3,598	-0,180	-0,029
3,800	-3,4	-3,441	-3,267	-0,150	-0,109
4,000	-3,06	-3,119	-2,978	-0,150	-0,091

	vdW residuals	Figure 8b (EPIC/P2E)	Figure 8b (EPIC/Bondi)	Figure 8b (Coulomb)
Distance	E_{vdw} Coulomb	$\Delta E(P,P)_{tot}$[a] EPIC/P2E	$\Delta E(P,P)_{tot}$[b] EPIC/Bondi	$\Delta E(P,P)_{tot}$[c] Coulomb
1,600	4,724	-9,305	-9,684	-9,340
1,800	0,829	-11,010	-10,210	-10,800
2,000	-0,648	-10,243	-9,932	-9,647
2,200	-1,072	-9,351	-9,922	-8,342
2,400	-1,072	-8,337	-8,478	-7,234
2,600	-0,924	-7,156	-7,858	-6,325
2,800	-0,753	-6,191	-6,364	-5,577
3,000	-0,607	-5,418	-5,540	-4,953
3,200	-0,492	-4,782	-4,751	-4,428
3,400	-0,398	-4,227	-4,229	-3,982
3,600	-0,332	-3,749	-3,805	-3,598
3,800	-0,283	-3,400	-3,378	-3,267
4,000	-0,232	-3,060	-3,078	-2,978

[a]Total energy includes a vdW term fitted on vdW residuals found in E_{vdw} EPIC/P2E given by E_{vdw} EPIC/P2E = 43248 (kcal/mol) *exp(-4.826 ($Å^{-1}$) * R) where R is the distance (see below) in Angstrom. [b]Total energy includes a vdW term fitted on vdW residuals found in E_{vdw} EPIC/Bondi given by E_{vdw} EPIC/Bondi = 199.41 (kcal/mol) *exp(-2.121 ($Å^{-1}$) * R) where R is the distance (see below) in Angstrom. [c]Total energy includes a vdW term fitted on vdW residuals found in Evdw Coulomb given by E_{vdw} Coulomb = 5252300 (kcal/mol) *exp(-8.701 ($Å^{-1}$) * R) where R is the distance (see below) in Angstrom.

Notation of table above

Symbol	Signification
DFT	B3LYP/6-311++G(3df,3pd) with BSSE
P,P	4-pyridone/4-pyridone dimer
B,Q	benzene/quinone complex
ΔE	interaction energy (energy dimer - energy monomer 1 - energy monomer 2)
elec	electrostatic
Coulomb	Coulomb electrostatic energy calculated with atomic partial charges derived by RESP fitting on the QM electrosatic potential
vdw	van der Waals repulsive term
Distance	closest distance between interacting atoms in the dimer in angstrom; for the B,Q complex, distance between a the benzene H and the quinone O; for the P,P dimer, distance between the H of the NH donor and the O of the O=C acceptor

3.6 Bibliography

1. Truchon, J.-F.; Nicholls, A.; Iftimie, R. I.; Roux, B.; Bayly, C. I. Accurate Molecular Polarizability based on Continuum Electrostatics. *J. Chem. Theory Comput.*, **2008,** *4*, 1480-1493.

2. Maple, J. R.; Cao, Y. X.; Damm, W. G.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A. A polarizable force field and continuum solvation methodology for modeling of protein-ligand interactions. *J. Chem. Theory Comput.*, **2005,** *1*, 694-715.

3. Schnieders, M. J.; Baker, N. A.; Ren, P.; Ponder, J. W. Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum. *J. Chem. Phys.*, **2007,** *126*, 124114.

4. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.*, **1993,** *97*, 10269-10280.

5. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.*, **2000,** *21*, 132-146.

6. Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.*, **2002,** *23*, 1623-1641.

7. Tan, Y. H.; Luo, R. Continuum treatment of electronic polarization effect. *J. Chem. Phys.*, **2007,** *126*, 094103.

8. Becke, A. D. Density-Functional Thermochemistry .3. the Role of Exact Exchange. *J. Chem. Phys.*, **1993,** *98*, 5648-5652.

9. Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.*, **1993,** *98*, 1372-1377.

10. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab-Initio Calculation of Vibrational Absorption and Circular-Dichroism Spectra Using Density-Functional Force-Fields. *J. Phys. Chem.*, **1994,** *98*, 11623-11627.

11. *Gaussian 03, Revision C.02*; Wallingford CT, USA, 2004

12. Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.*, **1984,** *80*, 3265-3269.

13. Simon, S.; Duran, M.; Dannenberg, J. J. How does basis set superposition error change the potential surfaces for hydrogen bonded dimers? *J. Chem. Phys.*, **1996,** *105*, 11024-11031.

14. Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors (Reprinted from Molecular Physics, vol 19, pg 553-566, 1970). *Mol. Phys.*, **2002,** *100*, 65-73.

15. Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Comput. Chem.*, **2001,** *22*, 608-640.

16. Richards, F. M. Areas, Volumes, Packing, and Protein-Structure. *Annu. Rev. Biophys. Bioeng.*, **1977,** *6*, 151-176.

17. Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.*, **1964,** *68*, 441-451.

18. Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.*, **1984,** *5*, 129-145.

19. Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. *J. Chem. Theory Comput.*, **2005,** *1*, 153-168.

20. Elking, D.; Darden, T.; Woods, R. J. Gaussian induced dipole polarization model. *J. Comput. Chem.*, **2007,** *28*, 1261-1274.

21. Connolly, M. L. Solvent-Accessible Surfaces of Proteins and Nucleic-Acids. *Science*, **1983,** *221*, 709-713.

22. Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B*, **2007,** *111*, 2242-2254.

23. Mobley, D. L.; Graves, A. P.; Chodera, J. D.; Shoichet, B. K.; Dill, K. A. Predictive absolute binding free energy calculations for an engineered binding site in T4 Lysozyme. *Biophys. J.*, **2007,** 368A-368A.

24. Williams, D. E.; Abraha, A. Site charge models for molecular electrostatic potentials of cycloalkanes and tetrahedrane. *J. Comput. Chem.*, **1999,** *20*, 579-585.

25. Williams, D. E. Failure of Net Atomic Charge Models to Represent the Van-Der-Waals Envelope Electric-Potential of N-Alkanes. *J. Comput. Chem.*, **1994,** *15*, 719-732.

26. Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. *J. Med. Chem.*, **2008,** *51*, 769-779.

27. The B3LYP/cc-pVTZ ESP based bond charge corrections were obtained through private communications with Cooper, M. D. and Bayly, C. I.

28. Lopes, P. E. M.; Lamoureux, G.; Roux, B.; MacKerell, A. D. Polarizable empirical force field for aromatic compounds based on the classical drude oscillator. *J. Phys. Chem. B*, **2007,** *111*, 2873-2885.

29. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X. X.; Murphy, R. B.; Zhou, R. H.; Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *J. Comput. Chem.*, **2002,** *23*, 1515-1531.

30. Anisimov, V. M.; Vorobyov, I. V.; Lamoureux, G.; Noskov, S.; Roux, B.; MacKerell, A. D. CHARMM all-atom polarizable force field parameter development for nucleic acids. *Biophys. J.*, **2004,** *86*, 415A-415A.

31. Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R. H.; Berne, B. J.; Friesner, R. A. Fluctuating charge, polarizable dipole, and combined models: Parameterization from ab initio quantum chemistry. *J. Phys. Chem. B*, **1999,** *103*, 4730-4737.

32. Jorgensen, W. L.; Severance, D. L. Aromatic-aromatic interactions: free energy profiles for the benzene dimer in water, chloroform, and liquid benzene. *J. Am. Chem. Soc.*, **1990,** *112*, 4768-4774.

33. Caldwell, J. W.; Kollman, P. A. Cation-Pi Interactions - Nonadditive Effects Are Critical in Their Accurate Representation. *J. Am. Chem. Soc.*, **1995,** *117*, 4177-4178.

34. Tsuzuki, S.; Yoshida, M.; Uchimaru, T.; Mikami, M. The Origin of the Cation/pi; Interaction: The Significant Importance of the Induction in Li+ and Na+ Complexes. *J. Phys. Chem. A*, **2001,** *105*, 769-773.

35. Gallivan, J. P.; Dougherty, D. A. Cation-pi interactions in structural biology. *P. Natl. Acad. Sci. U. S. A.*, **1999,** *96*, 9459-9464.

36. Roux, B. Nonadditivity in Cation Peptide Interactions - A Molecular-Dynamics and Ab-Initio Study of Na+ in the Gramicidin Channel. *Chem. Phys. Lett.*, **1993,** *212*, 231-240.

37. Ma, J. C.; Dougherty, D. A. The cation-pi interaction. *Chem. Rev.*, **1997,** *97*, 1303-1324.

38. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J. Am. Chem. Soc.*, **1984,** *106*, 765-784.

39. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A 2Nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.*, **1995,** *117*, 5179-5197.

40. Guo, H.; Gresh, N.; Roques, B. P.; Salahub, D. R. Many-body effects in systems of peptide hydrogen-bonded networks and their contributions to ligand binding: A

comparison of the performances of DFT and polarizable molecular mechanics. *J. Phys. Chem. B*, **2000,** *104*, 9746-9754.

41. Guo, H.; Salahub, D. R. Cooperative hydrogen bonding and enzyme catalysis. *Angew. Chem. Int. Edit.*, **1998,** *37*, 2985-2990.

42. Chen, Y. F.; Dannenberg, J. J. Cooperative 4-pyridone H-bonds with extraordinary stability. A DFT molecular orbital study. *J. Am. Chem. Soc.*, **2006,** *128*, 8100-8101.

43. Harder, E.; Kim, B. C.; Friesner, R. A.; Berne, B. J. Efficient simulation method for polarizable protein force fields: Application to the simulation of BPTI in liquid. *J. Chem. Theory Comput.*, **2005,** *1*, 169-180.

44. Schropp, B.; Tavan, P. The Polarizability of Point-Polarizable Water Models: Density Functional Theory/Molecular Mechanics Results. *J. Phys. Chem. B*, **2008,** *112*, 6233-6240.

45. Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.*, **2002,** *23*, 1244-1253.

46. Prabhu, N. V.; Zhu, P. J.; Sharp, K. A. Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method. *J. Comput. Chem.*, **2004,** *25*, 2049-2064.

47. Im, W.; Beglov, D.; Roux, B. Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.*, **1998,** *111*, 59-75.

48. Lu, Q.; Luo, R. A Poisson--Boltzmann dynamics method with nonperiodic boundary condition. *J. Chem. Phys.*, **2003,** *119*, 11035-11047.

4 Approches intégrées fondées sur un continuum diélectrique pour traiter la polarisabilité moléculaire et la phase condensée: indice de réfraction et la solvatation implicite

Résumé. L'idée d'utiliser un continuum diélectrique à l'intérieur d'une molécule pour modéliser précisément la polarisabilité moléculaire est prolongée pour inclure un plus grand nombre de molécules bioorganiques et la phase condensée. Les rayons atomiques de polarisation et le diélectrique interne (ε_{in}) ont été optimisés pour reproduire des tenseurs de polarisabilité quantique B3LYP/aug-cc-pVTZ provenant d'un ensemble de 707 molécules. Les erreurs moyennes sur la polarisabilité isotropique et sur l'anisotropie sont de 2.6 % et 5.2 % respectivement. Il est montré que les rayons de contact habituels des applications Poisson-Boltzmann et un diélectrique interne faible ne sont pas appropriés et requièrent une révision substantielle. Pour reproduire l'anisotropie de la polarisabilité, la constante diélectrique interne doit être supérieure à 6. En réinterprétant le lien théorique entre ε_{in} et l'indice de réfraction expérimental (n), cette étude montre, avec 23 molécules organiques couvrant toutes les valeurs de n, que même avec ε_{in}=24, les indices de réfraction obtenus démontrent une bonne corrélation avec l'expérience (pente de 1.00, ordonnée à l'origine de 0.05 et R = 0.95). La nouvelle méthodologie utilisée ici pour calculer un indice de réfraction quasi macroscopique montre que l'application d'EPIC à la phase condensée conduit à un comportement souhaitable. Bien que le but premier en développant EPIC était d'inclure la polarisabilité pour des calculs avec solvant explicite, nous élargissons également le modèle pour inclure la polarisabilité pour les calculs avec solvant implicite. Cela demande d'utiliser une fonction diélectrique douce à 3 régions qui permet de passer de la polarisation diélectrique à l'intérieur de la molécule au diélectrique continu du solvant. L'ajustement et la validation de ce modèle sont faits avec 485 énergies libres d'hydratation. Avec 8 rayons atomiques pour décrire la cavité du solvant et une seule tension de surface,

l'erreur positive moyenne est de 1.1 kcal/mol et produit un coefficient de corrélation de 0.9, ce qui valide l'utilisation d'EPIC dans la phase condensée.

Mise en contexte. Les résultats encourageants obtenus aux chapitres 2 et 3 laissent une avenue complètement inexplorée: la phase condensée. Puisque la niche que nous voulons donner à EPIC est précisément pour les simulations en phase condensée, il nous faut aborder la question. Dans ce chapitre, nous examinerons la polarisabilité d'agrégats composés de milliers de molécules et vérifierons qu'ils concordent avec l'indice de réfraction expérimental, une mesure de la polarisation électronique macroscopique. Parce que les méthodes Poisson-Boltzmann dans les applications biomoléculaires prennent de l'ampleur, nous intégrerons notre modèle et nos concepts aux calculs avec solvants implicites qui représentent plusieurs facettes de la phase condensée. Ceci mettra à l'épreuve les charges atomiques, déterminées avec la méthode du chapitre 3, et le modèle polarisable du chapitre 2 puisque la polarisation coopérative du soluté et du solvant nécessite une bonne balance des différentes composantes électrostatiques. Cette intégration demandera l'élaboration d'un nouveau type de fonctions diélectriques.

Pour être en mesure de faire une validation adéquate sur les points mentionnés ci-haut, nous devrons poursuivre un de nos objectifs principaux, c'est-à-dire la généralisation de nos paramètres pour la polarisabilité à une grande partie de la chimie bioorganique. Nous porterons une attention particulière dans ce chapitre à la généralisation des paramètres.

Avec la reproduction de l'indice de réfraction expérimental, nous répondons à la question du sens physique soulevée par une grande constante diélectrique interne. De plus, nous montrons que le potentiel électrostatique permanent (charges atomiques partielles) n'est pas seulement approprié pour le calcul d'interactions électrostatiques entre dimères, mais aussi pour la réponse moyenne du solvant. Nous démontrons également que les rayons atomiques reliés à la polarisabilité et ceux utilisés habituellement pour les calculs en solvants implicites ont des rôles très distincts à jouer.

Integrated Continuum Dielectric Approaches to treat Molecular Polarizability and the Condensed Phase: Refractive Index and Implicit Solvation

Jean-François Truchon[a,b], Anthony Nicholls[c] , Benoît Roux[d], Radu I. Iftimie[a] and Christopher I. Bayly[b]*

[a]Département de chimie, Université de Montréal, C.P. 6128 Succursale centre-ville, Montréal, Québec, Canada H3C 3J7 [b]Merck Frosst Canáda Ltd., 16711 TransCanada Highway, Kirkland, Québec, Canada H9H 3L1 [c]OpenEye Scientific Software, Inc., Santa Fe, New Mexico 87508 [d]Institute of Molecular Pediatric Sciences, Gordon Center for Integrative Science, University of Chicago, Illinois 929 East 57th Street, Chicago, Illinois 60637

Abstract

The idea of using a dielectric continuum inside a molecule to accurately model molecular polarizability is extended to include a larger spectrum of bioorganic molecules and the condensed phase. Atomic polarization radii and an internal dielectric (ε_{in}) were fitted to reproduce ab initio B3LYP/aug-cc-pVTZ polarizability tensors taken from a dataset of 707 molecules. The average unsigned error on the isotropic polarizability and anisotropy are 2.6% and 5.2% respectively. It is shown that usual Poisson-Boltzmann contact radii and a low internal dielectric are not appropriate and require major revision. To account for the anisotropy of polarizability, the internal dielectric (ε_{in}) constant needs to be larger than 6.0. Re-interpreting the theoretical link between ε_{in} and the experimental refractive index (n), this study shows, with a set of 23 organic molecules spanning the entire range of n, that even with ε_{in}=24 the obtained refractive indices can correlate well with experiment (slope of 1.00, intercept of 0.05 and R = 0.95). The novel methodology

used here to calculate a macroscopic-like refractive index shows that the application of the EPIC parameterization to condensed phase leads to suitable behavior. Although the primary goal in developing EPIC was to include polarizability in explicit solvent calculations, we also extend the model to work with implicit solvent. This requires the use of a 3-zone smooth dielectric function to transition from the polarization dielectric inside the molecules to the dielectric continuum of the solvent. The parameterization and validation of this model are performed against 485 experimental free energies of hydration. Using 8 solvent cavity atomic radii and a single surface tension an average unsigned error of 1.1 kcal/mol and a correlation coefficient of 0.9 are obtained, validating the use of the EPIC model in the condensed phase.

4.1 Introduction

The newly introduced treatment of electronic polarization by an internal continuum (EPIC) was shown to be accurate in reproducing experimental and density functional (DFT) molecular polarizability tensors with a remarkably small number of adjustable parameters[1]. Moreover, the high accuracy found when computing intermolecular interaction energies, in which the appropriate treatment of electronic polarization is crucial, opens up the possibility of using EPIC to include polarizability in force fields[2]. This led us to propose the use of EPIC to embed polarizability in all-atom-explicit-solvent calculations. EPIC uses continuum dielectric electrostatic theory to account for the way electronic density polarizes under the presence of an external electric field that can come from either other molecules in explicit condensed phase calculations, or the reaction field in an implicit solvent calculation. In contrast with the point inducible dipoles[3-5] or the Drude's oscillator models[6,7] that use the atomic nuclear positions as polarizable centers, EPIC employs a polarizability density that induces a dipole density, normally referred to as polarization, throughout the molecule volume as a response to the local electric field. In a recent study, Schropp and Tavan[8] proposed that the use of single centers in point inducible dipole polarizable calculations was responsible for the large difference between the best condensed phase atomic polarizability and the best vacuum phase atomic polarizabilities previously noticed[9,10]. Other studies, based on Quantum Mechanical (QM) assessment, suggest that the polarizability in condensed phase should only be slightly reduced[11]. The idea of using a continuum dielectric to account for electronic polarization was first formulated by Sharp *et al.*[12], but was not further pursued until Tan and Luo[13] optimized the internal dielectric of solutes to produce the electrostatic potential in the context of Poisson-Boltzmann calculations with different implicit solvents. In their two studies[13,14], they do not attempt to give a detailed molecular polarizability description, but rather focus on the shift in dipole moments when a solute is put in different solvent. It is difficult to decouple the solvent polarization from the solute polarization and the cooperative polarization when calculations

in implicit solvent are done. The current study uses previously developed techniques[1,2] to separate the charge fitting from the polarizability fitting by optimizing separately an *electronic volume* on quantum mechanics (QM) polarizability tensors for molecules in vacuum, as was done originally with other polarizable models[3,5,15]. Curiously, we found that in order to accurately reproduce the polarizability tensors of even challenging molecules, the atomic radii needed to be much smaller than the van der Waals (vdW) contact radii usually used in implicit solvent calculations (e.g. Bondi radii[16]). At the same time, the internal dielectric needed to be surprisingly high in order to reproduce the anisotropy of the polarizabilities. While that work allowed for a systematic way of adjusting a dielectric function to account for electronic polarization, it raised two issues: the abnormally high internal dielectric of 14 seems questionable and the small radii made implicit solvent calculations impractical. Regarding the first issue, the dielectric inside the molecule is closely related to the refractive index squared ($\varepsilon_\infty=n^2$) of the pure liquid, which adopts values between 1.7 and 2.9 for organic liquids, far below our large values. Regarding the second issue, if such small atomic radii were used to define the molecular cavity in solvent, the free energy of charging would become unrealistically negative e.g. in Poisson-Boltzmann (PB) calculations. In this work, we specifically address both issues and demonstrate the physical soundness of the approach. An important change from our previous work is the use of a smooth dielectric boundary to represent both the solute and the solvent polarization. We present a newly designed dielectric function with 3 zones (3-zone dielectric) that permits the use of EPIC for implicit solvent calculations. We show that describing the dielectric function this way better reflects the underlying physical principles involved in solvation than the usual 2-zone dielectric (i.e. inside and outside the cavity).

Another question that we examine is the ability to optimize the EPIC parameters in a general and robust way with few parameters on a larger variety of chemical functionality than in earlier work. For this purpose, we have formed a large database of QM molecular polarizability tensors for 707 diverse bioorganic molecules (or a total of 4242 polarizabilities) along with their optimized molecular geometries (c.f. Annexe V). As will

be outlined below, this dataset contains a large variety of chemical functional groups representing a significant component of bioorganic chemistry. This substantially enlarged parameterization of the polarizable EPIC model is then used for the calculation of refractive indices and hydration free energies. The validity of both the internal dielectric function and the 3-zone dielectric function is assessed with the independent fit of the solvent cavity atomic radii (which define the third zone of the function) on 485 experimental free energies of hydration.

In the remainder of this article, section 2 presents the theoretical basis and methods employed, where we present the 3-zone dielectric function for implicit solvent calculations and we review the polarizability tensor calculation. This is followed by the theoretical background for the calculation of the refractive index. A theoretical layout for free energy of hydration calculations and computational details related to quantum calculations close this section. Section 3 describes the chemical datasets used in section 4 where the results and their analysis are presented. Section 4 closes with a 3rd zone dielectric optimization on experimental hydration free energies, leading into the conclusons.

4.2 Theory and Methods

4.2.1 3-Zone dielectric in implicit solvents

The dielectric function in continuum approaches is fundamental as it is modulating all sources of polarization. In this work, we move away from our previous use of vdW envelope surfaces[17] toward a smooth functional form based on a sum of atomic Gaussians which has previously proven successful[18,19] in PB applications. Although useful, the hard dielectric boundary often leads to numerical problems: iterative convergence failure, slower convergence, strong dependency on orientation and translation, and unstable force evaluations[18,20]. The use of smooth solute/solvent dielectric boundary was shown to

improve over the hard boundary on all these aspects. More specifically, the molecular dielectric function used in the present work is given by

$$\varepsilon(\vec{r}) = \varepsilon_{in} - (\varepsilon_{in} - \varepsilon_{ext}) \exp(-A \cdot f_{in}(\vec{r})) \tag{4.1}$$

where ε_{in} is the dielectric constant inside the molecular volume and ε_{ext} the dielectric value outside. The dielectric here is expressed as a permittivity relative to the vacuum permittivity. The exponential behaves as a switching function that is turned on or off depending on the value of a molecular 'density' function $f_{in}(\vec{r})$. The A parameter modulates the steepness of the switching function. The details of the dielectric are then incorporated into the 'density' function

$$f_{in}(\vec{r}) = \sum_{i=1}^{Natoms} p \cdot \exp\left(-k \frac{\left|\vec{r}_i - \vec{r}\right|^2}{\sigma_i^2} \right) \tag{4.2}$$

The summation runs over all atoms and a 3-dimensional Gaussian defines the radial extent of the atomic volume; σ_i are atomic radii and r_i their positions. The σ_i will be the subject of an extensive parameterization in the next sections. The constant k is set to 2.3442 and p to 2.7 following the Grant *et al.* recommendation[18]. Eq. (4.1) can be conceptually understood in terms of electronic density that would have a constant susceptibility (polarizability density) inside and drops rapidly as the density vanishes as shown in Figure 4.1.

The main methodological novelty proposed in this work is the 3-zone dielectric for the coupling of EPIC with implicit solvation. When atomic radii are optimized on QM-based molecular polarizability tensors, their resulting small size prevents their use to define the cavity formed by the solute in implicit solvent calculations. Indeed, it presents a dilemma: on the one hand, accurate solute polarization requires atomic radii far smaller than accepted contact radii. On the other hand, the solvent boundary for implicit solvation

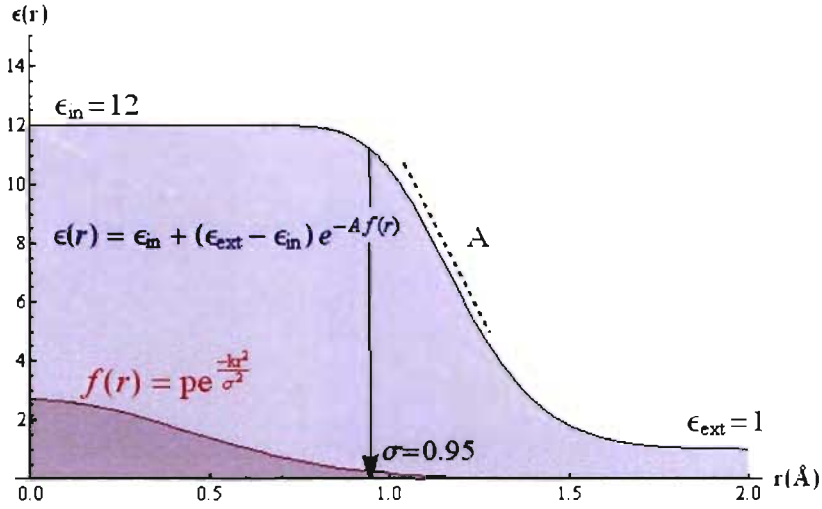


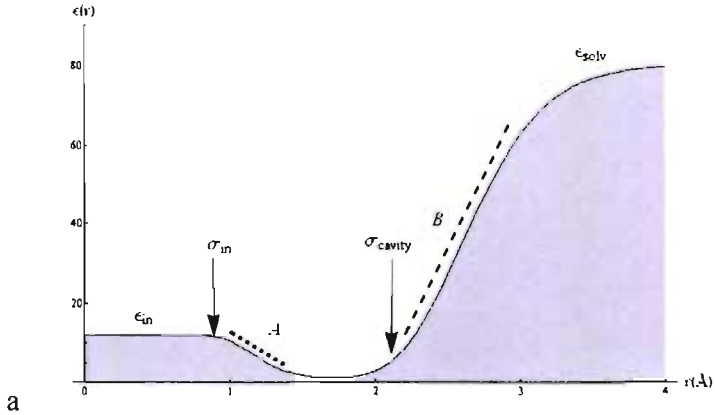
Figure 4.1. This figure shows the smooth dielectric function used in this work for a single atom with σ = 0.95Å, ε_{in} = 12, ε_{ext} = 1 and A = 10.0 (Cl of the G1-12 set). Starting from the center of the atom (r = 0), the dielectric (blue curve) stays constant until the 'density', expressed with a sum of Gaussians (pink curve), reaches a certain small value that causes the dielectric to smoothly transition to the external dielectric value. The steepness and the position of the switching region depends on the value of the A parameter. The sum-of-Gaussians density expression is explained in equation 2 (see text).

requires atomic radii as large or larger than contact radii. The resolution to this dilemma is found in challenging the assumption that the atomic radii for solute polarization and for the solvent boundary should be the same. There is no underlying physical reason why the polarization response of an atom in a molecule would be uniform all the way out to its contact radius; on the contrary our QM model for molecules tells us the electron density (the source of electronic polarization) drops exponentially in moving from an atomic nucleus towards the contact surface of the molecule. We believe that it is more reasonable to think that the radial extent of the electronic polarization can be different from the vdW radius used for the solvent cavity. The idea presented here is that both kinds of smooth surfaces could be simultaneously used: one for solute polarization, formed with the smaller atomic polarization radii, and one for solvent polarization, defined with the solvent cavity atomic radii. In between the two surfaces is a transition region of low dielectric since it

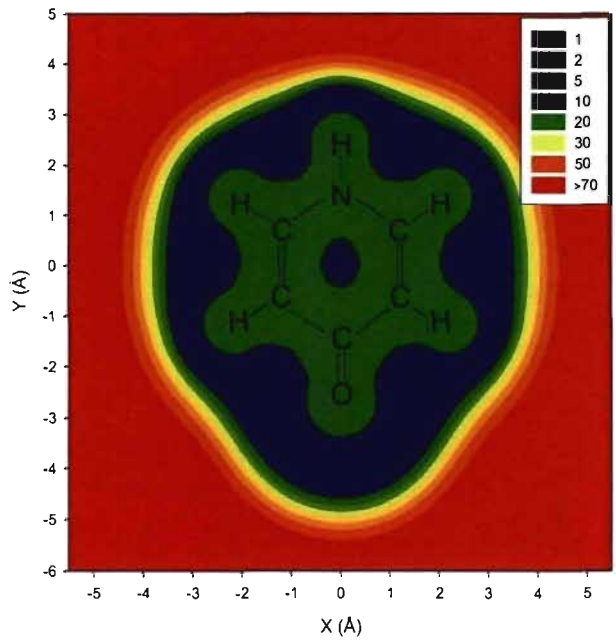
describes where the solvent and the solute electrons are both at a minimum. This leads to a 3-zone dielectric function to which we give the form

$$\varepsilon(\vec{r}) = \varepsilon_{in} + (\varepsilon_{trans} - \varepsilon_{in})\exp\left[-A \cdot f_{in}(\vec{r})\right] + (\varepsilon_{solv} - \varepsilon_{trans})\exp\left[-B \cdot f_{solv}(\vec{r})\right] \tag{4.3}$$

where ε_{in} is the dielectric constant inside the molecular cavity, ε_{solv} the bulk solvent dielectric constant (80 for water), and ε_{trans} the dielectric constant in the zone of transition between the solute and the solvent. For the smooth inner dielectric boundary, A has the same meaning as in eq. (4.1) and $f_{in}(\vec{r})$ is given by eq. (4.2). The additional exponential term, for the outer dielectric boundary (with solvent), is a switching function that turns on when a second Gaussian sum ($f_{solv}(\vec{r})$) becomes sufficiently small. The $f_{solv}(\vec{r})$ term is also given by eq. (4.2) with the difference that the atomic radii are larger as they define the solvent cavity. The B parameter is responsible for the steepness of the cavity boundary, but with a sufficiently large value it has the effect of moving the position of the boundary as if the radii were scaled. The radial behaviour of the 3-zone dielectric is illustrated in Figure 4.2 for a single atom (a) and for the 4-pyridone molecule (b), both with typical parameters. In eq.(4.3), it is important to set $\varepsilon_{trans} = 1$ when the first zone of the dielectric function is fitted on molecular polarizability since the shape of the dielectric function needs to drops to one in order to present the same ability to polarize. Also, if the atomic partial charges are fitted with DRESP, a change in the first zone boundary would also change the ability of the dielectric to form the full internal polarization taken into account during the charge fitting process.



a



b

Figure 4.2. The 3-zone dielectric function allows an accurate description of both the solute polarization and the solvent polarization within the EPIC approach. (a) The radial component of the dielectric for a single atom (G1-12 aromatic carbon) is shown together with the polarization (σ_{in}) and the solvent cavity (σ_{cavity}) atomic radii. Each plateau of the dielectric function defines a zone. The intermediate zone corresponds to the solute/solvent contact distance. (b) The resulting dielectric function is also shown in the ring plane of 4-pyridone (b) when applying the G2-12 parameters.

4.2.2 Molecular polarizability tensor

In this section, we review the methodology previously developed to calculate molecular polarizability tensor with a finite difference Poisson solver[1] and we summarize how the parameters involved are optimized in this work.

4.2.2.1 Method

Our formulation of electronic polarization based on continuum electrostatics allows the calculation of induced multipolar moments by considering the bound charge density, which results from the polarizability density of the media (from the *bound* electrons in our case). A formula to calculate the bound charge density is[21]

$$\begin{aligned}\frac{\rho^b(\vec{r})}{\varepsilon_0} &= -\vec{\nabla}\left(\left[\varepsilon(\vec{r})-1\right]\vec{E}(\vec{r})\right) \\ &= -\vec{\nabla}\cdot\vec{P}(\vec{r})\end{aligned} \tag{4.4}$$

where ρ^b is the bound charge density and $\vec{E}(\vec{r})$ the total electric field. Physically, ρ^b is a consequence of the formation of dipoles at each point in space due to the electric field (the polarization $\vec{P}(\vec{r})$ or dipole density). The bound charge density can be thought of as an induced charge density from the dielectric polarization that appears where the dielectric varies, as an excess of charge builds due to the head or tail of the dipole density. Although the polarization occurs everywhere the dielectric is greater than one, the bound charge density appears in regions of spaces where $\varepsilon(\vec{r})$ varies, such as the dielectric boundary of a molecule. Eq (4.4) is useful since it transforms the locally induced dipoles into a scalar value, the bound charge density, which can be used more easily as done below. In eq.(4.4), $\varepsilon(\vec{r})-1$ plays the role of a local polarizability density, also called the electric susceptibility, and $\vec{P}(\vec{r})=(\varepsilon(\vec{r})-1)\vec{E}(\vec{r})$ corresponds to the induced dipole density (polarization). The analogy with the point inducible dipole model, a different polarizable model, is obvious since, in that case, the atomic induced dipole is given by $\vec{\mu}(\vec{r}_i)=\alpha_i\vec{E}(\vec{r}_i)$ where $\vec{\mu}(\vec{r}_i)$, α_i and

$\vec{E}(\vec{r}_i)$ are the dipole induced at the atomic position $\vec{r}_i$, the atomic polarizability and the electric field at $\vec{r}_i$. Here, the polarization is more smoothly distributed over the molecular volume. Eq. (4.4) is intrinsic to the definition of Poisson's equation.

A classical example, for which an analytical solution exists, is the dielectric sphere in vacuum experiencing an external electric field. In this case the mathematics show that bound charges appear on the surface of the sphere with opposite charge sign on both hemispheres, resulting in an induced potential equivalent to an ideal induced dipole moment aligned with the external field located at the center of the sphere. The induced dipole moment is proportional to the external electric field and the sphere polarizability α_{sphere} is given by the Clausius-Mossoti equation

$$\alpha_{sphere} = \left(\frac{\varepsilon_{sphere} - 1}{\varepsilon_{sphere} + 2} \right) R_{sphere}^3 \tag{4.5}$$

where R_{sphere} is the sphere radius. For a molecular system, the analytical solution is unknown and we use a finite difference algorithm to solve Poisson's equation numerically with a uniform electric field in the form of a voltage clamp applied by means of the boundary conditions. More precisely, a uniform electric field in the z direction can be produced with a null potential on one side of the grid boundary and the value $-E_{ext} \times Lz$ on the opposite side, where Lz is the box size in the z direction and Eext the magnitude of the applied field. On the four other sides, parallel to the field, the grid boundary potential is simply calculated as a linear interpolation along the z direction: $\varphi(z\text{-}z0) = -(z\text{-}z0) \times E_{ext}$. As with the dielectric sphere in vacuum, a molecular dielectric volume responds linearly to the applied field (given an isotropic dielectric function) and the proportionality constant is the molecular polarizability tensor. The field is applied in three orthogonal directions to build the polarizability tensor, which depends on the orientation of the molecule:

$$\bar{\alpha} = \begin{bmatrix} \frac{\mu_{x,x}}{E_{ext}} & \frac{\mu_{x,y}+\mu_{y,x}}{2E_{ext}} & \frac{\mu_{x,z}+\mu_{z,x}}{2E_{ext}} \\ & \frac{\mu_{y,y}}{E_{ext}} & \frac{\mu_{y,z}+\mu_{z,y}}{2E_{ext}} \\ & & \frac{\mu_{z,z}}{E_{ext}} \end{bmatrix} \tag{4.6}$$

where $\mu_{x,y}$ is the x component of the induced dipole moment when an external electric field of magnitude E_{ext} is applied in the y direction. Some experimental values are available for the eigenvalues of this tensor in vacuum (ε_{ext}= 1); also, the polarizability tensor can be calculated using approaches based on quantum mechanics (QM) methods such as density functional theory.

The induced dipole moment is calculated analogously to the sphere dielectric system, integrating the bound charge density over space. From eq. (4.4) (or simply from Gauss's law), one can show that

$$\rho^b(\vec{r}) = -\rho^f(\vec{r}) + \varepsilon_0 \vec{\nabla} \cdot \vec{E}(\vec{r}) \tag{4.7}$$

In the present context, there is no free charge density $\rho^f(\vec{r})$ (from atomic partial charges, for instance) and as such the bound charge density, induced only by the external uniform electric field, is given by the divergence of the field. With a finite difference solver, the total charge (bound and free charges) can be calculated by integrating over each differential volume element (grid cube) which leads to bound charges on grid points. This can be done simply by calculating

$$\begin{aligned} \frac{q_{ijk}}{\varepsilon_0} &= \frac{q^b_{ijk} + q^f_{ijk}}{\varepsilon_0} \\ &= -\left(\frac{h_y h_z}{h_x}\right)\left(\varphi_{i+1jk} + \varphi_{i-1jk} - 2\varphi_{ijk}\right) - \left(\frac{h_x h_z}{h_y}\right)\left(\varphi_{ij+1k} + \varphi_{ij-1k} - 2\varphi_{ijk}\right) - \left(\frac{h_x h_y}{h_z}\right)\left(\varphi_{ijk+1} + \varphi_{ijk-1} - 2\varphi_{ijk}\right) \end{aligned} \tag{4.8}$$

where q_{ijk}, q^b_{ijk} and q^f_{ijk} are the total charge, the bound charge and the free charge inside the volume element associated with the *ijk* grid point, φ_{ijk} and φ_{ijk-1} the electrostatic potential at

the (x,y,z) and $(x,y,z\text{-}dz)$ grid points respectively. The grid spacing in x, y and z are given by h_x, h_x and h_z. The grid free charge q^f_{ijk} are zero for this calculation and, in general, it is given by the atomic partial charges as distributed on the grid. Finally, the total dipole moment is given by

$$\vec{\mu} = \sum_{i,j,k}^{Grid} \vec{r}_{ijk}\, q_{ijk} \tag{4.9}$$

With the free charges equal to zero (no atomic partial charge), the dipole calculated is then the induced dipole and the only contributor is the bound charge density. More generally, any molecular electric moment can be calculated with analogs to eq. (4.9). The overall procedure to calculate the polarizability tensor requires three solutions from the numerical solver. The calculation does not involve atomic partial charges (free charges) which allows them to be fit independently (although this must still be done in the context of the molecular dielectric).

4.2.2.2 Computational details

The finite difference Poisson calculations were performed with a modified version of the OpenEye Inc. ZapTK[22]. The distance between two grid points was set to 0.35 Å and the grid boundary was at least 5 Å away from the surface defined by the polarization radii. Atomic charges of $\pm 0.001e$ were assigned randomly on the atoms as the grid energy was used to determine the convergence of the algorithm set to 0.000001 k_BT. The results were not sensitive to these small charges. Atom typing was assigned via SMARTS[23-25] with the OpenEye Inc. OEchem toolkit[26].

4.2.2.3 Optimization of the polarizabilities

The atomic radii were optimized in order to minimize a chi-square function using a Levenberg-Marquardt algorithm as implemented in scipy[27], a scientific Python library. The error was defined as the difference between the 6 components of the polarizability tensor obtained with B3LYP and EPIC

$$\chi^2 = \sum_{i}^{molecules} \sum_{\substack{k=xx,xy,xz \\ yy,yz,zz}} (\alpha_{k,i}^{EPIC} - \alpha_{k,i}^{QM})^2 \tag{4.10}$$

where $\alpha_{xy,i}$ is one of the six-independent polarizability tensor elements of molecule i either under optimization (EPIC) or from the QM target values. By using the six independent tensor elements, we included both the magnitude and the direction of the polarizability in a natural way[28]. We optimized the cube of the polarization radii because their contribution to the polarizability grows with the atomic volume (c.f. eq. (4.5)). For analysis purposes, we also defined the average polarizability (eq. (4.11)) and the anisotropy of the polarizability tensor (eq. (4.12)) below

$$\alpha_{avg} = (\alpha_1 + \alpha_2 + \alpha_3)/3 \tag{4.11}$$

$$\Delta\alpha = \sqrt{\frac{(\alpha_1 - \alpha_2)^2 + (\alpha_1 - \alpha_3)^2 + (\alpha_2 - \alpha_3)^2}{2}} \tag{4.12}$$

where $\alpha_1 \leq \alpha_2 \leq \alpha_3$ are the eigenvalues of the polarizability tensor. The polarizability anisotropy is significantly harder to fit than the average polarizability. We defined the error in the average polarizability (eq. (4.13)) and anisotropy (eq. (4.14)) for a set of molecules as

$$\delta_{avg} = \frac{1}{N} \sum_{i}^{N} \frac{\left| \alpha_{i,avg}^{QM} - \alpha_{i,avg} \right|}{\alpha_{i,avg}^{QM}} \tag{4.13}$$

$$\delta_{aniso} = \frac{1}{N} \sum_{i}^{N} \frac{\left| \Delta\alpha_i^{QM} - \Delta\alpha_i \right|}{\alpha_{i,avg}^{QM}} \tag{4.14}$$

where N is the total number of molecules considered and QM corresponds to the target value. Finally, the relative root-mean-square deviation (RRMS) of the tensor was defined as

$$RRMS = \frac{\sum_{i}^{molecules} \sum_{\substack{k=xx,xy,xz, \\ yy,yz,zz}} (\alpha_{k,i}^{EPIC} - \alpha_{k,i}^{QM})^2}{\sum_{i}^{molecules} \sum_{\substack{k=xx,xy,xz, \\ yy,yz,zz}} \left(\alpha_{k,i}^{QM}\right)^2} \tag{4.15}$$

and constituted a single metric for the overall fitness of the optimized polarizability tensors. If the RRMS was calculated for a single molecule, the summations on the molecules in the numerator and the denominator were simply omitted.

4.2.3 Refractive index calculations

4.2.3.1 Theory

The dielectric constant of an isotropic material at the high frequency limit (ε_∞) is related to the material refractive index[29] n by

$$n^2 = \varepsilon_\infty \tag{4.16}$$

where n is usually measured with the D line of the sodium spectrum at 589 nm (n_D). The ε_∞ corresponds to the material's dielectric constant solely due to the electronic polarization since the frequency of the visible light is too high for nuclear relaxation to contribute. Typically, a pure liquid of an organic compound will have a refractive index between 1.3 and 1.7 leading to a ε_∞ between 1.7 and 2.9. Since the work of Debye and Onsager[12,13,30,31], it has become a dogma that the interior dielectric (ε_{in}) of a solute cavity in implicit solvent models should be close to the experimental ε_∞ in order to capture the dipole moment change due to the cooperative solute-solvent polarization. It is when we seek for accuracy in solute polarization that we found the generally accepted relation $\varepsilon_\infty = \varepsilon_{in}$ to badly fail[1]. A way to reconcile this puzzling finding is by computing the macroscopic refractive index that corresponds to what is measured instead of assuming it is the same as the *internal refractive index* (quoting Onsager[30]). The Clausius-Mossoti equation relates the polarizability of a sphere to its interior dielectric. Since ε_∞ and n are macroscopic intensive quantities, their measurement should not depend on the size of the studied sample, given that it is large enough to exhibit a macroscopic behavior, the worst case being the use of a single molecule. It is not to say that Onsager's use of the Clausius-Mossoti equation with

the radius of a single molecule was not justified. In fact, he was primarily interested in the molecular polarizability (α_{mol}) and used the formula

$$\alpha_{mol} = \left(\frac{n_D^2 - 1}{n_D^2 + 2}\right)\frac{3v}{4\pi n} \tag{4.17}$$

where v is the volume of the liquid sphere considered and n the number of molecules it contains. In eq. (4.17), the rightmost factor corresponds to the cube of an effective single spherical molecule radius. It is however understood that the same molecular polarizability is obtained as long as the v/n factor is preserved and is therefore size independent with the key assumption that ε_∞ is filling the space uniformly, i.e. that it is a spatially averaged value. In order to calculate the refractive indices for the general case where the internal dielectric is not uniformly distributed in the liquid, we generated pure liquid configurations from molecular dynamics (MD) simulations at room temperature and cut out spherical clusters (or droplets) from individual snapshots. We maintained the v/n ratio by fixing the density to experiment and calculating the droplet effective ε_{in} with the formula

$$n^2 = \frac{R_{droplet}^3 + 2\alpha_{droplet}}{R_{droplet}^3 - \alpha_{droplet}} \tag{4.18}$$

here $R_{droplet}$ and $\alpha_{droplet}$ are the droplet radius and polarizability. We assigned the dielectric function on all molecules and applied the procedure outlined above to calculate the droplet polarizability and thereby access the droplet refractive index. Averaging the droplet refractive index over many droplets yields an approximation of the bulk refractive index.

4.2.3.2 Computational details

To obtain the liquid phase droplets, molecular dynamic simulations, using the AMBER 8.0 package, were performed on 3375 molecules (15x15x15) in a cubic box. The NVT ensemble and periodic boundary conditions allowed the density to be fixed to the experimental value. The temperature was set to 20°C to match the experimental conditions used to report refractive indices and maintained constant with the Berendsen's weak

coupling algorithm[32] with the kinetic energy adjusted every 1 ps. The short-range non-bonded interaction cutoff was set to 8.0 Å and long range interactions computed with particle mesh Ewald[14,33] using the default Amber 8.0 setup. The molecules were charged with AM1-BCC[34,35] and the Generalized Amber Force Field (GAFF)[36] was used. The SHAKE procedure[37] was used to fix all bond lengths to hydrogen.

The initial liquid box was generated by positioning the molecules on a cubic lattice, randomly oriented with the Marsaglia[38] quaternions method. The system was first minimized until the root-mean-square (RMS) of the gradient is less than 0.1 kcal/mol/Å. This was followed by a 8 ps annealing phase integrated by steps of 1 fs, during which the non-bonding interactions were gradually turned on and the temperature increased from 0 K to 40 K and decreased to 0 K. The system was then heated over 20 ps up to 293.15 K with a 2 fs integration time step. Following a 1 ns equilibration, 50 evenly-spaced snapshots were written over a 2 ns production run. Each of the liquid boxes for a given molecule was then wrapped in the primary cell. A sphere with a diameter set to 85% of the box length formed a liquid droplet when picking all molecules with an atom lying inside the sphere. The droplet radius was then determined by considering the position of the outermost non-hydrogen atoms. The precise definition of the radius is not unique and we have verified, for example, that using the experimental density to calculate the radius of the corresponding ideal sphere gives refractive indices within ±0.01 of those obtained by the chosen algorithm. Also, this model assumes a perfectly spherical object, ignoring the dimples formed because of the finite size of the spheres. The relatively large size of the droplet and the averaging over 50 independent configurations reduced the effect of this approximation.

The solution to Poisson's equation in the presence of the voltage clamp boundary conditions was obtained on a rectangular grid sized to encompass the full droplet plus half its radius on each side of the droplet. The target grid spacing was set to be 0.5Å. The smooth dielectric functions (eq. (4.1)), fitted on the molecular polarizability tensors only, were assigned together with the matching atomic polarization radii, internal dielectric ε_{in}

and A parameter. The external dielectric was always set to the vacuum value $\varepsilon_{ext} = 1$. The convergence criteria for the ZapTk solver was based on the grid energy and set to 0.0001 kT. This convergence criterion required the assignement of atomic charges that we choose to be $\pm 0.001e$ were randomly placed on half the atoms, keeping an overall neutral system. Given the strength of the external field applied, this was not perceptibly affecting the answer.

4.2.4 Free energy of hydration

4.2.4.1 Theory

Implicit solvent models are commonly used to incorporate the effects of solvation in molecular models as a mean field[39-43]. These models considerably reduce the computational burden needed to sample the solvent configurational space when each atom of the solvent are explicitly simulated. An important validation for solvation models comes from the experimental free energy of hydration (ΔG_{hyd}) that consists in the chemical potential difference for the transfer of a solute from vacuum to bulk solvent. The computational evaluation of ΔG_{hyd} is separated into two processes. Firstly, the non-polar free energy of hydration (ΔG_{np}) comes from the formation of the solute-shaped cavity in the bulk solvent that causes a reorganization of the solvent molecules, and non-polar interactions between the solute and the solvent. Secondly, the electrostatic free energy of hydration (ΔG_{elec}) results from the electrostatic work necessary to place the solute charge density in the solute cavity, involving interactions between solute and solvent charge densities and their response to one another. This results in the equation

$$\Delta G_{hyd} = \Delta G_{elec} + \Delta G_{np} \tag{4.19}$$

The longstanding use of implicit solvent to evaluate ΔG_{elec} is based on a high continuum dielectric solvent region that gets polarized by a static solute electric field. While the solute cavity is traditionally formed with a molecular surface with a discrete transition of the

dielectric function at the solute-solvent boundary, we chose a smooth boundary transition as explained earlier. The solute cavity volume and shape is determined by atomic radii. For a given set of charges, atomic radii that are too small exaggerate the affinity of the solute for water, while radii that are too large will have the opposite effect. The calculation of ΔG_{elec} is normally done with a non-polarizable solute or, if the cavity is assigned a $\varepsilon_{in} > 1$, the very significant screening of the atomic partial charges requires a special treatment that was not done until recently[2]. For non-polarizable solutes, knowing that water increases the dipole moment of solvated molecules often by as much as 15%, the atomic charges should not be fit on a gas phase QM ESP. For this reason, the charges are often generated from RESP[44] or AM1-BCC[34,35] that are known to be sufficiently overpolarized compared to the gas phase.

In the 3-zone dielectric model that we propose in this article (c.f. Figure 4.2a and eq.(4.3)), the first zone should accurately account for the solute polarizability, which allows for the use of vacuum phase atomic charges obtained independently of the internal dielectric function. The second zone located between the internal dielectric and the solvent is set to vacuum and the transition to the full implicit solvent model of the third zone needs to be parameterized. Following the suggestion of Grant *et al.*[18] for their non-polarizable 2-zone dielectric function, we fixed the B parameter in eq. to 11.8, which leaves the solvent cavity atomic radii to be fitted on the experimental free energy of hydration. However, in order to compare the calculated ΔG_{hyd} to experiment, we needed to use existing values or methods for ΔG_{np}. Fortunately, converged molecular dynamics free energy[45,46] calculations based on free energy perturbation (FEP) calculations are available for each compound from our hydration free energy dataset. We feel this is the best achievable theoretical estimation of ΔG_{np}, so this is our preferred estimation in current study. However, since this is not very useful for prospective evaluations of ΔG_{hyd}, due to the heavy computational demands for such FEP calculations, we also tested a surface area based model that calculates ΔG_{np} as

$$\Delta G_{np} = \gamma \times S \tag{4.20}$$

where γ is a surface tension and S the surface area of the molecule as defined by a solvent accessible surface[17] created with a 1.4 Å rolling probe and the Bondi radii[16]. This crude approximation has been proven useful and it can be improved by atom typing the γ[47] or by using some treatment of the dispersion energy[48-51] instead; in this work a single value of γ was fitted for each model.

4.2.4.2 Computational details

The atomic partial charges responsible for the permanent electrostatic potential (ESP) were determined by a least-squares-fit on the QM ESP calculated on a face-centered-cubic grid of points. Following Jakalian *et al*[34], the grid spacing was set to 0.5 Å and the grid points were positioned around the molecule in a volume formed by two vdW surfaces, each built with Bondi radii scaled by a factor of 1.4 and 2.0. The dielectric has the effect of scaling down by a factor of $1/\varepsilon_{in}$ the charges; this is partly compensated by the bound charges appearing from the internal polarization. Hence, the least-squares-fit requires a Poisson solver in order to capture the overall effect, which depends on the shape of the dielectric boundary. It is noteworthy that the charge fitting process is independent of the EPIC polarizability model and, as such, can be fit after the solute dielectric parameters are optimized. The details of the procedure, called DRESP, can be found elsewhere[2].

A finite difference Poisson solver was written to allow the implementation of the 3-zone dielectric model. Here is a brief description of the algorithms implemented. We use successive over-relaxation (SOR) and a Gauss-Seidel iterative scheme[52,53] where the over-relaxation parameter w is estimated by

$$w = \frac{2}{1+\sqrt{1-\lambda_{\max}^2}}$$
$$\lambda_{\max} = 1 - \frac{\pi^2}{2(n-1)^2} \qquad (4.21)$$

where n the number of grid points in one of the dimension of the grid[52]. This crude estimate of the spectral radius of the A matrix in the finite difference form of the Poisson's equation used (see Appendix of ref. 18), was sufficient to reduce by a factor of approximately 30 the number of Gauss-Seidel steps necessary.

The free charges of the system were assigned on the grid with a quadratic inverse interpolation scheme[18] that has the advantage of conserving the dipole moment, has a continuous first derivative and is more robust to the effects of rotation and translation. The same interpolation rule is used to calculate the potential in between grid points. In our calculations, we use a convergence criteria base on grid energy defined as the sum of the electrostatic potential times the distributed free charges on the grid. This convenient criterion is directly related to the energy in an absolute way and thus ensures that relative energies are also converged. The boundary conditions, in energy calculations, were determined with a Coulomb potential.

The ΔG_{elec} was computed by taking the grid charge energy difference between a solution obtained in vacuum ($\varepsilon_{ext} = 1$) and another solution in water ($\varepsilon_{ext} = 80$) from the resulting Poisson's equation and calculated with

$$\Delta G_{elec} = \frac{1}{2} \sum_{i}^{Atoms} q_i \left(\varphi(\vec{r}_i)^{water} - \varphi(\vec{r}_i)^{vacuum} \right) \tag{4.22}$$

where q_i is the atomic partial charge of atom i, $\varphi(\vec{r}_i)^{vacuum}$ is the interpolated electrostatic potential at atom i position $\vec{r}_i$. The grid spacing for the solver was set to 0.35 Å and the minimum distance between the solute internal radii and the grid boundary was set to 7 Å. In those cases where the solute was non-polarizable, ε_{in} was set to one. Finally, the parameters (solvent cavity atomic radii and surface tension) were adjusted with the same Levenberg-Marquardt algorithm used for the fit to the polarizability tensor. All parameters were simultaneously optimized.

4.2.5 Quantum calculations

The B3LYP exchange-correlation functional[54,55] was used for all DFT quantum calculations of this work within the Gaussian 03 software[56]. All molecular structures of this work were initially relaxed with B3LYP and the 6-31G(d,p) basis set[57-59]. Property calculations required larger basis sets for accuracy. The electrostatic potential values were obtained with B3LYP and the 6-311++G(3df,3pd) extended triple zeta basis set[57-59]. The molecular polarizability tensor computations used the aug-cc-pVTZ basis set[60], as it was shown to lead to accurate results[61]. The implemented method in Gaussian 03 to calculate the molecular polarizability tensor is the Coupled Perturbed Hartree Fock (CPHF) method[62]. The Hartree-Fock calculations performed to fit water-adapted atomic partial charges were also performed with the Gaussian 03 software with the 6-31G(d,p) basis set.

4.3 Datasets

In this work, we made extensive use of three kinds of data: B3LYP/aug-cc-pVTZ polarizability tensors, free energies of hydration and refractive indices. A total of five datasets were created.

4.3.1 Polarizability training dataset (PTD)

A training dataset was used to optimize the internal radius in order to match B3LYP polarizability tensors. To this end, we made use of the previously published training datasets[1] and added new molecules for a total of 265 polarizability tensors. In this dataset, many neutral functional groups are represented: alkanes, alkenes, alkynes, halogens (bromo, fluoro, chloro), alcohols, thiols, amines, ethers, thioether, nitriles, aldehydes, ketones, esters, thioesthers, amides, acids, ureas, imines, amidines, sulfones, sulfoxides, sulfonamides, heteroaromatics, hydrazines, hydroxamic acids, N-oxides, pyridones and peptides. In addition, charged functional groups were also included with the sole purpose of

covering charged side chains in amino acids. They were: carboxylates, guanidiniums, imidazoliums and ammoniums. The strength of the polarizability training dataset is in the wide coverage of functional groups, but its clear weakness is the lack of polyfunctional molecules. To get to this level of coverage would require calculations on a great many more larger molecules, and consequently an enormous amount of computational power. The intention in this paper is to assess whether a small and reasonably general first set of parameters can adequately treat a wide variety of bioorganic small molecules in addition to most biomolecules.

4.3.2 Polarizability validation dataset

The polarization validation dataset is composed of the previously published validation sets[1] and 401 molecules from the hydration free energy dataset (below) not included in the polarizability training dataset. In addition, a few special molecules such as neutral and charged peptides, melamine, sugars, etc. were added, giving a total of 442 datapoints.

4.3.3 Polarizability dataset

The polarizability dataset is the combination of the validation and training datasets, making available all 707 polarizability tensors together with the molecule coordinates (see Supporting Information).

4.3.4 Hydration free Energy Dataset

This dataset is built from a compilation of 504 experimental free energies of hydration of neutral molecules recently published with the corresponding ΔG_{np} and ΔG_{chg} from Molecular Dynamics based absolute free energy calculations[45]. We took the published dataset, eliminated the iodine- and phosphorus-containing compounds and formed a dataset

of 485 molecules on which we could fit the solvent part of the dielectric function (eq.(4.3)) and the surface tension coefficient (γ).

4.3.5 Refractive indices dataset

The refractive indices dataset contains 23 small organic molecules (c.f. Figure 4.5) that are liquids at 20°C, for which the density and the refractive indices are taken from the CRC Handbook of Chemistry and Physics[63]. They span a variety of functional groups and most of the entire spectrum of refractive indices measured for bioorganic molecules.

4.4 Results and discussion

4.4.1 Polarizability tensor

This work follows the precedent of ref. 1 in fitting atomic polarization radii and a single inner dielectric constant to QM molecular polarizability tensors to produce an accurate EPIC model of electronic polarization. In this section, we generalize the parameterization to account for most of the biomolecules and a significantly wider spectrum of bioorganic functional groups. In contrast to our previous work, we use a smooth dielectric function as described earlier and a single internal dielectric (ε_{in}) value.

4.4.1.1 Choice of ε_{in} and A parameters

It was previously shown that a more accurate polarizable model was obtained when different ε_{in} were fitted for alkanes and aromatics. However, the single-ε_{in} model performed as well as the multi-ε_{in} model and DFT against experimental directional polarizabilities. Furthermore, in another study[2] that examined the local electronic polarization, the same single-ε_{in} model was only slightly worse than the multi-ε_{in} model. In this work we pursue the single-ε_{in} model because it greatly simplified the Poisson solver implementation and the robust parameterization for a wide spectrum of bioorganic chemistry.

Before the global parameterization of polarizability atomic radii, a range-finding study was performed with a smaller training set examining which combination of ε_{in} and A (c.f. eq. (4.1)) is best to use for extending the EPIC parameterization previously initiated[1]. We used a set of 13 alkanes (set g in ref. 1) including methane, propane, cyclopropane, butane (cis, trans), hexane (cis, trans), and neopentane,. together with a set of 10 heteroaromatic molecules (set a in ref. [1]). We formed the two-dimensional grid of ε_{in} and A pairs and optimized four radii (hydrogen, alkane carbon, aromatic carbon and aromatic nitrogen) for each point of the grid. The polarizability tensor RRMS deviation from QM for this dataset at each (ε_{in}, A) pair is shown as an iso-contour plot in Figure 4.3. It is clear that in order to fit a general dielectric function, a sufficiently large ε_{in} is needed. Also, the flatness of the error surface allows for multiple equivalent choices, a potential advantage if other criteria become more stringent in the development of the polarizable model. As shown by red circles in Figure 4.3, four starting points were selected for further examination: G1-24 (ε_{in}=24, A=4.188), G1-12 (ε_{in}=12, A=10), G1-9 (ε_{in}=9, A=5) and G1-4 (ε_{in}=4, A=10). In the case of G1-24 only, the A parameter was relaxed to a value of 4.18. The G1-12 seems slightly superior to the G1-9. Finally, while the G1-4 parameter set showed the worst RRMS, it was still a good case for having a small value of ε_{in}, picked by Tan and Luo[13] as being optimal. Each of the G1 ε_{in} and A choices was fixed in the global parameterization of atomic polarization radii described below. Finally, Figure 4.3 shows that making a poor selection of (ε_{in}, A), in particular having $\varepsilon_{in}< 6$, cannot be redeemed by adjusting either the radii or the A parameter.

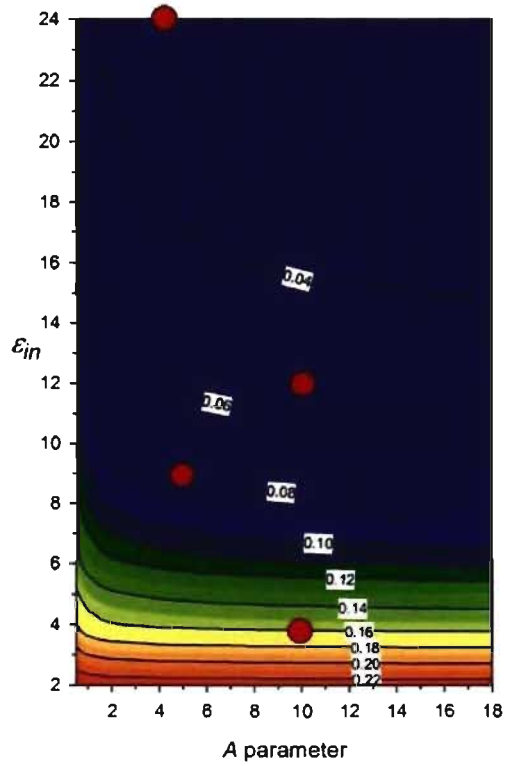


Figure 4.3. The iso-contour plot of the RRMS error between B3LYP/aug-cc-pVTZ and EPIC polarizability tensors are shown as a function of the ε_{in} and A parameters of eq. (4.1). This RRMS surface was generated from a simultaneous fit of the H, alkyl C, aromatic C and aromatic N atomic polarization radii on training set of 11 aromatic and 14 alkane molecules against their B3LYP polarizabilities. It shows that in order for a single dielectric model to fit the polarizabilities of these two chemical classes to within 10% error, the ε_{in} needs to be sufficiently large (>6). Deviations in the anisotropy of the polarizability are the main source of error for lower values of ε_{in}.

4.4.1.2 The optimized polarization radii

The parameterization of the four G1 sets on the 265 molecules of the polarizability training dataset proceeded as described in the Method section. The ε_{in} and A values were fixed and the atomic polarization radii σ_i were adjusted to optimize the fit to the B3LYP polarizability tensors. The atom typing of the radii was a primary concern and we aimed at

minimizing the number of radii fitted to reduce the fitting complexity, ensuring a better generalization of the chemistry. Each non-symmetric molecule produced 6 data points from their polarizability tensor; structurally symmetric molecules produced fewer data points. The number of fitted parameters was kept small compared to the number of associated data points to prevent over-fitting. The determination of the atom typing was done iteratively by hand: first, the polarizability training dataset was designed in terms of chemical functional-group classes. Adjustable parameters were added gradually with new molecules having unmet chemical functionalities. Often, the addition of a new chemical functionality class led to one or two additional parameters. We also merged atom types when the radii values were similar and the fitness metrics (χ^2, δ_{avg} and δ_{aniso}) were not significantly affected. For example, the alkane H and C radii were the first to be fitted. This was followed by aromatic C, H and N. It was determined early that a single aromatic and alkyl atom type for C and H could be utilized. Then the alcohol oxygen radius, halogen radii, alkene carbon and alkyne carbon radii were individually fitted. The final stage was a global simultaneous fit of all radii with all the molecules of the polarizability training dataset. Because of its special importance as a solvent, water was treated separately with its own special O and H radii.

The resulting polarization radii are given in Table 4.1 for the four G1 parameter sets. It is important to note that the ordering of atom types in Table 4.1 matters since the atom typing was done in the given order (from top to bottom) to deal with the issue of a particular atom falling in more than one category (H for instance). The first observation is that all polarization radii are significantly smaller than vdW contact radii such as Bondi[16], Pauling[64] or Parse[40] often used in Poisson-Boltzmann approaches. Instead of being a contradiction, this finding unveils the two different natures of the physical phenomena described. On the one hand, the polarization radii aim at calibrating how the electrons polarize in reaction to an external field created, for example, by an interacting molecule. On the other hand the vdW radii determine the position of the repulsive molecular wall towards solvent. It is also expected that the larger the ε_{in}, the smaller the radii: to maintain the over-all polarization the dielectric must increase as the radii decrease. This is illustrating a

Table 4.1. Reported optimal polarization radii (σ_{in}) and atom typing for the four G1 sets defining the internal dielectric (c.f. eq.(4.1)).

SMARTS[23-25]	Typical functional groups	Radius (Å)			
	Model name:	G1-4	G1-9	G1-12	G1-24
Fixed parameters					
ε_{in}		4	9	12	24
A		10	5	10	4.19
H					
[H]	all H	0.83	0.65	0.55	0.52
C					
[CX4]	alkanes	0.78	0.79	0.67	0.62
[c,CX2,CX3]	aromatic, sp, sp^2	1.25	1.02	0.87	0.78
N					
[n,NX1,NX3,$(Nc),$(NN)]	aromatic, nitriles, sp^3, aniline, hydrazine,	1.09	0.89	0.76	0.69 0.74[a]
[$(N[C,S]=*)]	amides, amidines, sulfonamides	0.89	0.77	0.64	0.58
[$(N=C)]	imine, amidine	1.07	0.93	0.81	0.76
[$([#7]~[OX1])]	N-oxides, nitro	0.00	0.79	0.68	0.59
O					
[$([OX2]([H])[#6,#7]),o,$([OD2]([CX4,c])[CX4]), $(O=[c,C,S])]	alcohols, furan, hydroxamic acids, ethers, ketones, aldehydes, amides, sulfones	0.88	0.73	0.63	0.60
[$(OC=[O,N])]	Esters, carboxylic acids	0.68	0.55	0.46	0.46
[$([OX1]~[#7])]	N-oxide, nitro	1.08	0.89	0.77	0.74
Others					
[S,s]	All sulfur atoms	1.44	1.22	1.06	1.01
[F]		0.77	0.62	0.53	0.51
[Cl]		1.30	1.09	0.95	0.91
[Br]		1.47	1.24	1.07	1.03
Water	Special fit				
[$([OX2]([H])[H]]		0.93	0.86	0.76	0.75
[$([H][OX2][H])]		0.64	0.45	0.36	0.31
Charged atoms					
[$([H][#7+]),$([H][#7][#6]=[#7+][H]), $([H][#7][#6]=[#7+]),$([H][n+]~c~n), $([H]n~c~[n+])]	proton in guanidiniums, amidiniums, ammoniums, pyridiniums	0.44	0.43	0.37	0.01
[$([O-]C=O),$(O=C[O-])]	O in carboxylates	1.20	1.02	0.88	0.85
[$([NX4+]),$([#7+]=C-N),$(N-C=[N+]))]	N in ammoniums, guanidiniums, amidiniums,	0.00	0.34	0.39	0.52
[$([n+]~c~n),$([n]~c~[n+])]	N in imidazoliums	0.00	0.00	0.00	0.42

general feature of the model that produces larger polarizabilities when either the 'electronic volume', decided by the radii, or the internal dielectric increase. The sort of relationship involved is given in eq. (4.5) above for a hard sphere and elsewhere for a diatomic[1].

It is also interesting to compare polarization radii between elements and between the different chemical environments. First, it is remarkable that the carbon atom can be split into only two atom types: sp^3 and others. The smaller sp^3 carbon radius implies that carbon makes a much smaller contribution to the overall polarizability when sp^3 hybridized than when pi electrons are present, i.e. in the sp or sp^2 hybridization states. This can be rationalized by the presence of π^* molecular orbitals, the different number of connected H atoms, and the difference in the molecule shape and the related anisotropy.

The nitrogen atoms were subdivided into four atom types amongst which two encompass almost all instances in the datasets. The first of these is a general nitrogen type assigned to amines, nitriles, hydrazines or anilines for example. The smaller second major nitrogen radius makes amide, amidine or sulfonamide nitrogen less polarizable. Surprisingly, the more specific nitro and N-oxide nitrogen radius, in the G1-4 set, has a radius of zero. The dielectric on this nitrogen atom is only slightly smaller than ε_{in} because of the large bound oxygen radii and the short N-O bond, typically 1.2 Å, which allows dielectric from the oxygen to spread over onto the nitrogen. It is also interesting to note that in the G1-24 set, there was a gain in accuracy when the nitrile nitrogen had its own radius.

The oxygen atom behavior can mainly be accounted for by two adjustable radii types, which was a significant advantage in the fitting process – the N-oxide and nitro functional groups still being an exception. Another interesting result is the large radius of the sulfur atom that is comparable to the bromine radius. However, it is not to say that the polarizability contribution of sulfur is equivalent. In fact, the bromine bonds are longer and hence offer a larger polarizable volume. This argument is also useful to explain why the fluorine radius is smaller than the hydrogen radius. For example, the model predicts a

polarizability for tetrafluoromethane of 18 a.u. compared to 17 a.u. for methane, and a polarizability of 76 a.u. for hexafluorobenzene compared to 70 for benzene, all in close agreement with B3LYP. Because of water's special importance as a solvent, both the oxygen radius and the hydrogen radius were optimized to exactly match the B3LYP polarizability tensor. These various behaviours of the atomic polarization radii underscore their difference in nature and purpose from vdW contact radii, which is why they must be treated differently.

Finally, charged species pose a special challenge that we decided to address specifically for charged side chains in proteins: Arg, Lys, Asp, Glu and His. Further generalization of the radii for charged species while retaining the same level of accuracy in the polarization tensor would require a more extensive parameterization. One reason for this is the expected reduction in polarizability on the neighbor atoms through the strong induction caused by the charged site. On the other hand, the electrostatic interactions around charged centers will be dominated by the monopole (i.e. the distribution of the charge itself), so high accuracy in the effects of polarization may become less important than with neutral species.

4.4.1.3 Polarizability tensors

The G1 parameterizations clearly showed the capacity of EPIC to produce accurate polarizabilities with a minimum of atom types. The choice of ε_{in} and A combinations made based on the very small range-finding subset showed the same behavior in the polarizability training dataset, the polarizability validation dataset and their combination (polarizability dataset), made of 265, 442 and 707 molecules respectively. Table 4.2, which summarizes the errors, shows the accuracy of the obtained models. The G1-24 dataset has an unsigned average error of 2% on the average polarizability (eq. (4.13)) and a 5% error on the anisotropy of the tensor (eq. (4.14)). With the point inducible dipole polarizable models, such a low level of error was obtained only when anisotropic atomic polarizabilities were

fitted[5,15,65-67], making their generalization very challenging. The other G1 models are worse, and as predicted from the range-finding study results shown in Figure 4.3, the G1-4 set is inadequate to reproduce the directional difference in the polarizability (the large δ_{aniso} values in Table 4.2). That the error obtained on both the training dataset and the validation dataset was similar indicates that our radii are not overfit. Finally, the three directional polarizabilities (eigenvalues of the tensor) obtained for the 707 molecules (2121 data points) are compared to the corresponding B3LYP values in Figure 4.4 for three representative G1 sets. The excellent correlation is obvious for the G1-24 and G1-12, and deteriorates in the G1-4 EPIC model (the Pearson correlation coefficients are 0.99, 0.99, 0.96 and the slopes 0.97, 1.02 and 1.20 respectively). An apparent outlier is the α_3 (longitudinal polarizability) of (3E)-hexa-1,3,5-triene for which B3LYP gives a value of 176 a.u. compared to the EPIC value of 125 a.u. For this specific molecule, Sekino *et al.*[68] showed that B3LYP greatly overestimates the α_3 value of acetylene chains. Their better estimate, based on very accurate CCSD and MP2 QM results, predicts a value of ~135 a.u., close to the EPIC value. Another remarkable discrepancy between EPIC and B3LYP is observed in Figure 4.4 for the α_3 of 1,4-dioxidopyrazine (doubly oxidized nitrogen on pyrazine) that is predicted to be 103 a.u. by the G1-12 model versus 129 a.u. by B3LYP. A similar observation can be made for 4-nitroaniline. Although we have not found better estimates for these molecules, they most certainly constitute a challenge both for classical and *ab initio* polarizability calculations.

Table 4.2. Error obtained with the optimized polarization radii of the G1 sets when EPIC molecular polarizability tensors are compared to B3LYP for different molecule datasets.

Model[a]	δ_{avg}[b] (%)	δ_{aniso}[b] (%)	RRMS[b] (%)
Polarizability training dataset: 265 molecules			
G1-4	5.0	20.9	12.7
G1-9	3.2	9.1	6.7
G1-12	2.9	5.3	5.0
G1-24	2.3	5.2	4.4
Polarizability validation dataset: 442 molecules			
G1-4	4.0	18.2	12.3
G1-9	2.7	7.6	6.7
G1-12	2.6	5.1	5.3
G1-24	2.1	5.4	4.6
Polarizability dataset: 707 molecules			
G1-4	4.4	19.2	12.4
G1-9	2.9	8.2	6.7
G1-12	2.7	5.2	5.2
G1-24	2.2	5.4	4.6

[a]Model using the parameters given in Table 4.1.
[b]c.f. section 4.2.2.3

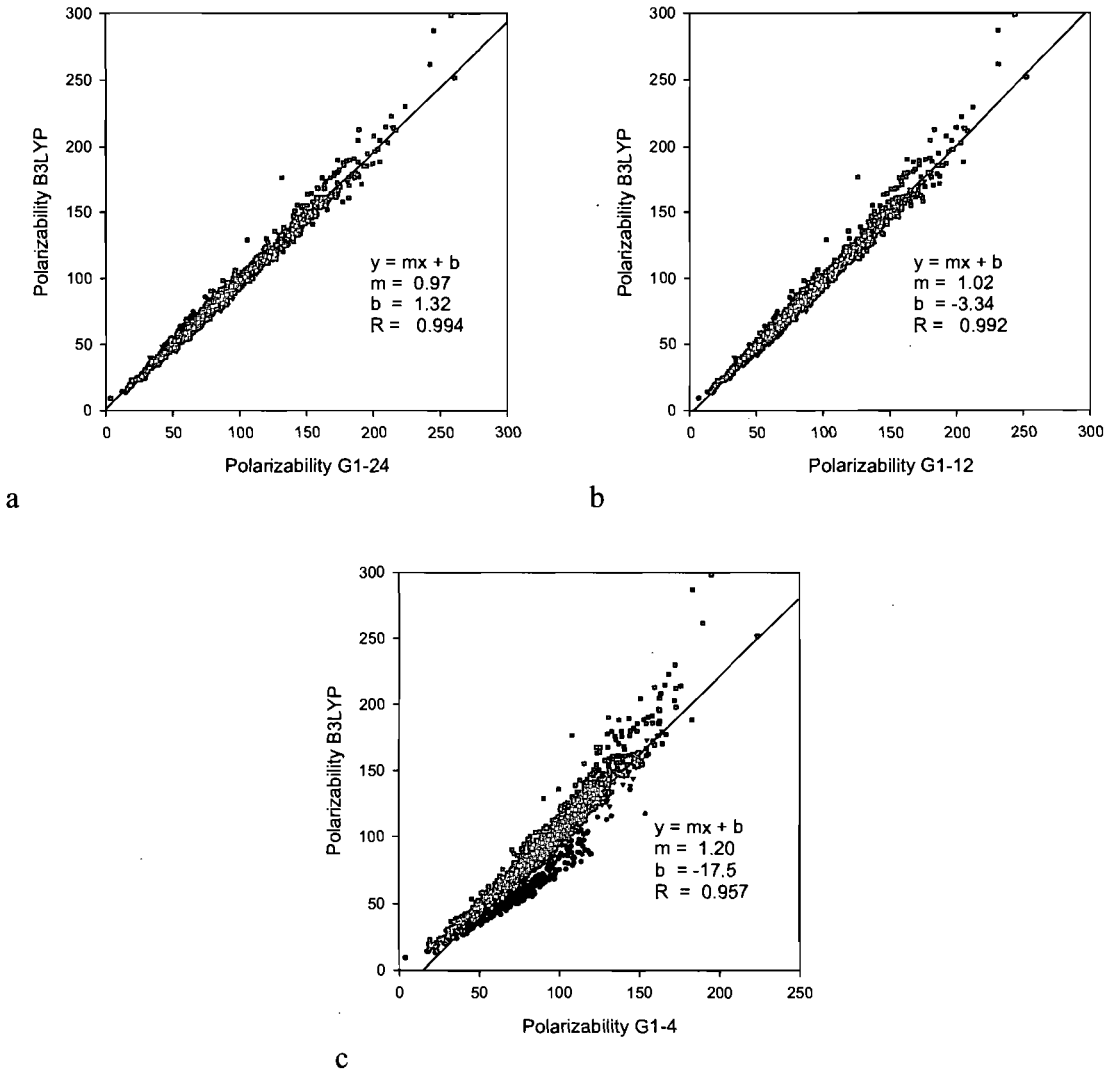


Figure 4.4. Correlation graph between the B3LYP/aug-cc-pVTZ directional polarizabilities (α_1 black circles, α_2 red triangles and α_3 green squares in a.u.) for three G1 dielectric parameter sets (c.f. Table 4.1). Each figure shows the data for 707 molecules for a total of 2121 points. From these figures, it is clear that a small number of parameters (optimized on 265 molecules) can generalize well. A large ε_{in} = 24 (a) produces the best fit, a medium range ε_{in} =12 (b) produces slightly larger discrepancies and a small ε_{in} = 4 (c) produces significantly larger deviations, in keeping with the results of the range-finding study on the small dataset.

4.4.2 Refractive indices

In the previous subsection, we have developed dielectric functions that predict remarkably well, relative to QM, the polarizabilities of a single molecule in the gas phase. In this section we present the *macroscopic* refractive index calculations and the corresponding effective high frequency limit dielectric (ε_∞). In a previous publication[1], we proposed that the vacuum of the intermolecular spacing may be sufficient to reduce the effective macroscopic ε_∞ resulting from the high intramolecular ε_{in} obtained in the optimization to polarizability tensors. Here we use a theoretical approach to verify this hypothesis. Another important point addressed by the refractive index calculation is the transferability of the dielectric function from the gas phase to the condensed phase.

As explained in further detail in the Theory and Method sections, we form liquid droplets containing thousands of molecules from snapshots obtained by MD simulations and calculate the effective ε_∞ by the use of the Clausius-Mossoti equation. The small range spanned by experimental refractive indices makes this test somewhat stringent. Figure 4.5 shows the correlation between the results obtained with three representative EPIC parameterizations and experiment; G1-9 is omitted here and for the remainder of the article because the results are so similar to those of G1-12. The first observation is the close agreement between the magnitudes of the ε_∞ values. This clearly demonstrates that the effective ε_∞ of the liquid droplets have the appropriate value in spite of the high ε_{in}, confirming our hypothesis. Figure 4.6 provides a visual explanation for the apparent mismatch between the small effective ε_∞ compared to the productives ε_{in}. This figure shows the molecular dielectric inside a CCl_4 droplet when it is sliced through its center. The G1-24, G1-12 and G1-4 models have a quite variable low-dielectric intermolecular space. The coloring scheme of the dielectric function (eq. (4.1)) assigns red when $\varepsilon(r) = \varepsilon_{in}$ and dark blue when $\varepsilon(r) = 1$. The low dielectric intermolecular space increases with ε_{in} as the atomic radii decrease. It is striking that these three parameterizations produce the same refractive

index, the same molecular polarizability and this in spite of the very different ε_{in}. Of course, if ε_{in} is further reduced, the whole droplet will be filled with a uniform dielectric (as the atomic radii increase and start to overlap) and the simultaneous prediction of the molecular polarizability and the refractive index becomes compromised.

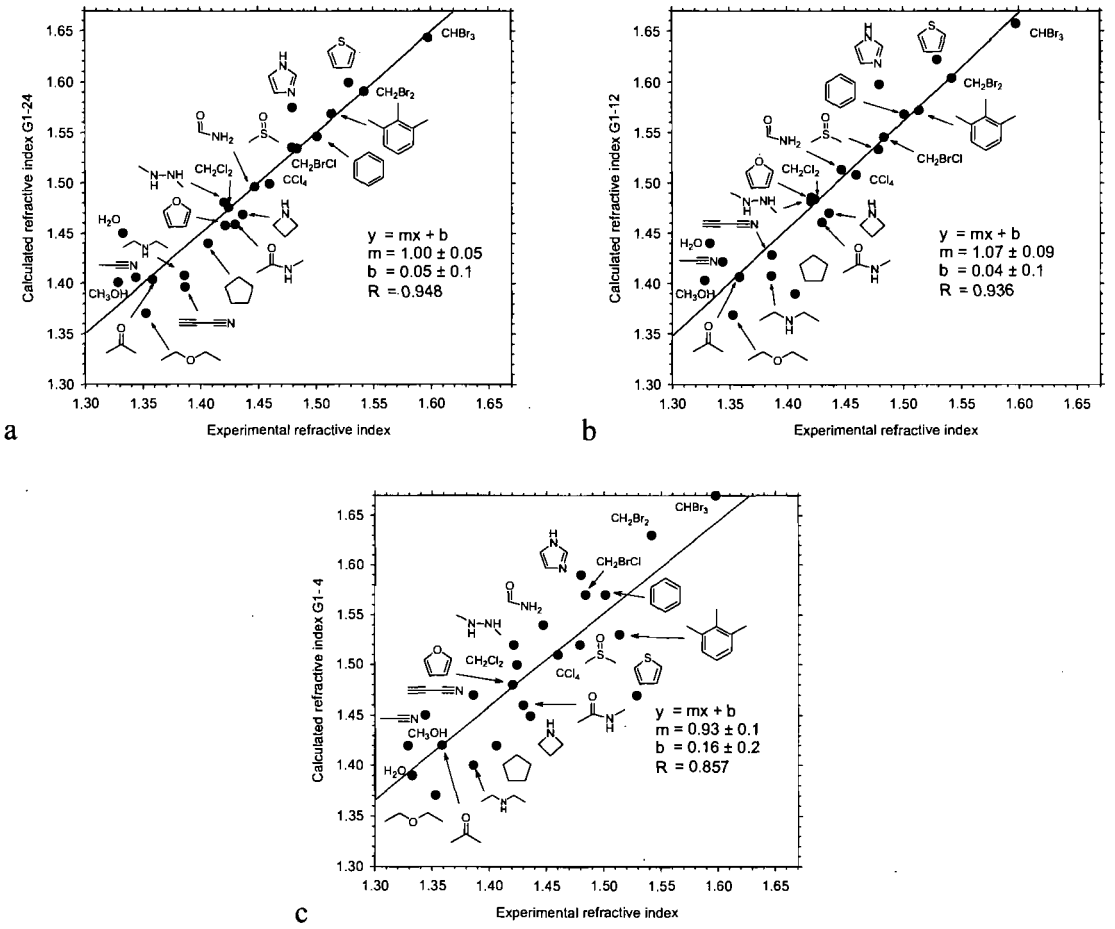


Figure 4.5. The calculated refractive indices (n) of 23 organic molecules are compared to experiment. Three dielectric parameter sets are used a) G1-24 b) G1-12 and c) G1-4 (Table 4.1). For each set, the pre-optimized radii can be found in Table 4.1. The reported refractive indices (n) were obtained by polarizing a liquid droplet formed by carving spheres from periodic MD liquid simulation snapshots. The Clausius-Mossoti equation leads to $n^2 = \varepsilon_\infty$ close to experiment, in spite of the large ε_{in}. The predicted values are systematically higher than experiment, which can be explained by potential artifacts or a polarizability shift when passing from vacuum to condensed phase (see text). As with the polarizabilities, the predictions deteriorate with decreasing ε_{in}, in keeping with the results of the range-finding study on the small dataset.

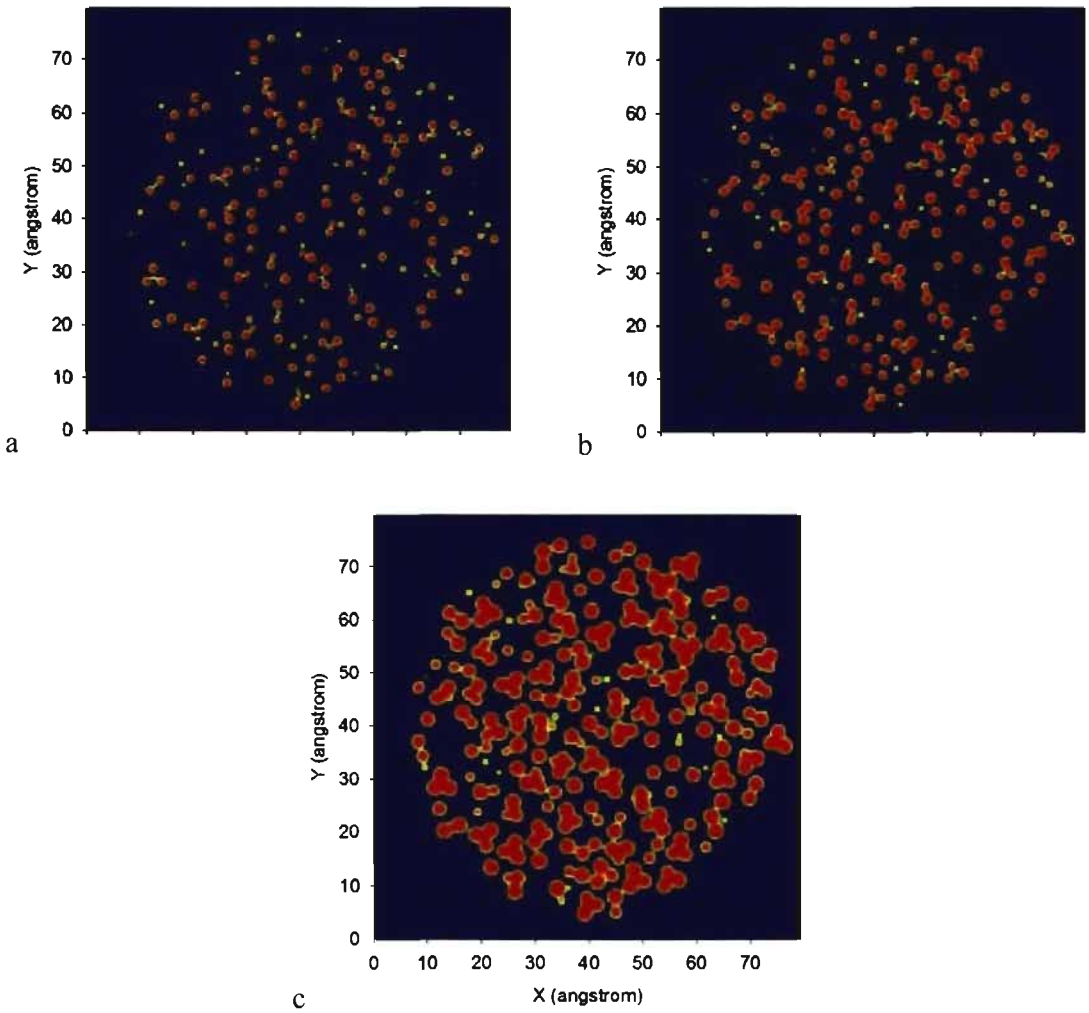


Figure 4.6. One of the 50 CCl_4 droplets is cut in its center and three dielectric functions (eq. (4.1)) are plotted: a) G1-24 b) G1-12 and c) G1-4. The red color is associated with $\varepsilon(r)=\varepsilon_{in}$ and blue to $\varepsilon(r)=1$.

Also noticeable in Figure 4.5 is that the correlation with experiment follows the previous assessment of the models based on molecular polarizabilities: the G1-24 parameterization (Figure 4.4a), has a R = 0.95, slightly better than the G1-12 (b) with R =0.94, which is in turn significantly better than the G1-4 correlation with R = 0.86 (Figure 4.4c). However, Figure 4.5 shows a 0.05 systematic overestimation of the refractive indices which could

correspond to a small overpolarization, a result not reflected in the gas-phase polarization tensors. The source for this deviation is not clear, but we have several hypotheses. Firstly, the Clausius-Mossoti equation is valid for a perfect sphere whereas we are dealing with an imperfect surface created by nanoscopic droplets. Secondly, we have verified that an underestimation of the droplet radius by 3-4% (1Å in a range of 25-35Å) could systematically shift the calculated refractive indices by 0.05. Thirdly, it is also possible that the liquid phase polarizability may be truly smaller than the predicted gas phase polarizability since a drop of 11% of the polarizability could explain the 0.05 shift. This would be in agreement with other studies that found similar phenomena[8,11,69] and based their reasoning on the increased Pauli exchange repulsion from the closer contact of the molecules in condensed phase. However, the magnitude of this effect differs considerably from study to study.

4.4.3 Hydration free energies

The calculation of hydration free energies is aimed at assessing whether the dielectric polarization model capable of accurately reproducing gas-phase polarizability tensors can be used "as is" in implicit solvent calculations. Because of the difference in nature and behaviour between the atomic polarization radii and the atomic cavity radii used for the solute-solvent boundary, the 3-zone dielectric model is required. The hydration free energies were calculated with the G1-24, G1-12 and G1-4 models found in Table 4.3. Each of the solute models was used to optimize the solvent cavity atomic radii (referred to as cavity radii and noted G2 in what follows) referred to in the second Gaussian summation in eq. (4.3). We decided to set B = 11.8 in all calculations, following the Grant *et al.*[18] suggestion as it was found to make the Bondi radii[16] optimally reproducing the hard dielectric boundary results with the same smooth boundary as used in this work. The results reported in Table 4.3 are split into two main categories based on the method used to approximate ΔG_{np}. The surface area (SA) based method follows eq. (4.20) and required the

optimization of the surface tension parameter (γ). The main effort here is however concentrated with the use of ΔG_{np} by converged free energy perturbation (FEP) calculations[45,46]. This is the first category of results that we examine below.

4.4.3.1 Results with FEP-based non-polar term

Two main classes of solute are studied as reported in Table 4.3. Firstly, we set $\varepsilon_{in} = \varepsilon_{trans} = 1$ in eq. (4.3), effectively turning eq. (4.3) into a 2-zone dielectric function with a non-polarizable solute as defined previously by Grant *et al.*[18]. For the non-polarizable solute model, we used water-polarized static atomic charges as given by ESP-fitting to HF-6-31G(d,p) wavefunction (G2-HF). These charge sets are positive controls, following the traditional approaches for non-polarizable force fields and which have been shown to produce the right degree of static polarization of the solute in water[70]. The G2-ε_{in}, with ε_{in} = 4, 12 and 24 (c.f. Table 4.1), has polarizable solutes assigned charges fitted to the B3LYP/6-311++G(3df,3pd) ESP known to reproduce the gas phase dipole moment of the molecules, being usually between 10% and 20% smaller than what is normally expected in water. Polarizability should rightly compensate for the use of gas phase charges.

It is quite interesting to observe in Table 4.3 that by allowing the cavity radii to optimize in each model, the same level of error over the 485 experimental free energies of hydration is obtained for the G2-HF, G2-4, G2-12 and G2-24 solute models. The average unsigned error (AUE) compared to experiment is 1 kcal/mol with a standard deviation of 1 kcal/mol. The Pearson correlation coefficient (R) is around 0.89 in all these G2 models. The relative root-mean-square deviation (RRMS) obtained is 0.35 and the average signed error (AE) is found to be between -0.15 kcal/mol and -0.18 kcal/mol. These errors can be compared to the Rizzo *et al.*[47] results, on almost the same dataset (460 neutral molecules included in the 485 that we use), that produce an AUE of 1.47 kcal/mol with RESP charges and R = 0.88. The reported numbers of Rizzo *et al.* were obtained with a SA evaluation of ΔG_{np} that allow them to subsequently optimize 14 atom typed surface tensions (γ), which improved the AUE to 1 kcal/mol while R = 0.89. For comparison, in the current study, we fit 8 atomic

Table 4.3 Solvent cavity atomic radii (σ_{cavity}) and γ for the 3-zone dielectric model optimized on 485 experimental free energy of hydration with different G1-ε_{in} solute models and ΔG_{np} sources

Model[a]:	G2-HF	G2-4	G2-12	G2-24	G2-12SA	
		Solute				
Charges[b]	HF[c]	B3LYP[d]	B3LYP[d]	B3LYP[d]	B3LYP[d]	
ε_{in} [e]	1	4	12	24	12	
A [e]		10	10	4.19	10	
Ref. Table 4.1		G1-4	G1-12	G1-24	G1-12	
ΔG_{np}	FEP[f]	FEP	FEP	FEP	SA[g]	
B	11.8	11.8	11.8	11.8	11.8	
	Optimized implicit solvent parameters					Bondi
H[h]	0.98	0.95	0.97	1.02	0.98	1.20
C	1.95	2.03	2.02	1.95	2.01	1.70
N	1.74	1.74	1.74	1.68	1.69	1.55
O	1.81	1.79	1.78	1.75	1.75	1.52
S	2.60	2.27	2.29	2.33	2.41	1.80
F	2.09	2.09	2.08	2.05	2.49	1.47
Cl	2.38	2.36	2.47	2.41	2.46	1.75
Br	2.18	2.23	2.46	2.45	2.63	1.85
γ [i]					6.8	
AUE[j]	1.06	0.99	1.04	1.08	1.13	
Stdev[k]	1.00	0.96	0.99	1.00	0.90	
RMS[l]	1.45	1.38	1.44	1.47	1.45	
R[m]	0.89	0.90	0.90	0.89	0.88	
RRMS[n]	0.34	0.33	0.34	0.35	0.34	
AE[o]	-0.18	-0.17	-0.17	-0.17	0.02	

[a] Tag names for each of the optimized solvent cavity radii. [b]Atomic partial charges from an ESP-fit or a DRESP fit on the given quantum method. [c] Pre-polarized charges from HF/6-31G(d,p) [d]Vacuum charges from B3LYP/6-311++G(3df,3pd) [e]A and ε_{in} of eq. (4.3) for the solute internal dielectric. The atomic radii used in the internal dielectric are given in Table 4.1.[f]ΔG_{np} from free energy perturbation[45] [g]ΔG_{np} calculated using the surface area (eq. (4.20)) with the γ term optimized [h]Cavity atomic radii are given in angstrom [i]Non-polar surface tension from eq. (4.20) in cal/Å^2 [j]Average unsigned error in kcal/mol [k]Standard deviation of the unsigned error [l]Root-mean-square deviation in kcal/mol [m]Pearson correlation coefficient [n]Relative root mean square deviation [o]Average signed error in kcal/mol: experiment – calculated.

radii. In addition, the recent work of Mobley *et al.*[46] using Bondi radii and the single γ fitted by Rizzo *et al.* on an almost identical dataset to ours obtained a root-mean-square

deviation of 2.05 kcal/mol. Finally, in a different article, Mobley *et al.*[45] obtained a RMS of 1.26 kcal/mol and R = 0.89 with explicit-solvent converged FEP calculations. The FEP based ΔG_{np} used in this work comes from this latter study. Our results are comparable to or better than most other studies. We attribute the small errors to the optimization of the radii, not necessarily to the quality of the solute model. We can however examine the fitted cavity radii with the different solute models to understand the effects of the electrostatic model on the solute cavity size.

The level of solute polarization brought by the polarizable solute models (G2-4 to G2-24) seems similar to what is obtained with the G2-HF solute model. This can be assessed by comparing the atomic radii and the cross-validation error showed in Table 4.4 where the G1-12 solute model is used with the different G2 radii sets. The level of error produced when ε_{in} =4, 12 and 24 or with the G2-HF cavity radii is similar, the G2-4 being the worst. The cross-validation results of Table 4.4 also show the transferability of the third zone dielectric parameters given that the solute has the physically appropriate electrostatic behavior. A possible advantage of the polarizable solute model is when the solvation free energies are computed relative to a solvent much less polar than water (e.g. a non-polar solvent or a non-polar binding site in a protein). In this case, the HF based charges may not be appropriate[71]. The fitted radii of Table 4.3 are significantly different from the contact Bondi radii reported in the last column. Firstly, the H radius is a little smaller than the usual 1.1 Å contact radius in all cases (the Bondi radius of 1.2 Å for H was recognized to be a little too large and was revised to be 1.1 Å[72]). The carbon radius obtained here is much larger than the Bondi radius and makes the C-H bonds behave like a united atom model. In this perspective the carbon radius size obtained here is similar to the Nina *et al.*[73] carbon radius they calculated by looking at MD water charge density in explicit solvent simulations. For the other elements, we also find larger radii than Bondi, in agreement with a recent study by Nicholls *et al.*[31]

Table 4.4. Effects of using different solvent cavity radii set (Table 4.3) with the G1-12 solute model (Table 4.1) on ΔG_{hyd}

	G2-12	G2-24	G2-4	G2-HF
AUE[a]	1.04	1.08	1.17	1.10
Stdev[b]	0.99	1.03	1.19	1.01
R[c]	0.90	0.90	0.86	0.89
RRMS[d]	0.34	0.35	0.39	0.35
AE[e]	-0.17	0.16	-0.23	0.15

[a]Average unsigned error in kcal/mol
[b]Standard deviation on the AUE in kcal/mol
[c]Pearson correlation coefficient
[d]Relative root-mean-square deviation
[e]Average signed error in kcal/mol: experiment - calculated

The larger cavity radii can be rationalized by considering the difference between contact radii (Bondi) and the cavity radii needed in implicit solvent calculations. The former is defined by crystal contacts between neighboring molecules and the latter defines where the mean solvent charge density appears. In terms of the 3-zone dielectric model, the contact radii would be located in the middle of the second zone where the electronic density should be minimal given that the dielectric goes to one (no electrons to be polarized on the solute side). This is supported by the fact that Fermi repulsion between the solvent molecule electrons and the solute electrons reduces the total electronic density exactly in the contact zone to its minimum. In Figure 4.2a, the contact radius of an aromatic carbon atom would become 1.7 Å, exactly the Bondi radius value. Similarly the middle of the blue area in Figure 4.2b defines the contact line between solvent molecules and the solute.

Although we claim here that having cavity radii larger than Bondi radii may be physically motivated, it is not possible at this stage to know if this effect should be as large as we find. In particular, the fluorine radii in Table 4.3 are surprisingly large. This was also

found by Nicholls *et al.*[31] where their optimal fluorine radius was 2.4 Å. Knowing that fluorine is particularly hydrophobic, this may be just another peculiar behavior of this atom. The Cl and Br radii difference in the G2-4, G2-12 and G2-24 sets uncover a drawback of using a small ε_{in}. Because the polarization radius of Cl and Br are larger in the G1-4 than in the other EPIC parameterizations, the transition zone shown in Figure 4.2 cannot reach $\varepsilon(r) = 1$ (in the case of Br, it only decreases to $\varepsilon(r) = 3$) and as a result the full polarizability coming from the halogen atom is not reached as the solvent cuts into the first zone dielectric function. This prevents enough solute bound charge density from building up.

A last point to mention in regard to the 3-zone dielectric function is its potential advantage in reducing the occurrence of the reentrant surface problem that often brings a lot of fluctuation in energy or force computations in proteins, for instance. The problem is the artificial formation of a cavity with high dielectric inside a protein due to the irregularity of molecular surfaces. The large size of the atomic cavity radii in the G2 sets and the use of a smooth dielectric function should contribute to create a sufficiently deep buffer of low dielectric and make implicit solvent models more stable. Indeed, the smoothness of the surface around 4-pyridone observed in Figure 4.2b looks like a solvent accessible surface[17]. This entire question is however left for future research.

4.4.3.2 Results with surface area-based non-polar term

Although the use of the very computationally intense FEP-based ΔG_{np} may be more physically grounded, the obtained models cannot practically be used in a prospective manner due to the heavy computational demands for such FEP calculations. For this reason, we also optimized the cavity radii and the surface tension with the G1-12 solute models. In these calculations the solvent accessible surface area was calculated with the Bondi radii and kept constant. The results are reported in Table 4.3. The error levels reported are comparable to those obtained with ΔG_{np} from FEP calculation. The G2-12/SA model gives error levels a little larger than the G2-12: AUE = 1.13 kcal/mol with a standard deviation of

0.90 kcal/mol, R = 0.88, RRMS = 0.34, RMS = 1.45 kcal/mol and AE = 0.02 kcal/mol. It is comforting that the optimal surface tensions of the two optimizations are close to each other. The radii obtained for the G2-12/SA fit are similar to the G2-12 fit except for S, F, Cl and Br. It is possible that the hydrophobicity of these atoms is overestimated by the single surface tension term used with the Bondi radii to determined ΔG_{np}.

4.5 Conclusion

The EPIC approach to molecular polarizability has been parameterized to include many more chemical functional groups than a previous effort[1]. This required generating a dataset of 707 B3LYP/aug-cc-pVTZ molecular polarizability tensors. The ability of EPIC to account for both the average polarizability and the anisotropy of the tensor was remarkable given that the optimization of only 14 parameters (excluding water and charged species) led to a relative unsigned error in the average polarizability and anistropy of 2.6% and 5.2% respectively (G1-12). An example of the parsimony of the atom typing is that a single radius parameter was sufficient for aromatic, nitrile, amine, aniline or hydrazine types of nitrogen. Obtaining the same level of error with both the validation and training datasets suggests that overfitting is not an issue. With previous polarizable models, such as point-inducible dipoles, this level of accuracy could only be attained with added complexity such as anisotropic polarizable centers or molecule-specific Thole screening parameters[5,15,28,65-67].

We found that the anisotropy could only be reproduced accurately if the interior dielectric constant was higher than 9.0. Above this value, almost any interior dielectric can also work well as long as the atomic polarization radii are appropriately adjusted. The need for a high interior dielectric raised the question of the physical soundness of the model as $\varepsilon_{in}=\varepsilon_{\infty}=n^2$ has become a dogma in the implicit solvent literature[13,30,31,39,40,74-76] whenever the

dielectric constant is used to replace electronic response. The conceptual flaw of this equality comes from the fact that the interior dielectric (ε_{in}) is not uniformly distributed whereas the refractive index (n) comes from a macroscopic measurement assuming a uniform ε_∞ in space. To verify that the optimized models agree with experimental refractive indices, we devised a new protocol to calculate a liquid refractive index from microscopic simulations. To this end, 23 organic molecules spanning the entire range of bioorganic-molecule-like refractive indices state were simulated by molecular dynamics in the liquid state and ε_∞ was calculated with the Clausius-Mossoti relation. The obtained refractive indices now come from the effective polarization of liquid configurations having intramolecular high dielectric with low-dielectric interstices. The results show a good correlation for all three G1-ε_{in} parameter sets. The highest interior dielectric (ε_{in} = 24) gave the best correlation with a slope of 1.00, an intercept of 0.05 and a correlation coefficient of 0.95. It is interesting to note that the polarizability anisotropy may play a role since the G1-4 parameters (ε_{in}=4) gave the poorest correlation and was also the worst model for polarizability anisotropy. Nevertheless, these results indicate that, when coupled with the appropriate radii, many choices of ε_{in} can give results in good agreement with the experimental refractive indices.

To use the EPIC polarizable electrostatic model with implicit solvent, we have developed a smoothed-boundary 3-zone dielectric function that works with the internal dielectric continuum model. The three zones are the internal dielectric constant ε_{in}, a transition zone that goes to the vacuum dielectric (ε_{trans}=1), and a third zone defined by the molecular cavity boundary, i.e. where the dielectric function reaches the bulk liquid dielectric. With this function, keeping the first zone fixed at the atomic polarization radii and ε_{in} determined for the gas phase polarizabilities, only the molecular cavity boundary needs to be parameterized. A dataset of 485 experimental free energy of hydration was used to optimize the solvent cavity radii, one per element, with different charge models. The resulting level of error was smaller than found in previous implicit solvent studies with a typical average unsigned error of 1 kcal/mol with a standard deviation of about 1 kcal/mol

and a Pearson correlation coefficient of 0.9. Atomic charge sets fitted from the unpolarized gas phase B3LYP density coupled with EPIC polarization led to cavity radii comparable to those obtained with the polar-condensed-phase-like HF/6-31G(d,p) charges. The low sensitivity of the optimal cavity radii resulting from the fit with different polarizable solutes (the different G1-ε_{in}) further supports the generality of the approach. These results clearly show that EPIC can lead to accurate description of solute polarization in implicit solvent. The anisotropy of the molecular polarizability does not seem to play an important role in fitting experimental hydration free energies. However, when considering intermolecular interactions, such as in an enzyme active site, the heterogeneity of the environment and of the interactions may require an accurate directional polarizability. An important example of this is in cation-π interactions[2,77]

The proposed global optimization scheme involves several independent layers. The polarizability part is fitted on uncharged QM molecular polarizability tensors. The charges are added with the DRESP fit on *ab initio* electrostatic potentials calculated on a grid, as usual, except that here the QM method can be systematically improved since gas phase charges are needed. For implicit solvation, solvent-related radii are obtained from a fit to experimental hydration free energies. Flexibility and transferability has been demonstrated for each stage. The ease with which we could fit the polarizability of so many functional groups makes us believe that the further extension of the parameterization and atom typing should be straightforward given more data. Moreover, the decoupling of the fitted polarization from the fitted charges, as well as the physical soundness of each step, makes the above parameterization scheme even more robust and general than is possible for two-body additive force fields.

This work partly addresses the question of applicability of EPIC in polar condensed phase. The calculations of the refractive indices were well behaved and show the high values for ε_{in} are not unphysical. It is not clear if the slightly larger calculated polarizabilities of the droplets were due to a change in polarizability when going from gas

phase to condensed phase. Also, the level of electronic induction seen in the 3-zone implicit solvent calculations suggests that both solute and solvent polarization in polar media is well modeled. To confirm those findings, it would be interesting to perform explicit atoms simulations with the EPIC model and Poisson's equation.

The EPIC approach to polarizability has shown unprecedented accuracy and flexibility on many accounts for such a simple model. Although the optimized parameters are unconventional compared to traditional Poisson-Boltzmann applications, it is for sound physical reasons that even clarify aspects of the implicit solvent approaches. In this paper and the two previous ones[1,2], EPIC was shown to be a powerful tool to include the effects of electronic polarization in molecular mechanics type calculations, especially appropriate to biomolecular force fields.

4.6 Acknowledgment

This work was made possible by the computational resources of the réseau québécois de calcul haute performance (RQCHP). The authors are grateful to OpenEye Inc. for free academic licenses. R. I. I. acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC). J.-F. T. is supported by NSERC through a Canada graduate scholarship (CGS D) and by Merck & Co. through the MRL Doctoral Program. I. B. R. is supported by NIH grant GM072558.

4.7 Supporting Information

DFT and EPIC polarizabilities are available for all 707 molecules examined together with the optimized coordinates of the molecules used to calculate the polarizability tensor. Calculated and experimental refractive indices are provided. The detailed free energies of hydration can also be found. This information is available free of charge via the Internet at http://pubs.acs.org.

4.8 Bibliography

1. Truchon, J.-F.; Nicholls, A.; Iftimie, R. I.; Roux, B.; Bayly, C. I. Accurate Molecular Polarizability based on Continuum Electrostatics. *J. Chem. Theory Comput.*, **2008,** *4*, 1480-1493.

2. Truchon, J.-F.; Nicholls, A.; Grant, J. A.; Iftimie, R. I.; Roux, B.; Bayly, C. I. Extending the Polarizable Continuum Dielectric Model to Account for Electronic Polarization in Intermolecular Interactions. *Submitted*, **2008,**

3. Applequist, J.; Carl, J. R.; Fung, K. K. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.*, **1972,** *94*, 2952-2960.

4. Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, **1976,** *103*, 227-249.

5. Miller, K. J. Calculation of the Molecular Polarizability Tensor. *J. Am. Chem. Soc.*, **1990,** *112*, 8543-8551.

6. Lamoureux, G.; Allouche, D.; Souaille, M.; Roux, B. Incorporating electronic polarizability in molecular dynamics: A model based on drude oscillators. *Biophys. J.*, **2000,** *78*, 330A-330A.

7. Lamoureux, G.; Roux, B. Polarizable force-field based on drude oscillators: Molecular dynamics simulation and parametrization. *Biophys. J.*, **2001,** *80*, 328A-328A.

8. Schropp, B.; Tavan, P. The Polarizability of Point-Polarizable Water Models: Density Functional Theory/Molecular Mechanics Results. *J. Phys. Chem. B*, **2008,** *112*, 6233-6240.

9. Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.*, **2003,** *119*, 5185-5197.

10. Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. *J. Chem. Theory Comput.*, **2005,** *1*, 153-168.

11. Morita, A. Water polarizability in condensed phase: Ab initio evaluation by cluster approach. *J. Comput. Chem.*, **2002,** *23*, 1466-1471.

12. Sharp, K.; Jean-Charles, A.; Honig, B. A Local Dielectric-Constant Model for Solvation Free-Energies Which Accounts for Solute Polarizability. *J. Phys. Chem.*, **1992,** *96*, 3822-3828.

13. Tan, Y. H.; Luo, R. Continuum treatment of electronic polarization effect. *J. Chem. Phys.*, **2007,** *126*, 094103.

14. Tan, Y. H.; Tan, C. H.; Wang, J.; Luo, R. Continuum polarizable force field within the Poisson-Boltzmann framework. *J. Phys. Chem. B*, **2008,** *112*, 7675-7688.

15. Birge, R. R. Calculation of Molecular Polarizabilities Using An Anisotropic Atom Point Dipole Interaction-Model Which Includes the Effect of Electron Repulsion. *J. Chem. Phys.*, **1980,** *72*, 5312-5319.

16. Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.*, **1964,** *68*, 441-451.

17. Richards, F. M. Areas, Volumes, Packing, and Protein-Structure. *Annu. Rev. Biophys. Bioeng.*, **1977,** *6*, 151-176.

18. Grant, J. A.; Pickup, B. T.; Nicholls, A. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Comput. Chem.*, **2001,** *22*, 608-640.

19. Prabhu, N. V.; Zhu, P. J.; Sharp, K. A. Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method. *J. Comput. Chem.*, **2004,** *25*, 2049-2064.

20. Im, W.; Beglov, D.; Roux, B. Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.*, **1998,** *111*, 59-75.

21. Griffiths, D. J. *Introduction to Electrodynamics;* 3rd ed.; Prentice-Hall Inc.: Upper Saddle River, NJ, 1999.

22. *Zap Toolkit,* version 2.2.1; Santa Fe, NM, USA, 2007

23. Weininger, D. Smiles .3. Depict - Graphical Depiction of Chemical Structures. *J. Chem. Inf. Model.*, **1990,** *30*, 237-243.

24. Weininger, D.; Weininger, A.; Weininger, J. L. Smiles .2. Algorithm for Generation of Unique Smiles Notation. *J. Chem. Inf. Model.*, **1989,** *29*, 97-101.

25. Weininger, D. Smiles, A Chemical Language and Information-System .1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.*, **1988,** *28*, 31-36.

26. *OEChem Toolkit,* version 2.2.1; Santa Fe, NM, USA, 2007

27. *SciPy: Open Source Scientific Tools for Python,* 2001

28. Elking, D.; Darden, T.; Woods, R. J. Gaussian induced dipole polarization model. *J. Comput. Chem.*, **2007,** *28*, 1261-1274.

29. Bottcher, C. J. F. *Theory of Electric Polarization;* 2 ed.; Elsevier Scientific Publishing Company: New York, 1973; Vol. 1.

30. Onsager, L. Electric Moments of Molecules in Liquids. *J. Am. Chem. Soc.*, **1936,** *58*, 1486-1493.

31. Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. *J. Med. Chem.*, **2008,** *51*, 769-779.

32. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to An External Bath. *J. Chem. Phys.*, **1984,** *81*, 3684-3690.

33. 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*, 10089-10092.

34. Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.*, **2002,** *23*, 1623-1641.

35. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.*, **2000,** *21*, 132-146.

36. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.*, **2004,** *25*, 1157-1174.

37. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of A System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.*, **1977,** *23*, 327-341.

38. Marsaglia, G. Choosing a Point from the Surface of a Sphere. *The Ann. of Math. Stat.*, **1972,** *43*, 645-646.

39. Warwicker, J.; Watson, H. C. Calculation of the Electric-Potential in the Active-Site Cleft Due to Alpha-Helix Dipoles. *J. Mol. Biol.*, **1982,** *157*, 671-679.

40. Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J. Phys. Chem.*, **1994,** *98*, 1978-1988.

41. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, **1998,** *102*, 3586-3616.

42. Roux, B.; Simonson, T. Implicit solvent models *Biophysical Chemistry*, **1999,** *78*, 1-20.

43. Mobley, D. L.; Dill, K. A.; Chodera, J. D. Treating entropy and conformational changes in implicit solvent Simulations of small molecules. *J. Phys. Chem. B*, **2008,** *112*, 938-946.

44. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *J. Phys. Chem.*, **1993,** *97*, 10269-10280.

45. Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.*, **2009,** *5*, 350-358.

46. Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. *J. Phys. Chem. B*, **2007,** *111*, 2242-2254.

47. Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. Estimation of absolute free energies of hydration using continuum methods: Accuracy of partial, charge models and optimization of nonpolar contributions. *J. Chem. Theory Comput.*, **2006,** *2*, 128-139.

48. Gallicchio, E.; Levy, R. M. AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.*, **2004,** *25*, 479-499.

49. Pitera, J. W.; van Gunsteren, W. F. The importance of solute-solvent van der Waals interactions with interior atoms of biopolymers. *J. Am. Chem. Soc.*, **2001,** *123*, 3163-3164.

50. Choudhury, N.; Pettitt, B. M. Local density profiles are coupled to solute size and attractive potential for nanoscopic hydrophobic solutes. *Molecular Simulation*, **2005,** *31*, 457-463.

51. Wagoner, J. A.; Baker, N. A. Assessing implicit models for nonpolar mean solvation forces: The importance of dispersion and volume terms. *Proc. Natl. Acad. Sci. U. S. A.*, **2006,** *103*, 8331-8336.

52. Nicholls, A.; Honig, B. A Rapid Finite-Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson-Boltzmann Equation. *J. Comput. Chem.*, **1991,** *12*, 435-445.

53. Varga, R. S. *Matrix Iterative Analysis;* 2 ed.; Springer: New York, 2000.

54. Becke, A. D. Density-Functional Thermochemistry .3. the Role of Exact Exchange. *J. Chem. Phys.*, **1993,** *98*, 5648-5652.

55. Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.*, **1993,** *98*, 1372-1377.

56. *Gaussian 03, Revision C.02*; Wallingford CT, USA, 2004

57. Kendall, R. A.; Dunning, J.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.*, **1992,** *96*, 6796-6806.

58. Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.*, **1984,** *80*, 3265-3269.

59. Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. Efficient Diffuse Function-Augmented Basis-Sets for Anion Calculations .3. the 3-21+G Basis Set for 1St-Row Elements, Li-F. *J. Comput. Chem.*, **1983,** *4*, 294-301.

60. Woon, D. E.; Dunning, J. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.*, **1993,** *98*, 1358-1371.

61. Hammond, J. R.; Kowalski, K.; deJong, W. A. Dynamic polarizabilities of polyaromatic hydrocarbons using coupled-cluster linear response theory. *J. Chem. Phys.*, **2007,** *127*, 144105.

62. Dykstra, C. E.; Jasien, P. G. Derivative Hartree-Fock Theory to All Orders. *Chem. Phys. Lett.*, **1984,** *109*, 388-393.

63. *Handbook of Chemistry and Physics;* 83 ed.; CRC Press: New York, 2002.

64. Pauling, L. *The Nature of the Chemical Bond;* Cornell University Press: Ithaca: New York, 1960.

65. Shanker, B.; Applequist, J. Polarizabilities of nitrogen heterocyclic molecules from atom monopole dipole interaction theory. *J. Phys. Chem.*, **1996,** *100*, 3879-3881.

66. Bode, K. A.; Applequist, J. A new optimization of atom polarizabilities in halomethanes, aldehydes, ketones, and amides by way of the atom dipole interaction model. *J. Phys. Chem.*, **1996,** *100*, 17820-17824.

67. Harder, E.; Anisimov, V. M.; Whitrield, T.; MacKerell, A. D.; Roux, B. Understanding the dielectric properties of liquid amides from a polarizable force field. *J. Phys. Chem. B*, **2008,** *112*, 3509-3521.

68. Sekino, H.; Maeda, Y.; Kamiya, M.; Hirao, K. Polarizability and second hyperpolarizability evaluation of long molecules by the density functional theory with long-range correction. *J. Chem. Phys.*, **2007,** *126*, 014107.

69. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. *J. Phys. Chem. A*, **2004,** *108*, 621-627.

70. Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.*, **1984,** *5*, 129-145.

71. Eksterowicz, J. E.; Miller, J. L.; Kollman, P. A. Calculation of chloroform/water partition coefficients for the N-methylated nucleic acid bases. *J. Phys. Chem. B*, **1997,** *101*, 10971-10975.

72. Rowland, R. S.; Taylor, R. Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der Waals radii. *J. Phys. Chem.*, **1996,** *100*, 7384-7391.

73. Nina, M.; Beglov, D.; Roux, B. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B*, **1997,** *101*, 5239-5248.

74. Jayaram, B.; Liu, Y.; Beveridge, D. L. A modification of the generalized Born theory for improved estimates of solvation energies and pK shifts. *J. Chem. Phys.*, **1998,** *109*, 1465-1471.

75. Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Modeling electrostatic effects in proteins. *Biochimica et Biophysica Acta-Proteins and Proteomics*, **2006,** *1764*, 1647-1676.

76. Green, D. F.; Tidor, B. Evaluation of a initio charge determination methods for use in continuum solvation calculations. *J. Phys. Chem. B*, **2003,** *107*, 10261-10273.

77. Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of protein–ligand binding free energy by using a polarizable potential. *Proceedings of the National Academy of Sciences*, **2008,** *105*, 6290-6295.

5 Conclusion

L'objectif principal de cette thèse était de valider un nouveau cadre méthodologique pour inclure une composante électrostatique polarisable générale et précise dans un champ de force biomoléculaire. La méthode EPIC, qui modélise la polarisation électronique par un continuum diélectrique intramoléculaire, s'est avérée meilleure que les méthodes existantes, c'est-à-dire les dipôles atomiques induits et les charges fluctuantes. En ce qui a trait au modèle des charges fluctuantes, leur complète incapacité à modéliser la composante de la polarisabilité perpendiculaire aux plans ou aux liens chimiques, en fait une méthode très peu générale. Par contre, notre méthode se compare plus justement avec celle des dipôles atomiques induits sur plusieurs comptes.

Premièrement, la polarisabilité moléculaire a pu être obtenue avec un nombre considérablement inférieur de paramètres et une complexité réduite. En particulier, l'anisotropie de la polarisabilité est naturellement modélisée sur une panoplie de molécules reconnues difficiles pour ces modèles: aromatiques, diatomiques, alcanes. Avec seulement 14 rayons atomiques optimisés sur un ensemble de 265 tenseurs de polarisabilité, nous avons pu démontrer que la précision de la méthode EPIC est comparable à celle des calculs de structure électronique avancés. Les fonctionnalités chimiques couvertes par notre paramétrisation globale sont: alcanes, alcènes, alcynes, halogènes (bromo, fluoro, chloro), alcools, thiols, amines, éthers, thioéthers, nitriles, aldéhydes, cétones, esters, thioesters, amides, acides carboxyliques, urées, imines, amidines, sulfones, sulfoxides, sulfonamides, hétéroaromatiques, hydrazines, acides hydroxamiques, N-oxides, nitro, pyridones, etc. Ces groupements fonctionnels incluent les acides aminés, les acides nucléiques, les sucres, les lipides, les acides gras, etc. L'anisotropie de la polarisabilité est certainement une quantité à bien reproduire étant donné le type de groupements fonctionnels qui composent les

molécules biologiquement actives que l'on souhaiterait étudier. Même pour des molécules aussi simples que des amides moléculaires, il a été récemment démontré que l'anisotropie est essentielle pour reproduire la constante diélectrique des liquides.

Deuxièmement, puisque la réponse au champ électrique ne provient pas seulement de la position des atomes, mais est distribuée sur le volume électronique, les cas où la perturbation électrique a une certaine symétrie sont bien mieux traités qu'avec les méthodes polarisables traditionnelles. Cela a été démontré clairement avec deux exemples. Il suffit de penser à un anion (Cl^-) dans l'eau, à un soluté qui subit un champ électrique complexe de la première couche de solvant ou à deux ponts-H formés par la partie amide d'un substrat lié à un site actif enzymatique, pour comprendre que cet avantage pourrait être considérable pour les applications typiques d'un champ de force.

Troisièmement, l'approche EPIC ne souffre pas de la catastrophe de la polarisabilité qui est inhérente aux modèles polarisables avec dipôles induits sur les atomes. Ce problème, qui demande une correction *ad hoc* par une fonction d'atténuation dite de Thole, est causé par le fait que la polarisabilité ne change pas quand deux points polarisables se rapprochent. Or, avec EPIC, quand deux atomes liés se rapprochent, le volume polarisable diminue par le recouvrement des fonctions de diélectriques, et donc la polarisabilité totale diminue. La correction de Thole appliquée aux dipôles induits sur les atomes est en partie responsable de la difficulté à généraliser cette approche puisqu'elle devient spécifique à une molécule lorsque l'on a besoin d'obtenir une bonne anisotropie.

Il n'était toutefois pas suffisant de bien calculer des tenseurs de polarisabilité. En effet, le potentiel électrostatique induit est l'élément nécessaire dans un champ de force polarisable. Pour s'assurer que la méthode, optimisée sur des polarisabitlitées moléculaires, fonctionne aussi quand la perturbation n'est pas uniforme, le potentiel électrostatique induit par une perturbation locale non uniforme a été calculé au Chapitre 3. Bien que la polarisabilité moléculaire soit précisément calculée avec certains rayons atomiques ajustés pour la méthode EPIC, l'introduction de la polarisabilité dans un champ de force nécessite

surtout que le potentiel électrostatique induit par une perturbation locale soit précis. Cela a été vérifié au chapitre 3 et les résultats sont concluants.

La question d'établir une méthodologie générale, flexible et fiable pour l'ensemble de la chimie bioorganique a beaucoup influencé notre travail. Tout d'abord, comme mentionné plus haut, EPIC nous a permis de modéliser précisément la polarisabilité de beaucoup de molécules composées de groupements fonctionnels variés avec peu de paramètres. En ce qui concerne la génération des charges partielles atomiques nécessaires pour fixer le potentiel électrostatique permanent, nous avons établi une méthode qui découple complètement l'optimisation des paramètres de polarisabilité de l'ajustement des charges atomiques. Ceci est différent de tout ce que l'on a pu relever dans la littérature. Le découplage est habituellement rendu difficile par la polarisation intramoléculaire qui modifie le potentiel électrostatique produit par la molécule. Notre formalisme, qui repose sur des fonctions de Green faciles à calculer numériquement, permet ce découplage. Plus encore, nous avons démontré que l'on peut transférer les charges atomiques obtenues d'une façon traditionnelle en calculant des fonctions de Green appropriées. Ces processus d'obtention des charges atomiques ou du transfert de charges existantes ne montraient pas de perte au niveau de la précision. Pour résumer, nous avons simplifié le problème de la paramétrisation charges/polarisabilité et pouvons revaloriser les modèles de charges atomiques déjà existants dans le contexte d'un champ de force polarisable qui reposerait sur l'approche EPIC. Il faudra toutefois que les charges utilisées avec EPIC n'aient pas été prépolarisées.

Avec un modèle polarisable précis et paramétrisé sur un large éventail de groupements chimiques et une façon générale pour introduire les charges partielles atomiques, nous avons tous les ingrédients du terme électrostatique pour construire un nouveau champ de force polarisable. Cela a été démontré par deux exemples qui nécessitent un bon traitement de la polarisation électronique: une interaction cation-π et un pont-H formé dans le dimère de 4-pyridone. Il restait à démontrer que le modèle polarisable obtenu

puisse aussi être transféré du cas avec une molécule isolée en phase gazeuse au cas de la phase condensée. Le fait que nous ayons été en mesure de reproduire des indices de réfraction macroscopiques à l'aide d'agrégats sphériques contenant quelques milliers de molécules montre un bon degré de transférabilité. Toutefois, une légère surestimation des indices de réfraction peut être l'indication que certains petits ajustements sont nécessaires.

Nous avons donc accompli les objectifs principaux que nous nous étions fixés. Nos travaux nous ont également permis de faire avancer les connaissances et les méthodes sur plusieurs points qui dépassent nos objectifs premiers.

Premièrement, nous avons mis au point trois techniques qui se sont avérées importantes pour l'avancement de nos travaux. Nous rapportons pour la première fois une approche pour le calcul de la polarisabilité moléculaire en phase gazeuse avec un algorithme aux différences finies. Cela nous a conduits à développer également une méthode théorique pour calculer l'indice de réfraction d'une substance liquide. Une autre nouveauté méthodologique est l'approche DRESP qui a permis d'établir un protocole rigoureux pour déterminer les charges atomiques optimales en présence d'un diélectrique interne supérieur à un.

Deuxièmement, les travaux de cette thèse ont changé considérablement certains dogmes reliés à la signification des paramètres utilisés dans les approches largement utilisées de Poisson-Boltzmann (PB). Nous avons montré que l'égalité $\varepsilon_{in} = \varepsilon_{\infty}$ était arbitraire et avons proposé une façon cohérente pour déterminer ε_{in} qui se fonde sur la réalité physique représentée par le diélectrique interne, c'est-à-dire la polarisation électronique. En effet, pour obtenir des valeurs raisonnables pour le tenseur de polarisabilité moléculaire, il a fallu que $\varepsilon_{in} \gg \varepsilon_{\infty}$. Bien que ce résultat en surprenne plus d'un, nos calculs des indices de réfraction macroscopiques à partir de notre fonction de diélectrique microscopique démontrent que les bons ingrédients physiques sont présents.

Troisièmement, nos travaux ont permis de clarifier le rôle joué par les rayons atomiques qui définissent différentes surfaces. Nous avons obtenu des rayons atomiques optimaux différents de ce que la littérature nous offrait. La taille des rayons atomiques optimaux qui définissent le *volume de polarisation électronique* est plus petite dans nos travaux que celle traditionnellement utilisée dans les calculs PB. D'autre part, nos rayons atomiques, qui décrivent la cavité formée par le soluté dans le solvant, sont plus grands que ceux utilisés dans les calculs PB. Cela peut sembler surprenant étant donné le pouvoir prédictif des énergies libres d'hydratation de ces approches, tel que rapporté dans la littérature. En réalité, nous avons montré que les paramètres utilisés dans les calculs avec solvant implicite, peuvent compenser l'un pour l'autre et donc, que ce pouvoir prédictif n'est pas gage d'un bon fondement physique. Nous avons résolu cette apparente contradiction en élaborant une fonction de diélectrique à trois zones qui distingue le phénomène de polarisation électronique (première zone) et de polarisation du solvant (troisième zone). La zone de transition a une polarisabilité nulle, ce qui peut être compris comme un endroit où la densité électronique du soluté et du solvant est très faible. Le lien avec le concept de répulsion de Fermi est évident, ce qui nous permet de dire que cette zone transitoire peut être définie comme la zone de contact soluté/solvant. Le centre de cette deuxième zone correspond assez fidèlement aux rayons dits de contact (comme les rayons de Bondi) utilisés dans les calculs de PB. L'aspect rassurant du modèle élaboré est que chacune des composantes peut être optimisée pour reproduire uniquement le phénomène physique associé: les rayons de polarisation vs polarisabilité, charges atomiques vs potentiel électrostatique en phase gazeuse, rayons de cavité vs énergie libre d'hydratation. L'aspect d'autant plus encourageant est, qu'avec cette approche, nous avons été capables de reproduire les énergies libres d'hydratation expérimentales de 485 molécules avec une erreur RMS entre 1.2 et 1.4 kcal/mol, ce qui constitue même une amélioration par rapport aux précédents de la littérature avec calculs PB qui était de 2 kcal/mol. La méthodologie utilisée nous permet de croire que cette exactitude accrue est en partie due à un meilleur modèle physique pour la polarisation électronique, le potentiel électrostatique permanent et

le bon niveau de polarisation du solvant – composantes importantes pour un champ de force polarisable.

Quatrièmement, nous avons proposé initialement au chapitre 2 que l'intérieur d'une molécule devrait avoir des régions avec des constantes diélectriques internes différentes pour rendre des polarisabilités moléculaires précises. Nous avons montré que l'utilisation d'une frontière diélectrique lisse (*smooth*) permet d'utiliser une seule constante diélectrique pour l'intérieur de molécules sans perdre trop de précision.

Les résultats obtenus dans cette thèse ouvrent la voie à beaucoup d'avenues pour les travaux futurs. D'abord, il y a le développement d'un champ de force à partir du terme électrostatique polarisable élaboré aux chapitres précédents. En effet, les termes intramoléculaires et intermoléculaires de Lennard-Jones devront être revus. En particulier, les potentiels des angles de torsion devraient nécessiter un effort considérable. Il est possible de croire que l'inclusion de la polarisabilité réduira la complexité des potentiels de torsion. Fort heureusement, il y a déjà une mécanique connue pour optimiser tous ces termes.

L'idée d'utiliser EPIC pour décrire complètement l'électrostatique dans une simulation Monte Carlo ou de dynamique moléculaire n'a pas été validée, malgré que des résultats partiels qui n'ont pas été présentés dans cette thèse, laissent croire que cela serait possible.

Puisque la méthode EPIC requiert une solution numérique, il sera important d'améliorer la vitesse des calculs numériques. L'implémentation de la méthode par différences finies faites pour les travaux de cette thèse est environ 1000 fois plus lente que le simple calcul des interactions électrostatiques par la loi de Coulomb. Avec une approche *multigrid*, nous avons pu vérifier que ce facteur n'est plus que de 10 fois, ce qui rivalise avec les approches polarisables rapides comme les oscillateurs de Drude. Il est de plus possible de tirer avantage de la parallélisation efficace des approches multigrid. Il faudra

s'assurer de mettre au point des méthodes numériques rapides et fiables pour le calcul des forces électrostatiques.

Le but ultime de l'élaboration d'un champ de force plus précis et général demeure éloigné et nécessite encore un effort considérable. Comme la première étape dans un processus de développement de champ de force est la partie électrostatique, nous avons franchi un bon pas. En effet, les étapes restantes sont moins incertaines puisque la méthodologie est connue. L'implémentation de la solution à l'équation Poisson numériquement rapide qui calcule les forces, l'ajout des termes de vdW et des énergies internes dans un champ de force nous permettra d'utiliser les concepts développés dans cette thèse pour les appliqués à toutes sortes de problèmes importants tant au niveau industriel que fondamental.

Annexe I

Déclaration des coauteurs

Nom de l'étudiant : Jean-François Truchon
Nom du programme : Philosophiæ Doctor (Ph.D.) en chimie

Auteurs : Jean-François Truchon, Anthony Nicholls, Radu I. Iftimie, Benoît Roux et Christopher I. Bayly
Titre : Accurate molecular polarizabilities based on continuum electrostatics
Journal : Journal of Chemical Theory and Computing
Volume : 4
Page : 1480
Année : 2008

À titre de coauteur de l'article identifié ci-dessus, je suis d'accord pour que Jean-François Truchon inclue cet article dans sa thèse de doctorat qui a pour titre Modéliser la polarisation électronique par un continuum diélectrique intramoléculaire : Vers un champ de force polarisable pour la chimie bioorganique.

Anthony Nicholls

Coauteur	Signature	Date

Radu I. Iftimie

Coauteur	Signature	Date

Benoît Roux

Coauteur	Signature	Date

Christopher Bayly

Coauteur	Signature	Date

Nom de l'étudiant : Jean-François Truchon
Nom du programme : Philosophiæ Doctor (Ph.D.) en chimie

Auteurs : Jean-François Truchon, Anthony Nicholls, J. Andrew Grant, Radu I. Iftimie, Benoît Roux et Christopher I. Bayly
Titre : Extending the polarizable continuum dielectric model to account for electronic polarization in intermolecular interactions
Journal : soumis à Journal of Computational Chemistry

À titre de coauteur de l'article identifié ci-dessus, je suis d'accord pour que Jean-François Truchon inclue cet article dans sa thèse de doctorat qui a pour titre Modéliser la polarisation électronique par un continuum diélectrique intramoléculaire : Vers un champ de force polarisable pour la chimie bioorganique.

Anthony Nicholls

Coauteur	Signature	Date

J. Andrew Grant

Coauteur	Signature	Date

Radu I. Iftimie

Coauteur	Signature	Date

Benoît Roux

Coauteur	Signature	Date

Christopher Bayly

Coauteur	Signature	Date

Nom de l'étudiant : Jean-François Truchon
Nom du programme : Philosophiæ Doctor (Ph.D.) en chimie

Auteurs : Jean-François Truchon, Anthony Nicholls, Benoît Roux, Radu I. Iftimie et Christopher I. Bayly
Titre :

Journal : soumis à Journal of Chemical Theory and Computation

À titre de coauteur de l'article identifié ci-dessus, je suis d'accord pour que Jean-François Truchon inclue cet article dans sa thèse de doctorat qui a pour titre Modéliser la polarisation électronique par un continuum diélectrique intramoléculaire : Vers un champ de force polarisable pour la chimie bioorganique.

Anthony Nicholls		
Coauteur	Signature	Date
Benoît Roux		
Coauteur	Signature	Date
Radu I. Iftimie		
Coauteur	Signature	Date
Christopher Bayly		
Coauteur	Signature	Date

Annexe II

Dérivation de l'équation de Poisson

Dans cette annexe, nous obtiendrons l'équation de Poisson en partant du concept de densité de polarisabilité.

Définissons le champ vectoriel $\vec{P}(\vec{r})$, que nous appellerons polarisation, qui correspond au dipôle volumique local, la densité dipolaire induite pour chaque élément de volume dV, et calculons le moment dipolaire associé à chaque point de l'espace avec

$$d\vec{\mu}^{ind}(\vec{r}) = \vec{P}(\vec{r})dV \tag{II. 1}$$

ce qui nous permet de calculer le moment dipolaire total pour la molécule ou le système étudié

$$\vec{\mu}^{ind} = \int \vec{P}(\vec{r})dV \tag{II. 2}$$

Maintenant, reprenons l'idée de la réponse linéaire au champ électrique et définissons une densité de polarisabilité χ, aussi appelée la susceptibilité électrique, qui donne lieu à la polarisation

$$\vec{P}(\vec{r}) = \chi(\vec{r}) \cdot \vec{E}(\vec{r}) \tag{II. 3}$$

Ici $\vec{E}(\vec{r})$ est le champ électrique total à chaque point de l'espace. De façon plus générale, la susceptibilité pourrait être un tenseur. Nous aurions également pu inclure d'autres termes dans le développement en série de Taylor comme la densité de moment quadrupolaire induit par le champ électrique. Par contre, ces complications mathématiques ne semblent pas nécessaires, et nous les évitons.

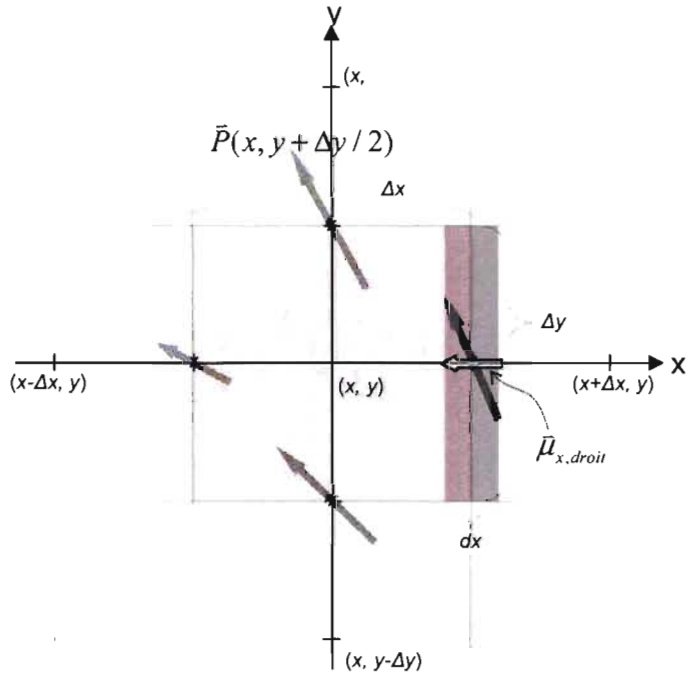


Figure II.1. Discrétisation de la polarisation d'un milieu diélectrique.

Le concept de la susceptibilité électrique n'est pas nouveau dans cette thèse. Ce qui est différent est le contexte de son utilisation. En fait, nous utilisons la susceptibilité électrique pour décrire la réponse électronique intramoléculaire, c'est-à-dire l'effet du déplacement des électrons qui subissent un champ électrique externe. Cela fait contraste avec l'usage plutôt macroscopique de la susceptibilité électrique en électrostatique. En mécanique quantique, ceci se traduit par une différence de densité électronique $\delta\rho(r)$ causée par l'application d'un champ électrique qui perturbe la densité électronique déjà présente $\rho(r)$. Relions à présent les éq. (II. 1), (II. 2) et (II. 3) à $\delta\rho(r)$. Ce faisant, nous obtiendrons le potentiel électrostatique induit, ce qui donne accès, à tout le moins par la loi de Coulomb, à la nouvelle énergie du système. Pour développer une stratégie mathématique, jetons un coup d'œil à la Figure II.1. Au centre, un carré centré à (x,y) possède une surface ΔS =

$\Delta x \Delta y$*. Commençons par rappeler qu'un dipôle peut se représenter avec un vecteur portant des charges de signe opposé à ses extrémités. Le dipôle d'un tel arrangement est alors

$$\bar{\mu} = q \times \bar{l} \tag{II. 4}$$

où $\bar{l}$ est un vecteur parallèle au dipôle qui porte une charge négative à son extrémité d'origine et une charge positive à son point final. Cette formule est valide pour un dipôle ponctuel également, mais dans la mesure où $|\bar{l}|$ est petit. À la Figure II.1, nous pouvons imaginer que le champ de polarisation qui traverse les parois de l'élément de surface ΔS, posera une extrémité dipolaire positive d'un côté et une extrémité négative de l'autre. Ainsi, la densité surfacique de charge produite à l'intérieur de ΔS par le champ de polarisation $\vec{P}(x, y)$ permettra d'évaluer $\delta\rho(x,y)$. Il faut réaliser que seulement la polarisation située sur la ligne contour du carré contribuera puisque les points à l'intérieur du carré voient leurs charges dipolaires s'annuler. Premièrement, trouvons le moment dipolaire produit par la surface du rectangle centré sur le côté droit de ΔS et de largeur dx

$$\begin{aligned} \bar{\mu}_{droit} &= \left[\int_{y-\Delta y/2}^{y+\Delta y/2} \vec{P}(x+\Delta x/2, y) dy \right] dx \\ &= \left[\vec{P}(x+\Delta x/2, y) \Delta y \right] dx \end{aligned} \tag{II. 5}$$

où $\bar{\mu}_{droit}$ est le dipôle produit par le rectangle d'intégration, de dimension $dx\Delta y$. Ici, nous supposons que les dimensions choisies sont suffisamment petites pour que le vecteur de polarisation soit constant sur la surface d'intégration. Si nous décomposons $\bar{\mu}_{droit}$ en $\mu_{x,droit}$ et $\mu_{y,droit}$, il apparaît que la composante en y n'apportera aucune charge à l'intérieur de ΔS. Ainsi, nous pouvons nous concentrer uniquement sur $\mu_{x,droit}$ qui a une composante nulle en y. De plus, reprenant l'éq. (II. 4), nous pouvons calculer la charge que porterait un vecteur

* L'utilisation de deux dimensions permet de simplifier les équations tout en demeurant général et valide pour 3 coordonnées spatiales.

dirigé en x de longueur dx de telle sorte que $\mu_{x,droit} = q_{droit}dx$, ce qui nous conduit à l'éq. (II. 6):

$$\begin{aligned} \frac{\mu_{x,droit}}{dx} &= P_x(x+\Delta x/2, y)\Delta y \\ \Rightarrow \frac{q_{droit}dx}{dx} &= -P_x(x+\Delta x/2, y)\Delta y \\ \Rightarrow \quad q_{droit} &= -P_x(x+\Delta x/2, y)\Delta y \end{aligned} \tag{II. 6}$$

Si P_x est positif et sort du carré, il contribue à la charge totale à l'intérieur de la surface ΔS avec charge négative et, par convention, cela fait apparaître une charge négative dans éq. (II. 6). De la même façon, on obtient la charge amenée dans le carré par le côté gauche:

$$q_{gauche} = P_x(x-\Delta x/2, y)\Delta y \tag{II. 7}$$

Nous pouvons alors calculer la contribution horizontale à la densité surfacique de charge

$$\Delta\rho_x = \frac{q_{gauche}+q_{droit}}{\Delta S} = -\frac{P_x(x+\Delta x/2, y)-P_x(x-\Delta x/2, y)}{\Delta x} \tag{II. 8}$$

La même démarche se fait pour le haut et le bas du carré permettant le calcul de la charge surfacique amenée par le champ de polarisation. L'étape suivante est de prendre la limite quand Δx et Δy tendent vers zéro. Le résultat d'un tel processus est connu en calcul différentiel: c'est la divergence du champ de polarisation. Cela nous conduit à

$$\begin{aligned} \delta\rho &= -\vec{\nabla}\cdot\vec{P}(\vec{r}) \\ &= -\vec{\nabla}\cdot\left(\varepsilon_0\chi(\vec{r})\vec{E}(\vec{r})\right) \end{aligned} \tag{II. 9}$$

où éq. (II. 3) a été utilisée. L'équation (II. 9) est une équation différentielle locale qui permet de calculer la densité de charge induite à chaque point (x,y) de l'espace, une fois que le champ électrique total est connu (qui lui-même a une composante induite). Or, comme le champ électrique total dépend également de la densité de charge permanente, définissons la densité de charge totale $\rho(\vec{r})$

$$\begin{aligned} \rho(\vec{r}) &= \rho^f(\vec{r}) + \delta\rho(\vec{r}) \\ &= \rho^f(\vec{r}) + \rho^b(\vec{r}) \end{aligned} \tag{II. 10}$$

avec l'exposant *f* signifiant *free* et *b bound*, ce qui suit une nomenclature déjà établie. L'étape finale de notre développement est d'utiliser la loi de Gauss qui s'applique à la densité de charge totale à chaque point de l'espace. Voici un rappel de l'équation de Gauss

$$\vec{\nabla}\cdot\vec{E}(\vec{r})=\frac{\rho(\vec{r})}{\varepsilon_0} \tag{II. 11}$$

avec ε_0 qui est la permittivité du vide. Nous pouvons réécrire l'équation de Gauss dans le contexte des éq. (II. 9) et (II. 10)

$$\begin{aligned}\varepsilon_0\vec{\nabla}\cdot\vec{E}(\vec{r})&=\rho^f(\vec{r})-\vec{\nabla}\cdot\left(\varepsilon_0\chi(\vec{r})\vec{E}(\vec{r})\right)\\ \vec{\nabla}\cdot\left[\varepsilon_0(\chi(\vec{r})+1)\vec{E}(\vec{r})\right]&=\rho^f(\vec{r})\end{aligned} \tag{II. 12}$$

Si l'on définit une fonction $\varepsilon(r)$ comme

$$\chi(\vec{r})=\varepsilon(\vec{r})-1 \tag{II. 13}$$

cela nous amène à l'équation de Poisson, la pierre angulaire de cette thèse

$$\vec{\nabla}\cdot\left[\varepsilon_0\varepsilon(\vec{r})\vec{E}(\vec{r})\right]=\rho^f(\vec{r}) \tag{II. 14}$$

En réécrivant pour le potentiel électrostatique $-\vec{\nabla}\varphi(\vec{r})=\vec{E}(\vec{r})$, une formulation équivalente de l'équation de Poisson est

$$\vec{\nabla}\cdot\left[\varepsilon_0\varepsilon(\vec{r})\vec{\nabla}\varphi(\vec{r})\right]=-\rho^f(\vec{r}). \tag{II. 15}$$

Annexe III

Où la densité de charge induite apparaît-elle?

Les charges liées (*bound*) ou induites n'apparaissent qu'aux régions de l'espace où la fonction de diélectrique varie. Par exemple, au chapitre 1, le calcul du potentiel électrostatique pour le problème de la sphère diélectrique sous l'influence d'un champ électrique uniforme prédit qu'une densité de charge apparaît à la surface de la sphère, l'unique endroit où la fonction de diélectrique varie. Ceci est général et en voici la démonstration. Utilisons d'abord les éq. (II. 9) et (II. 13), pour écrire les égalités suivantes:

$$\begin{aligned}\rho^b(\bar{r}) &= -\bar{\nabla}\cdot\bar{P}(\bar{r}) \\ &= -\bar{\nabla}\cdot\left[\varepsilon_0\left(\varepsilon(\bar{r})-1\right)\bar{E}(\bar{r})\right] \\ &= -\varepsilon_0\bar{E}(\bar{r})\cdot\bar{\nabla}\varepsilon(\bar{r}) - \varepsilon_0\varepsilon(\bar{r})\bar{\nabla}\cdot\bar{E}(\bar{r}) + \varepsilon_0\bar{\nabla}\cdot\bar{E}(\bar{r}) \\ &= -\varepsilon_0\bar{E}(\bar{r})\cdot\bar{\nabla}\varepsilon(\bar{r}) + \varepsilon_0\left(1-\varepsilon(\bar{r})\right)\bar{\nabla}\cdot\bar{E}(\bar{r}) \qquad \text{(III. 1)}\end{aligned}$$

Supposons maintenant que la fonction diélectrique $\varepsilon(\bar{r})$ ne varie pas en $\bar{r}$, c'est-à-dire que $\bar{\nabla}\varepsilon(\bar{r}) = 0$, et éq. (III. 1) s'écrit

$$\begin{aligned}\rho^b(\bar{r}) &= \varepsilon_0\left(1-\varepsilon(\bar{r})\right)\bar{\nabla}\cdot\bar{E}(\bar{r}) \\ &= \left(1-\varepsilon(\bar{r})\right)\left(\rho^b(\bar{r})+\rho^f(\bar{r})\right) \\ &= \varepsilon(\bar{r})\rho^b(\bar{r}) + \rho^b(\bar{r}) - \varepsilon(\bar{r})\rho^f(\bar{r}) \qquad \text{(III. 2)}\end{aligned}$$

où l'on a fait usage de la loi de Gauss $\varepsilon_0\bar{\nabla}\cdot\bar{E} = \rho = \rho^b + \rho^f$. L'éq. (III. 2) se simplifie pour donner éq. (III. 3).

$$\rho^b(\bar{r}) = \left(\frac{1-\varepsilon(\bar{r})}{\varepsilon(\bar{r})}\right)\rho^f(\bar{r}) \qquad \text{(III. 3)}$$

L'éq. (III. 3) signifie que la densité de charge induite sera nulle localement si a) la fonction de diélectrique a un gradient nul (ne varie pas localement) et la densité de charge libre est nulle b) le diélectrique local est 1. Dans le cas où la fonction de diélectrique ne varie pas

localement, mais que la densité de charge libre n'est pas nulle, la densité de charge totale est donnée par

$$\rho(\vec{r}) = \rho^f(\vec{r}) + \rho^b(\vec{r})$$
$$= \frac{\rho^f(\vec{r})}{\varepsilon(\vec{r})} \qquad \text{(III. 4)}$$

où l'on a combiné éq. (III. 3) et éq. (II. 10). Donc, l'effet d'écrantage dont il est fait mention au Chapitre 3 provient de deux phénomènes. Premièrement, les charges atomiques partielles qui sont des charges libres sont situées dans des régions où la fonction diélectrique est une constante. Ainsi, leur effet est divisé par ε_{in} ce qui devrait nécessiter des charges 12 fois plus grandes pour produire le même potentiel électrostatique. La cause en est la création d'une densité locale induite locale qui écrante la charge de beaucoup selon éq. (III.3) Deuxièmement, la variation de la fonction de diélectrique à la surface de polarisation induit une densité de charge de signe opposée à celle donnée par éq. (III.3) puisque le gradient du diélectrique est négatif. Ceci compense pour l'effet 1/ ε_{in} de telle sorte que pour le cas d'une charge centrée dans un diélectrique sphérique (modèle de Born), le potentiel électrostatique à l'extérieur de la sphère demeure le même que si ε_{in} = 1 comme le montrent les équations analytiques. Cela est indépendant de la taille de la sphère. Par ailleurs, dans le cas où la charge dans la sphère est décentrée vers la droite, les équations analytiques de Tanford-Kirkwood* prévoient que le potentiel du côté gauche, à l'extérieur de la sphère, sera moindre que si ε_{in} = 1. Par contre, le facteur n'est pas aussi grand que 1/ ε_{in}. Ceci montre que dans le cas d'une cavité de forme générale, cet effet se fera sentir de façon non évidente.

---

* Tanford, C. et Kirkwood, J.G. Theory of protein titration curves. I. General equations for impenetrable spheres. *Journal of the American Chemical Society*, **79**, (1957) 5333-5347

Annexe IV

Unités des calculs par différences finies

La conversion des unités en électrostatique pose souvent des difficultés. Dans les calculs numériques, il faut utiliser des grandeurs raisonnables pour les variables qui ne sont pas en S.I. Dans cette annexe, les unités utilisées dans le programme écrit pour cette thèse qui résout l'équation de Poisson par la méthode des différences finies sont présentées. Posons d'abord que l'unité de charge est celle du proton, l'unité de distance le Å, le diélectrique est relatif à la permittivité du vide et l'unité d'énergie est le *RT* (énergie associée à la température).

Le potentiel électrostatique. L'unité d'énergie est le *RT* (*T* est la température et R la constante des gaz parfaits). Définissons d'abord le *RT* et sa conversion en d'autres unités:

$1RT = 8.314472\ \mathrm{JK^{-1}mol^{-1}} \times 298.15\mathrm{K} \times 0.001\ \mathrm{kJ/J} = 2.478960\ \mathrm{kJ/mol}$

$= 2.478960\ \mathrm{kJ/mol} / (4.184\ \mathrm{kJ/kcal})$

$= 0.5924856\ \mathrm{kcal/mol}$

Le système d'équations linéaires résolues en différences finies est le suivant

$$\phi_{ijk}\left[\left(\frac{h_y h_z}{h_x}\right)\left(\varepsilon_{i-1/2\,jk}+\varepsilon_{i+1/2\,jk}\right)+\left(\frac{h_x h_z}{h_y}\right)\left(\varepsilon_{ij-1/2k}+\varepsilon_{ij+1/2k}\right)+\left(\frac{h_x h_y}{h_z}\right)\left(\varepsilon_{ijk-1/2}+\varepsilon_{ijk+1/2}\right)\right]$$
$$-\phi_{i-1\,jk}\left(\frac{h_y h_z}{h_x}\right)\varepsilon_{i-1/2\,jk}-\phi_{i+1\,jk}\left(\frac{h_y h_z}{h_x}\right)\varepsilon_{i+1/2\,jk}-\phi_{ij-1k}\left(\frac{h_x h_z}{h_y}\right)\varepsilon_{ij-1/2k}-\phi_{ij+1k}\left(\frac{h_x h_z}{h_y}\right)\varepsilon_{ij+1/2k}$$
$$-\phi_{ijk-1}\left(\frac{h_x h_y}{h_z}\right)\varepsilon_{ijk-1/2}-\phi_{ijk+1/2}\left(\frac{h_x h_y}{h_z}\right)\varepsilon_{ijk+1/2}=q_{ijk}$$

avec ϕ_{ijk} le potentiel électrostatique au point de grille *ijk*, h_x le pas de grille suivant l'axe des x, $\varepsilon_{i+1/2\,jk}$ la fonction de diélectrique évaluée à mi-chemin entre les points de grille *ijk* et

$i+1jk$, et q_{ijk} la charge libre totale dans le volume couvert par le point de grille ijk (il ne s'agit pas de la densité de charge, mais de son intégrale sur le volume V_{ijk}). Supposons que nous voulions le potentiel en SI (*J/C*) alors que les unités utilisées sont celles mentionnées plus haut. Il faudrait alors multiplier le potentiel obtenu numériquement* par 180.95126:

$$[\varphi]_{SI} = \frac{[Q]_E}{[L]_E[\varepsilon]_E} \times \frac{\mathbf{C}(C/e)}{10^{-10}(m/Angstrom)\mathbf{E}(C^2J^{-1})}$$
$$= \frac{[Q]_E}{[L]_E[\varepsilon]_E} \times 180.95126(J/C)$$

avec **C** = 1.602176E-19 la constante de Coulomb et **E** = 8.8541878E-12 la permittivité du vide. Si l'on voulait que le programme exprime naturellement le potentiel obtenu en *J/C*, nous pourrions multiplier les charges q_{ijk} par 180.95126 lors de l'initialisation tout en conservant le système unitaire choisi pour les autres quantités. Toutefois, nous préférons que le potentiel obtenu soit plutôt en RT/e où e est la charge d'un proton. La constante multiplicatrice des charges devrait plutôt être 7042.9316:

$$[\varphi]_{RT/e} = [\varphi]_{SI}(J/C) \times 2478.960^{-1}(RTmol/J) \times \mathbf{C}(C/e) \times \mathbf{N}_a(mol^{-1})$$
$$= [\varphi]_{SI} \times 38.921705(RT/e)$$
$$[\varphi]_{RT/e} = \frac{[Q]_E}{[L]_E[\varepsilon]_E} \times 7042.9316(RT/e)$$

Cette multiplication des charges nous amène à être prudents lorsque l'on doit utiliser des charges obtenues par calcul.

Conversion des charges induites. Plus particulièrement, le calcul numérique des charges induites (liées) est fait à partir du potentiel obtenu aux points de grille qui est en *RT/e*. Pour convertir les charges obtenues en unités de *e*, regardons l'équation aux différences finies utilisée pour calculer la charge induite au point de grille *ijk*:

---

* La notation avec crochets pour $[A]_B$ signifie la valeur numérique de la propriété A exprimée dans les unités B. Quand les unités sont celles utilisées dans le programme (comme la charge du proton ou la distance en Å), on appelle ce système d'unités E. Les lettres en caractère gras sont des constantes universelles en SI. Les unités sont données entre parenthèses.

$$\frac{q_{ijk}^{f}+q_{ijk}^{b}}{\varepsilon_0}=-\left(\frac{h_y h_z}{h_x}\right)\left(\phi_{i+1jk}+\phi_{i-1jk}-2\phi_{ijk}\right)-\left(\frac{h_x h_z}{h_y}\right)\left(\phi_{ij+1k}+\phi_{ij-1k}-2\phi_{ijk}\right)-\left(\frac{h_x h_y}{h_z}\right)\left(\phi_{ijk+1}+\phi_{ijk-1}-2\phi_{ijk}\right)$$

qui se simplifie dans le cas où $h = h_x = h_y = h_z$:

$$q_{ijk}^{b}=-q_{ijk}^{f}+6h\varepsilon_0\left(\phi_{ijk}-\frac{\phi_{i+1jk}+\phi_{i-1jk}+\phi_{ij+1k}+\phi_{ij-1k}+\phi_{ijk+1}+\phi_{ijk-1}}{6}\right).$$

La charge en SI se calcule comme suit

$$[Q]_{SI}=[L]_{SI}[\varphi]_{SI}\times(m)(J/C)\times\mathbf{E}\times(C^2/N/m^2)$$

et l'on peut changer les unités de distance et de potentiel avec

$$\begin{aligned}[Q]_{SI}&=[L]_E[\varphi]_{RT/e}\mathbf{E}/38.921705(J/C)^{-1}(RT/e)\times10^{-10}(m/Angstrom)\\&=[L]_E[\varphi]_{RT/e}\times2.274871519\times10^{-23}(C)\\[Q]_E&=[L]_E[\varphi]_{RT/e}\times2.274871519\times10^{-23}(C)\times\mathbf{C}^{-1}(e/C)\\&=[L]_E[\varphi]_{RT/e}\times1.41986325\times10^{-4}(e)\\[Q]_E&=[L]_E[\varphi]_{RT/e}/7042.9316(e)\end{aligned}$$

Il faut donc diviser le résultat obtenu par 7042.9316, ce qui est complètement cohérent avec la transformation des charges libres qui est faite initialement.

Calcul de la polarisabilité. Si le potentiel est calculé en unités de *RT/e* et les distances en Å, le champ électrique sera en *RT/e/Å*. Si on veut que le champ électrique soit uniforme, on peut fixer le potentiel à une borne à zéro et il faut trouver le potentiel à l'autre borne:

$$\frac{\Delta\phi}{\Delta L}=E\Rightarrow\phi(L)=E\cdot\Delta L$$

Ici ΔL est la longueur totale de la grille dans la direction où le champ électrique est appliqué. Donc, les valeurs numériques assignées au potentiel sur les frontières de la grille sont en *RT/e* et ne nécessitent aucune conversion puisque n'importe quelle valeur est acceptable étant donné que la réponse donnée par l'équation de Poisson est linéaire. Il est donc maintenant plus important de convertir la polarisabilité obtenue en un système d'unités connues. Prenons la formule pour le dipôle

$$\bar{\mu} = \sum_{m=0}^{M} q_m \bar{r}_m = \alpha \bar{E}$$

qui calcule le dipôle produit par les charges q_m positionnées en $\bar{r}_m$. Nous pouvons alors calculer la polarisabilité en SI à partir des unités utilisées dans le programme

$$[\alpha]_{SI} = \frac{[Q]_{SI}[L]_{SI}}{[E]_{SI}}$$

$$[\alpha]_{SI} = \frac{[Q]_E[L]_E}{[E]_E}(e)(A)(RT)^{-1}(e)(\mathring{A}) \times \mathbf{C}^2(C^2e^{-2})$$
$$\times 10^{-20}(m^2\mathring{A}^{-2}) \times 2478.960^{-1}(RT)(J^{-1}mol) * \mathbf{N}_a(mol^{-1})$$

$$[\alpha]_{SI} = \frac{[Q]_E[L]_E}{[E]_E} 6.23594365 \times 10^{-38}(C^2m^2/J)$$

où $\mathbf{N}_a$ est la constante d'Avogadro. La conversion vers un système d'unités plus approprié aux calculs moléculaires, c'est-à-dire les $\mathring{A}^3$, peut alors être faite comme suit

$$[\alpha]_{m^3} = \frac{1}{4\pi\varepsilon_0}[\alpha]_{SI}$$

$$[\alpha]_{m^3} = \frac{[Q]_E[L]_E}{[E]_E} 6.23594365 \times 10^{-38}(C^2m^2J^{-1})/4/\pi/\mathbf{E}(C^{-2}Jm)$$
$$= \frac{[Q]_E[L]_E}{[E]_E} \times 5.604586652 \times 10^{-28}(m^3)$$

$$[\alpha]_{\mathring{A}^3} = \frac{[Q]_E[L]_E}{[E]_E} \times 5.604586652 \times 10^{-28}(m^3) \times 10^{30}(\mathring{A}^3m^{-3})$$
$$= \frac{[Q]_E[L]_E}{[E]_E} \times 560.4586652(\mathring{A}^3)$$

Donc, si le moment dipolaire calculé est donné en $e\mathring{A}$ et que le potentiel imposé est en $RT/e/\mathring{A}$, on multiplie le quotient dipôle sur champ par 560.45866 pour obtenir des $\mathring{A}^3$. Le système d'unité utilisé dans cette thèse est plutôt le bohr^3 (b_0^3) (unités atomiques). Comme 1 bohr mesure 0.52917725 Å, la conversion est: $[\alpha]_{b_0^3} = [\alpha]_{\mathring{A}^3} / 0.148185$.

Annexe V

Polarisabilités moléculaires B3LYP

Ici, les polarisabilités moléculaires pour les 707 molécules dont il a été question au Chapitre 4 sont données. Le nom IUPAC est suivi des valeurs propres du tenseur de polarisabilité, de la polarisabilité moyenne et de l'anisotropie. Les unités de polarisabilité sont le bohr^3. Dans le tableau ci-bas, si une molécule se retrouve sur plusieurs lignes, c'est qu'elle y est représentée dans plus d'une conformation.

Nom	FORMULE	α_1	α_2	α_3	α_{avg}	$\Delta\alpha$
acetic acid	C2H4O2	26.9	37.7	40.0	34.9	6.9
cyclopropanecarboxylic acid	C4H6O2	47.2	53.1	67.0	55.8	10.2
3-oxopropanoic acid	C3H4O3	43.8	47.1	55.3	48.7	10.1
benzoic acid	C7H6O2	53.2	98.1	123.1	91.5	29.7
pyrazine-2-carboxylic acid	C5H4O2N2	46.3	84.3	111.8	80.8	26.3
methanol	C1H4O1	20.4	20.9	24.1	21.8	2.7
propan-2-ol	C3H8O1	43.5	47.2	49.2	46.6	3.3
2-methylpropan-2-ol	C4H10O1	56.8	59.4	60.1	58.8	2.7
glycerol	C3H8O3	49.1	54.7	62.9	55.6	11.3
glycerol	C3H8O3	52.4	54.6	58.0	55.0	4.4
propane-1,3-diol	C3H8O2	44.9	49.3	59.8	51.3	10.9
propane-1,3-diol	C3H8O2	48.6	49.3	54.9	50.9	6.3
cyclohexane-1,3,5-triol	C6H12O3	74.7	92.1	96.0	87.6	15.0
phenol	C6H6O1	47.2	85.5	95.1	75.9	22.0
pyridine-2,4-diol	C5H5O2N1	45.6	86.7	96.9	76.4	22.7
formaldehyde	C1H2O1	13.2	18.4	22.9	18.2	2.3
acetaldehyde	C2H4O1	24.4	31.5	37.5	31.1	5.0
propanedial	C3H4O2	34.7	44.3	56.7	45.2	9.1
benzaldehyde	C7H6O1	50.6	93.8	119.5	88.0	26.7

furan-2-carbaldehyde	C5H4O2	40.6	66.3	99.9	68.9	19.6
butane	C4H10	49.7	52.6	59.0	53.8	7.1
hexane	C6H14	72.1	74.3	88.8	78.4	13.7
cyclobutane	C4H8	44.8	51.5	51.5	49.3	2.0
cyclohexane	C6H12	65.7	76.0	76.0	72.5	7.3
cyclopentane	C5H10	54.6	63.6	63.7	60.6	4.5
cyclopropane	C3H6	33.9	38.8	38.8	37.2	0.8
2,3-dimethylbutane	C6H14	68.8	81.3	84.8	78.3	12.5
isobutane	C4H10	49.5	56.6	56.6	54.2	5.3
methane	C1H4	17.2	17.2	17.2	17.2	0.1
neopentane	C5H12	66.5	66.5	66.5	66.5	0.2
propane	C3H8	38.5	40.8	46.1	41.8	4.7
butane	C4H10	48.7	51.6	62.7	54.3	9.5
hexane	C6H14	68.8	73.8	96.7	79.8	20.1
ethylene	C2H4	23.0	25.7	36.3	28.3	5.3
2,3-dimethylbut-2-ene	C6H12	60.7	78.6	101.4	80.3	18.3
2-methylprop-1-ene	C4H8	42.5	54.3	65.3	54.0	10.3
(E)-pent-2-ene	C5H10	55.1	59.5	88.9	67.8	18.1
(E)-but-2-ene	C4H8	42.4	49.2	72.3	54.6	14.7
cyclopenta-1,3-diene	C5H6	42.8	65.4	68.5	58.9	12.6
1,2-divinylbenzene	C10H10	75.1	149.6	168.3	131.0	39.7
formamide	C1H3O1N1	20.4	28.5	36.7	28.5	5.9
N-methylacetamide	C3H7O1N1	40.5	53.6	63.8	52.6	13.0
pyrrolidin-2-one	C4H7O1N1	46.4	60.3	70.2	59.0	9.1
1-pyrrolidin-1-ylethanone	C6H11O1N1	65.8	86.9	100.4	84.4	17.9
N-(4-pyridyl)acetamide	C7H8O1N2	62.6	108.6	147.6	106.3	37.9
1H-pyrrole-3-carboxamide	C5H6O1N2	50.5	87.4	104.9	80.9	26.6
N-methylpyrimidine-2-carboxamide	C6H7O1N3	59.6	101.3	135.9	98.9	35.1
2-formamidoacetamide	C3H6O2N2	43.3	67.4	80.6	63.8	21.5
propan-2-imine	C3H7N1	38.1	51.2	56.7	48.7	8.9
N-methylpropan-2-imine	C4H9N1	48.0	61.4	79.1	62.8	13.4
N-methylacetamidine	C3H8N2	44.8	63.3	67.0	58.4	13.9
acetamidine	C2H6N2	34.9	48.9	55.1	46.3	10.1
N,N,N'-trimethylacetamidine	C5H12N2	67.1	85.7	101.2	84.7	15.3

2-methyl-1,4,5,6-tetrahydropyrimidine	C5H10N2	60.9	81.2	90.7	77.6	16.3
1-methoxyethanimine	C3H7O1N1	41.3	56.2	61.1	52.9	11.9
1-methylsulfanylethanimine	C3H7N1S1	51.7	74.9	85.3	70.6	17.2
1-phenylethanimine	C8H9N1	63.6	112.0	142.9	106.2	35.1
N-phenylpropan-2-imine	C9H11N1	85.8	111.4	166.0	121.1	33.8
ammonia	H3N1	14.0	14.0	15.9	14.7	1.5
N-methylmethanamine	C2H7N1	35.5	38.9	43.8	39.4	6.7
N,N-dimethylmethanamine	C3H9N1	49.6	54.4	54.4	52.8	7.6
1-methylpiperazine	C5H12N2	74.6	82.5	88.8	82.0	15.4
propane-1,2,3-triamine	C3H11N3	62.0	72.1	75.9	70.0	12.7
aniline	C6H7N1	51.4	89.0	107.6	82.6	24.5
pyrimidin-2-amine	C4H5N3	44.0	74.9	97.3	72.1	21.1
N-ethylaniline	C8H11N1	71.4	109.6	150.7	110.6	37.4
benzene-1,3,5-triamine	C6H9N3	63.7	129.2	129.2	107.4	33.9
N,N-dimethylaniline	C8H11N1	71.6	114.5	145.9	110.7	34.6
methylammonium	C1H6N1	19.4	19.4	22.6	20.5	1.1
ethylammonium	C2H8N1	29.7	31.5	36.7	32.6	2.7
propylammonium	C3H10N1	39.8	42.1	53.6	45.2	4.0
dimethylammonium	C2H8N1	29.1	30.5	36.1	31.9	3.8
trimethylammonium	C3H10N1	38.6	45.4	45.4	43.1	5.6
tetramethylammonium	C4H12N1	54.2	54.2	54.2	54.2	0.2
1,2,3,4,5,6-hexamethylbenzene	C12H18	97.2	176.0	176.2	149.8	45.9
tetralin	C10H12	79.6	124.4	145.4	116.5	32.5
7-oxabicyclo[3.3.0]octa-5,8-diene	C7H8O1	59.9	88.8	98.4	82.4	18.4
2,4,6-triethyl-1,3,5-triazine	C9H15N3	103.7	147.1	148.0	132.9	23.3
4-[(1R,2R)-2-(2-furyl)cyclopropyl]-3H-imidazole	C10H10O1N2	95.7	126.3	185.3	135.8	43.7
4,5-dipropylpyridazine	C10H16N2	102.6	152.8	157.4	137.6	30.6
1,3-dioxolan-2-one	C3H4O3	35.8	43.8	58.3	46.0	7.8
dimethyl carbonate	C3H6O3	39.9	48.4	66.4	51.6	15.7
methyl N-methylcarbamate	C3H7O2N1	43.1	53.8	73.5	56.8	16.8
oxazolidin-2-one	C3H5O2N1	39.3	50.0	62.7	50.7	8.8
formate	C1H1O2	24.3	39.8	42.3	35.5	6.4
acetate	C2H3O2	35.8	52.2	58.4	48.8	7.5

propanoate	C3H5O2	46.0	63.9	70.9	60.3	12.1
cyclopropanecarboxylate	C4H5O2	56.5	66.0	80.4	67.6	11.3
3-oxopropanoate	C3H3O3	55.7	57.0	69.9	60.9	9.6
benzoate	C7H5O2	62.2	111.6	132.2	102.0	33.1
pyrazine-2-carboxylate	C5H3O2N2	53.9	96.7	122.6	91.1	29.3
methyl formate	C2H4O2	26.9	35.8	41.8	34.8	7.3
methyl acetate	C3H6O2	37.3	47.6	56.6	47.2	11.0
tetrahydropyran-2-one	C5H8O2	53.9	66.5	78.5	66.3	9.7
2-pyridyl acetate	C7H7O2N1	70.0	93.3	127.6	96.9	30.8
methyl furan-3-carboxylate	C6H6O3	53.4	84.6	110.0	82.7	25.7
methyl benzoate	C8H8O2	63.3	108.1	143.8	105.1	35.1
methoxymethane	C2H6O1	31.8	32.5	39.7	34.7	6.2
2-methoxyethoxyethane	C5H12O2	68.1	74.8	93.0	78.6	16.3
2-methoxyethoxyethane	C5H12O2	67.8	69.4	96.1	77.8	21.4
1,4-dioxane	C4H8O2	52.9	55.9	65.5	58.1	8.2
1,3,5-trioxane	C3H6O3	44.9	51.9	51.9	49.6	6.3
dimethoxymethoxymethane	C4H10O3	59.2	71.0	74.4	68.2	10.9
ethoxyethane	C4H10O1	52.7	56.8	71.5	60.3	12.5
anisole	C7H8O1	57.6	94.2	117.7	89.9	27.8
7,9-dioxa-3,5-diazabicyclo[4.3.0]nona-2,4,10-triene	C5H4O2N2	47.8	78.9	106.3	77.6	21.9
acetylene	C2H2	19.7	19.7	32.1	23.9	6.0
but-2-yne	C4H6	39.6	39.6	75.8	51.7	15.6
prop-1-ynylbenzene	C9H8	65.6	102.2	179.6	115.8	38.7
hepta-1,5-diyne	C7H8	63.9	88.3	104.7	85.7	18.8
1-methoxybut-2-yne	C5H8O1	53.8	56.6	103.5	71.3	24.3
(E)-hex-2-en-4-yne	C6H8	53.7	62.3	135.6	83.9	29.5
pent-3-ynenitrile	C5H5N1	46.8	57.2	92.2	65.4	20.6
2-ethynylpyrimidine-5-carbonitrile	C7H3N3	55.6	83.8	190.0	109.8	42.0
but-2-ynamide	C4H5O1N1	42.4	57.8	94.3	64.9	20.7
carbon tetrabromide	C1Br4	102.0	102.0	102.0	102.0	0.2
bromoform	C1H1Br3	61.4	91.0	91.0	81.1	14.4
1,1,1-tribromopropane	C3H5Br3	96.5	107.4	112.2	105.4	7.0
2,2,3,3-tetrabromobutane	C4H6Br4	115.6	144.0	153.2	137.6	15.3

1,1,2,2-tetrabromocyclopentane	C5H6Br4	135.9	138.5	154.5	143.0	7.8
1,3-dibromobenzene	C6H4Br2	71.6	121.3	160.7	117.8	36.7
1,2,3,4,5,6-hexabromobenzene	C6Br6	117.2	251.7	251.8	206.9	69.4
2,4,5-tribromo-6-methyl-pyrimidine	C5H3N2Br3	86.8	154.0	195.0	145.3	43.1
carbon tetrachloride	C1Cl4	72.0	72.0	72.0	72.0	0.1
chloroform	C1H1Cl3	45.5	64.8	64.8	58.4	9.6
1,1,1-trichloropropane	C3H5Cl3	77.5	82.4	88.0	82.7	6.8
2,2,3,3-tetrachlorobutane	C4H6Cl4	96.6	109.3	115.7	107.2	6.7
1,1,3,3-tetrachloropropane	C3H4Cl4	87.9	93.6	110.2	97.3	17.3
1,1,2,2-tetrachlorocyclopentane	C5H6Cl4	109.6	113.1	118.2	113.6	3.7
1,3-dichlorobenzene	C6H4Cl2	60.4	107.9	135.5	101.3	31.2
1,2,3,4,5,6-hexachlorobenzene	C6Cl6	89.4	198.1	198.1	161.9	54.0
2,4,5-trichloro-6-methyl-pyrimidine	C5H3N2Cl3	71.2	131.7	158.9	120.6	35.7
tetrafluoromethane	C1F4	19.7	19.7	19.7	19.7	0.1
fluoroform	C1H1F3	18.1	19.6	19.6	19.1	0.1
1,1,1-trifluoropropane	C3H5F3	39.9	41.0	46.4	42.4	4.3
2,2,3,3-tetrafluorobutane	C4H6F4	52.0	54.1	58.2	54.7	5.9
1,1,3,3-tetrafluoropropane	C3H4F4	39.9	43.0	47.4	43.4	4.2
1,3-difluorobenzene	C6H4F2	43.6	82.5	85.2	70.4	19.4
1,2,3,4,5,6-hexafluorobenzene	C6F6	42.7	88.9	89.0	73.5	20.4
2,4,5-trifluoro-6-methyl-pyrimidine	C5H3N2F3	46.2	84.7	92.7	74.5	20.0
benzofuran	C8H6O1	56.6	100.7	133.7	97.0	30.3
2,6-dioxabicyclo[3.3.0]octa-3,7,9-triene	C6H4O2	46.9	71.5	107.9	75.4	23.6
furan	C4H4O1	35.2	53.9	59.1	49.4	11.9
isoxazole	C3H3O1N1	30.8	48.6	53.4	44.3	10.6
1,2,4-oxadiazole	C2H2O1N2	26.5	42.0	47.7	38.8	9.6
oxazole	C3H3O1N1	30.6	47.8	53.4	43.9	10.8
1,3-dihydroimidazol-2-one	C3H4O1N2	36.2	58.5	67.8	54.2	14.7
1-(1H-pyrrol-2-yl)-3H-imidazol-2-one	C7H7O1N3	65.9	115.0	156.0	112.3	42.0
1H-pyridin-2-one	C5H5O1N1	43.3	78.3	94.9	72.2	19.6
quinolizin-4-one	C9H7O1N1	64.9	136.4	175.8	125.7	40.1
1H-pyridin-4-one	C5H5O1N1	43.7	75.7	96.9	72.1	19.9
1H-pyrimidin-4-one	C4H4O1N2	39.8	68.2	90.7	66.2	17.7
3,7-dihydropurine-2,6-dione	C5H4O2N4	51.9	109.0	126.4	95.8	31.4

4H-1,2,4-triazole	C2H3N3	29.8	49.4	52.6	44.0	11.0
2,6-diazabicyclo[3.3.0]octa-3,7,9-triene	C6H6N2	54.6	87.4	121.0	87.7	26.3
1H-imidazole	C3H4N2	34.3	56.0	58.8	49.7	12.1
1H-indole	C8H7N1	60.1	108.2	143.1	103.8	33.1
1H-pyrazole	C3H4N2	34.5	55.8	58.1	49.5	12.1
1H-pyrrole	C4H5N1	39.4	62.1	64.8	55.4	13.7
benzothiophene	C8H6S1	65.8	119.8	157.4	114.4	34.3
2,6-dithiabicyclo[3.3.0]octa-3,7,9-triene	C6H4S2	65.4	113.0	145.9	108.1	29.8
isothiazole	C3H3N1S1	40.5	64.5	74.1	59.7	13.3
1,2,4-thiadiazole	C2H2N2S1	36.5	59.1	67.0	54.2	11.9
thiazole	C3H3N1S1	40.8	66.9	71.9	59.8	13.2
thiophene	C4H4S1	45.0	72.0	78.4	65.1	15.3
benzene	C6H6	45.1	82.6	82.6	70.1	18.3
naphthalene	C10H8	66.3	126.0	173.2	121.8	38.7
phenylbenzene	C12H10	88.2	139.6	212.2	146.7	51.7
phthalazine	C8H6N2	59.1	116.6	150.5	108.7	33.4
2,5,7,10-tetrazabicyclo[4.4.0]deca-2,4,6,8,10-pentaene	C6H4N4	52.6	95.1	155.0	100.9	32.6
3,4,8,9-tetrazabicyclo[4.4.0]deca-2,4,7,9,11-pentaene	C6H4N4	107.1	130.0	96.5	68.9	0.2
pyridine	C5H5N1	41.1	74.0	78.7	64.6	16.7
pyrimidine	C4H4N2	37.3	67.9	71.1	58.8	15.0
quinoline	C9H7N1	62.7	118.5	167.3	116.2	36.9
1,2,4-triazine	C3H3N3	34.5	60.9	67.8	54.4	13.8
N-hydroxyformamide	C1H3O2N1	23.0	36.4	40.5	33.3	7.5
N-hydroxy-N-methyl-acetamide	C3H7O2N1	44.0	62.7	66.7	57.8	12.9
3-hydroxy-1,3-oxazinan-4-one	C4H7O3N1	53.3	76.1	78.6	69.3	14.2
1-hydroxy-3,4-dihydroquinolin-2-one	C9H9O2N1	75.1	129.7	170.2	125.0	38.4
pyridine-3-carbohydroxamic acid	C6H6O2N2	57.1	98.6	137.1	97.6	33.6
1,3-dihydroimidazole	C3H5N2	27.5	48.4	51.3	42.4	10.5
5-ethyl-1,3-dihydroimidazole	C5H9N2	49.5	69.3	86.5	68.4	16.0
3,5-dimethyl-1H-imidazole	C5H9N2	46.6	70.5	89.7	68.9	20.6
4-methyl-1,3-dihydroimidazole	C4H7N2	37.1	59.2	71.1	55.8	14.5
2,4,5-trimethyl-1,3-dihydroimidazole	C6H11N2	55.9	88.7	103.9	82.8	21.1

acetone	C3H6O1	34.6	47.2	47.6	43.1	7.3
pentane-2,4-dione	C5H8O2	57.3	72.8	77.9	69.3	14.1
cyclopentanone	C5H8O1	51.2	64.6	69.6	61.8	7.9
cyclohexane-1,4-dione	C6H8O2	59.3	76.6	90.0	75.3	13.1
1-pyrazin-2-ylethanone	C6H6O1N2	53.4	92.9	120.2	88.8	27.9
hydrogen cyanide	C1H1N1	14.4	14.4	23.1	17.3	3.7
acetonitrile	C2H3N1	24.7	24.7	41.0	30.2	7.4
pentanedinitrile	C5H6N2	52.4	62.6	94.2	69.7	23.6
pentanedinitrile	C5H6N2	59.4	64.6	79.2	67.8	12.1
cyclopentanecarbonitrile	C6H9N1	68.1	69.6	82.7	73.5	8.4
benzonitrile	C7H5N1	51.1	87.8	125.6	88.2	27.7
terephthalonitrile	C8H4N2	57.5	93.5	177.6	109.5	39.3
pyrimidine-2-carbonitrile	C5H3N3	44.5	73.7	112.5	76.9	24.9
1-nitrobutane	C4H9O2N1	59.7	68.5	82.9	70.4	15.3
1-nitropropane	C3H7O2N1	47.2	58.6	66.9	57.6	10.6
2-nitrophenol	C6H5O3N1	54.8	108.8	122.6	95.4	30.6
nitrobenzene	C6H5O2N1	50.1	98.3	119.0	89.1	29.9
3-nitroaniline	C6H6O2N2	56.0	114.6	136.0	102.2	34.9
1-oxidopyridine	C5H5O1N1	43.2	79.4	103.0	75.2	21.0
1,4-dioxidopyrazine	C4H4O2N2	41.6	76.2	129.1	82.3	23.5
trimethyl-oxido-ammonium	C3H9O1N1	57.5	57.6	58.2	57.7	1.3
methylsulfonylmethane	C2H6O2S1	53.2	54.7	59.1	55.7	5.4
methylsulfinylmethane	C2H6O1S1	47.1	58.7	60.4	55.4	6.3
1-ethylsulfonylpropane	C5H12O2S1	80.8	83.9	116.6	93.8	26.2
1-[(R)-ethylsulfinyl]propane	C5H12O1S1	76.9	86.9	118.4	94.1	26.0
methylsulfonylbenzene	C7H8O2S1	80.4	111.2	140.7	110.8	28.0
[(S)-methylsulfinyl]benzene	C7H8O1S1	78.2	111.1	143.0	110.8	28.4
thiolane 1,1-dioxide	C4H8O2S1	68.3	72.9	81.1	74.1	7.3
thiolane 1-oxide	C4H8O1S1	64.2	74.1	82.2	73.5	7.8
N-methylmethanesulfonamide	C2H7O2N1S1	58.6	61.9	70.6	63.7	8.5
N-methylmethanesulfinamide	C2H7O1N1S1	53.5	64.0	77.4	65.0	12.8
N-(methylsulfamoyl)methanamine	C2H8O2N2S1	66.0	69.6	78.8	71.5	8.8
N-methylsulfinamoylmethanamine	C2H8O1N2S1	62.7	68.7	85.8	72.4	12.9
thiazolidine 1,1-dioxide	C3H7O2N1S1	64.9	67.9	77.6	70.1	7.6

(1R)-thiazolidine 1-oxide	C3H7O1N1S1	60.8	69.1	77.9	69.3	8.4
N-methylpyrimidine-5-sulfonamide	C5H7O2N3S1	83.3	105.5	134.7	107.8	22.3
N-methylpyrimidine-5-sulfinamide	C5H7O1N3S1	84.1	98.4	147.2	109.9	30.3
N-pyrazin-2-ylmethanesulfonamide	C5H7O2N3S1	77.1	110.5	148.7	112.1	32.1
N-pyrazin-2-ylmethanesulfinamide	C5H7O1N3S1	80.6	98.9	160.9	113.5	35.4
methylsulfanylformaldehyde	C2H4O1S1	38.2	49.9	70.7	52.9	11.0
1-methylsulfanylethanone	C3H6O1S1	48.5	65.3	82.9	65.5	15.6
tetrahydrothiopyran-2-one	C5H8O1S1	64.5	87.4	99.0	83.6	14.6
1-(2-pyridylsulfanyl)ethanone	C7H7O1N1S1	70.1	115.2	176.5	120.6	43.2
3-furyl-methylsulfanyl-methanone	C6H6O2S1	63.7	101.6	141.7	102.4	34.2
methylsulfanyl-phenyl-methanone	C8H8O1S1	73.5	123.7	179.9	125.7	44.0
methylsulfanylmethane	C2H6S1	43.2	50.0	59.0	50.7	7.7
2-methylsulfanylethylsulfanylethane	C5H12S2	87.2	99.2	156.7	114.4	40.0
1,4-dithiane	C4H8S2	75.2	97.5	97.6	90.1	15.2
1,3,5-trithiane	C3H6S3	78.7	109.9	110.0	99.6	17.2
bis(methylsulfanyl)methylsulfanylmethane	C4H10S3	98.1	132.2	132.2	120.9	19.9
ethylsulfanylethane	C4H10S1	63.1	70.0	98.4	77.2	20.9
methyldisulfanylmethane	C2H6S2	62.3	65.6	92.2	73.4	11.9
ethyldisulfanylethane	C4H10S2	81.6	85.5	135.6	100.9	26.0
methyldisulfanylethane	C3H8S2	72.0	75.4	113.7	87.0	19.1
methylsulfanylbenzene	C7H8S1	77.1	103.8	140.0	107.0	26.9
7,9-dithia-3,5-diazabicyclo[4.3.0]nona-2,4,10-triene	C5H4N2S2	69.0	117.8	150.5	112.4	31.0
methanethiol	C1H4S1	34.4	34.4	44.2	37.7	5.4
propane-2-thiol	C3H8S1	55.2	62.7	71.4	63.1	8.6
2-methylpropane-2-thiol	C4H10S1	71.9	72.0	82.5	75.5	5.5
propane-1,2,3-trithiol	C3H8S3	89.8	102.2	122.1	104.7	21.6
propane-1,3-dithiol	C3H8S2	70.4	76.1	110.9	85.8	25.2
benzenethiol	C6H6S1	59.3	95.6	129.0	94.6	28.7
pyridine-2,4-dithiol	C5H5N1S2	69.5	116.7	159.1	115.1	35.3
N-acetylacetamide	C4H7O2N1	47.3	70.5	82.9	66.9	19.7
1,3-dimethylurea	C3H8O1N2	46.8	60.1	78.3	61.8	18.8
hexahydropyrimidin-2-one	C4H8O1N2	52.9	71.1	80.7	68.3	12.8
2-oxopiperidine-1-carbaldehyde	C6H9O2N1	62.5	91.1	101.3	85.0	19.2

(E)-but-2-enal	C4H6O1	38.6	50.9	87.1	58.9	17.1
N-[(E)-prop-1-enyl]acetamide	C5H9O1N1	54.7	72.5	122.9	83.4	29.3
1,3-dihydropyrrol-2-one	C4H5O1N1	40.3	64.1	69.5	58.0	13.5
[(Z)-1-methylprop-1-enyl] acetate	C6H10O2	69.1	84.3	100.5	84.6	17.9
1,2-dimethylprop-1-enylsulfanylformaldehyde	C6H10O1S1	83.1	98.9	130.0	104.0	18.7
(E)-pent-3-en-2-one	C5H8O1	48.8	65.6	96.3	70.2	21.8
hydrazine	H4N2	22.0	22.2	26.9	23.7	3.9
1,1,2-trimethylhydrazine	C3H10N2	57.0	60.9	70.0	62.6	11.0
phenylhydrazine	C6H8N2	58.7	96.7	124.2	93.2	29.0
1H-pyrrol-3-ylhydrazine	C4H7N3	54.5	79.4	97.5	77.1	22.7
piperazin-1-amine	C4H11N3	71.2	79.8	86.4	79.1	14.5
N-methylaminoformamide	C2H6O1N2	40.2	51.5	61.2	50.9	10.8
diaziridine	C1H4N2	27.7	31.7	32.2	30.5	2.4
adamantane	C10H16	108.8	108.8	108.9	108.8	0.4
bicyclo[2.2.2]octane	C8H14	89.7	89.8	91.4	90.3	0.4
octane	C8H18	92.0	96.5	119.6	102.7	23.3
decalin	C10H18	99.9	118.2	132.8	116.9	20.7
ethane	C2H6	27.9	27.9	32.1	29.3	2.3
isopropylcyclopropane	C6H12	65.1	75.8	82.8	74.6	8.1
1,2,3,3a,4,5,6,6a-octahydropentalene	C8H14	82.0	92.7	100.4	91.7	10.1
tert-butylcyclobutane	C8H16	90.1	95.9	108.0	98.0	10.1
octane	C8H18	88.7	95.8	132.3	105.6	31.9
(3E)-hexa-1,3,5-triene	C6H8	51.7	66.5	176.3	98.2	34.0
ammonium	H4N1	9.3	9.3	9.3	9.3	0.1
oxatriazole	C1H1O1N3	23.8	38.4	43.0	35.1	8.4
1-(2-furyl)pyrrole	C8H7O1N1	64.3	104.7	157.8	108.9	39.0
2-(2-furyl)furan	C8H6O2	60.8	98.1	167.5	108.8	36.9
difuro[3,2-b:2',3'-d]furan	C8H4O3	57.7	89.1	167.5	104.8	37.4
benzo[1,2-b:3,4-b':5,6-b"]trifuran	C12H6O3	78.5	185.8	185.8	150.0	49.9
1,7-dihydropurin-6-one	C5H4O1N4	50.1	102.4	120.7	91.1	28.6
2,4-dihydro-1,2,4-triazol-3-one	C2H3O1N3	31.7	52.8	62.4	48.9	13.0
1-(1,3,4-oxadiazol-2-yl)-3H-imidazol-2-one	C5H4O2N4	54.4	97.9	132.5	94.9	35.0

7H-purine	C5H4N4	48.1	88.8	118.4	85.1	27.7
1-(2-furyl)pyridin-2-one	C9H7O2N1	87.3	114.2	160.8	120.8	37.1
2H-pyridazin-3-one	C4H4O1N2	39.5	70.7	89.2	66.5	18.6
3H-pyrimidin-4-one	C4H4O1N2	39.3	72.5	85.5	65.8	17.4
1H-pyrazin-2-one	C4H4O1N2	39.6	72.0	91.1	67.6	18.0
1-(2-thienyl)pyridin-4-one	C9H7O1N1S1	92.8	122.8	214.6	143.4	44.5
7,9-dihydro-3H-purine-2,6,8-trione	C5H4O3N4	53.7	114.2	139.1	102.4	34.8
1-(1H-pyrrol-2-yl)pyrrole	C8H8N2	79.3	103.1	151.7	111.4	35.7
2-(1H-pyrrol-2-yl)-1H-pyrrole	C8H8N2	68.3	112.2	179.7	120.1	43.6
4,7-dihydro-1H-dipyrrolo[3,2-b:2',3'-d]pyrrole	C8H7N3	69.0	114.6	188.1	123.9	42.2
1H-tetrazole	C1H2N4	26.8	45.6	46.7	39.7	9.4
4,7-dihydro-1H-benzo[1,2-b:3,4-b':5,6-b"]tripyrrole	C12H9N3	89.8	214.1	214.1	172.7	58.8
anthracene	C14H10	87.2	170.0	298.4	185.2	64.6
phenanthrene	C14H10	86.8	179.7	261.8	176.1	61.2
phenazine	C12H8N2	81.3	156.6	287.2	175.0	60.2
3-phenylpyridazine	C10H8N2	75.8	131.5	204.8	137.4	50.4
5-phenylpyrimidine	C10H8N2	79.9	126.9	196.3	134.4	49.1
pyrazine	C4H4N2	37.4	66.0	75.9	59.7	15.4
pyrido[2,1,6-de]quinolizine	C12H9N1	80.5	212.8	212.8	168.7	50.0
quinazoline	C8H6N2	59.0	113.9	154.5	109.1	34.1
quinoxaline	C8H6N2	59.4	111.1	163.7	111.4	36.2
1,3,5-triazine	C3H3N3	33.8	62.4	62.4	52.9	13.2
hydroxylamine	H3O1N1	17.2	17.7	22.9	19.3	3.6
2-azaniumylacetate	C2H5O2N1	33.8	44.8	52.7	43.8	6.5
(2R)-2,3,4,5-tetrahydropyrrole-2-carboxylic acid	C5H10O2N1	56.3	69.6	79.8	68.6	12.1
(2S)-2-azaniumyl-3-hydroxy-propanoate	C3H7O3N1	48.5	63.2	68.9	60.2	12.2
(2S)-2-azaniumylpropanoate	C3H7O2N1	45.0	59.4	62.1	55.5	10.1
(2R)-2-azaniumyl-4-methylsulfanyl-butanoate	C5H11O2N1S1	92.2	95.1	120.6	102.6	23.5
(2R)-2-azaniumyl-3-(4-hydroxyphenyl)propanoate	C9H11O3N1	87.8	131.5	171.2	130.2	43.9

(3R)-3-(azaniumylamino)-4-hydroxy-4-oxo-butanoate	C4H8O4N2	66.2	80.3	99.8	82.1	16.1
(2R)-2-azaniumyl-3-carbamoyl-propanoate	C4H8O3N2	68.8	74.9	87.3	77.0	9.3
[(1S,2S)-1-carboxy-2-hydroxy-propyl]ammonium	C4H10O3N1	59.3	66.9	73.0	66.4	10.7
(2R)-2-amino-3-methyl-butanoate	C5H10O2N1	81.3	94.8	102.2	92.8	14.3
(2R)-2-azaniumyl-3-phenyl-propanoate	C9H11O2N1	85.6	128.7	157.5	123.9	40.1
(2S)-2-amino-3-(1,3-dihydroimidazol-5-yl)propanoic acid	C6H10O2N3	65.5	98.6	127.5	97.2	27.7
2-aminoacetic acid	C2H5O2N1	34.3	46.1	51.7	44.0	9.7
(2S)-pyrrolidine-2-carboxylic acid	C5H9O2N1	63.1	76.0	87.1	75.4	15.0
(2S)-2-amino-3-hydroxy-propanoic acid	C3H7O3N1	53.8	62.0	67.1	61.0	8.0
(2S)-2-aminopropanoic acid	C3H7O2N1	49.1	58.6	62.1	56.6	10.2
(2R)-2-amino-4-methylsulfanyl-butanoic acid	C5H11O2N1S1	86.6	100.5	120.2	102.5	22.0
(2S)-2-amino-3-(4-hydroxyphenyl)propanoic acid	C9H11O3N1	96.9	127.1	166.2	130.1	34.8
(2R)-2-aminobutanedioic acid	C4H7O4N1	63.9	71.5	85.9	73.8	17.6
(2S)-2-amino-3-carbamoyl-propanoic acid	C4H8O3N2	70.4	77.4	90.3	79.4	15.6
(2R,3R)-2-amino-3-hydroxy-butanoic acid	C4H9O3N1	60.8	76.4	83.2	73.5	15.8
(2S)-2-amino-3-phenyl-propanoic acid	C9H11O2N1	94.6	124.2	153.2	124.0	30.5
(2S)-2-amino-3-(1H-imidazol-4-yl)propanoic acid	C6H9O2N3	79.8	105.0	124.9	103.2	26.9
1,1,1-trichloroethane	C2H3Cl3	64.1	73.9	73.9	70.6	2.6
1,1,1-trifluoro-2,2,2-trimethoxy-ethane	C5H9O3F3	75.4	81.5	85.3	80.7	5.8
(2S)-1,1,1-trifluoropropan-2-ol	C3H5O1F3	43.6	47.6	50.4	47.2	4.9
1,1,1-trimethoxyethane	C5H12O3	73.2	80.7	84.1	79.3	6.9
1,1,2-trichloroethane	C2H3Cl3	62.3	72.2	74.2	69.6	7.5
1-acetoxyethyl acetate	C6H10O4	78.2	88.8	103.0	90.0	17.2
1,1-dichloroethane	C2H4Cl2	46.8	57.5	66.8	57.0	8.5
1,1-dichloroethylene	C2H2Cl2	38.9	63.0	65.5	55.8	12.4
1-ethoxyethoxyethane	C6H14O2	78.0	89.6	102.3	90.0	16.2
1,1-difluoroethane	C2H4F2	28.7	30.3	31.7	30.2	1.6

1,2,3-trimethylbenzene	C9H12	71.9	127.1	131.4	110.1	32.2
1,2,4-trimethylbenzene	C9H12	72.3	117.9	143.7	111.3	34.0
2-acetoxyethyl acetate	C6H10O4	76.0	87.7	109.7	91.1	18.8
1,2-dibromoethane	C2H4Br2	58.0	60.9	106.4	75.1	22.0
1,2-dichlorobenzene	C6H4Cl2	60.2	111.3	128.3	99.9	30.1
1,2-dichloroethane	C2H4Cl2	47.5	56.6	64.6	56.2	9.6
(2S)-1,2-dichloropropane	C3H6Cl2	61.8	68.6	74.7	68.4	7.7
2-ethoxyethoxyethane	C6H14O2	79.7	88.1	106.5	91.4	18.8
1,2-dimethoxyethane	C4H10O2	57.3	61.2	78.8	65.8	15.2
ethylene glycol	C2H6O2	36.5	39.5	40.5	38.8	3.0
mesitylene	C9H12	72.7	131.2	131.3	111.7	31.9
1,3-dichloropropane	C3H6Cl2	61.6	65.0	78.5	68.4	11.0
1,3-dimethylnaphthalene	C12H12	84.3	159.9	208.2	150.8	48.7
1,4-dichlorobenzene	C6H4Cl2	60.4	96.3	148.4	101.7	34.5
1,4-dichlorobutane	C4H8Cl2	71.6	79.1	92.9	81.2	15.6
1,4-dimethylpiperazine	C6H14N2	85.3	92.4	110.6	96.1	22.8
1,4-dimethylnaphthalene	C12H12	84.0	172.9	190.0	148.9	46.6
1-bromo-2-chloro-ethane	C2H4Cl1Br1	52.1	55.0	91.5	66.2	18.6
1-bromo-2-methyl-propane	C4H9Br1	68.3	70.2	85.5	74.7	10.4
1-bromobutane	C4H9Br1	63.4	68.3	99.0	76.9	20.4
1-bromoheptane	C7H15Br1	93.3	100.8	154.5	116.2	39.2
1-bromohexane	C6H13Br1	83.4	90.5	135.4	103.1	32.0
1-bromooctane	C8H17Br1	103.3	112.5	172.1	129.3	44.7
1-bromopentane	C5H11Br1	73.4	78.7	118.1	90.1	26.0
1-bromopropane	C3H7Br1	53.4	56.3	82.0	63.9	15.2
2-chloro-1,1,1-trifluoro-ethane	C2H2F3Cl1	39.1	40.4	52.3	43.9	6.8
1-chlorobutane	C4H9Cl1	57.4	62.4	86.2	68.7	17.2
1-chlorohexane	C6H13Cl1	80.0	88.1	112.0	93.4	23.3
1-chloropentane	C5H11Cl1	69.6	78.0	93.8	80.5	17.0
1-chloropropane	C3H7Cl1	48.5	53.4	63.4	55.1	8.6
1-ethylnaphthalene	C12H12	88.0	164.2	191.2	147.8	42.5
1-methylimidazole	C4H6N2	44.1	67.7	76.6	62.8	16.5
1-methylpyrrole	C5H7N1	49.3	73.9	84.0	69.1	18.2
1-methylcyclohexene	C7H12	68.1	87.0	100.8	85.3	17.5

1-methylnaphthalene	C11H10	75.2	148.3	182.1	135.2	42.3
naphthalen-1-ol	C10H8O1	68.2	137.6	175.9	127.2	40.7
naphthalen-1-amine	C10H9N1	72.4	149.8	180.4	134.2	42.8
1-nitropentane	C5H11O2N1	70.0	78.5	101.5	83.3	20.8
2,2,2-trifluoroethanol	C2H3O1F3	33.6	34.4	37.7	35.3	2.8
2,2,4-trimethylpentane	C8H18	94.5	100.4	114.1	103.0	15.4
2,2,5-trimethylhexane	C9H20	105.8	110.8	128.6	115.0	19.4
2,2-dimethylbutane	C6H14	75.4	75.8	83.3	78.2	7.2
2,3,4-trimethylpentane	C8H18	92.8	100.4	112.9	102.0	15.8
2,3-dimethylbuta-1,3-diene	C6H10	55.8	79.7	106.0	80.5	21.6
2,3-dimethylnaphthalene	C12H12	84.0	149.5	222.6	152.0	51.0
2,3-dimethylphenol	C8H10O1	65.3	119.1	123.2	102.5	30.5
2,3-dimethylpyridine	C7H9N1	59.7	100.3	115.8	91.9	26.2
2,4-dimethylphenol	C8H10O1	65.5	112.7	132.0	103.4	31.3
2,4-dimethylpyridine	C7H9N1	60.2	100.3	117.1	92.6	26.7
2,5-dimethylphenol	C8H10O1	65.5	107.9	136.6	103.4	33.4
2,5-dimethylpyridine	C7H9N1	60.1	93.2	125.3	92.9	29.5
(2S,5S)-2,5-dimethyltetrahydrofuran	C6H12O1	69.4	75.8	91.3	78.8	11.1
2,6-dimethylaniline	C8H11N1	69.7	128.4	129.4	109.2	32.8
2,6-dimethylnaphthalene	C12H12	84.6	145.9	229.7	153.4	54.4
2,6-dimethylphenol	C8H10O1	65.6	117.2	125.5	102.7	30.7
2,6-dimethylpyridine	C7H9N1	60.5	101.9	115.9	92.8	26.7
2-bromo-2-methyl-propane	C4H9Br1	70.5	70.5	87.4	76.1	6.6
2-bromopropane	C3H7Br1	53.9	61.2	75.9	63.7	9.9
2-butoxyethanol	C6H14O2	77.4	83.6	109.2	90.1	23.5
2-chloro-1,1,1-trimethoxy-ethane	C5H11O3Cl1	77.3	96.6	103.0	92.3	17.5
2-chloro-2-methyl-propane	C4H9Cl1	64.8	64.8	74.5	68.0	3.0
2-chloroaniline	C6H6N1Cl1	58.8	107.8	124.9	97.2	29.6
(2S)-2-chlorobutane	C4H9Cl1	58.2	72.0	73.3	67.8	10.4
2-chlorophenol	C6H5O1Cl1	54.8	98.1	118.8	90.6	27.7
2-chloropropane	C3H7Cl1	48.1	55.2	63.9	55.7	7.4
2-chloropyridine	C5H4N1Cl1	49.3	82.2	109.8	80.4	24.6
1-chloro-2-methyl-benzene	C7H7Cl1	61.7	108.2	125.4	98.4	29.0
2-ethoxyethanol	C4H10O2	58.6	61.5	75.3	65.1	11.8

2-ethylpyrazine	C6H8N2	60.3	86.9	112.6	86.6	22.9
2-ethylpyridine	C7H9N1	63.9	93.7	117.0	91.5	25.1
1-ethyl-2-methyl-benzene	C9H12	76.2	118.7	133.9	109.6	28.4
2-fluorophenol	C6H5O1F1	46.5	85.7	95.1	75.8	22.2
2-isobutylpyrazine	C8H12N2	84.2	111.0	136.5	110.6	24.5
1,1,1,2-tetramethoxyethane	C6H14O4	92.1	95.8	104.8	97.6	9.1
2-methoxyaniline	C7H9O1N1	63.8	116.7	125.2	101.9	30.7
2-methoxyethanamine	C3H9O1N1	49.8	53.1	68.1	57.0	15.1
2-methoxyethanol	C3H8O2	47.5	49.8	59.7	52.3	8.9
2-methoxyphenol	C7H8O2	59.5	104.6	120.6	94.9	29.4
2-methylbut-2-ene	C5H10	51.7	64.7	85.3	67.2	15.3
isoprene	C5H8	46.9	63.2	98.1	69.4	17.9
(2R)-2-methylbutan-1-ol	C5H12O1	63.7	71.9	76.5	70.7	9.6
2-methylbutan-2-ol	C5H12O1	66.1	69.2	76.6	70.6	8.2
isopentane	C5H12	61.5	65.8	69.8	65.7	7.1
2-methylhexane	C7H16	83.8	85.3	104.3	91.1	16.9
2-methylpent-1-ene	C6H12	64.9	77.9	94.5	79.1	17.4
2-methylpentan-2-ol	C6H14O1	76.0	79.8	94.5	83.4	14.0
(3R)-2-methylpentan-3-ol	C6H14O1	73.6	81.7	93.6	83.0	14.6
isohexane	C6H14	69.8	76.9	90.6	79.1	14.9
2-methylpropan-1-ol	C4H10O1	52.7	59.8	63.8	58.8	8.6
2-methylpyrazine	C5H6N2	47.2	77.2	97.1	73.8	20.7
2-methylpyridine	C6H7N1	50.9	84.4	100.8	78.7	22.7
(2S)-2-methyltetrahydrofuran	C5H10O1	58.8	64.0	75.1	66.0	9.0
2-methylthiophene	C5H6S1	54.1	87.0	95.6	78.9	20.3
naphthalen-2-ol	C10H8O1	68.3	129.3	189.3	129.0	43.1
naphthalen-2-amine	C10H9N1	72.3	134.3	204.9	137.1	46.0
2-nitroaniline	C6H6O2N2	55.8	122.4	134.8	104.3	34.0
2-nitropropane	C3H7O2N1	51.8	56.1	62.3	56.7	5.1
1-methyl-2-nitro-benzene	C7H7O2N1	59.0	118.5	130.9	102.8	33.3
2-phenylethanol	C8H10O1	70.8	103.4	132.0	102.0	29.9
2-propoxyethanol	C5H12O2	67.7	72.3	93.2	77.7	18.3
3,3,3-trimethoxypropanenitrile	C6H11O3N1	86.6	88.2	105.0	93.2	9.3
3,3-dimethylbutan-2-one	C6H12O1	71.7	81.1	83.9	78.9	8.8

3,3-dimethylpentane	C7H16	84.7	91.9	92.7	89.7	7.1
3,4-dimethylphenol	C8H10O1	65.1	111.2	132.6	102.9	31.5
3,4-dimethylpyridine	C7H9N1	59.3	102.5	111.9	91.2	26.2
3,5-dimethylphenol	C8H10O1	65.6	118.8	126.3	103.6	30.6
3,5-dimethylpyridine	C7H9N1	59.7	98.1	118.7	92.1	27.9
1-(3-pyridyl)ethanone	C7H7O1N1	56.5	98.8	126.0	93.8	31.7
3-chloroaniline	C6H6N1Cl1	58.8	105.8	129.9	98.2	30.4
3-chlorophenol	C6H5O1Cl1	54.8	99.9	119.2	91.3	27.6
3-chloroprop-1-ene	C3H5Cl1	46.4	47.2	75.2	56.2	13.4
3-chloropyridine	C5H4N1Cl1	48.9	82.0	109.4	80.1	25.0
3-hydroxybenzonitrile	C7H5O1N1	53.3	96.5	133.6	94.4	31.1
pyridine-3-carbonitrile	C6H4N2	47.4	80.3	120.3	82.7	27.2
3-ethylphenol	C8H10O1	69.9	107.5	130.7	102.7	28.7
3-ethylpyridine	C7H9N1	63.8	92.8	116.9	91.2	24.9
pyridine-3-carbaldehyde	C6H5O1N1	46.9	85.9	113.2	82.0	25.9
3-hydroxybenzaldehyde	C7H6O2	52.7	100.6	128.6	94.0	30.3
3-methoxyaniline	C7H9O1N1	63.7	105.2	138.5	102.5	33.7
3-methoxyphenol	C7H8O2	59.6	101.3	125.8	95.6	30.3
3-methyl-1H-indole	C9H9N1	69.0	126.4	156.5	117.3	37.8
3-methylbut-1-ene	C5H10	58.3	61.3	80.8	66.8	11.9
3-methylbutan-1-ol	C5H12O1	64.4	69.0	78.1	70.5	10.4
3-methylbutan-2-one	C5H10O1	59.6	67.8	73.9	67.1	10.7
3-methylbutanoic acid	C5H10O2	65.3	67.7	82.3	71.7	12.9
(3S)-3-methylheptane	C8H18	90.2	96.9	125.8	104.3	27.1
(3R)-3-methylhexane	C7H16	81.8	88.9	100.8	90.5	14.0
3-methylpentane	C6H14	69.9	75.7	90.0	78.5	15.2
3-methylpyridine	C6H7N1	50.4	83.8	100.7	78.3	22.9
3-nitrophenol	C6H5O3N1	52.2	106.5	127.1	95.3	33.0
1-methyl-3-nitro-benzene	C7H7O2N1	59.3	114.2	136.1	103.2	34.3
3-phenylpropan-1-ol	C9H12O1	84.6	112.9	147.8	115.1	32.1
1-(4-pyridyl)ethanone	C7H7O1N1	56.7	103.1	120.2	93.3	30.4
4-bromophenol	C6H5O1Br1	60.4	97.4	140.7	99.5	32.7
1-bromo-4-methyl-benzene	C7H7Br1	67.7	103.1	153.6	108.1	36.5
4-chloroaniline	C6H6N1Cl1	58.7	95.7	140.8	98.4	32.8

4-chlorophenol	C6H5O1Cl1	54.7	92.4	126.8	91.3	29.4
4-hydroxybenzonitrile	C7H5O1N1	53.2	90.4	142.5	95.4	31.9
pyridine-4-carbonitrile	C6H4N2	47.5	84.6	114.9	82.3	26.0
4-ethylphenol	C8H10O1	69.8	103.1	135.7	102.9	30.1
4-ethylpyridine	C7H9N1	63.9	96.3	112.8	91.0	23.4
1-ethyl-4-methyl-benzene	C9H12	77.1	108.9	147.8	111.3	33.6
4-fluorophenol	C6H5O1F1	46.2	85.1	95.6	75.6	22.1
pyridine-4-carbaldehyde	C6H5O1N1	47.0	90.4	107.3	81.6	24.9
4-hydroxybenzaldehyde	C7H6O2	52.7	96.2	138.6	95.8	30.9
1-isopropyl-4-methyl-benzene	C10H14	90.7	117.9	162.4	123.7	35.1
4-methoxyaniline	C7H9O1N1	63.6	101.6	141.7	102.3	35.3
4-methyl-1H-imidazole	C4H6N2	43.9	67.5	78.0	63.1	16.7
1-(p-tolyl)ethanone	C9H10O1	69.5	114.8	161.5	115.3	40.1
4-methylbenzaldehyde	C8H8O1	59.9	102.7	147.5	103.4	33.9
(2R)-4-methylpentan-2-ol	C6H14O1	76.4	80.7	92.6	83.2	12.7
4-methylpentan-2-one	C6H12O1	69.5	78.6	91.4	79.8	17.1
4-methylpyridine	C6H7N1	50.5	87.5	96.7	78.3	22.0
4-propylphenol	C9H12O1	80.0	112.7	156.2	116.3	36.4
4-nitroaniline	C6H6O2N2	55.8	103.6	163.1	107.5	37.0
4-nitrophenol	C6H5O3N1	52.1	100.7	140.3	97.7	33.7
4-tert-butylphenol	C10H14O1	92.5	125.8	163.1	127.2	35.7
(E)-1,2-dichloroethylene	C2H2Cl2	39.0	46.7	86.0	57.2	18.6
(E)-hept-2-ene	C7H14	77.3	80.6	123.6	93.8	29.2
(E)-hex-2-enal	C6H10O1	61.6	71.7	124.0	85.8	27.4
(E)-oct-2-enal	C8H14O1	84.0	92.7	160.4	112.4	38.7
4-methoxy-N,N-dimethyl-benzamide	C10H13O2N1	97.2	133.1	202.9	144.4	49.6
N,N,4-trimethylbenzamide	C10H13O1N1	94.9	130.0	187.1	137.3	44.2
N,N-dimethylformamide	C3H7O1N1	40.1	54.7	64.9	53.2	10.2
N-methyl-N-(2,2,2-trifluoroethyl)aniline	C9H10N1F3	83.0	126.5	160.9	123.4	37.3
N-methylaniline	C7H9N1	61.3	99.6	129.4	96.8	30.5
4-methylmorpholine	C5H11O1N1	69.3	79.0	82.3	76.9	12.3
1-methylpiperidine	C6H13N1	74.9	83.4	87.0	81.8	9.3
(Z)-1,2-dichloroethylene	C2H2Cl2	39.1	54.4	73.4	55.6	13.3
(Z)-pent-2-ene	C5H10	54.9	60.6	86.0	67.2	15.2

acenaphthene	C12H10	80.2	161.8	185.7	142.6	42.8
1-phenylethanone	C8H8O1	60.3	106.2	132.9	99.8	32.7
isopropenylbenzene	C9H10	70.9	115.1	149.9	112.0	35.3
azetidine	C3H7N1	42.2	47.2	49.3	46.2	3.5
benzamide	C7H7O1N1	58.2	104.0	126.4	96.2	31.4
trifluoromethylbenzene	C7H5F3	56.1	91.8	104.7	84.2	23.6
phenylmethanol	C7H8O1	60.6	94.4	111.7	88.9	23.0
bromomethylbenzene	C7H7Br1	77.9	104.3	145.8	109.4	28.5
chloromethylbenzene	C7H7Cl1	70.9	98.9	129.9	99.9	25.8
1-chloro-2-(2-chloroethoxy)ethane	C4H8O1Cl2	72.1	94.3	95.0	87.1	15.9
bromobenzene	C6H5Br1	58.4	94.4	127.6	93.5	28.6
bromoethane	C2H5Br1	43.4	46.1	63.3	50.9	9.5
bromomethane	C1H3Br1	33.1	33.1	46.7	37.6	6.1
bromo-trifluoro-methane	C1F3Br1	35.3	35.3	47.4	39.3	5.9
but-1-ene	C4H8	45.6	48.5	68.6	54.3	12.1
but-1-yne	C4H6	40.3	43.6	67.3	50.4	12.2
buta-1,3-diene	C4H6	37.6	46.5	90.2	58.1	17.4
butan-1-ol	C4H10O1	52.1	56.5	68.6	59.1	11.2
(2S)-butan-2-ol	C4H10O1	53.2	57.6	66.2	59.0	9.6
butanenitrile	C4H7N1	45.4	48.6	72.4	55.5	14.9
butyric acid	C4H8O2	47.4	59.7	71.5	59.5	15.8
butanal	C4H8O1	45.0	56.8	64.5	55.4	12.6
chlorobenzene	C6H5Cl1	52.8	89.4	114.0	85.4	25.6
chloro-difluoro-methane	C1H1F2Cl1	27.9	29.5	38.3	31.9	4.6
chloroethane	C2H5Cl1	37.2	39.9	52.2	43.1	7.3
chloroethylene	C2H3Cl1	31.0	37.6	57.6	42.1	10.4
chloro-fluoro-methane	C1H2F1Cl1	26.8	28.4	38.0	31.1	4.7
chloromethane	C1H3Cl1	26.6	26.7	37.2	30.2	4.9
(1R,2S)-1,2-dimethylcyclohexane	C8H16	87.0	98.2	104.5	96.6	11.1
cyclohepta-1,3,5-triene	C7H8	60.7	96.4	99.8	85.7	18.8
cycloheptanol	C7H14O1	79.4	90.6	97.4	89.1	11.4
cyclohexanol	C6H12O1	68.7	79.1	85.1	77.6	10.3
cyclohexanone	C6H10O1	62.7	76.4	82.9	74.0	10.0
cyclohexene	C6H10	59.0	73.2	83.2	71.8	12.0

cyclohexanamine	C6H13N1	75.9	82.6	86.0	81.5	7.9
cyclopentanol	C5H10O1	60.9	66.4	68.0	65.1	3.9
cyclopentene	C5H8	48.5	61.3	69.3	59.7	9.0
decan-1-ol	C10H22O1	111.5	121.5	177.0	136.7	48.0
decan-2-one	C10H20O1	104.8	122.5	169.9	132.4	46.9
2-isopropylsulfanylpropane	C6H14S1	89.5	93.6	123.2	102.1	20.2
1-butoxybutane	C8H18O1	92.6	100.6	141.1	111.4	35.4
N-butylbutan-1-amine	C8H19N1	96.2	105.4	147.2	116.2	39.5
1-propoxypropane	C6H14O1	72.6	77.9	106.0	85.5	23.9
1-propylsulfanylpropane	C6H14S1	82.8	92.3	135.8	103.7	32.6
N-propylpropan-1-amine	C6H15N1	76.2	83.2	111.5	90.3	27.6
dibromomethane	C1H2Br2	47.5	52.1	79.0	59.5	13.6
dichloromethane	C1H2Cl2	35.8	40.0	56.9	44.2	9.0
diethyl propanedioate	C7H12O4	87.5	95.0	129.0	103.8	27.6
diethyl butanedioate	C8H14O4	87.0	108.6	154.4	116.7	43.2
N-ethylethanamine	C4H11N1	56.3	62.1	76.9	65.1	16.0
2-isopropoxypropane	C6H14O1	75.8	81.9	94.7	84.1	12.4
N-isopropylpropan-2-amine	C6H15N1	82.2	84.4	100.0	88.9	15.5
methoxymethoxymethane	C3H8O2	45.7	49.7	58.4	51.3	8.8
acetamide	C2H5O1N1	30.8	43.7	46.6	40.4	8.5
ethanethiol	C2H6S1	44.6	46.8	60.0	50.5	9.5
ethanol	C2H6O1	31.7	34.1	37.1	34.3	3.0
ethyl acetate	C4H8O2	47.7	59.7	72.5	60.0	13.7
ethyl butanoate	C6H12O2	67.8	80.7	107.0	85.2	25.0
ethyl formate	C3H6O2	37.5	46.6	58.7	47.6	10.3
ethyl pentanoate	C7H14O2	77.8	91.2	125.3	98.1	32.2
ethoxybenzene	C8H10O1	67.9	104.8	138.2	103.6	33.9
ethyl propanoate	C5H10O2	57.6	69.2	90.0	72.3	20.3
ethanamine	C2H7N1	35.7	38.5	43.9	39.4	4.4
ethylbenzene	C8H10	67.8	100.2	122.5	96.8	25.6
fluorobenzene	C6H5F1	44.3	82.0	84.3	70.2	19.2
fluoromethane	C1H3F1	16.9	16.9	18.5	17.4	0.3
(2R)-2-bromo-2-chloro-1,1,1-trifluoro-ethane	C2H1F3Cl1Br1	53.4	63.6	77.7	64.9	10.3

hept-1-ene	C7H14	77.8	81.0	120.1	93.0	29.1
hept-1-yne	C7H12	70.7	77.2	120.0	89.3	29.2
heptan-1-ol	C7H16O1	82.1	89.2	121.2	97.5	28.7
heptan-2-one	C7H14O1	75.0	90.1	115.3	93.4	29.2
heptan-4-one	C7H14O1	74.9	88.5	114.8	92.8	30.1
heptanal	C7H14O1	75.1	90.6	115.7	93.8	30.2
hex-1-ene	C6H12	67.7	70.0	102.2	80.0	22.7
hex-1-yne	C6H10	60.6	67.0	101.1	76.3	22.8
hexa-1,5-diene	C6H10	67.0	79.7	88.0	78.2	11.8
1,1,2,3,3,3-hexafluoroprop-1-ene	C3F6	34.3	43.1	54.8	44.1	8.8
hexan-1-ol	C6H14O1	72.1	78.5	103.1	84.6	22.0
hexan-2-one	C6H12O1	64.9	78.9	97.8	80.5	23.4
(3S)-hexan-3-ol	C6H14O1	73.3	80.2	97.9	83.8	18.3
hexanal	C6H12O1	65.1	79.3	98.4	80.9	23.4
hexanoic acid	C6H12O2	67.5	81.6	106.1	85.1	26.7
hydrogen sulfide	H2S1	24.8	24.9	25.6	25.1	1.2
indane	C9H10	70.0	111.8	131.5	104.4	27.9
isopentyl acetate	C7H14O2	84.5	90.2	116.3	97.0	23.1
isopentyl formate	C6H12O2	74.6	79.5	99.2	84.4	17.4
isobutyl acetate	C6H12O2	70.1	83.4	101.1	84.9	19.2
isobutyl formate	C5H10O2	60.0	71.5	85.8	72.4	14.7
isobutylbenzene	C10H14	91.2	124.3	146.3	120.6	27.8
2-methylpropanal	C4H8O1	46.8	57.2	61.4	55.1	8.0
(2R)-2-chloro-2-(difluoromethoxy)-1,1,1-trifluoro-ethane	C3H2O1F5Cl1	54.5	64.1	70.5	63.0	9.1
isopropyl acetate	C5H10O2	62.0	69.9	84.2	72.0	13.0
isopropyl formate	C4H8O2	51.7	58.0	69.2	59.6	9.5
cumene	C9H12	81.6	108.9	136.7	109.1	26.4
m-cresol	C7H8O1	56.5	98.0	114.7	89.7	27.1
m-xylene	C8H10	63.6	106.9	122.8	97.8	28.3
methanesulfonyl chloride	C1H3O2S1Cl1	52.4	54.8	69.2	58.8	7.0
2,2-dichloro-1,1-difluoro-1-methoxy-ethane	C3H4O1F2Cl2	60.4	76.5	86.0	74.3	14.5
methyl 2-chloroacetate	C3H5O2Cl1	46.4	57.2	78.8	60.8	18.6

methyl 2-cyanoacetate	C4H5O2N1	44.7	56.2	82.3	61.1	20.8
methyl cyclohexanecarboxylate	C8H14O2	90.8	99.2	115.4	101.8	17.0
1-cyclohexylethanone	C8H14O1	86.8	95.6	113.9	98.8	19.7
methyl cyclopropanecarboxylate	C5H8O2	57.3	61.9	86.6	68.6	15.6
1-cyclopropylethanone	C5H8O1	54.6	61.4	76.1	64.0	11.2
methoxyethane	C3H8O1	42.4	44.2	55.9	47.5	9.6
methylsulfanylethane	C3H8S1	53.2	59.9	78.4	63.8	13.9
methyl hexanoate	C7H14O2	77.6	90.4	125.9	98.0	34.7
2-methoxypropane	C4H10O1	53.2	58.3	66.3	59.3	9.2
methyl octanoate	C9H18O2	97.5	112.2	162.2	124.0	46.3
methyl 4-methoxybenzoate	C9H10O3	75.5	119.3	188.7	127.9	48.5
methyl 4-nitrobenzoate	C8H7O4N1	68.6	124.6	185.2	126.1	49.4
methyl pentanoate	C6H12O2	67.6	79.6	107.9	85.0	27.4
methyl propanoate	C4H8O2	47.3	57.8	73.0	59.4	16.3
1-methoxypropane	C4H10O1	52.4	55.5	71.8	59.9	13.9
2-methoxy-2-methyl-propane	C5H12O1	67.3	68.5	76.2	70.7	6.5
methyl 2,2,2-trifluoroacetate	C3H3O2F3	40.1	49.4	57.5	49.0	10.8
methyl 2,2-dimethylpropanoate	C6H12O2	74.6	82.2	93.6	83.5	12.2
methanamine	C1H5N1	24.8	25.5	29.4	26.6	3.6
methylcyclohexane	C7H14	75.4	85.9	95.0	85.4	11.9
methylcyclopentane	C6H12	64.6	73.5	81.7	73.3	8.9
morpholine	C4H9O1N1	58.4	61.7	69.0	63.0	9.0
butane-1-thiol	C4H10S1	66.0	70.6	89.5	75.4	16.3
butyl acetate	C6H12O2	67.9	81.5	107.3	85.6	23.9
hexanamide	C6H13O1N1	73.7	86.2	112.7	90.9	27.8
butan-1-amine	C4H11N1	56.3	60.7	75.4	64.1	15.3
butylbenzene	C10H14	91.7	119.4	158.8	123.3	36.1
decane	C10H22	108.6	117.6	168.7	131.6	45.0
heptane	C7H16	78.8	85.1	114.1	92.6	25.0
heptan-1-amine	C7H17N1	86.0	93.5	128.8	102.8	31.1
hexan-1-amine	C6H15N1	76.0	82.7	110.7	89.8	25.3
nonane	C9H20	98.7	106.8	150.2	118.6	37.2
octan-1-amine	C8H19N1	96.2	104.4	146.4	115.6	39.5
pentane	C5H12	58.8	63.1	79.0	67.0	14.2

pentyl acetate	C7H14O2	77.9	92.2	125.6	98.5	29.8
pentan-1-amine	C5H13N1	66.3	71.4	93.0	76.9	21.5
pentylbenzene	C11H16	101.7	129.2	178.8	136.6	43.0
propane-1-thiol	C3H8S1	54.6	57.8	78.1	63.5	14.3
propyl acetate	C5H10O2	57.8	70.2	90.1	72.7	19.0
propyl butanoate	C7H14O2	77.8	91.0	125.4	98.1	31.9
propyl formate	C4H8O2	47.7	58.3	74.6	60.2	15.2
propyl propanoate	C6H12O2	67.7	79.9	107.7	85.1	25.4
propan-1-amine	C3H9N1	45.9	49.3	59.6	51.6	9.6
propylbenzene	C9H12	78.0	109.7	142.9	110.2	32.5
propylcyclopentane	C8H16	84.8	93.7	117.0	98.5	20.5
1-nitroethane	C2H5O2N1	34.2	46.9	53.6	44.9	9.3
nitromethane	C1H3O2N1	24.0	35.4	39.3	32.9	5.7
non-1-ene	C9H18	98.8	102.3	156.1	119.0	40.0
nonan-1-ol	C9H20O1	102.0	111.0	157.4	123.5	40.9
o-cresol	C7H8O1	56.5	101.1	110.7	89.5	26.2
2-methylaniline	C7H9N1	60.6	106.4	120.5	95.8	28.5
o-xylene	C8H10	63.2	107.8	120.1	97.0	27.8
oct-1-ene	C8H16	88.8	91.4	137.7	106.0	34.2
oct-1-yne	C8H14	80.6	89.6	136.5	102.3	34.3
octan-1-ol	C8H18O1	92.1	100.3	139.1	110.5	32.8
octan-2-one	C8H16O1	84.9	100.8	133.4	106.4	35.3
octanal	C8H16O1	85.1	101.3	134.0	106.8	36.0
p-cresol	C7H8O1	56.3	94.3	118.5	89.7	27.6
4-methylaniline	C7H9N1	60.4	97.7	132.0	96.7	31.0
p-xylene	C8H10	63.6	100.1	130.4	98.1	30.7
pent-1-ene	C5H10	55.9	59.8	85.6	67.1	17.2
pent-1-yne	C5H8	50.5	54.0	85.6	63.4	18.5
penta-1,4-diene	C5H8	52.8	62.3	84.2	66.5	17.5
pentan-1-ol	C5H12O1	62.1	67.2	86.0	71.8	16.4
(2R)-pentan-2-ol	C5H12O1	63.5	68.4	83.3	71.7	14.9
pentan-2-one	C5H10O1	54.9	68.3	80.1	67.7	17.6
pentan-3-ol	C5H12O1	64.4	66.5	82.1	71.0	13.5
pentan-3-one	C5H10O1	54.5	67.2	79.4	67.1	18.5

pentanal	C5H10O1	55.1	68.7	80.5	68.1	17.3
pentanenitrile	C5H9N1	55.5	61.4	88.0	68.3	19.6
pentanoic acid	C5H10O2	57.5	70.8	88.4	72.3	21.5
phenyl formate	C7H6O2	59.7	93.1	116.3	89.7	27.5
[(1R)-1,2,2-trifluoroethoxy]benzene	C8H7O1F3	70.0	105.6	135.7	103.8	31.8
piperazine	C4H10N2	63.6	68.2	72.6	68.1	10.8
piperidine	C5H11N1	65.0	72.1	74.4	70.5	8.9
prop-2-en-1-ol	C3H6O1	39.3	41.4	58.6	46.4	8.9
propan-1-ol	C3H8O1	41.9	44.7	52.9	46.5	6.9
propanenitrile	C3H5N1	35.1	38.2	54.6	42.6	9.6
propionic acid	C3H6O2	37.2	48.9	54.6	46.9	10.4
prop-1-ene	C3H6	32.9	38.1	52.9	41.3	9.1
propanal	C3H6O1	34.7	46.6	47.5	42.9	7.4
prop-1-yne	C3H4	29.9	29.9	52.8	37.5	10.1
pyrrolidine	C4H9N1	53.5	59.4	61.9	58.3	6.2
sec-butylbenzene	C10H14	97.6	118.1	147.1	120.9	22.7
styrene	C8H8	58.7	100.1	147.0	101.9	34.3
(2R)-2-bromo-1,1,1,2-tetrafluoro-ethane	C2H1F4Br1	45.6	47.7	62.9	52.1	8.8
tert-butylbenzene	C10H14	90.7	122.7	149.6	121.0	30.8
tetrahydrofuran	C4H8O1	47.7	52.7	58.5	53.0	4.5
tetrahydropyran	C5H10O1	59.5	65.6	70.8	65.3	7.2
toluene	C7H8	54.5	91.4	105.8	83.9	24.1
1,4-dimethylcyclohexane	C8H16	84.9	95.8	114.8	98.5	17.7
1,1,2-trichloroethylene	C2H1Cl3	46.9	71.1	93.9	70.7	18.4
N,N-diethylethanamine	C6H15N1	83.3	94.4	94.4	90.7	13.4
trimethoxymethylbenzene	C10H14O3	112.8	127.0	160.8	133.5	24.5
undecan-2-one	C11H22O1	114.7	133.5	188.3	145.5	55.2
(2R,3R,4S,5R,6R)-6-(hydroxymethyl)tetrahydropyran-2,3,4,5-tetrol	C6H12O6	85.1	107.0	111.8	101.3	21.0
(2S,3S,4R,5R)-5-(hydroxymethyl)tetrahydrofuran-2,3,4-triol	C5H10O5	78.8	81.0	88.9	82.9	5.7
(2R,4R,5R)-5-(hydroxymethyl)tetrahydrofuran-2,4-diol	C5H10O4	72.5	78.2	87.2	79.3	8.8