Solvatation implicite généralisée pour les systèmes finis et infinis périodiques

Dario Vassetti

▶ To cite this version:

Dario Vassetti. Solvatation implicite généralisée pour les systèmes finis et infinis périodiques. Chimie analytique. Université Paris sciences et lettres, 2021. Français. NNT : 2021UPSLC010 . tel-03956192

HAL Id: tel-03956192
https://pastel.hal.science/tel-03956192

Submitted on 25 Jan 2023

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

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

THÈSE DE DOCTORAT

DE L'UNIVERSITÉ PSL

Préparée à l'École nationale supérieure de chimie de Paris

Generalized implicit solvation for finite and infinite periodic systems
Solvatation implicite généralisée pour les systèmes finis et infinis périodiques

Soutenue par

Dario Vassetti

Le 29/10/2021

École doctorale n°388

Chimie Physique et Chimie Analytique de Paris-Centre

Spécialité

Chimie Physique

Composition du jury :

M. Alexis MARKOVITS Professeur, Sorbonne Université , France	*Président*
M. Frédéric CASTET Professeur , Université de Bordeaux, France	*Rapporteur*
M. Michel RERAT Professeur , Univ. de Pau et des pays de l'Adour, France	*Rapporteur*
Mme Silvia CASASSA Professeur, Université de Turin, Italie	*Examinateur*
M. Frédéric LABAT Maître de Conférences, Chimie ParisTech, France	*Directeur de thèse*

Acknowledgments

First, I would like to thank all members of the dissertation committee for their time, and for having accepted to review and examine my thesis work.

I would then like to thank my supervisor, Dr. Frédéric Labat, for the opportunity to work at Chimie ParisTech. In particular, for your infinite patience, help, and advice throughout these three years of PhD. You have been a great supervisor, giving me the time to explore the various topics and you have always been available for my many doubts and questions.

I would also like to thank both past and present members of the CTM group which have accompanied me in these last three years of my life. The pleasant and friendly work environment has made the years fly by fast, and helped more than words can tell.

Even if far, I would like to thank my friends in Florence (especially the ones that kept me company for many years, and many hours, on the bench outside the library), as without you I probably would not have endured those many years of university, and I would also like to thank all the people and friends I meet here at Chimie ParisTech.

I would also like to greatly thank my family, especially my mother and my sister, for their constant support throughout the years. You have allowed me too finish my studies without worries and is thank too you that I could have this experience in Paris.

Finally, I would like to thank Christelle for her love and support throughout these years of PhD. Thank you for your patience (especially during my long monologues and complaints about society), and the wonderful time and experiences we have passed together.

Contents

Resume de these en francais xi

General context xiii

Méthodologie xvii

Principaux résultats et conclusions obtenus xxiii

Context and Thesis Outline 1

1 Electronic Structure methods 5
- 1.1 Schrödinger Equation . . . 5
  - 1.1.1 The Born-Oppenheimer Approximation . . . 6
- 1.2 Hartree-Fock . . . 6
  - 1.2.1 Limits of the Hartree-Fock approach . . . 10
- 1.3 Density Functional Theory . . . 10
  - 1.3.1 Exchange-Correlation Energy Functionals . . . 13
- 1.4 Periodic Systems . . . 16
  - 1.4.1 Localized Basis Sets . . . 18
  - 1.4.2 Plane Wave Basis Sets . . . 19
  - 1.4.3 Surface Modeling . . . 19

2 Implicit Solvation Models: State of the Art 23
- 2.1 Solvation Energy . . . 23
- 2.2 Electrostatic Contribution . . . 25
  - 2.2.1 The Electrostatic Problem: Poisson Equation . . . 25
    - 2.2.1.1 Self-Consistent Reaction Field Formalism . . . 27
    - 2.2.1.2 Poisson-Boltzmann . . . 28
    - 2.2.1.3 Atomic Point Charge Approximation . . . 30
  - 2.2.2 Overview of Numerical Methods . . . 30
    - 2.2.2.1 Boundary Element Method . . . 30
    - 2.2.2.2 Finite-Difference Method . . . 32
  - 2.2.3 Cavities Definition . . . 34
  - 2.2.4 Born and Generalized Born Model . . . 37
  - 2.2.5 Multipolar Expansion Methods . . . 38
  - 2.2.6 Apparent Surface Charge Formalism . . . 40
    - 2.2.6.1 D-PCM . . . 41
    - 2.2.6.2 C-PCM and COSMO . . . 42
    - 2.2.6.3 IEF-PCM . . . 43

2.3 Non-Electrostatic Contribution 44
    2.3.1 Cavitation 45
    2.3.2 Dispersion 46
    2.3.3 Repulsion 46
    2.3.4 SASA based models 47
        2.3.4.1 Cavity Dispersion and Solvent structural effects: CDS . . . 49
        2.3.4.2 SCCS and SCCS-P 50
        2.3.4.3 xESE and uESE 51
2.4 Periodic Implicit Solvation 52
    2.4.1 Smooth Cavity Periodic Models 53
    2.4.2 Periodic C-PCM 54

3 Implicit solvation in Crystal 57
3.1 Electrostatic Contribution 57
    3.1.1 Finite-Difference Generalized Poisson 58
        3.1.1.1 Atomic Charge Models 62
    3.1.2 Apparent Surface Charges 65
    3.1.3 Solvent Excluded Surface 66
    3.1.4 Self-Consistent Reaction Field 68
3.2 Non-Electrostatic Contribution 71
    3.2.1 Cavity, Dispersion, and Solvent Structural Effects Model 71

4 Analytical calculation of the SASA and its nuclear gradients by stereographic projection 75
4.1 Methodology 77
    4.1.1 SASA calculation 77
    4.1.2 SASA nuclear gradients calculation 81
    4.1.3 Algorithm structure 82
4.2 Computational details 83
4.3 Results and discussion 84
    4.3.1 Total and atomic SASA 85
        4.3.1.1 Mobley test set 85
        4.3.1.2 42 proteins test set 85
        4.3.1.3 SASA size-extensivity for periodic systems 85
        4.3.1.4 HCOOH/NiO (100) 86
    4.3.2 SASA gradients 88
        4.3.2.1 Mobley test set 88
        4.3.2.2 Helix model of poly-glycine and HCOOH/NiO (100) 89
4.4 Conclusions 90

5 Calculating solvation energies with FDPB: effects of the atomic charge and non-electrostatic models 91
5.1 Review of the Methods 92
5.2 Computational Details 93
5.3 Results 94
    5.3.1 ΔG_{ne} contribution 94
    5.3.2 ΔG_{el} contribution 96
        5.3.2.1 Atomic charge and level of theory effects 96

5.3.2.2 Basis set dependence 97
5.3.3 Combined effects on ΔG_{solv} 99
5.3.4 Improving CM5 performances: scaling atomic charges 100
5.3.5 Effects of the charge model on the SCRF performance 102
5.3.6 Effects of the charge model on charged molecules 103
5.4 Conclusions 104

6 Revised CDS model for FDPB calculations: parametrisation and extension to ionic solutes 107
6.1 Methods 107
6.1.1 Electrostatic Contribution 108
6.1.2 Non-Electrostatic Contribution 108
6.1.3 Solvation Energy Correction for Charged Species 110
6.1.4 Training Set and Parameters Optimization 111
6.1.4.1 Training set 111
6.1.4.2 CDS parameter optimization 111
6.1.4.3 Charged species correction and parameter optimization . . . 113
6.2 Computational Details 113
6.3 Results 115
6.3.1 Neutral solutes 115
6.3.1.1 Finite molecular solutes 115
6.3.1.2 Wetting of TiO_2 anatase (101) 120
6.3.2 Ionic solutes 123
6.3.2.1 Free energies of solvation 123
6.3.2.2 pK_a evaluation 125
6.4 Conclusions 127

7 Towards electrostatic forces in FDPB 129
7.1 Overview of dielectric boundary forces 130
7.2 Methods 131
7.2.1 Electric and Displacement fields calculation 133
7.3 Model Systems 134
7.3.1 Two-atoms System 134
7.3.2 Three-atoms System 136
7.4 Molecular Systems 137

8 Conclusions and Perspectives 139

A Partial derivatives of the curve integrals for SASA gradients 145
A.1 Evaluation of $\frac{\partial I^i_{j,\lambda}}{\partial\left(a^i_j,b^i_j,c^i_j,d^i_j\right)}$ 145
A.2 Evaluation of $\frac{\partial I^i_{j,\lambda}}{\partial\left(a^i_k,b^i_k,c^i_k,d^i_k\right)}$ and $\frac{\partial I^i_{j,\lambda}}{\partial\left(a^i_l,b^i_l,c^i_l,d^i_l\right)}$ 151

B Effects of the grid spacing on the solvation energy 153
B.1 1-2 dichloroethane: grid spacing effects 153
B.2 Ions: grid spacing effects 154

C Ions: Effects of the charge and non-electrostatic models 157
C.1 ΔG_{ne} distributions for ions 158
C.2 Effects of the charge model on the SCRF performance 159

List of Acronyms

List of the main acronyms used throughout this thesis:

AO	Atomic Orbital
ASC	Apparent Surface Charges
BEM	Boundary Element Method
BGP	Boundary Grid Points
BZ	Brillouin Zone
CDS	Cavity, Dispersion, Solvent structural effects
CM5	Charge Model 5
COI	Circle of Intersection
DFT	Density Functional Theory
ECP	Effective Core Potential
ESP	ElectroStatic Potential fitted atomic charges
FD	Finite-Difference
FDM	Finite-Difference Method
FDPB	Finite-Difference Poisson-Boltzmann
FEM	Finite-Element Method
GBA	Generalized Born Approximation
GGA	Generalized Gradient Approximation
GH	Global Hybrid
GTF	Gaussian Type Function
HF	Hartree-Fock
KS	Kohn-Sham
HPA	Hirshfeld Population Analysis
IEF	Integral Equation Formalism
LDA	Local Density Approximation
MD	Molecular Dynamics
MO	Molecular Orbital
MPE	Multipolar Expansion
MUD	Mean Unsigned Difference
MUE	Mean Unsigned Error
PAW	Projector Augmented Wave
PBC	Periodic Boundary Conditions
PCM	Polarizable Continuum Model
QM	Quantum-Mechanics
RF	Reaction Field
RMSE	Root Mean Square Error
RUE	Relative Unsigned Error
SASA	Solvent-Accessible Surface Area
SCF	Self-Consistent Field
SCRF	Self-Consistent Reaction Field
SES	Solvent-Excluded Surface
SMx	Solvation Model x (x=6,8,12,D)
O-SOR	Optimal-Successive Over Relaxation
SP	Stereographic Projection
SPT	Scaled Particle Theory
SR	Shrake and Rupley algorithm
STF	Slater Type Orbital
vdW	van der Waals

List of Publications

Publications

[3] Can Oğuz, I, Vassetti, D, Labat, F. "Assessing the performances of different continuum solvation models for the calculation of hydration energies of molecules, polymers and surfaces: a comparison between the SMD, VASPsol and FDPB models." , Theor. Chem. Acc., invited article for a special collection devoted to Young Investigators in the field of Theoretical and Computational Chemistry, 140, 99 (2021)

[2] Vassetti, D, Labat, F. "Evaluation of the performances of different atomic charge and non-electrostatic models in the finite-difference Poisson–Boltzmann approach.", Int. J. Quantum Chem. 2021; 121:e26560.

[1] Vassetti, D, Civalleri, B, Labat, F. "Analytical calculation of the solvent-accessible surface area and its nuclear gradients by stereographic projection: A general approach for molecules, polymers, nanotubes, helices, and surfaces", J. Comput. Chem. 2020; 41: 1464–1479.

Accepted for publication

[1] Vassetti, D, Can Oğuz, I, Labat, F. "Generalizing continuum solvation in Crystal to non-aqueous solvents: implementation, parametrization and application to molecules and surfaces.", submitted for publication to J. Chem. Theory Comput.

Publications in development

[3] Vassetti, I, Civalleri, B, Labat, F. "Solvation forces in self-consistent reaction-field finite-difference approaches: implementation and application to finite and infinite periodic systems".

[2] Vassetti, I, Labat, F. "A simple non-electrostatic model based on the Solvent-Accessible Surface Area and the solvent dielectric constant for FDPB- and IEF- based electrostatics"

[1] Vassetti, D, Labat, F. "A new parametrization of the VASPsol continuum solvation model for non-aqueous solvents".

Resume de these en francais

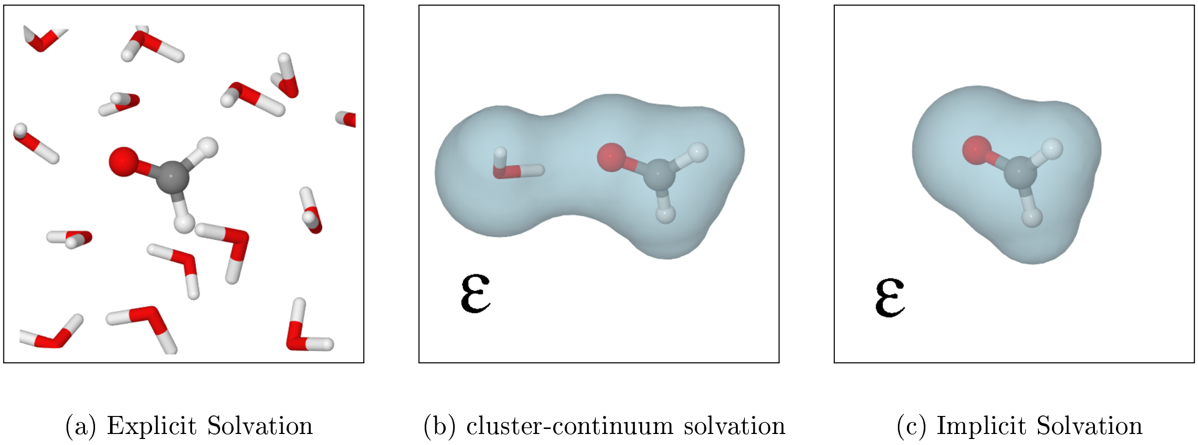
General Context

Les effets des solvants jouent un rôle fondamental dans de nombreuses propriétés physiques et chimiques de la matière [1, 2, 3, 4, 5]. Par exemple, les couleurs de certaines molécules peuvent être affectées par des déplacements de la position des bandes dans les spectres optiques [6, 7, 8], dus au fait que le solvant modifie la stabilité des états excités, et la réactivité chimique d'une molécule change sous l'effet catalytique du solvant [5, 9, 10] en raison du changement de stabilité d'une configuration d'état de transition. Ainsi, la prise en compte des effets du solvant et plus généralement de l'environnement dans les calculs de la mécanique quantique (MQ) est cruciale pour mieux modéliser, comprendre et prévoir les propriétés physiques et chimiques d'un système.

Du point de vue de la modélisation chimique quantique, il existe deux approches principales pour modéliser un solvant et prendre en compte les effets de solvatation, ces approches sont généralement appelées modèles de solvatation explicite et implicite [11, 12, 13, 14].

Les modèles de solvatation explicites sont conceptuellement simples, car les coordonnées atomiques composant une molécule de solvant sont traitées de manière explicite en tenant compte des interactions intermoléculaires directes et spécifiques qui se produisent entre un soluté et un solvant, comme la liaison hydrogène, ainsi que des interactions au sein du solvant lui-même. La Fig. 11a montre un de ces modèles pour une molécule générique dans l'eau. Cette approche de la modélisation du solvant dans un calcul QM est donc intuitive, et offre une image sensible d'un système solvaté en prenant pleinement en compte les complexités de l'environnement, et offre ainsi le niveau le plus détaillé de description des effets de solvatation. La prise en compte de ces complexités est malheureusement aussi l'inconvénient et le facteur limitant de ces approches explicites, surtout en MQ, car le coût de calcul augmente régulièrement avec le nombre d'atomes dans un système. Par exemple, les codes informatiques courants pour l'étude de la structure électronique avec des méthodes telles que la théorie de la fonction de densité (DFT) s'échelonnent avec la taille des atomes du système ($O(N^3)$) [15]. On peut alors voir comment l'inclusion de quelques molécules de solvant, même relativement petites, telles que l'eau, peut augmenter considérablement les temps de calcul, les rendant excessivement longs, voire impossibles à réaliser, en particulier dans le cas de solvants organiques plus grands en raison des nombreux degrés de liberté. De plus, les modèles explicites de solvant nécessitent l'échantillonnage d'un grand nombre de conformations soluté/solvant [16] pour avoir une description entièrement réaliste d'un système, et pour effectuer une analyse statistique afin d'obtenir les propriétés du système. Ceci aggrave encore le coût de calcul de ces modèles, les reléguant à la mécanique moléculaire (MM) [17, 18, 19], ou à des méthodes hybrides telles que QM/MM [20, 21, 22].

Une approche alternative au traitement explicite du solvant est celle suivie dans la solvatation implicite [23, 24, 25, 26]. Dans ces modèles de solvatation, la structure atomique et la densité de charge du solvant sont remplacées par un milieu continu, sans structure, polarisable et défini uniquement par sa permittivité relative statique ϵ. En raison de la nature sans structure de ce milieu diélectrique, une surface *ad hoc* (généralement appelée cavité du soluté, et représentée sur la Fig. 11c pour une molécule générique) doit être définie pour servir d'interface entre la région atomique du soluté caractérisée par $\epsilon = 1$ et le milieu diélectrique caractérisé par la permittivité relative du solvant. Le remplacement des degrés de liberté du solvant par ce milieu diélectrique réduit considérablement le temps de calcul nécessaire pour inclure les effets de solvatation dans un calcul QM et, en outre, supprime la nécessité d'échantillonner la conformation du soluté/solvant, car l'électrostatique globale du solvant est approximée par un effet de champ moyen décrit par la permittivité relative du solvant. En tant que tels, les modèles de solvatation implicites échangent des interactions intermoléculaires soluté/solvant spécifiques, telles que la liaison hydrogène, et d'autres interactions non électrostatiques, comme la dispersion et la répulsion, contre une efficacité de calcul permettant l'introduction d'effets de solvatation dans les calculs QM de grands systèmes à un coût de calcul raisonnable. Il est intéressant de noter que certaines des limitations des modèles de solvatation implicites peuvent être surmontées en considérant un nombre limité de molécules de solvant dans la cavité du soluté, ce qui donne lieu à une approche hybride qui porte généralement le nom de modèle cluster-continuum [27, 28, 29], et qui est illustrée à la Fig. 11b.



(a) Explicit Solvation (b) cluster-continuum solvation (c) Implicit Solvation

Figure 1: Représentations des modèles de solvatation (a) Explicite, (b) cluster-continuum, et (c) Implicite. En bleu clair, la cavité du soluté séparant la région atomistique du soluté et le milieu diélectrique caractérisé par la permittivité relative ϵ.

C'est pour ces raisons que des modèles de solvatation implicites ont été développés et implémentés dans de nombreux codes de QM courants [30, 31, 32, 33, 34, 35]. Au départ, ces modèles étaient principalement axés sur les systèmes moléculaires finis typiques de la chimie organique, puis sur les molécules d'intérêt biologique présentes dans l'eau. Les développements de ces modèles pour les systèmes périodiques, tels que les polymères et les surfaces, ont été moins nombreux. Cela s'explique en partie par les difficultés liées à la nature périodique du système qui complique les équations dans le formalisme de la MQ et dans le problème électrostatique, et par un manque général de données expérimentales disponibles, ce qui rend les validations des modèles de solvatation implicites périodiques plus difficiles que leurs homologues non périodiques, qui peuvent compter sur une grande variété de bases de données pour les tests et le développement [36, 37, 38].

L'objectif de cette thèse est donc de développer et d'implémenter un modèle de solvatation implicite pour les systèmes périodiques et non périodiques, dans le but de mieux modéliser les matériaux en incluant un environnement plus réaliste et plus proche des conditions expérimentales. L'implémentation et le développement sont faits dans CRYSTAL [31, 39], un programme de calcul de la structure électronique de molécules, de polymères, de nanotubes, d'hélices, de surfaces et de cristaux, tant au niveau de la théorie de Hartree-Fock (HF) que de la théorie de la fonction de densité (DFT), en utilisant des combinaisons linéaires de fonctions gaussiennes centrées sur les atomes.

Méthodologie

L'énergie libre de solvatation d'un soluté est définie comme le travail thermodynamique réversible de transfert d'une molécule entre un gaz idéal et un solvant à une pression et une température fixes. Elle a été utilisée pour évaluer la précision du modèle de solvatation implicite dans Crystal.

Pour une conformation donnée, l'énergie libre de solvatation ΔG_{solv} est généralement séparée en deux contributions représentant des phénomènes physiques différents, et appelées changements d'énergie libre électrostatique ΔG_{el} et non électrostatique ΔG_{ne}, ou simplement énergies :

$$\Delta G_{sol} = \Delta G_{el} + \Delta G_{ne}. \tag{1}$$

La contribution électrostatique est liée au travail nécessaire pour polariser le système soluté-solvant. D'un point de vue physique, ces effets de polarisation électronique et nucléaire soluté-solvant proviennent de la réponse électrostatique à courte et longue portée du solvant diélectrique à la densité de charge du soluté $\rho(\boldsymbol{r})$. Cette réponse est un champ électrique appelé champ réactionnel (RF) [72], qui est responsable de la polarisation de la densité de charge du soluté. À son tour, cette polarisation induit une autre polarisation dans le diélectrique, ce qui conduit à un processus autocohérent de polarisation mutuelle entre le soluté et le solvant, appelé champ réactionnel autocohérent (SCRF) [79].

La contribution non-électrostatique n'est pas liée à un seul effet, mais à un terme générique représentant plusieurs d'entre eux que l'on trouve dans la première sphère de solvatation, et qui ne peuvent pas être directement pris en compte en considérant le solvant comme un milieu diélectrique caractérisé uniquement par la constante diélectrique ϵ, comme mentionné précédemment. Les principaux effets sont la répulsion de Pauli à courte portée ΔG_{rep}, les interactions attractives de dispersion ΔG_{dis}, la liaison hydrogène ΔG_{hb}, et la cavitation ΔG_{cav}, qui est définie comme le travail nécessaire pour former une cavité à l'intérieur du milieu diélectrique dans lequel elle sera placée. En première approximation, chaque phénomène est traité indépendamment des autres et l'énergie non électrostatique peut être décomposée en une somme de chaque contribution :

$$\Delta G_{ne} = \Delta G_{rep} + \Delta G_{dis} + \Delta G_{hb} + \Delta G_{cav}. \tag{2}$$

Il est clair que les équations 1 et 2 sont une approximation, car ces phénomènes ne sont pas totalement indépendants les uns des autres. A la lumière de ceci, et du fait que les contributions électrostatiques et non-électrostatiques ne sont pas des observables physiques directes, contrairement à l'énergie de solvatation, la séparation permet un traitement simplifié capable de prédire les énergies de solvatation avec une précision bien supérieure à l'incertitude des données expérimentales pour la plupart des solutés neutres, et proche de l'incertitude expérimentale pour les espèces chargées.

Contribution électrostatique

La contribution électrostatique est calculée par une méthode SCRF [79] où les effets de polarisation de surface à l'interface soluté-solvant sont pris en compte par des charges de surface apparentes (ASC) projetées sur la cavité du soluté [73], et où le problème électrostatique sous-jacent est traité par une approche en différences finies (FD) pour résoudre l'équation de Poisson généralisée [289].

L'équation de Poisson généralisée relie la densité de charge du soluté $\rho(\boldsymbol{r})$ au potentiel $\phi(\boldsymbol{r})$ qu'elle génère, en tenant compte du milieu diélectrique par le biais de la permittivité relative $\epsilon(\boldsymbol{r})$, et est donnée par :

$$\nabla[\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r})] = -4\pi\rho(\boldsymbol{r}) = \sum_{i}^{atoms} q_i \delta(\boldsymbol{r} - \boldsymbol{r}_i), \tag{3}$$

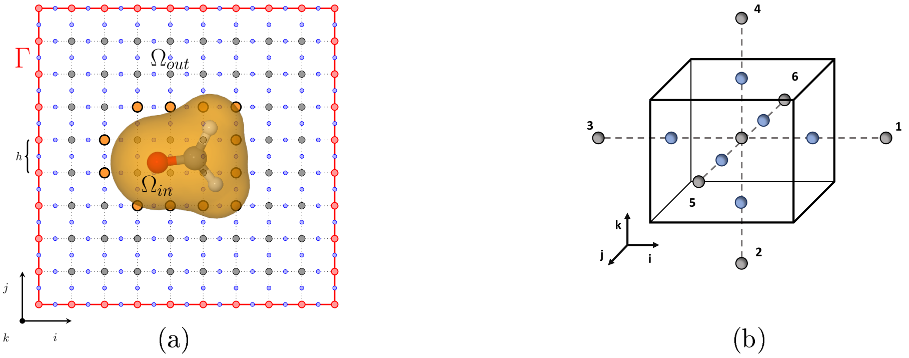
où la densité de charge du soluté $\rho(\boldsymbol{r})$ est ici discrétisée sous forme de charges ponctuelles atomiques q_i.

L'utilisation d'une cavité de soluté [169, 170] à surface exclue du solvant (SES) pour séparer la région atomistique Ω_{in}, où les atomes sont traités explicitement, de la région du solvant Ω_{out}, définit la permittivité relative comme une fonction échelon de la position :

$$\epsilon(\boldsymbol{r}) = \begin{cases} \epsilon_{in} & \text{if } \boldsymbol{r} \in \Omega_{in} \\ \epsilon_{out} & \text{if } \boldsymbol{r} \in \Omega_{out}, \end{cases} \tag{4}$$

où ϵ_{in} est égal à 1 dans la région atomique Ω_{in}, et ϵ_{out} est égal à la permittivité relative du solvant dans la région du solvant Ω_{out}, comme le montre la figure 2a.

Les solutions analytiques de l'équation de Poisson généralisée ne peuvent être obtenues que pour des géométries de cavités simples (par exemple, sphères et ellipsoïdes), alors que les méthodes numériques sont nécessaires pour les cavités de formes moléculaires. Ainsi, pour résoudre numériquement l'équation de Poisson généralisée, une méthode FD a été utilisée, dans laquelle les propriétés physiques du système sont cartographiées sur une grille avec un espacement h et les dérivées sont résolues numériquement aux nœuds de la grille, chacun étant caractérisé par trois entiers (i, j, k) définissant sa position dans l'espace. Un exemple de grille FD pour un système moléculaire fini est présenté à la Fig. 2a.



(a) (b)

Figure 2: (a) Exemple 2D d'une grille FD avec un espacement h. En gris, les nœuds de grille où l'équation de Poisson généralisée est résolue. En bleu, permittivité relative cartographiée entre les nœuds de la grille. En rouge, les points de la grille aux bords où les conditions aux limites sont imposées, et en orange les points de grille des limites à l'interface soluté-solvant. La région atomistique est indiquée par Ω_{in}, et la région du solvant par Ω_{out}. (b) Élément de volume pour une grille cubique FD.

La grille FD nécessite un certain nombre d'étapes préliminaires : (i) chargement de la grille, lors de laquelle les charges atomiques du soluté q_i sont distribuées sur la grille à l'aide d'un algorithme d'interpolation inverse quadratique [288], (ii) attribution des valeurs du diélectrique en chaque point de la grille, et (iii) imposition de conditions limites de Dirichlet ou d'Ewald périodiques aux nœuds de grille en bordure de domaine, en fonction de la périodicité du système. Le potentiel est ensuite résolu de manière itérative à l'aide de l'algorithme O-SOR (Optimal Successive Over Relaxation) [148, 149] en chaque nœud de la grille.

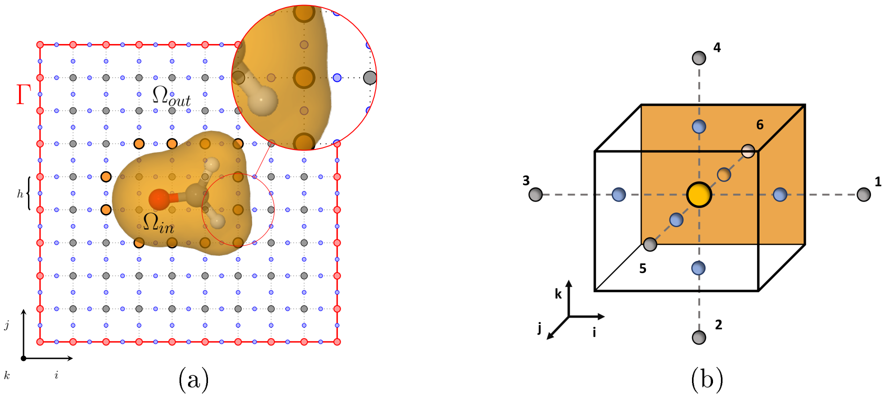
En utilisant comme exemple un système moléculaire fini et une grille FD avec un espacement h, le potentiel en chaque nœud de grille $\phi_{ijk} = \phi(i,j,k)$ est alors donné par :

$$\phi_{ijk}^{0D} = \frac{1}{\sum_i^6 \epsilon_i} \left(\sum_i^6 \epsilon_i \phi_i + 4\pi \frac{q_{ijk}}{h} \right), \tag{5}$$

et dépend de la charge du soluté au nœud de grille q_{ijk} et de la permittivité relative aux six nœuds de grille adjacents, comme le montre également l'illustration 2b.

Cette approche est générale et peut également être appliquée aux systèmes périodiques en une et deux dimensions, tels que des polymères et des surfaces, en utilisant des formules d'Ewald pour des systèmes en périodicités réduites[290], et en imposant des conditions de Dirichlet aux seules directions non périodiques. De plus, des grilles non cubiques peuvent également être considérées[289].

Les potentiels convergés aux nœuds de la grille sont ensuite utilisés pour calculer les valeurs des ASC afin de prendre en compte la polarisation de surface à l'interface soluté-solvant, et sont déterminése à partir des potentiels électrostatiques convergés aux points limites de la grille (BGP). Il s'agit de points spéciaux de la grille FD caractérisés par la présence d'au moins un point voisin diélectrique dans un milieu diélectrique différent, comme le montrent la Fig. 3.3a et la Fig.3.3b.



(a) (b)

Figure 3: (a) Focus sur un point de grille limite sur une grille de différences finies cubique en 2D, et (b) élément de volume élémentaire contenant un point de grille limite en son centre en orange. L'interface soluté-solvant est représentée par un plan orange.

En utilisant à nouveau comme exemple une grille cubique avec un espacement h, l'ASC q_{asc} correspondant à un BGP est obtenu à partir de la valeur du potentiel convergé ϕ_{BGP} à la position de la grille du BGP (i,j,k), et des potentiels ϕ_n aux nœuds de grille voisins $(n \in [1;6])$, selon [289] :

$$q_{asc} = -q_{BGP} + \frac{3h}{2\pi} \left(\phi_{BGP} - \frac{1}{6} \sum_n^6 \phi_n \right), \tag{6}$$

où la charge du soluté q_{BGP} à l'emplacement du BGP doit être retirée de l'ASC. Puisque seule la polarisation de surface, et non de volume, est prise en compte dans le formalisme FD-ASC, une procédure de renormalisation pour les ASC est également effectuée : [231].

Un potentiel d'interaction soluté-solvant V_{int} peut alors être défini à partir de l'ASC cartographié sur la surface exclue du solvant (SES) du soluté à partir du BGP. L'inclusion de ce potentiel d'interaction dans l'hamiltonien de la phase gazeuse du soluté H_{gas} définit un hamiltonien soluté-solvant :

$$H_{solv} = H_{gas} + V_{int}, \tag{7}$$

qui permet un traitement en mécanique quantique des effets de polarisation, et les effets de polarisation mutuelle soluté-solvant peuvent être pris en compte par une méthode itérative de champ réactionnel autocohérent (SCRF) [79, 97], représentée schématiquement en Fig. 2.3.

Le ΔG_{el} est obtenu par la différence entre l'énergie du système soluté-solvant en interaction complète $\langle\psi_s| H_{gas} + V_{int} |\psi_s\rangle$, et la somme de l'énergie de la phase gazeuse du soluté $\langle\psi_g| H_{gaz} |\psi_g\rangle$ et de l'énergie du champ réactionnel E_{RF}, selon :

$$\Delta G_{el} = \langle\psi_s| H_{gas} + V_{int} |\psi_s\rangle - \Big(\langle\psi_g| H_{gas} |\psi_g\rangle + E_{\mathrm{RF}} \Big), \tag{8}$$

où ψ_s et ψ_g sont les fonctions d'onde du système solvaté et en phase gaz, respectivement. L'énergie de champ réactionnel E_{RF} est donnée par une somme d'interactions coulombiennes entre le i^{ime} ASC q_i et le potentiel du soluté $\phi_s(\boldsymbol{r})$ évalué à la position de ces charges ASC $(\boldsymbol{r}_i^{asc})$, selon :

$$E_{RF} = \frac{1}{2} \langle\psi_s| V_{int} |\psi_s\rangle = \frac{1}{2} \sum_i^{N_{asc}} q_i \cdot \phi(\boldsymbol{r}_i^{asc}). \tag{9}$$

D'un point de vue physique, l'énergie de champ réactionnel E_{RF} correspond au travail nécessaire pour placer les ASC $\{q_i\}$, aux positions $\boldsymbol{r}_i^{asc}$, dans le champ généré par la distribution de charge du soluté, et est également égale au travail non réversible nécessaire pour polariser un diélectrique linéaire isotrope, et doit être enlevée pour obtenir le statut d'énergie libre [25, 23]. Pour un système périodique, les ASC $\{q_i\}$ sont calculées dans la maille unitaire de référence, et le potentiel du soluté ϕ est obtenu par les formules d'Ewald [289] aux positions spatiales des ASC.

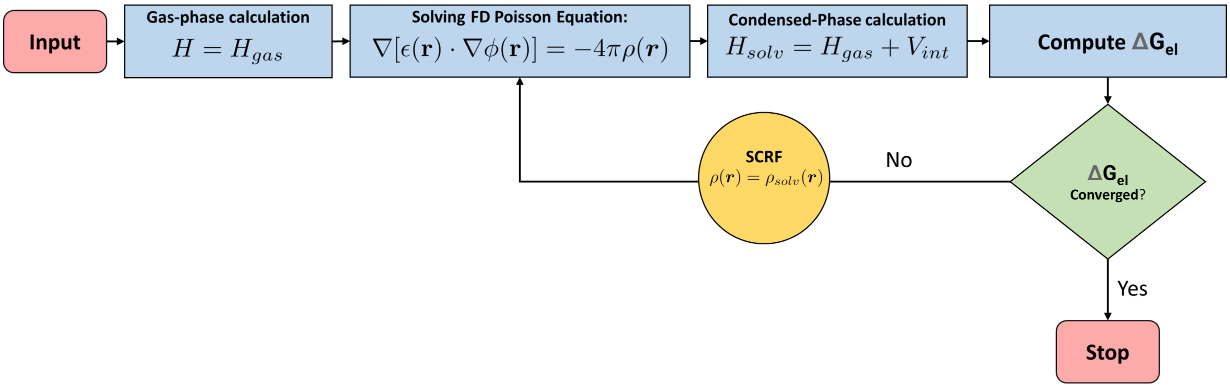

Input
Gas-phase calculation
$H = H_{gas}$
Solving FD Poisson Equation:
$\nabla[\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r})] = -4\pi\rho(\boldsymbol{r})$
Condensed-Phase calculation
$H_{solv} = H_{gas} + V_{int}$
Compute ΔG_{el}
SCRF
$\rho(\boldsymbol{r}) = \rho_{solv}(\boldsymbol{r})$
No
ΔG_{el} Converged?
Yes
Stop

Figure 4: Flow chart for the SCRF process.

Contribution non-électrostatique

L'énergie non électrostatique est liée à la surface accessible au solvant (SASA) [167, 168] du soluté A_m, et est donnée par la somme des surfaces des sphères centrées sur les atomes A_i, selon :

$$A(R_i; r_p) = \sum_{i}^{atoms} A_i(R_i + r_p). \tag{10}$$

où R_i sont les rayons dépendants des atomes des sphères, et r_p est le rayon d'une sonde sphérique de solvant ajoutée à chaque rayon atomique. Les contributions atomiques SASA sont obtenues par une approche analytique de projection stéréographique qui a été implémentée dans le code Crystal et généralisée aux systèmes périodiques [341]. Un exemple de SASA est présenté en Fig. 2.10.

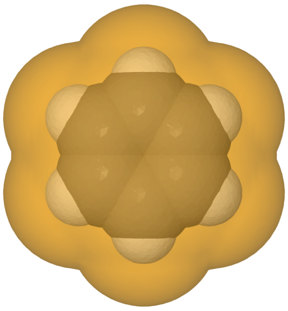


Figure 5: En orange, SASA d'une molécule de benzène, en noir et blanc les sphères dures de van der Waals pour les atomes de carbone et d'hydrogène, respectivement.

La SASA est un composant de base du modèle de Cavitation, Dispersion and Structural Effects (CDS) [81, 80], pour lequel l'énergie non électrostatique ΔG_{ne} est donnée par une contribution pondérée des SASA atomiques A_i et moléculaire A_m, selon :

$$\begin{cases} \Delta G_{ne}^{\mathrm{CDS}} = \sum_{i}^{atoms} \sigma_i A_i + \sigma_M A_M \\ \sigma_i = \left(\tilde{\sigma}_{z_i}^{n} n + \tilde{\sigma}_{z_i}^{\alpha} \alpha + \tilde{\sigma}_{z_i}^{\beta} \beta \right) + \sum_{j \neq i}^{atoms} \left(\tilde{\sigma}_{z_i z_j}^{n} n + \tilde{\sigma}_{z_i z_j}^{\alpha} \alpha + \tilde{\sigma}_{z_i z_j}^{\beta} \beta \right) T(\{z_j, R_{ij}\}) \\ \sigma_M = \tilde{\sigma}^{\gamma} \gamma + \tilde{\sigma}^{\phi^2} \phi^2 + \tilde{\sigma}^{\psi^2} \psi^2 + \tilde{\sigma}^{\beta^2} \beta^2 \end{cases} \tag{11}$$

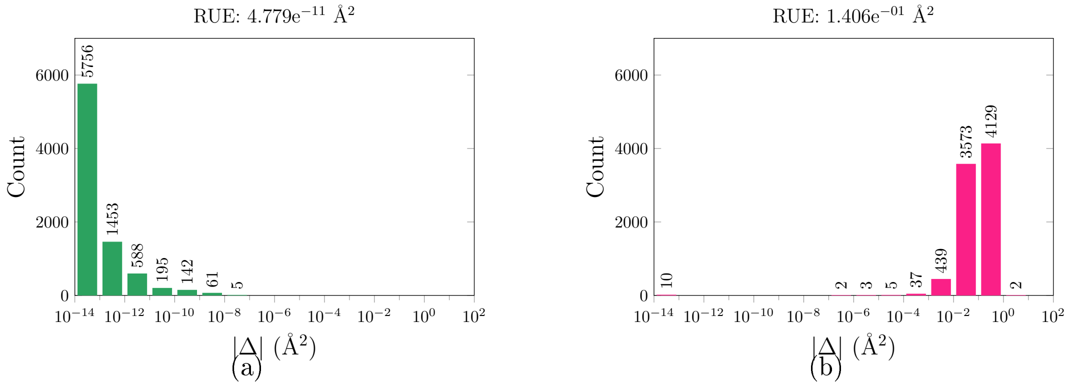
les poids sont des tensions superficielles atomiques σ_i et moléculaires σ_M (d'un point de vue des unités), et dépendent d'un ensemble de descripteurs empiriques du solvant $n, \alpha, \beta, \gamma, \psi, \phi$ et de paramètres optimisables.

En particulier, les descripteurs empiriques sont : l'indice de réfraction n (à 293K), les paramètres d'acidité α [311, 312, 313] et de basicité β [314] hydrogène d'Abraham, la tension superficielle macroscopique du solvant γ, le carré de la fraction d'atomes de solvant non hydrogène qui sont des atomes de carbone aromatiques ϕ^2, et le carré de la fraction d'atomes de solvant non hydrogène qui sont F, Cl ou Br ψ^2. Les paramètres optimisables sont plutôt donnés par des paramètres dépendants du numéro atomique $\tilde{\sigma}_{z_i}$ ainsi que des paramètres de paires d'atomes $\tilde{\sigma}_{z_i z_j}$, plus quatre paramètres globaux indépendants des atomes $\tilde{\sigma}$. En outre, une fonction de coupure $T(\{z_j, R_{ij}\})$ basée sur le numéro atomique et la distance de la paire d'atomes R_{ij} est utilisée pour définir les types d'atomes.

Principaux résultats et conclusions obtenus

Cette thèse contient un total de 216 pages divisées en huit chapitres. Tout au long de la thèse, l'énergie de solvatation a été utilisée pour évaluer les performances du modèle de solvatation, et elle a été considérée comme la somme des contributions électrostatique ΔG_{el} et non-électrostatique ΔG_{ne}. La contribution électrostatique a été prise en compte par une méthode SCRF utilisant un formalisme ASC, où le problème électrostatique sous-jacent est résolu numériquement par un schéma de Poisson généralisé FD, où la densité de charge du soluté est approximée par des charges ponctuelles atomiques. La contribution non-électrostatique représente les effets de solvatation restants qui ne sont pas pris en compte par le modèle électrostatique, et est calculée efficacement par un modèle basé sur la SASA du soluté, car la majorité des effets non-électrostatiques concertent principalement la première sphère de solvatation du soluté.

Dans le chapitre 4, une approche analytique pour calculer la SASA du soluté et ses gradients nucléaires basée sur une technique de projection stéréographique, valable àa la fois pour les systèmes finis et infinis périodiques, a été implémentée dans le code CRYSTAL [341]. Pour les systèmes finis, les SASA obtenues ont été comparées aux valeurs analytiques de référence [330] avec un excellent accord, comme le montre la Fig. 4.8, tandis que pour les systèmes périodiques infinis, des tests d'extensivité de taille ont été réalisés avec succès. L'exactitude des gradients analytiques a également été confirmée par l'excellent accord obtenu avec les gradients numériques et l'invariance translationnelle trouvée pour les systèmes finis et infinis périodiques. Les résultats globaux montrent que la technique de projection stéréographique



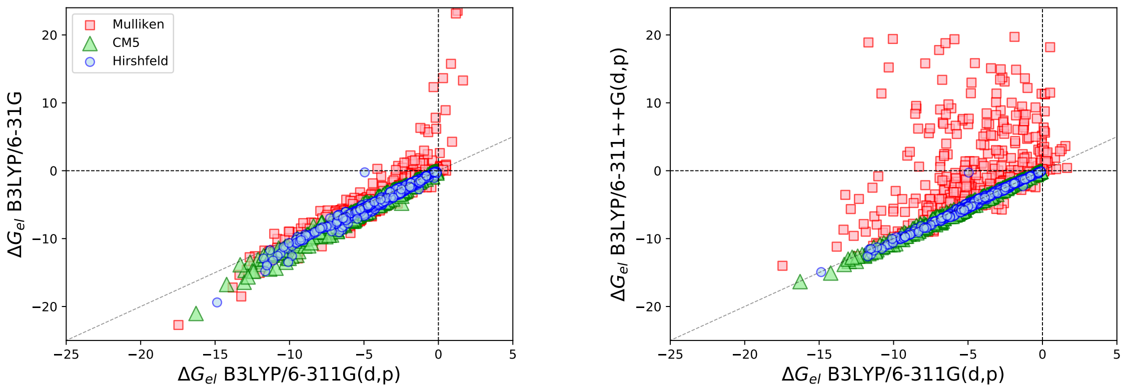
(a) (b)

Figure 6: Histogramme des erreurs relatives non signées (en Å^2) des SASA atomiques pour les 501 molécules de l'ensemble de test de Mobley [38], par rapport à une autre approche analytique de référence [330], obtenues soit avec l'algorithme analytique SP (à gauche), soit avec l'algorithme numérique SR (à droite). SASA atomique de la molécule .

est un moyen général, simple et efficace pour calculer la SASA [341], qui est un composant de base de nombreux modèles de solvatation non-électrostatiques précis et efficaces.

Dans le chapitre 5, comme le terme source de l'équation de Poisson généralisée est approximé par des charges ponctuelles atomiques, la procédure FDPB et le calcul de l'énergie électrostatique dépendent du modèle de charge atomique considéré. Les effets de différents modèles de charge et non électrostatiques sur l'énergie de solvatation ont été étudiés sur un ensemble d'essai de 501 molécules neutres et 112 molécules chargées dans l'eau [38]. Pour la contribution électrostatique, cinq modèles de charge différents ont été considérés : Mulliken [291], Hirshfeld (HPA) [292], Hirshfeld-I (HPA-I) [293, 294], CM5 [194], et CM5-I [191], calculés au niveau DFT avec la fonctionnelle B3LYP et au niveau HF avec les bases 6-31G, 6-311G(d,p) et 6-311++G(d,p). Pour le modèle non électrostatique, deux approches basées sur la sSASA du soluté ont été explorés: le modèle αSASA et le modèle CDS. Ce dernier utilise deux ensembles distincts de paramètres obtenus dans la littérature : CDS-CM5 et CDS-ESP, développés à l'origine pour un modèle de solvatation basé sur l'approche GBA [81].

Les valeurs de l'énergie électrostatique pour l'ensemble de test varient considérablement en fonction des modèles de charge et du niveau de théorie considérés. En moyenne, les énergies électrostatiques obtenues avec le modèle de charge Hirshfeld sont les plus faibles, tandis que les valeurs calculées avec Hirshfeld-I sont les plus élevées, les autres modèles de charge se situant entre les deux. Du point de vue du niveau de théorie, la méthode HF donne globalement des énergies électrostatiques plus négatives par rapport à B3LYP en raison des effets de surpolarisation, et le modèle de charge le moins dépendant de l'ensemble de base est, du plus bas au plus élevé: Hirshfeld<CM5<CM5-I<Hirshfeld-I«Mulliken, ce dernier modèle donnant des énergies de solvatation non physiques pour des ensembles de base plus importants, comme le montre également la figure 5.3. En outre, le modèle de charge Mulliken présente des problèmes techniques dans la procédure de convergence SCRF par rapport aux autres modèles de charge, augmentant le nombre total de cycles SCRF nécessaires à la convergence et nuisant à la performance du modèle.



(a) (b)

Figure 7: Comparaison entre les valeurs de ΔG_{el} calculées avec B3LYP/6-311G(d,p) et B3LYP/6-31G et B3LYP/6-311++G(d,p). Les données obtenues avec des charges de Mulliken (carrés rouges), Hirshfeld (cercles bleus) et CM5 (triangles verts) sont représentées.

En couplant la contribution électrostatique calculée avec les différents modèles de charge avec les modèles non électrostatiques αSASA [289, 191], CDS-ESP et CDS-CM5 pour les molécules neutres, seul un nombre limité de combinaisons a obtenu une erreur moyenne non signée (MUE) inférieure à 1,00 kcal/mol, notamment le modèle de charge Hirshfeld avec CDS-ESP. En général, la surestimation de l'énergie électrostatique par la méthode HF se traduit par une MUE plus élevée par rapport à B3LYP et de meilleures performances sont observées lors de l'utilisation d'ensembles de base incluant des fonctions de diffuses. La valeur de MUE élevée obtenue avec des charges CM5 en utilisant la paramétrisation CDS-CM5 souligne également la faible transférabilité des paramètres d'origine du modèle CDS obtenus avec une approche GBA à une approche FDPB.

Pour améliorer les performances des charges CM5 en utilisant un schéma FD et les paramètres CDS-CM5 par défaut, un facteur d'échelle a été appliqué [191]. La valeur du facteur d'échelle a été obtenue en minimisant l'erreur moyenne non signée (MUE) au niveau de théorie B3LYP/6-311G(d,p) sur les 501 molécules neutres considérées comme un ensemble de test, comme le montre la figure 5.4. Le processus a donné deux minima avec des MUE de 0,68 et 0,74 kcal/mol avec un facteur d'échelle de 0,52 et 0,73 pour les modèles CDS-CM5 et CDS-ESP, respectivement.

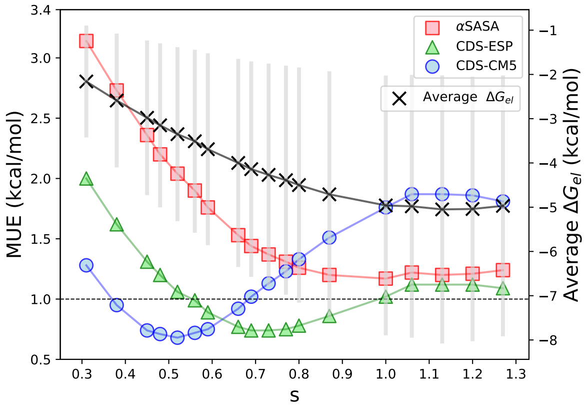


Figure 8: MUE du ΔG_{solv} comparé aux valeurs expérimentales et terme électrostatique moyen pour 501 molécules neutres en fonction du facteur d'échelle des charges CM5 s. Le ΔG_{el}, obtenu au niveau B3LYP/ 6-311G en utilisant différents facteurs d'échelle s, est combiné avec le ΔG_{ne} obtenu avec les modèles ∝SASA, CDS-ESP et CDS-CM5. La ligne en pointillés représente le seuil de 1,00 kcal/mol pour un MUE acceptable. La moyenne des ΔG_{el} (ligne noire pleine) et la dispersion de chaque distribution pour chaque valeur de s (lignes verticales grises) sont également indiquées.

Pour les espèces chargées, Mulliken et le modèle de charge 0,52·CM5 ont donné de piètres résultats, tandis que la performance des autres modèles de charge est fortement affectée par le niveau de théorie. Contrairement aux molécules neutres, l'effet de surpolarisation HF [80] contribue à un MUE plus faible par rapport à B3LYP pour les anions. De bons résultats sont obtenus en considérant la charge CM5 avec les paramètres CDS-CM5. En fait, cette combinaison atteint un MUE de seulement 2,62 kcal/mol, une valeur dans l'incertitude expérimentale de l'ensemble de test de 3,00 kcal/mol. En général, nous pouvons conclure que la surpolarisation et le modèle de charge avec une magnitude de charge plus élevée sont capables d'atteindre un MUE plus faible pour les espèces chargées négativement et devraient être préférés. D'autre part, les performances pour les cations sont bien en deçà de 3,00 kcal/mol pour toutes les combinaisons de modèles de charge et de modèles non électrostatiques, à l'exception des modèles de charge de Mulliken et de 0,52 · CM5.

Bien qu'une bonne précision ait été atteinte dans la prédiction des énergies de solvatation en utilisant les charges 0,52 ·CM5, les résultats sont limités aux molécules neutres, car les espèces chargées se sont révélées extrêmement sensibles aux variations des amplitudes de charge et fortement dépendantes du niveau de théorie. En outre, la mise à l'échelle n'a été vérifiée que pour le solvant aqueux. Cela a mis en évidence la nécessité de réoptimiser les paramètres du modèle CDS et de développer des corrections spécifiques aux espèces chargées, tout en essayant de minimiser la dépendance du modèle au niveau de la théorie.

Dans le chapitre 6, le modèle de solvatation implicite a été étendu aux solvants non aqueux, et le modèle CDS a été reparamétré pour les espèces neutres et étendu pour les espèces chargées en considérant un terme correctif supplémentaire basé sur le paramètre d'acidité d'Abraham du solvant. En particulier, la correction pour les espèces chargées est additive à Eq. 2, et est donnée par :

$$\Delta G_{ne}^{ion} = I(Q, C, k) \cdot \left(\frac{A\alpha}{(1+\alpha^2)} + B \right) A_M \tag{12}$$

où A, B et C sont des paramètres à optimiser, A_M est la SASA totale du soluté, α est l'acidité de liaison hydrogène d'Abraham du solvant et $I(Q, C, k)$ est une fonction choisie pour passer facilement des solutés neutres, chargés positivement et négativement sans discontinuité, définie comme suit :

$$I(Q, C, k) = \frac{|Q|}{1 + exp(k \cdot Q) + C}. \tag{13}$$

Les tests du modèle CDS reparamétré et étendu ont été effectués en considérant les énergies de solvatation pour des systèmes moléculaires finis, les énergies de transfert entre les solvants organiques et l'eau, les calculs de pK_a dans un solvant aqueux, et les calculs de structure de bandes sur un modèle slab d'une surface de TiO_2.

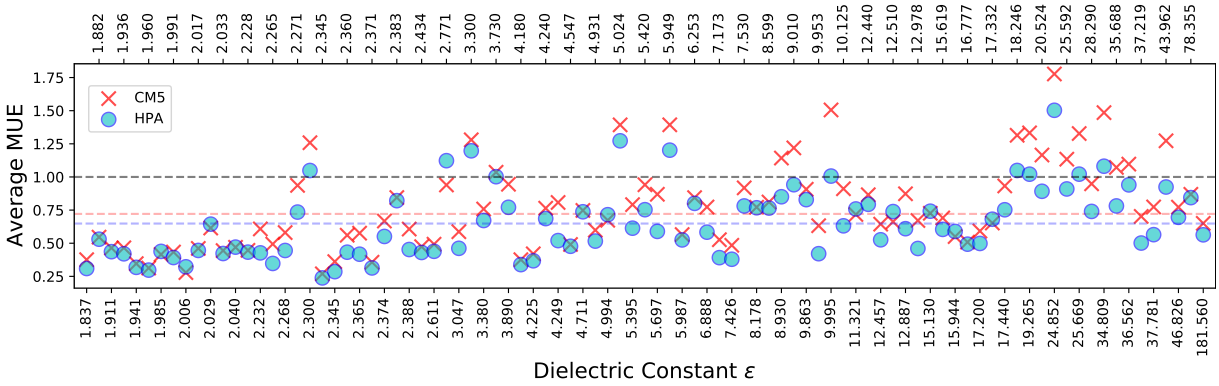


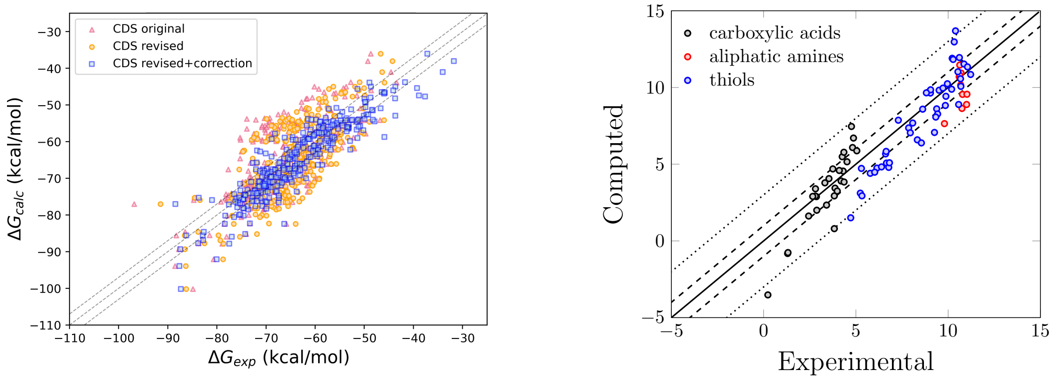
Figure 9: MUE moyenne (en kcal/mol) des énergies de solvatation pour chaque solvant donnée en fonction de la constante diélectrique du solvant (ϵ), en utilisant les charges CM5 (croix rouges) et HPA (points bleus). Les lignes en pointillés colorés indiquent le MUE moyen de chaque modèle de charge $0,72 \pm 0,11$ et $0,64 \pm 0,09$ kcal/mol pour les charges CM5 et HPA, respectivement. La ligne pointillée noire souligne le seuil de 1,00 kcal/mol visé pour les solutés neutres.

La reparamétrisation du modèle CDS est basée sur 10 niveaux de théorie donnés par la combinaison de 5 fonctionnelles (B3LYP, B3PW, mPW1PW, PBE et PBE0) et 2 ensembles de base (6-31G* et 6-311G**). De plus, les modèles de charge HPA et CM5 ont été considérés pour résoudre l'équation de Poisson généralisée en raison de leur plus faible dépendance au niveau de théorie, et de leur plus grande stabilité, par rapport aux autres modèles de charge [191]. Il en résulte un ensemble indépendant de paramètres pour chacun des deux modèles de charge considérés.

Pour le calcul des énergies de solvatation des systèmes moléculaires finis, la base de données MNSOL a été utilisée [36]. Cette base de données comprend 2523 espèces neutres, 91 solvants, et 327 espèces chargées dans 4 solvants, ainsi que 144 énergies libres de transfert entre l'eau et 14 solvants organiques. Pour tester davantage la robustesse du modèle CDS reparamétré et étendu, 83 calculs supplémentaires de pK_a [363] dans l'eau ont été effectués. Pour les systèmes périodiques, en raison du manque d'énergies de solvatation expérimentales fiables, les résultats du modèle FDPB ont été comparés avec le modèle de solvatation implicite de VASPsol en utilisant 3 solvants différents.

En commençant par les espèces neutres, de très bons résultats ont été obtenus pour 2523 combinaisons soluté/solvant neutre dans 91 solvants [36]. En fait, la plupart des MUE calculées se situent sous le seuil de 1,00 kcal/mol pour des résultats acceptables, comme le montre la Fig. 6.3,et il y a un bon accord global avec le modèle de solvatation SM12 utilisé comme référence. De plus, une très faible dépendance au niveau de théorie a été observée pour les 10 niveaux de théorie, justifiant le choix d'utiliser les charges CM5 et HPA. En considérant une moyenne sur les 10 niveaux de théorie, le MUE de l'ensemble des tests est de $0,71 \pm 0,01$ et $0,64 \pm 0,01$ kcal/mol avec les charges atomiques CM5 et HPA, respectivement, et sont comparables aux 0,62 kcal/mol rapportés par le modèle de solvatation SM12 au niveau B3LYP/6-31G* avec les charges atomiques CM5.

Des conclusions similaires aux molécules neutres peuvent être tirées en considérant les 144 énergies de transfert entre l'eau et 14 molécules organiques. Les valeurs MUE sont de $0,71 \pm 0,02$ et $0,86 \pm 0,01$ kcal/mol avec les charges atomiques CM5 et HPA, respectivement. Ces valeurs sont à nouveau en accord avec la valeur du modèle SM12 de 0,60 kcal/mol au niveau B3LYP/6-31G* avec les charges atomiques CM5.



(a) (b)

Figure 10: (a) Energies de solvatation totales calculées à l'aide des charges CM5 au niveau B3LYP/6-311++G(d,p) pour 327 espèces chargées par rapport aux énergies de solvatation expérimentales en utilisant la paramétrisation CDS originale et révisée, et le modèle étendu. (b) Comparaison entre les valeurs pK_a calculées et expérimentales pour 28 acides carboxyliques (en noir), 10 amines aliphatiques (en rouge) et 45 thiols (en bleu) dans l'eau, obtenues avec les charges CM5. Les lignes pointillées intérieures et extérieures représentent des erreurs de ±1,00 et ±3,00 unités pK_a, respectivement.

Pour les 327 espèces chargées dans quatre solvants différents [36], l'inclusion d'un terme correctif basé sur le paramètre d'acidité d'Abraham du solvant améliore considérablement le calcul des énergies de solvatation au niveau DFT. En fait, la MUE pour les anions est réduite de moitié par rapport à celles obtenues avec le modèle CDS original, tandis que la MUE pour les cations est légèrement réduite. Pour les cations et les anions, la MUE calculée est proche de l'incertitude expérimentale de 3,00 kcal/mol, comme le montre la figure 6.6a.

Le terme correctif a été testé plus en détails en effectuant des calculs de pK_a en solvant aqueux pour 83 molécules [363]. Les valeurs de MUE par rapport aux données expérimentales sont de 1,18 et 2,09 unités pK_a avec les charges atomiques CM5 et HPA, respectivement. Ces valeurs améliorent considérablement le modèle de solvatation SMD de référence utilisant le modèle CDS, pour lequel une valeur MUE de 4,37 pK_a unités a été rapportée [363]. En particulier, lorsqu'on utilise les charges CM5, la majorité des valeurs pK_a se situent dans un seuil de 3,00 pK_a unités pour lequel les résultats peuvent être considérés comme acceptables, comme le montre la Fig. 6.6b.

En considérant un modèle slab d'une surface d'anatase TiO_2, un très bon accord a également été trouvé pour les énergies de solvatation calculées et les changements de structure de bandes lors de la solvatation entre le modèle de Poisson en FD dans CRYSTAL et le modèle de solvatation implicite de VASPsol [281], comme le montre le Tab. 6.5. Lors du passage de la phase gazeuse à l'eau, à l'acétonitrile ou au toluène, des tendances très similaires peuvent être observées pour les énergies de solvatation, la stabilisation de la surface, et une légère ouverture de la bande interdite et un déplacement vers le haut des bandes calculées lors de la solvatation. Tous ces effets se sont avérés plus significatifs avec des solvants plus polaires.

	Water		Acetonitrile		Toluene	
	FDPB	VASPsol	FDPB	VASPsol	FDPB	VASPsol
ΔG_{el}	−10.37	−10.94	−8.84	−8.32	−3.27	−1.17
ΔG_{ne}	−0.69	+0.56	+0.52	+0.56	+0.22	+0.54
ΔG_{tot}	−11.06	−10.38	−8.32	−7.76	−3.05	−0.63
$\Delta\Delta G_{el}^{w/x}$	–	–	+1.53	+2.62	+7.10	+9.77

Table 1: Energies libres totales de solvatation calculées (ΔG_{tot}), composantes électrostatiques (ΔG_{el}) et non électrostatiques (ΔG_{ne}), d'un modèle slab d'une surafce de TiO_2 anatase (101) à 10 couches de Ti, solvaté dans de l'eau, de l'acétonitrile et du toluène implicites avec les modèles de solvatation FDPB et VASPsol. $\Delta\Delta G_{el}^{w/x}$ est la différence entre les données ΔG_{el} calculées dans l'eau et dans les autres solvants. Toutes les données sont en kcal/mol. Les charges atomiques CM5 ont été utilisées.

Ces résultats montrent que le modèle de Poisson en FD étendu et reparamétré est capable d'obtenir des énergies de solvatation précises pour une large gamme de solvants et de systèmes différents, tout en présentant une faible dépendance par rapport au niveau de théorie, ce qui démontre la robustesse du modèle. Après avoir obtenu des énergies de solvatation précises, les forces de solvatation pour l'optimisation de la géométrie d'un soluté en solvant sont nécessaires comme étape finale pour compléter le modèle.

Dans le chapitre 7, l'implémentation actuelle des forces de solvatation est détaillée. Elle est considérée comme une somme de contributions non électrostatiques et électrostatiques. Les forces non-électrostatiques précises et efficaces sont obtenues analytiquement à partir du gradient de l'énergie non-électrostatique du modèle CDS, tandis que les forces électrostatiques analytiques sont obtenues à partir d'une formulation basée sur les ASC développée par Cai et al. [153], en utilisant une approche intégrale du tenseur des contraintes de Maxwell.

Les forces électrostatiques analytiques ont été testées sur des systèmes modèles à deux et trois sphères chargeées, et se sont révélées qualitativement conformes aux résultats obtenus à partir des forces calculées numériquement et des résultats du modèle de solvatation implicite MIB [175], validant l'implémentation globale des forces.

Néanmoins, la conservation des forces reste un problème lors du passage à des systèmes moléculaires finis. Ceci peut être attribué à la discrétisation dépendante de la grille et aux erreurs numériques qui peuvent être liées à: (i) la construction numérique des portions réentrantes de la SES, (ii) la décomposition des forces s'exerçant sur la cavité aux deux centres atomiques les plus proches, et (iii) l'interpolation unilatérale des champs électriques à l'interface soluté/solvant. En tant que tel, pour obtenir des forces électrostatiques précises, il faut encore améliorer la formulation de Cai et al. [153] en diminuant sa dépendance à la grille FD par l'utilisation de techniques d'interpolation plus précises et en améliorant la construction des portions réentrantes de la SES.

Dans l'ensemble, le modèle de solvatation implicite s'est révélé efficace et robuste, capable de calculer efficacement les énergies de solvatation pour de nombreux systèmes dans différents environnements de solvant. Parallèlement à l'amélioration des forces électrostatiques, les travaux ont également commencé sur les applications du modèle à des systèmes d'intérêt chimique, tels que la réaction du monoxyde de carbone et de l'eau pour former de l'hydrogène et du dioxyde de carbone sur une plaque de Pt(111) (également connue sous le nom de réaction de déplacement du gaz vers l'eau) en considérant un solvant aqueux afin de tenir compte des effets environnementaux et de modéliser plus précisément la réaction.

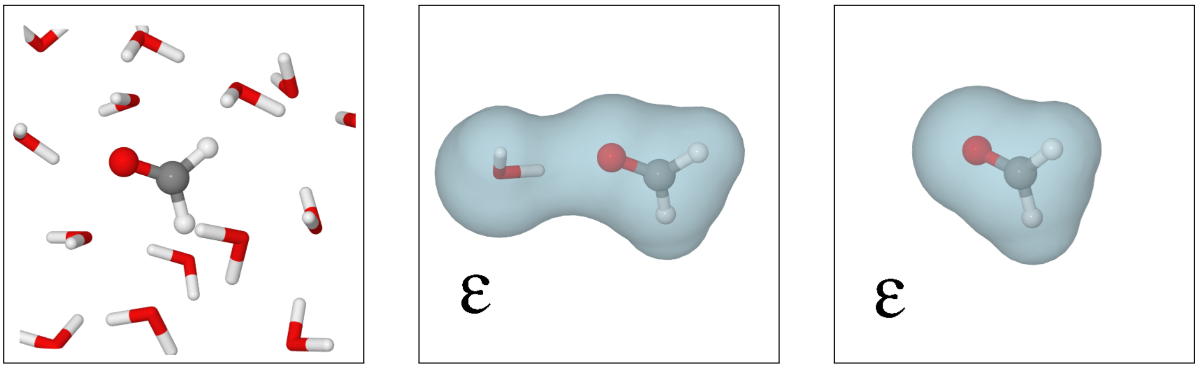
General Context

Solvent effects play a fundamental role in many physical and chemical properties of matter [1, 2, 3, 4, 5]. For instance, the colors of some molecules can be affected by shifts in band position in the optical spectra [6, 7, 8], due to the solvent altering the stability of excited states, and the chemical reactivity of a molecule changing under the catalytic effect of the solvent [5, 9, 10] due the stability change of a transition state configuration. As such, accounting for solvent and more in general environmental effects in Quantum-Mechanical (QM) calculations is crucial to better model, understand, and predict both physical and chemical properties of a system.

From the perspective of quantum chemical modeling two main approaches exist to model a solvent and take in account solvation effects, these approaches are usually referred to as explicit and implicit solvation models [11, 12, 13, 14].

Explicit solvation models are conceptually straightforward, as the atomic coordinates composing a solvent molecule are treated explicitly accounting for direct, and specific, intermolecular interactions that occur between a solute and a solvent, such as hydrogen bonding, as well as interactions within the solvent itself. Fig. 11a shows one such models for a generic molecule in water. This approach to modeling the solvent in a QM calculation is thus intuitive, and offers a sensible picture of a solvated system by fully taking in account the environment complexities, and thus offers the most detailed level of description of solvation effects. Accounting for these complexities is unfortunately also the downside and limiting factor of these explicit approaches, especially in QM, as the computational cost steadily increases with the number of atoms in a system. For example, common computer codes for the study of the electronic structure with methods such as Density Functional Theory (DFT) scale with the size of atoms of the system ($\mathrm{O}(N^3)$) [15]. One can then see how including even few, relatively small, solvent molecules such as water can dramatically increase calculation times, rendering them excessively time consuming if not impossible to perform, especially in the case of larger organic solvents due to the many degrees of freedom. In addition, explicit solvent models require the sampling of a high number of solute/solvent conformations [16] to have a fully realistic description of a system, and to perform statistical analysis to obtain system properties. This further aggravates the computational cost of such models, relegating them to Molecular Mechanics (MM) [17, 18, 19], or hybrid methods such as QM/MM [20, 21, 22].

An alternative approach to explicitly treating the solvent is the one followed in implicit solvation [23, 24, 25, 26]. In these solvation models the atomistic structure and charge density of the solvent is replaced by a continuous, structure-less, polarizable media defined uniquely by its static relative permittivity ϵ. Due to the structure-less nature of this dielectric media an *ad hoc* surface (usually referred to as solute cavity, and shown in Fig. 11c for a generic molecule) has to be defined which acts as interface between the solute atomistic region characterized by $\epsilon = 1$ and the dielectric media characterized by the solvent relative permittivity. The replacement of the solvent degrees of freedom with this dielectric media drastically lowers the computational time required to include solvation effects in a QM calculation and, in addition, removes the need of solute/solvent conformation sampling, as the solvent bulk electrostatics are approximated through a mean-field effect described by the solvent relative permittivity. As such, implicit solvation models trade specific intermolecular solute/solvent interaction such as hydrogen bonding, and other non-electrostatic interactions, like dispersion and repulsion, for computational efficiency allowing the introduction of solvation effects in QM calculations of large systems at a reasonable computational cost. Worth noting is that some of the limitations of implicit solvation models can be overcomed by considering a limited number of solvent molecules within the solute cavity giving rise to a hybrid approach which usually goes under the name of a cluster-continuum model [27, 28, 29], and shown in Fig. 11b.



(a) Explicit Solvation (b) cluster-continuum solvation (c) Implicit Solvation

Figure 11: Representations of (a) Explicit, (b) cluster-continuum, and (c) Implicit solvation models. In light blue the solute cavity separating the solute atomistic region and the dielectric media characterized by the relative permittivity ϵ.

For these reasons implicit solvations models have been developed and implemented in many of the mainstream QM codes [30, 31, 32, 33, 34, 35]. Initially focusing mostly on finite molecular systems typical of organic chemistry, and later also on molecules of biological interest in water, with fewer developments of such models for periodic systems, such as polymer and surfaces. Some reasons for this can be traced back to difficulties related to the periodic nature of the system complicating the equations in the QM formalism and in the electrostatic problem, and a general lack of available experimental data, making validations of periodic implicit solvation models more troublesome than their non-periodic counterparts, which may count on a wide variety of databases for both testing and development [36, 37, 38].

The aim of this thesis is then to develop and implement an implicit solvation model for both periodic and non-periodic systems, with the goal to better model materials by including a more realistic environment closer to the experimental conditions.Implementation and development is made in CRYSTAL [31, 39], a program for the electronic structure calculations of molecules, polymers, nanotubes, helices, surfaces, and crystals, at both the Hartree-Fock (HF) and Density Functional Theory (DFT) levels of theory using linear combinations of atom centred Gaussian functions.

Outline of the Thesis

This manuscript is organized into seven chapters:

- **Chapter I** will introduce the Quantum Mechanical methods for the study of the electronic structure of both molecular, and periodic systems used throughout this thesis. In particular, it will first focus on Hartree-Fock (HF) and Density Functional Theory (DFT) methods for finite molecular systems, and will then focus on their extension to periodic ones.

- **Chapter II** covers the State of the Art in implicit solvation, thus focusing on models of both practical and historical significance. Starting with the wide literature for finite molecular systems, and concluding with models for periodic systems. The solvation free energy of a solute and its main contributions will be used as a guideline within the chapter, due to its importance as a property that any solvation model should accurately predict.

- **Chapter III** will focus on the implicit solvation model in Crystal. In particular, it will discuss the various methods involved in the development of the model and their interconnections. First, considering methods related to the dielectric media and the solute cavity, and then methods to include non-electrostatic effects which are lost by considering the solvent uniquely as a dielectric media.

- **Chapter IV** covers the implementation, and extension to infinite periodic systems, of an analytical method to compute the Solvent-Accessibile Surface Area (SASA) of a solute and its nuclear gradients. The SASA is the starting point to efficiently model effects found in the first solvation shell which are not accounted for by approximating the solvent as a dieletric structureless media, as it will be shown in the following chapters.

- **Chapter V** explores the effects of atomic point charge models, and level of theory, on the computation of the electrostatic contribution to the total solvation energy for neutral and charged solutes in aqueous solvent. Furthermore, effects found in the first solvation shell, which are neglected by approximating the solvent as a structureless dielectric media, are reintroduced by considering computationally efficient non-electrostatic models based on the solute SASA.

- **Chapter VI** will focus on the reparametrization of the non-electrostatic model, introduced in chapter V, to nonaqueous solvents, and its extension to charged solutes. Furthermore, the implicit solvation model is validated considering a large set of neutral and charged solutes, transfer energies between organic and aqueous solvents, computation of pK$_a$'s in water, and calculations on periodic systems in both aqueous and nonaqueous environments.

- **Chapter VII** focuses on the current implementation of the electrostatic forces in Crystal, and will cover the first tests on various model systems used to validated the implementation; together with a comparison with other implicit solvation models.

Chapter 1

Electronic Structure methods

This chapter will overview the main non-relativistic Quantum Mechanical (QM) methods used throughout this thesis. Schrödinger equation and the Born-Oppenheimer approximation will be introduced first, followed by the Hartree-Fock (HF) and Density Functional Theory (DFT) methods for the study of the electronic structure of quantum systems. First focusing on their description for finite molecular systems, and then to their extension to the study of periodic systems such as surfaces.

1.1 Schrödinger Equation

The Schrödinger Equation is a linear partial differential equation that describes the quantum state, also referred to as wave function, of a quantum-mechanical system [40, 41]. The most general form is the time-dependent Schrödinger equation. This equation accounts for the time t evolution of a quantum system of N particles at a position $\boldsymbol{r} = \{\boldsymbol{r_1}, \boldsymbol{r_2}, \boldsymbol{r_3}, ..., \boldsymbol{r_N}\}$:

$$i\hbar \frac{\partial}{\partial t} |\psi(\boldsymbol{r}, t)\rangle = H |\psi(\boldsymbol{r}, t)\rangle , \tag{1.1}$$

where H is the Hamiltonian operator corresponding to the total energy of the system including both kinetic and potential energy, and $\psi(\boldsymbol{r}, t)$ is the time-dependent wave function which describes the state of the system.

The time-dependent Schrödinger equation can be simplified for stationary states and described by a time-independent wave function:

$$H |\psi(\boldsymbol{r})\rangle = E |\psi(\boldsymbol{r})\rangle , \tag{1.2}$$

where E corresponds to the total energy of the system. In this case the Hamiltonian H for a system of N electrons, and M nuclei is, in atomic units, given by:

$$H = \underbrace{-\frac{1}{2}\sum_{i}^{N} \nabla_i^2}_{T_e} \underbrace{-\frac{1}{2}\sum_{A}^{M} \frac{\nabla_A^2}{M_A}}_{T_N} \underbrace{-\sum_{i}^{N}\sum_{A}^{M} \frac{Z_A}{R_{iA}}}_{V_{eN}} + \underbrace{\sum_{i}^{N-1}\sum_{j>i}^{N} \frac{1}{R_{ij}}}_{V_{ee}} + \underbrace{\sum_{A}^{M-1}\sum_{B>A}^{N} \frac{Z_A Z_B}{R_{AB}}}_{V_{NN}} \tag{1.3}$$

where the indexes i and j are used for electrons, and A and B are used for nuclei, R_{iA} and R_{ij} are distances, M_A is the mass, and Z_A the charge of a nucleus A, respectively. The first two terms are the electron T_e and nuclear T_N kinetic energies, the third and the fourth are the electron-nuclear V_{eN} and electron-electron V_{ee} interaction potentials, and the last term is the nuclear-nuclear interaction potential V_{NN}.

1.1.1 The Born-Oppenheimer Approximation

The Born-Oppenheimer approximation allows to further simplify Schrödinger equation by decoupling the electronic and nuclear motions, and by considering the latter as stationary [40, 41].

Physically, this approximation is justified by the higher mass of the nuclei compared to the mass of electrons. This mass difference allows the electrons to rapidly adapt to any change in the nuclear configuration, as opposed to nuclei which are affected by an average electron distribution. Under these conditions the nuclei can be considered stationary, and the nuclear kinetic energy term T_N in the Hamiltonian can be neglected, while the nuclear-nuclear interaction potential V_{NN} becomes a constant.

As a result of considering the nuclei stationary the Hamiltonian can be reorganized as sum of an electronic Hamiltonian H_{el}, which depends on the electrons coordinates $\boldsymbol{r} = \{\boldsymbol{r_1}, \boldsymbol{r_2}, ..., \boldsymbol{r_N}\}$ and parametrically on the nuclear coordinates $\boldsymbol{R} = \{\boldsymbol{R_1}, \boldsymbol{R_2}, ..., \boldsymbol{R_N}\}$, and a nuclear Hamiltonian H_{Nuc} which depends uniquely on the nuclear coordinates:

$$H = \underbrace{T_e + V_{eN} + V_{ee}}_{H_{el}} + \underbrace{V_{NN}}_{H_{nuc}}, \tag{1.4}$$

while the wave function can be factorized, and written as a product of a electronic $\psi_{el}(\boldsymbol{r})$ and nuclear $\psi_{nuc}(\boldsymbol{R})$ wave function:

$$\psi(\boldsymbol{r}, \boldsymbol{R}) = \psi_{el}(\boldsymbol{r})\psi_{nuc}(\boldsymbol{R}). \tag{1.5}$$

Overall, the Born-Oppenheimer reduces the problem of solving Schrödinger equation to an electronic problem:

$$H_{el} \left|\psi_{el}(\boldsymbol{r})\right\rangle = E_{el} \left|\psi_{el}(\boldsymbol{r})\right\rangle, \tag{1.6}$$

and with the total energy of the system E, expressed as the sum of the nuclear E_{nuc} and electronic E_{el} energies:

$$E = E_{el} + E_{nuc}. \tag{1.7}$$

Eq. 1.6 can be solved exactly only for mono-electronic systems with no electronic repulsion. Instead, approximate solutions have to be used for multi-electronic systems, as the electronic wave function $\psi_{el}(\boldsymbol{r})$ cannot be further factorized due to the electron-electron interaction potential V_{ee}. Various methods have been developed to solve Eq. 1.6, and this thesis will focus on two methods in particular: Hartree-Fock (HF) and Density Functional Theory (DFT).

1.2 Hartree-Fock

Hartree-Fock is a method to approximate the wave function and the energy of a multi-electron system under the Born-Oppenheimer approximation [40, 41]. The initial assumption of the method is that the wave function of a N electron system can be factorized, meaning that the electronic wave function $\psi = \psi_{el}(\boldsymbol{r_1}, \boldsymbol{r_2}, ..., \boldsymbol{r_N})$ can be written as a product of mono-electronic wave functions $\phi(\boldsymbol{r}_i)$, called spatial orbitals, each dependent on the coordinates of the i^{th} electron, according to:

$$\psi(\boldsymbol{r_1}, \boldsymbol{r_2}, ..., \boldsymbol{r_N}) = \prod_{i}^{N} \phi(\boldsymbol{r}_i), \tag{1.8}$$

and the spatial orbitals $\phi(\boldsymbol{r})$ are assumed as orthonormal:

$$\langle\psi_i(\boldsymbol{r})|\psi_j(\boldsymbol{r})\rangle = \delta_{i,j} \tag{1.9}$$

where δ_{ij} is the Kronecker delta for which $\delta_{ij} = 0$ if $i \neq j$, and $\delta_{ij} = 1$ if $i = j$. The wave function of Eq. 1.8 corresponds to the wave function of a system of N non-interacting electrons, and is called Hartree product [42].

The Hartree product is not sufficient to accurately treat a system of electrons, as it does not account for the Pauli exclusion principle, for which the function of a system of fermions (in this case electrons) should be antisymmetric to the exchange of two electronic coordinates:

$$\psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = -\psi(\boldsymbol{r}_2, \boldsymbol{r}_1). \tag{1.10}$$

To correctly describe the multi-electron system both spin, and an antisymmetric wave function have to be introduced.

To account for spin the spatial orbital of each electron has to be supplemented by two orthonormal functions $\alpha(\omega)$ and $\beta(\omega)$ describing the two possible spin states of the electron, for which the spin coordinate ω is $+\frac{1}{2}$ for spin up (α) and $-\frac{1}{2}$ for spin down (β). As such, each electron depends both on spatial $\boldsymbol{r}$ and spin ω coordinates $\boldsymbol{x} = \{\boldsymbol{r}, \omega\}$, and for each spatial orbital in Eq. 1.8 two spin-orbitals $\chi(\boldsymbol{x})$ can be defined [40, 41]:

$$\chi(\boldsymbol{x}) = \begin{cases} \phi(\boldsymbol{r})\alpha(\omega) \\ \phi(\boldsymbol{r})\beta(\omega) \end{cases} \tag{1.11}$$

where the spin-orbitals are still orthonormal.

The antisymmetric behavior of the electronic wave function requires the definition of a functional form allowing for sign change upon the exchange of two electrons spatial and spin coordinates. This type of wave function is referred to as Slater determinant [43], and is defined as:

$$\psi(\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\boldsymbol{x}_1) & \chi_2(\boldsymbol{x}_1) & \cdots & \chi_N(\boldsymbol{x}_1) \\ \chi_1(\boldsymbol{x}_2) & \chi_2(\boldsymbol{x}) & \cdots & \chi_N(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\boldsymbol{x}_N) & \chi_2(\boldsymbol{x}_N) & \cdots & \chi_N(\boldsymbol{x}_N) \end{vmatrix}$$

where $\frac{1}{\sqrt{N!}}$ is a normalization constant. The Slater determinant satisfies Pauli exclusion principle as the interchange of two electrons coordinates corresponds to the exchange of two rows, and consequently to a sign change in the determinant. Furthermore, the determinant is consistent with the principle of indistinguishability of identical particles, as electrons occupy all χ_N spin-orbitals, and no one electron can be localized in a specific spin-orbital.

In contrast to the Hartree product, where electrons are fully independent from each other, the Slater determinant accounts only partially for correlation between electrons. In particular, it neglects the Coulomb correlation between electrons, but accounts for correlation between electrons with parallel spin, also known as electron exchange, which prevents two parallel-spin electrons from occupying the same region.

Considering a normalized wave function ($\langle\psi|\psi\rangle = 1$), the energy E_{el} of the system is given by:

$$E_{el} = \langle\psi| H_{el} |\psi\rangle, \tag{1.12}$$

where H_{el} is the electronic Hamiltonian defined in Eq. 1.4. This term can be rearranged to better suit the problem at hand by defining a set of mono-electronic terms h_i for each electron i:

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_A^M \frac{Z_A}{R_{iA}}, \tag{1.13}$$

where the first term is the electron kinetic energy T_e, and the second term is the electron-nuclear interaction potential V_{eN}. The electronic Hamiltonian H_{el} is then given by the sum of h_i over the N electrons and of the electron-electron repulsion V_{ee}:

$$H_{el} = \sum_i^N h_i + V_{ee} \tag{1.14}$$

Using this electronic Hamiltonian H_{el}, and the previously defined Slater determinant as wave function the energy E_{el}, from Eq. 1.12, assumes the expression [40]:

$$E_{el} = \textstyle\sum_i^N \langle \chi_i(1) | \, h_i \, | \chi_i(1) \rangle + \frac{1}{2} \sum_{i,j=1}^N [\langle \chi_i(1)\chi_j(2) | \, \frac{1}{R_{ij}} \, | \chi_i(1)\chi_j(2) \rangle - \langle \chi_i(1)\chi_j(2) | \, \frac{1}{R_{ij}} \, | \chi_j(1)\chi_i(2) \rangle] \tag{1.15}$$

where, due to the indistinguishability principle, the electrons coordinates can be replaced with dummy variables 1 and 2 for an electron one, and for an electron two. This equation can be written in a more compact form to simplify the notation:

$$E_{el} = \sum_i^N \langle i | \, h \, | i \rangle + \frac{1}{2} \sum_{i,j=1}^N [\langle ij | ij \rangle - \langle ij | ji \rangle] \tag{1.16}$$

where $\langle i | \, h \, | i \rangle$ is a one electron integral, and the two electron contribution include two seperate integrals [40, 41]: i) The Coulomb integral $\langle ij | ij \rangle$ accounts for the interaction between the charge density with itself, as the electrostatic interaction is between charges with the same sign, it is a repulsive term, and the contribution to the total energy is positive. ii) The exchange integral $\langle ij | ji \rangle$ arises from the antisymmetry of the Slater determinant, and as the integral is postive the contribution to the total energy will be negative, and thus stabilizing.

Following the variational principle [40, 41], which states that the expectation value E_{el} of a wave function is higher or equal to the exact ground state energy E_0 of the system:

$$E_{el} = \langle \psi | \, H_{el} \, | \psi \rangle \geq E_0, \tag{1.17}$$

the goal of the Hartree-Fock method is to determine the set of spin-orbitals which minimize the energy while staying orthonormal. This is accomplished by Lagrange's method of undetermined multipliers, by defining a functional $\mathcal{L}[\chi_i, \chi_2, ..., \chi_N] = \mathcal{L}[\chi]$, according to:

$$\mathcal{L}[\chi] = E_{el}[\chi] - \sum_{ij} \epsilon_{ij} [\langle i | j \rangle - \delta_{ij}], \tag{1.18}$$

where ϵ_{ij} are undetermined Lagrange multiplier, and by taking the first variation of $\delta\mathcal{L}[\chi]$ in respect to the spin-orbitals, and imposing it to be equal to zero:

$$\delta\mathcal{L}[\chi] = 0. \tag{1.19}$$

This mathematical condition is satisfied when the spin orbital χ_i obeys the following equation:

$$\left[h_i + \sum_j^N (J_j - K_j)\right]\chi_i(1) = \sum_j \epsilon_{ij}\chi_j(1), \tag{1.20}$$

where the term within the square brackets in Eq. 1.20 is a one-electron operator called Fock operator F_i:

$$F_i = h_i + \underbrace{\sum_j^N (J_j - K_j)}_{v_i^{HF}}. \tag{1.21}$$

Within the Fock operator h_i includes the kinetic energy, and nuclear attraction contributions, while the electron repulsion is included in the Hartree-Fock potential v_i^{HF} through the Coulomb J_j and exchange K_j operators, defined as:

$$J_j\chi_i(1) = \int \chi_j^*(2)\frac{1}{r_{12}}\chi_j(2)\chi_i(1)\,dx_2 \tag{1.22}$$

$$K_j\chi_i(1) = \int \chi_j^*(2)\frac{1}{r_{12}}\chi_i(2)\chi_j(1)\,dx_2. \tag{1.23}$$

The potential v_i^{HF} accounts only for an average repulsion experienced by the i^{th} electron due to the remaining electrons, meaning that the charge distribution is fixed and no direct correlation between electron pairs exists.

Nevertheless, Eq. 1.20 must hold for every spin orbital, and can be rewritten to obtain the canonical Hartree-Fock equation after diagonalization through an orthogonal transformation [40, 41]:

$$F_i\chi_i(1) = \epsilon_i\chi_i(1), \tag{1.24}$$

where ϵ_j are the diagonal elements of the Fock operator, also called orbital energies:

$$\delta_{ij}\epsilon_i = \langle\chi_j(1)|\,F\,|\chi_i(1)\rangle\,, \tag{1.25}$$

and which account for the kinetic, nuclear attraction, and electron repulsion energy of one electron in the spin orbital $\chi_j(1)$.

The Hartree-Fock equations are non-linear, as the potential v_i^{HF} defined in Eq. 1.21 for the i^{th} electron depends on the spin orbitals of the remaining electrons. This requires a Self-Consistent Field (SCF) method which iteratively optimizes the spin orbitals to minimize the energy: starting from an initial guess for the spin orbitals the mean field felt by the electrons is computed and the Hartree-Fock equations are solved giving new set of spin orbitals. These spin orbitals are used to compute a new mean field, and so on iteratively until self-consistency; meaning that the mean field, and the total electronic energy E_{el}, does not change below a predefined threshold. The final spin orbitals are used within the Slater determinant, and according to the variational principle, the corresponding energy E_{el} will be an upper bound to the true ground state energy of the system E_0.

A further approximation introduced in the HF method is expanding the spatial orbitals $\{\phi_i\}$ as a Linear Combination of Atomic Orbitals (LCAO) $\{\varphi_i\}$, according to:

$$\phi_i = \sum_\mu c_{\mu i}\varphi_\mu, \tag{1.26}$$

where $\{c_{\mu i}\}$ are coefficients of the linear expansion to be minimized through the HF method, and the spatial orbitals ϕ_i ($i \in [1;\mu]$) are usually referred to as molecular orbitals (MO). The summation over μ in Eq. 1.26 ideally runs to infinity, in what is called a complete basis set, which would allow for the most accurate description of the system. In practice, the MOs are expanded to a limited number of basis sets functions, in a compromise between computational cost and accuracy. The basis set is formed by atomic orbitals (AO) $\{\varphi_i\}$ which are atom centered mono-electronic wave functions, and are typically expressed as Slater Type Functions (STF), or through Gaussian Type Functions (GTF), as described in 1.4.1 on page 18 for both molecular and periodic systems. The LCAO approximation recasts Eq. 1.24 into the Roothaan-Hall equation:

$$\mathbf{FC} = \boldsymbol{\epsilon}\mathbf{SC}, \tag{1.27}$$

where $\mathbf{F}$ is the Fock matrix with elements $F_{\mu\nu} = \langle\varphi_\mu| F |\varphi_\nu\rangle$ (the subscript i to the Fock operator F has been neglected as electrons are indistinguishable), $\mathbf{S}$ is the AO overlap matrix with elements $S_{\mu\nu} \langle\varphi_\mu|\varphi_\nu\rangle$, and $\mathbf{C}$ is the square matrix of coefficients. In contrast to the Hartree-Fock equations the Roothaan-Hall equations have a matrix-form and can be solved using standard linear algebra techniques efficiently.

1.2.1 Limits of the Hartree-Fock approach

In the Hartree-Fock method the correlation between the electrons with parallel spin states is fully taken in account through the exchange term, while the correlation between two electrons with opposite spin states is neglected. This means that the probability of finding two electrons with opposite spin states occupying the same position in space is not zero, as it should be. Then, in the limits of the non-relativistic Born-Oppenheimer approximation, the Hartree-Fock ground state energy E_0^{HF} will be higher then the exact ground state energy $E_0^{Ex.}$ of the system. The missing correlation energy E_0^{corr} can then be defined as the difference between the exact and the Hartree-Fock energy:

$$E_0^{corr} = E_0^{Ex.} - E_0^{HF} < 0, \tag{1.28}$$

in the limit of a complete basis set. To overcome, the limitations of the Hartree-Fock method in accounting for the missing electron correlation post-HF and Density Functional Theory (DFT) methods have been developed. The latter method, has been widely used throughout this thesis.

1.3 Density Functional Theory

Compared to the Hartree-Fock method (HF) the Density Functional Theory (DFT) [44, 45] describes the system not via the multi electron wave function $\psi(\boldsymbol{x}) = \psi(\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N})$, but via the electron density $\rho(\boldsymbol{r})$. The advantage is that while the multi electron wave function $\psi(\boldsymbol{x})$ is dependent on 4N coordinates (three spatial coordinates $\boldsymbol{r}$, and one spin coordinate ω) for each electron in the system, the electron density $\rho(\boldsymbol{r})$ depends uniquely on three spatial coordinates $\boldsymbol{r}$, independently from the size of the system. Furthermore, the electron density $\rho(\boldsymbol{r})$ is a direct physical observable compared to $\psi(\boldsymbol{x})$, which is instead merely a mathematical object. Nevertheless, a relation also exists between the multi electron wave function $\psi(\boldsymbol{x})$ and the electron density $\rho(\boldsymbol{r})$, and is given by:

$$\rho(\boldsymbol{r}) = \int_{\boldsymbol{r}_2} ... \int_{\boldsymbol{r}_N} \psi^*(\boldsymbol{x})\, \psi(\boldsymbol{x}) d\boldsymbol{r}_N, \tag{1.29}$$

with the integration over $\boldsymbol{r}$ giving the total number of electrons N of the system:

$$N = \int_{\boldsymbol{r}} \rho(\boldsymbol{r})\, d\boldsymbol{r}. \tag{1.30}$$

In DFT, the detailed knowledge of the electron density is sufficient to determine all physical properties of a system. In fact, within a non-relativistic Born-Oppenheimer approximation the Hamiltonian H, as defined in Eq. 1.1, can be determined by $\rho(\boldsymbol{r})$. This is demonstrated for the ground state properties of a N electron system in an external potential v_{ext} by the two Hohenberg-Kohn (HK) theorems [46, 44, 45]. In particular, the two theorems state that:

Theorem 1. In the case of a non-degenerate ground state, different Hamiltonians cannot have in common the same ground state electron density $\rho(\boldsymbol{r})$. This allows to define the ground state energy as a unique functional of the electron density $\rho(\boldsymbol{r})$:

$$E = E[\rho(\boldsymbol{r})]. \tag{1.31}$$

Theorem 2. If the electron density $\rho(\boldsymbol{r})$ is the actual ground state density of a system, the functional $E[\rho(\boldsymbol{r})]$ gives the lowest possible energy E. It is thus possible to find the value of the energy of the ground state by minimizing the functional $E[\rho(\boldsymbol{r})]$.

Under these conditions the energy of a system can be written as:

$$E[\rho(\boldsymbol{r})] = T_e[\rho(\boldsymbol{r})] + V_{ee}[\rho(\boldsymbol{r})] + \int \rho(\boldsymbol{r}) v_{ext}(\boldsymbol{r})\, d\boldsymbol{r} \tag{1.32}$$

where $T_e[\rho(\boldsymbol{r})]$ is the electrons kinetic energy, $V_{ee}[\rho(\boldsymbol{r})]$ is the electron-electron interaction potential, and the last term corresponds to the nuclear-electron interaction potential when no further potential other than the nuclear one is considered. The V_{ee} potential can be further decomposed in the sum of two contributions:

$$V_{ee}[\rho(\boldsymbol{r})] = J[\rho(\boldsymbol{r})] + K[\rho(\boldsymbol{r})] \tag{1.33}$$

where $J[\rho(\boldsymbol{r})]$ is the classical Coulomb repulsion potential, given by:

$$J[\rho(\boldsymbol{r})] = \int \frac{\rho(\boldsymbol{r})\rho(\boldsymbol{r'})}{|\boldsymbol{r} - \boldsymbol{r'}|}\, d\boldsymbol{r'}, \tag{1.34}$$

and the $K[\rho(\boldsymbol{r})]$ term accounts for the quantum nature of the interacting electrons. Unfortunately, both the kinetic T_e and $K[\rho(\boldsymbol{r})]$ are not know exactly, but their determination would allow to exactly determine the functional $E[\rho(\boldsymbol{r})]$ and solve the multi electron problem analytically from the electron density $\rho(\boldsymbol{r})$.

Nevertheless, both the density and the energy of a system of N interacting electrons subject to an external potential V_{ext} can be determined. This was rigorously done by Kohn and Sham [46, 44, 45] by introducing a virtual system of N non interacting electrons with the exact same electron density $\rho(\boldsymbol{r})$ of the interacting system under the influence of an external potential v_{KS} now compensating for the missing electron-electron interaction contributions.

In the Kohn-Sham method the exact energy of the system $E[\rho(\boldsymbol{r})]$ is rewritten according to:

$$E[\rho(\boldsymbol{r})] = T_v[\rho(\boldsymbol{r})] + J[\rho(\boldsymbol{r})] + \int \rho(\boldsymbol{r}) v_{ext}(\boldsymbol{r})\, d\boldsymbol{r} + \underbrace{T_i[\rho(\boldsymbol{r})] + K[\rho(\boldsymbol{r})]}_{E_{xc}[\rho(\boldsymbol{r})]} \tag{1.35}$$

where the kinetic energy T_e has been decomposed in the sum of a kinetic energy of the virtual system $T_v[\rho(\boldsymbol{r})]$, and a reminder $T_i[\rho(\boldsymbol{r})]$ which accounts for the difference with the interacting electron-electron system. Both $T_i[\rho(\boldsymbol{r})]$ and $K[\rho(\boldsymbol{r})]$ are regrouped in an additional term know as the exchange-correlation energy functional $E_{xc}[\rho(\boldsymbol{r})]$ which should account for exchange effects such as in the Hartree-Fock method, and correlation between electron pairs. This latter term is given as various approximations, which will be discussed in Sec. 1.3.1. The variation of the exact energy (Eq. 1.35) with respect to the electron density $\rho(\boldsymbol{r})$, in accordance to the variational principle, is:

$$\frac{\partial E[\rho(\boldsymbol{r})]}{\partial \rho(\boldsymbol{r})} = \frac{\partial T_v[\rho(\boldsymbol{r})]}{\partial \rho(\boldsymbol{r})} + \frac{\partial J[\rho(\boldsymbol{r})]}{\partial \rho(\boldsymbol{r})} + v_{ext}(\boldsymbol{r}) + \frac{\partial E_{xc}[\rho(\boldsymbol{r})]}{\partial \rho(\boldsymbol{r})}, \tag{1.36}$$

while in the virtual system invoked by Kohn and Sham the variation is given by:

$$\frac{\partial E[\rho(\boldsymbol{r})]}{\partial \rho(\boldsymbol{r})} = \frac{\partial T_v[\rho(\boldsymbol{r})]}{\partial \rho(\boldsymbol{r})} + v_{KS}(\boldsymbol{r}) \tag{1.37}$$

where $v_{KS}(\boldsymbol{r})$ is the previously mentioned external potential compensating for the electron-electron interaction contributions missing in the virtual system. Eq. 1.36 and Eq. 1.37 are mathematically equivalent when the potential $v_{KS}(\boldsymbol{r})$ is simply set to:

$$v_{KS}(\boldsymbol{r}) = \frac{\partial J[\rho(\boldsymbol{r})]}{\partial \rho(\boldsymbol{r})} + v_{ext}(\boldsymbol{r}) + \frac{\partial E_{xc}[\rho(\boldsymbol{r})]}{\partial \rho(\boldsymbol{r})}, \tag{1.38}$$

and this equivalence is a necessary condition to have the same electron density $\rho(r)$ between the virtual system invoked by Kohn and Sham and the real interacting system of N electrons.

As no explicit expression for the kinetic energy $T_v[\rho(\boldsymbol{r})]$ valid for non homogeneous ($\rho(\boldsymbol{r})$ wise) system is available, and to obtain the electron density $\rho(\boldsymbol{r})$, the Kohn-Sham method then introduces N Schrödinger equations. Each equation described by one electron orbitals $\phi_i(\boldsymbol{r})$ (i=1,...,N) for non interacting electrons in the potential $v_{KS}(\boldsymbol{r})$, according to:

$$\left[-\frac{1}{2}\boldsymbol{\nabla}_i^2 + v_{KS}(\boldsymbol{r}) \right] \phi_i(\boldsymbol{r}) = \epsilon_i \phi_i(\boldsymbol{r}). \tag{1.39}$$

The corresponding electron density $\rho(\boldsymbol{r})$ is given by the sum of the square moduli of the occupied orbitals:

$$\rho(\boldsymbol{r}) = \sum_i |\phi_i(\boldsymbol{r})|^2. \tag{1.40}$$

Eq. 1.39 is also known as Kohn-Sham equation and represents a way to find the ground state electron density, and energy, of a system instead of minimizing $E[\rho(\boldsymbol{r})]$ directly.

As both $J[\rho(\boldsymbol{r})]$ and E_{xc} in $v_{HK}(\boldsymbol{r})$ depend on the electron density $\rho(r)$, which in turn depends on $\phi_i(\boldsymbol{r})$, the problem is non-linear, and is iteratively solved starting from a set of trial orbitals $\phi_i(\boldsymbol{r})$ until self-consistency of the electron density $\rho(\boldsymbol{r})$, just like in the Hartree-Fock method. The corresponding ground state energy of a system is then given by Eq. 1.36 using the self-consistent electron density $\rho(r)$, with the exception of the kinetic energy of the virtual system $T_v[\rho(\boldsymbol{r})]$ which is given by [44]:

$$T_v[\rho(\boldsymbol{r})] = -\sum_i \phi_i^*(\boldsymbol{r})\boldsymbol{\nabla}^2\phi_i^*(\boldsymbol{r}). \tag{1.41}$$

In case the potential $v_{KS}(\boldsymbol{r})$ is defined as in Eq. 1.38 the ground state electron density of the non interacting electron system will be exactly the same as the electron density in the interacting electron system. As such, if the exchange-correlation E_{xc} was known, solving the Kohn-Sham equations self-consistently would allow to obtain the exact energy of the ground state. The difference between the Hatree-Fock method is that, while Hartree-Fock is an approximate theory, the DFT Kohn-Sham method is an exact theory limited by the fact that E_{xc} has an unknown form. Over the years a significant effort has been put to develop new approximations for the E_{xc} term to increase the accuracy of DFT.

1.3.1 Exchange-Correlation Energy Functionals

The exchange-correlation energy E_{xc} in Eq. 1.35 accounts for the contributions to the energy E for which no exact expressions are available [44, 45], and can be generically expressed as:

$$E_{xc}[\rho(\boldsymbol{r})] = E_{Exact}[\rho(\boldsymbol{r})] - \left(T_v[\rho(\boldsymbol{r})] + J[\rho(\boldsymbol{r})] + \int \rho(\boldsymbol{r})v_{ext}(\boldsymbol{r})\, d\boldsymbol{r}\right), \tag{1.42}$$

where $E_{Exact}[\rho(\boldsymbol{r})]$ is the exact energy functional of a system. The E_{xc} term then accounts for purely quantum effects, mostly focusing on the exchange $E_x[\rho(\boldsymbol{r})]$ and correlation $E_c[\rho(\boldsymbol{r})]$ contributions, treated in two separate terms:

$$E_{xc}[\rho(\boldsymbol{r})] = E_x[\rho(\boldsymbol{r})] + E_c[\rho(\boldsymbol{r})], \tag{1.43}$$

and where the electron kinetic energy term $T_i[\rho(\boldsymbol{r})]$, accounting for the kinetic energy difference between the interacting and non interacting electron systems, can be neglected as it tends to be small [44].

Overall, many approximation for the $E_{xc}[\rho(\boldsymbol{r})]$ have been developed, and they are divided in different families according to their complexity and accuracy. In the following, we will cover the Local Density Approximation (LDA) [47], due to its importance in developing more advanced exchange-correlation functionals, and then we will treat the two main families of functionals employed in this thesis: Generalized Gradient Approximation (GGA) [48], and Global Hybrid Functionals (GH). Furthermore, we will mention the Range-Separated Hybrid (RSH) functionals.

The Local Density Approximations the LDA approximations, introduced by Kohm and Sham [47], are a class of exchange-correlation E_{xc} energy functional which depend on the electron density $\rho(\boldsymbol{r})$ of the system.

For spin-unpolarized systems the LDA approximation is generally given by:

$$E_{xc}^{LDA}[\rho(\boldsymbol{r})] = \int \rho(\boldsymbol{r})\epsilon_{xc}(\rho(\boldsymbol{r}))\, d\boldsymbol{r}, \tag{1.44}$$

where ϵ_{xc} is the exchange-correlation energy from the homogeneous electron gas [49] of charge density $\rho(\boldsymbol{r})$. As separate expressions for the correlation E_c and the exchange terms E_x are used (Eq. 1.43), simple analytical forms are available for the homogeneous electron gas of both terms [45]. Albeit, for the correlation term E_c expressions are available only for the limit cases of weak and strong correlation. This has led to many possible formulations for this term to take in account systems with moderate correlation [50, 51, 52] .

An extension of LDA for spin polarized systems also exists, and goes under the name of Local Spin-Density Approximation (LSDA) [47, 50] . For LSDA the electron density $\rho(\boldsymbol{r})$ is given by the sum of spin densities for α and β electrons:

$$\rho(\boldsymbol{r}) = \rho_\alpha(\boldsymbol{r}) + \rho_\beta(\boldsymbol{r}), \tag{1.45}$$

and the exchange-correlation energy for the homogeneous electron gas $\epsilon_{xc}(\rho(\boldsymbol{r}))$ in Eq. 1.44 becomes function of both: $\epsilon_{xc}(\rho_\alpha(\boldsymbol{r}), \rho_\beta(\boldsymbol{r}))$. As in LDA the exchange term E_x is known [53] in LSDA, but approximation are needed for the correlation term E_c. The LDA and LSDA approximations have been improved by introducing a dependency on the exchange-correlation functional on the gradient of the electron density.

Generalized Gradient Approximation (GGA) The GGA family of exchange-correlation functionals takes in account the gradient of the electron density $\boldsymbol{\nabla}\rho(\boldsymbol{r})$ [48]. This is justified as any real electron system is non-homogeneous, and variation of the density in space should be taken into account. The exchange-correlation E_{xc} functional for GGA then takes the form:

$$E_{xc}^{GGA}[\rho(\boldsymbol{r})] = \int \rho(\boldsymbol{r})\epsilon_{xc}(\rho(\boldsymbol{r}), \boldsymbol{\nabla}\rho(\boldsymbol{r}))\, d\boldsymbol{r}. \tag{1.46}$$

These functionals can be divided in two classes [54]: i) functional whose construction is based on the uniform electron gas [49], and ii) semi empirical functional containing parameters fitted to particular classes of systems. Some examples of popular functionals are the non-empirical PW91, developed by Perdew and Wang in 1991 [55] and PBE by Perder, Burke, Ernzerhof [56]. Instead, some examples of popular empirical functionals are: LYP by Lee, Yang, and Par [57], or BLYP when including Becke exchange [58]. More accurate functionals are obtained when including fractions of the exact Hartree-Fock Exchange [45] in a GGA functional, in what is called hybrid functionals.

Global Hybrid (GH) Functionals GH are another class of E_{xc} functionals [59] that account for a fraction of the exact Hartree-Fock exchange E_x^{HF} expressed in terms of the Kohn-Sham orbitals $\phi(\boldsymbol{r})$ instead of the electron density $\rho(\boldsymbol{r})$:

$$E_x^{HF} = -\frac{1}{2}\sum_{ij} \langle\phi_i(1)\phi_j(2)| \frac{1}{r_{ij}} |\phi_j(1)\phi_i(2)\rangle\,. \tag{1.47}$$

Some examples of widely used hybrid E_{xc} functionals, also used in this thesis, are:

- **B3LYP** is a functional based on the Becke exchange (B) [58] and the LYP [57] correlation mentioned for the GGA functionals. The exchange-correlation functional for B3LYP takes the form:

  $$E_{xc}^{B3LYP} = (1-a)E_x^{LSDA} + aE_x^{HF} + bE_x^{B} + (1-c)E_c^{LSDA} + cE_c^{LYP} \tag{1.48}$$

  where $a = 0.20$, $b = 0.72$, and $c = 0.81$ are three empirically determined parameters to reproduce a set of molecular properties (ionization potentials, atomistic energies, and electronic affinities) for a wide range of molecules [50].

- **B3PW** construction is very similar to the above mentioned B3LYP functional. The main difference between the two functionals is the use of the Perdew and Wang correlation (PW) [55] in place of the LYP correlation [57]. The exchange-correlation functional is:

$$E_{xc}^{B3PW} = (1-a)E_x^{LSDA} + aE_x^{HF} + bE_x^{B} + (1-c)E_c^{LSDA} + cE_c^{PW}. \tag{1.49}$$

- **PBE0** as the name suggests the PBE0 functional [60] is based on the PBE GGA functional [56], where a fraction of Hartree-Fock exchange E_x^{HF} has been used:

$$E_{xc}^{PBE0} = E_c^{PBE} + 0.25(E_x^{HF} - E_x^{PBE}). \tag{1.50}$$

  Here the scaling for the exchange part of 0.25 is not obtained empirically, as the above mentioned functionals, and its justification lies on arguments from perturbation theory [60].

Range-Separated Hybrid (RSH) functionals The RSH functionals are a subgroup of hybrid functionals in which the fraction of Hartree-Fock exchange within the functional depends on the distance between the electrons. Usually, within the RSH functionals the Coulomb potential is given as a sum of a short- and long-range contributions treated separately. An example of RSH is the Heyd–Scuseria–Ernzerhof (HSE) [61] screened hybrid functional.

- **HSE** the Coulomb potential for the exchange term in the HSE functional is separated in a short- and long-range contributions in the following way:

$$\frac{1}{r} = \frac{1-erf(\omega r)}{r} + \frac{erf(\omega r)}{r}, \tag{1.51}$$

  where ω is an adjustable parameter related to the separation range. The HSE exchange-correlation functional itself is given by:

$$E_{xc}^{HSE} = a\underbrace{E_x^{HF}(\omega)}_{sr} + (1-a)\underbrace{E_x^{PBE}(\omega)}_{sr} - \underbrace{E_x^{PBE}(\omega)}_{lr} + E_c^{PBE}, \tag{1.52}$$

  where "sr" and "lr" stand for short- and long-range, respectively. The parameter a is the same as in PBE0 (a=0.25). For the limit of $\omega = 0$ the HSE functional reduces to PBE0, instead for $\omega \to \infty$ the functional reduces to PBE.

This class of functional corrects the slowly decaying Hartree-Fock (distance wise) exchange term for metals and narrow band gap semiconductors [61].

1.4 Periodic Systems

A central concept for the study of a periodic system, such as a crystal, is the definition of an infinite lattice, which can be seen as an evenly distributed collection of points (lattice points) along three non coplanar directions extending infinitely [62, 63]. For such lattice a set of basis vectors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, called primitive lattice vectors, can be defined. Each basis vector can be seen as lying on one of the three non coplanar planes. A vector $\mathbf{R}$ joining two lattice points is instead a lattice vector, and is defined as:

$$\mathbf{R}_n = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \tag{1.53}$$

where n_1,n_2, and n_3 are integers. The basis vectors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, forming a parallelepiped, can be used to define a unit cell, which, when containing only one lattice point, coincides with the primitive cell. All other cells within the periodic system are obtained by translation of the unit cell through the lattice vectors $\mathbf{R}$ and are uniquely defined by a set n_i (i=1,2,3) of integers. An example of a periodic lattice, lattice vector, and basis vectors for a 2D system is shown in Fig. 1.1

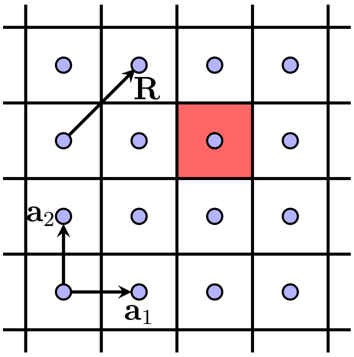


Figure 1.1: Example of a periodic lattice in two-dimensions. Blue dots represent lattice points, and $\mathbf{a}_1$, $\mathbf{a}_2$ form the basis vectors of the translation vector $\mathbf{R}_n = n_1\mathbf{a}_1 + n_2\mathbf{a}_2$. R_n is here represented with $n_1 = n_2 = 1$ in the direct lattice. In red a unit cell.

A useful tool for the study of periodic systems is the definition of a reciprocal lattice [62, 63], which depends on a set of basis vectors $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ that follow an orthogonality rule with the basis vectors of the direct lattice $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, given by:

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}, \tag{1.54}$$

where δ_{ij} is equal to zero if $i \neq j$, and equal to 1 if $i = j$. Similarly to the direct lattice, the reciprocal lattice vectors are defined as a linear combination of the basis vectors:

$$\mathbf{K}_m = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3 \tag{1.55}$$

where m_1,m_2, and m_3 are integers. A unit cell can also be defined, called the first Brillouin Zone (BZ), which is the equivalent of the primitive cell in the direct lattice, but in reciprocal space. The BZ can be visualized as the volume within the closed surface traced by orthogonal planes which pass through the midpoint of lines connecting one reciprocal lattice point to its nearest neighbors, as shown in Fig. 1.2

In particular, for the study of periodic crystal the potential energy must be a periodic function with the same lattice periodicity of the crystal [62, 63]. Meaning, that for any translation by any lattice vector $\mathbf{R} = \mathbf{R}_n$ the value of the potential does not change:

$$V(\boldsymbol{r} + \mathbf{R}) = V(\boldsymbol{r}). \tag{1.56}$$

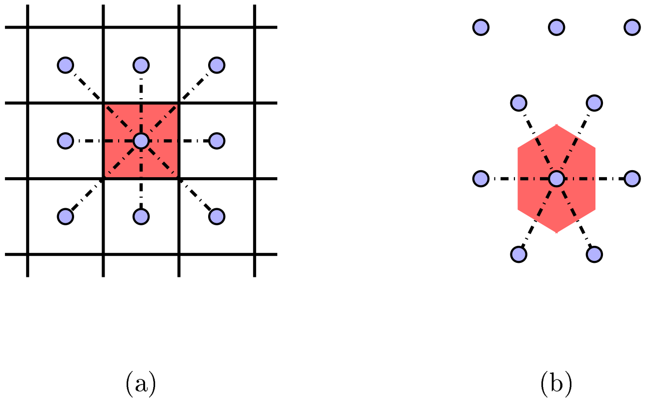


Figure 1.2: Example of the first Brillouin zone for (a) square lattice and (b) hexagonal lattice. Dashed lines represent the path connecting one reciprocal lattice point to its nearest neighbor. In red the area within a closed surface traced by planes orthogonal to the dotted lines midpoints.

Under these conditions, the Schrödinger equation (1.2 on page 5) is also invariant under translation:

$$H(\boldsymbol{r}+\mathbf{R})\psi(\boldsymbol{r}+\mathbf{R}) = E\psi(\boldsymbol{r}+\mathbf{R}), \tag{1.57}$$

and the total energy E of the system does not change $E = E(\boldsymbol{r}+\mathbf{R})$. The wave function for such periodic system are called Bloch functions, and follow the Bloch theorem [62, 63], which states that the solutions of the Schrödinger equation, considering a periodic potential, is a plane wave modulated by a periodic function μ:

$$\psi(\boldsymbol{r};\mathbf{k}) = e^{i\mathbf{k}\cdot\boldsymbol{r}}\mu(\boldsymbol{r};\mathbf{k}), \tag{1.58}$$

the periodic function $\mu(\boldsymbol{r};\mathbf{k})$ has the same periodicity as the direct lattice:

$$\mu(\boldsymbol{r};\mathbf{k}) = \mu(\boldsymbol{r}+\mathbf{R};\mathbf{k}), \tag{1.59}$$

and $\mathbf{k}$ is called wave vector, with the wave function parametrically depending on $\mathbf{k}$. As such, a solution exists for each wave vector $\mathbf{k}$.

The wave vector can be interpreted as a continuous variable in the reciprocal lattice. This if the periodicity is preserved through Periodic Boundary Conditions (PBC), and by defining the wave vector in the reciprocal lattice basis vectors $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ as:

$$\mathbf{k} = \frac{n_1}{N_1}\mathbf{b}_1 + \frac{n_2}{N_2}\mathbf{b}_2 + \frac{n_3}{N_3}\mathbf{b}_3 \tag{1.60}$$

where n_1, n_2, and n_3 are integers, and N_1, N_2, and N_3 are the number of cells within the lattice along each direction and for $N_1, N_2, N_3 \rightarrow \infty$.Furthermore, Bloch functions have particular translational properties in the reciprocal space, for which [62, 63]:

$$\psi(\boldsymbol{r}+\mathbf{R};\mathbf{h}) = e^{i(\mathbf{k}+\mathbf{K})\cdot\mathbf{R}}\psi(\boldsymbol{r};\mathbf{k}+\mathbf{K}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi(\boldsymbol{r};\mathbf{h}), \tag{1.61}$$

where $\mathbf{h} = \mathbf{k}+\mathbf{K}$ is given by the translation of $\mathbf{k}$ by any reciprocal lattice vector $\mathbf{K}$. This allows to limit the resolution of the Hartree-Fock and Kohm-Sham equations to the first Brillouin zone.

In analogy with molecular orbitals [64] Crystaline orbitals [62] can be developed according to the Linear Combination of Atomic Orbitals (LCAO) approximation [40, 41, 44], for which the wave function $\psi_n(\boldsymbol{r};\mathbf{k})$ is given by:

$$\psi_n(\boldsymbol{r};\mathbf{k}) = \sum_{\mu} c_{\mu}^{\mathbf{k}} \phi_{\mu}(\boldsymbol{r};\mathbf{k}), \tag{1.62}$$

where $c_{\mu}^{\mathbf{k}}$ are coefficients of the linear combination, and $\phi_{\mu}(\boldsymbol{r};\mathbf{k})$ are one electron Bloch functions which can be expressed trough a localized basis sets, or plane waves.

1.4.1 Localized Basis Sets

Localized basis sets are given by atomic orbitals (AO) [41, 40], which are atom centered functions $\chi_{\mu}(\boldsymbol{r})$ in the primitive cell, and the corresponding Bloch functions are obtained by considering the AO in different cells $\chi_{\mu}^{\mathbf{R}}(\boldsymbol{r} - \boldsymbol{r}_{\mu})$, according to [62]:

$$\phi_{\mu}(\boldsymbol{r};\mathbf{k}) = \frac{1}{\sqrt{N}} e^{i\mathbf{k}\cdot\mathbf{R}} \chi_{\mu}^{\mathbf{R}}(\boldsymbol{r} - \boldsymbol{r}_{\mu}). \tag{1.63}$$

Two main type of functions are used to described the AO χ_{μ}:

- **Slater Type Functions (STF):** this kind of AO are expressed by the following function obtained from the exact solutions of the Schrödinger equation for the hydrogen atom [41]:

  $$\chi^{STF}(\boldsymbol{r}) = N Y_l^m(\theta,\varphi) \boldsymbol{r}^{n-1} e^{-\zeta \boldsymbol{r}} \tag{1.64}$$

  where ζ is a parameter affecting the exponential radial function, and how contracted the orbital is. The indexes n, l, and m are quantum numbers, N is a normalization constant, and $Y_l^m(\theta,\varphi)$ are spherical harmonics functions. This class of function provide an accurate description of the wave function near the nuclei, as the expected cusp like behavior is reproduced. Nevertheless, STF integrals are computationally expensive compared to Gaussian Type Functions (GTF).

- **Gaussian Type Functions (GTF):** The advantage over STF is given by the Gaussian product theorem, allowing for products of multi center Gaussian functions to be easily integrated. In fact, the product of two GTF centered on different atoms is still a Gaussian function centered along the axis connecting them. This behavior greatly simplifies solving two electrons integrals seen in the Hartree-Fock and Kohn-Sham methods. GTF are given by [41]:

  $$\chi^{GTF}(\boldsymbol{r}) = N Y_l^m(\theta,\varphi) x^i y^j z^k e^{-\zeta \boldsymbol{r}^2} \tag{1.65}$$

  where i, j, and k are integers, and x, y, z are Cartesian coordinates.

  Contracted Gaussian Type Functions (CGTF), given by a linear combination of primitive functions χ_{μ}^{GTF}, are considered to overcome the poor description of the wave function close to the nuclei:

  $$\chi^{CGTF}(\boldsymbol{r}) = \sum_{\mu} d_{\mu r} \chi_{\mu}^{CGTF}(\boldsymbol{r}) \tag{1.66}$$

  where $d_{\mu r}$ are the contraction coefficients. The use of CGTF allows to reproduce the cusp like behavior of STF. Furthermore, even if a higher number of functions are considered the use of GTF is still more computationally efficient compared to STF.

In general, the use of larger AO basis sets allows for a more accurate description of the system, but at the price of increased computational cost. As such, basis sets are chosen as a compromise between computational cost and accuracy. In particular, for atoms with larger atomic numbers the computational cost can be reduced by treating the core and the valence electrons separately. This is justified by the core electrons being less involved in forming chemical bonds, and their accurate description is less needed to fully describe the system. The use of a pseudopotential [65] optimized to reproduce the Coulomb potential generated by the tightly bounded core electrons is then common. Depending on the number of electrons small core, or large core, effective core pseudopotentials (ECP) can be distinguished.

Typically, AO are suitable for the description of crystal where covalent chemical bonds are present, but the description of free electrons, as found typically in metals, is harder to achieve, and an increase in the number of AO may lead to numerical instabilities [62]. Instead, for systems with free electrons plane waves are typically more suitable [62, 66]. Most calculations in this thesis have been performed with localized GTF basis set.

1.4.2 Plane Wave Basis Sets

Plane waves can also be used to expand the crystalline orbitals [66], and are commonly preferable to localized AO when considering metallic systems with free electrons. They are given by:

$$\psi(\boldsymbol{r};\mathbf{k}+\mathbf{K}_n)=\frac{1}{\sqrt{V}}e^{i(\mathbf{k}+\mathbf{K}_n)\cdot\boldsymbol{r}} \tag{1.67}$$

where $\mathbf{K}_n$ is a generic translation vector for the reciprocal lattice, and V is the volume of the system. The orbitals are expanded in a finite number of plane waves, and the total number of plane waves in the expansion is obtained by considering a kinetic energy cutoff E_{cut} that satisfy the following relation [66]:

$$(\mathbf{k}+\mathbf{K}_n)\leq E_{cut}. \tag{1.68}$$

The energy cutoff E_{cut} is chosen to guarantee an accurate representation of the wave functions and of the system.

In this thesis some calculations with plane waves have been performed using the VASP program [34]. In particular, plane waves with the Projected Augmented Wave (PAW) [67, 66] method, a generalization of the pseudopotential seen in localized basis sets.

The PAW method addresses the issue of rapidly oscillating valence wave functions close to nuclei arising from the orthogonality requirements with the core electrons states [67]. This is done by smoothing the wave functions, as it is more convenient from the computational point of view, especially in the case of grid-based methods for which a very fine mesh is otherwise required. Furthermore, PAW provides a way to calculate all electron properties from the smooth wave functions.

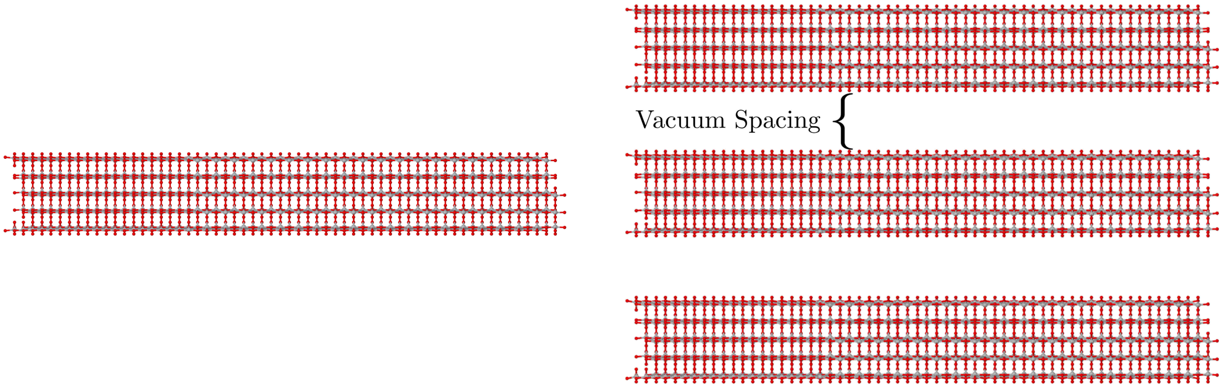
1.4.3 Surface Modeling

The study of surfaces is of great interest in material modeling as many chemical and physical processes take place at the interface between the surface and the environment. A surface can be obtained by cutting an infinite crystal, as the ones previously defined, through one of its crystalline planes.

The crystalline planes are specified through a set of three integers (h, k, l) called Miller indexes [62, 63, 68], and each set of indexes gives a lattice plane orthogonal to the reciprocal lattice vector:

$$\mathbf{K}_{hkl} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3 \tag{1.69}$$

where $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ are the basis vectors, and the (h, k, l) are above mentioned Miller indexes. Then, by cutting the infinite periodic crystal through the crystalline plane defined by (h, k, l) two semi infinite crystal are formed with an infinite number of atoms along the direction orthogonal to the surface. In particular, in this thesis a slab model has been used to model surfaces [69, 62] when using CRYSTAL [31] and a multi slab model when using VASP [34].



(a) (b)

Figure 1.3: Examples of TiO$_2$ anatase (101) slabs generated with (a) a slab model with 2D periodic boundary conditions in the surface plane, and (b) Multi slab model with 3D periodic boundary conditions.

Slab Model this model consists of only N finite layers parallel to the chosen crystalline plane, as shown in Fig 1.3a. As such, the slab is periodically repeated only in the directions orthogonal to the surface normal, and two dimensional periodic boundary conditions are needed.

To ensure an accurate description of the surface an appropriate number of layers for the slab model has to be chosen. This can be done by computing the surface energy E_s until its convergence with the number of layers N in the slab. The surface energy E_s for a N layers slab, is given by the energy per unit area A necessary to form the surface from the bulk crystal, and is given by the following formula [69, 62]:

$$E_s = \frac{E_N - N \cdot E_{bulk}}{2A}, \tag{1.70}$$

where A is the primitive cell surface area, E_N is the energy of a N layer slab, and E_{bulk} is the energy of the primitive unit cell of the bulk material. Alternatively, for metals or small band semiconductor, a more stable formula can be used, given by [69, 62]:

$$E_s = \frac{E_N - N \cdot [E_N - E_{N-1}]}{2A}. \tag{1.71}$$

An acceptable slab model is then obtained when the difference in surface energies between different layers slabs, calculated with one of the previously defined equations, is below a predetermined threshold ϵ:

$$E_{s,n} - E_{s,n-1} \leq \epsilon, \tag{1.72}$$

and at this point, a sufficiently accurate description of an ideal surface is reached.

Multi Slab Model in this model the system is periodic in all three spatial dimensions, as shown in Fig 1.3b. This is done by separating each N layer slab by a certain distance called vacuum spacing. Both the slab thickness N and the vacuum spacing have to be optimized in this model until convergence of the surface energy as defined in Eq. 1.70 or Eq. 1.71 to obtain an accurate description of the system. In particular, vacuum spacing is necessary to prevent any interactions and spurious stabilization effects between the various slabs along the periodic direction perpendicular to the surface, resulting in an incorrect description of the surface to be modelled.

Chapter 2

Implicit Solvation Models: State of the Art

Implicit solvation models have a long history dating back to 1920 with the first models by Born [70], and later by Kirkwood [71], and Onsager [72]. These first models considered simple spherical or ellipsoid solute cavities and focused their treatment on electrostatic effects. Since then the field has seen an impressive growth, especially from 1970-1980s with the first Polarizable Continuum Model (PCM) from Tomasi and co-workers [73, 74, 75, 76, 77, 23] which allowed for a more accurate, and computationally efficient inclusion of solvation effects within a QM framework, overcoming part of the limitations due to the limited computational power of the time. Still today the field remains in constant development with many technical and scientific improvements [23, 24, 25, 26] to the electrostatic treatment and the inclusion of non-electrostatic effects with great improvement in accuracy. In this chapter we will focus on the main historical models that have been developed and their progresses and evolution using a central quantity to compute for any solvation model: the solvation energy [78], defined as the reversible thermodynamic work of transferring a molecule between an ideal gas and a solvent at fixed pressure and temperature.

2.1 Solvation Energy

In implicit solvation models the solvation energy ΔG_{sol} of a solute in a fixed conformation is usually separated between two contributions accounting for different physical phenomena and referred to as electrostatic ΔG_{el} and non-electrostatic ΔG_{ne} free energies changes [64], or simply energies:

$$\Delta G_{sol} = \Delta G_{el} + \Delta G_{ne}. \tag{2.1}$$

The electrostatic contribution to the solvation energy is related to the work needed to polarize the solute-solvent system. Physically, these electronic and nuclear solute-solvent polarization effects arise from both short-range and long-range electrostatic response of the dielectric solvent to the solute charge density $\rho(\mathbf{r})$. This response is an electric field, usually referred to as Reaction-Field (RF) [72], which is responsible for the solute charge density polarization. In turn, this polarization induces further polarization in the dielectric which leads to a self-consistent process of mutual polarization between the solute and the solvent called Self-Consistent Reaction Field (SCRF) [79].

The non-electrostatic contribution is instead not related to a single effect but acts as a wild-card term accounting for several of them which can be found in the first solvation shell, and that cannot be directly accounted for by considering the solvent as a dielectric media characterized uniquely by the dielectric constant ϵ, as previously mentioned. The main effects are short-range Pauli repulsion ΔG_{rep}, attractive dispersion interactions ΔG_{dis}, hydrogen bonding ΔG_{hb} between the solute and the solvent, and cavitation ΔG_{cav}, which is defined as the work required to form a cavity within the dielectric media where the solvent will be placed within. In a first approximation, each phenomena is treated independently from each other and the non-electrostatic energy can be further decomposed as a sum of each contribution, as schematically shown in Fig. 2.1:

$$\Delta G_{ne} = \Delta G_{rep} + \Delta G_{dis} + \Delta G_{hb} + \Delta G_{cav} \tag{2.2}$$

Clearly, Eq. 2.1 and 2.2 are an approximation, as these phenomena are not fully independent from each other. For example hydrogen bonding is affected by electrostatics and both by repulsion and dispersion effects. Another example is cavitation which affects both repulsion and electrostatics, as cavity size plays a major role on the latter. In light of this, and the fact that the electrostatic and non-electrostatic contributions are not direct physical observables, as opposed to the solvation energy, the separation of each physical phenomena allows for a simplified treatment capable of predicting solvation energies well within the experimental uncertainty for most neutral solutes, and close to the experimental uncertainty for charged solutes [80, 81, 82, 83, 37, 84].

In the following sections we will use the above mentioned separation of the solvation energy in an electrostatic and non-electrostatic contribution as a guideline to review the computational mechanisms and theoretical framework of implicit solvation models. We will first focus on the electrostatic formalism, and then we will proceed to describe the non-electrostatic contribution and its importance to obtain an accurate description of solvation effects and energies.

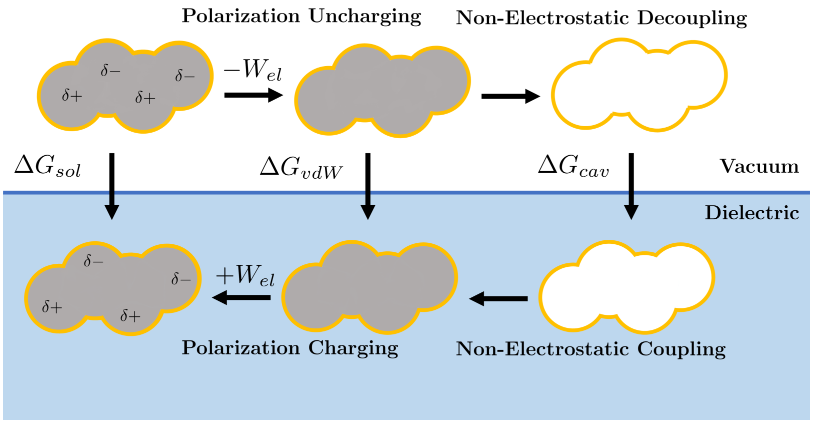


Figure 2.1: Thermodynamic cycle depicting the solvation free energy decomposition into contributions [85]. W_{el} stands for the electrostatic work needed to polarize the dielectric media. The cavitation contribution ΔG_{cav} has been considered separate from ΔG_{ne} (gray) to point out the cavity (yellow) formation within the dielectric media (blue).

2.2 Electrostatic Contribution

In this section we will discuss the main methods developed to obtain the electrostatic contribution to the solvation energy. First, we will focus on Poisson's equation which forms the mathematical backbone of implicit solvation models, followed by numerical methods for its solution, the main definitions of a solute cavity, and we will conclude with main solvation models. In particular, the main families of solvation models: namely the Born [70] and it's generalization to any cavity type known as Generalized Born Approximation (GBA) [86, 87], the MultiPolar Expansion (MPE)[88], and the Apparent Surface Charge (ASC) [73] will be briefly described. It should be noted that the latter two model families are based on analytical solutions of the generalized Poisson equation (with approximations in the case of the GBA model), while the former requires a numerical resolution of this equation.

2.2.1 The Electrostatic Problem: Poisson Equation

The main assumption of most implicit solvation models is to consider the solvent as an isotropic linear dielectric [26] for which the solvent displacement field $\mathbf{D}(\mathbf{r})$, arising from the macroscopic polarization $\mathbf{P}(\mathbf{r})$ of the media, is proportional to the applied electric field $\mathbf{E}(\mathbf{r})$ generated by a given solute charge density $\rho(\mathbf{r})$:

$$\mathbf{D}(\boldsymbol{r}) = \epsilon(\boldsymbol{r})\mathbf{E}(\boldsymbol{r}) = \mathbf{E}(\boldsymbol{r}) + 4\pi\mathbf{P}(\boldsymbol{r}) \tag{2.3}$$

where for a isotropic dielectric $\epsilon(\boldsymbol{r})$ is a scalar value function of the position $\boldsymbol{r}$, while for anisotropic materials $\epsilon(\boldsymbol{r})$ can be replaced by a 3D tensor [89].

Then, under the assumption of the dielectric displacement being proportional to the applied electric field the differential form of Maxwell first equation:

$$\nabla \cdot \mathbf{D}(\boldsymbol{r}) = 4\pi\rho(\boldsymbol{r}) \tag{2.4}$$

can be rearranged to obtain the generalized Poisson's equation [90] by considering the relation between the electric field and the negative gradient of the potential $\mathbf{E}(\boldsymbol{r}) = -\nabla\phi(\boldsymbol{r})$:

$$\nabla[\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r})] = -4\pi\rho(\boldsymbol{r}). \tag{2.5}$$

This differential equation forms the theoretical backbone of implicit solvation models and relates the solute charge density $\rho(\boldsymbol{r})$ to the potential it generates $\phi(\boldsymbol{r})$ taking in account the dielectric media through the above mentioned relative permittivity $\epsilon(\boldsymbol{r})$ function of the spatial coordinates.

The replacement of explicit solvent molecules with a dielectric media removes the clear boundary between the the solute and the solvent which arises from Pauli repulsion and requires the introduction of an arbitrary defined surface Γ which acts as the interface between the solute region where atoms are treated explicitly Ω_{in}, and the region where the solvent is replaced by the dielectric media Ω_{out}, also shown in Fig. 2.2. The introduction of this surface then sets $\epsilon(\boldsymbol{r})$ as a step function [26] with values ϵ_{in} and ϵ_{out}, respectively if $\boldsymbol{r}$ is within the inner Ω_{in} or outer Ω_{out} region, according to:

$$\epsilon(\boldsymbol{r}) = \begin{cases} \epsilon_{in} & \text{if } \boldsymbol{r} \in \Omega_{in}, \\ \epsilon_{out} & \text{if } \boldsymbol{r} \in \Omega_{out} \end{cases} \tag{2.6}$$

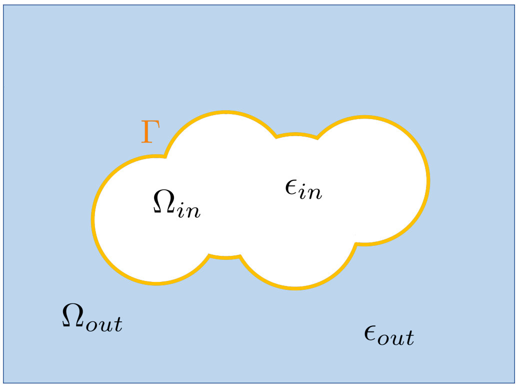


Figure 2.2: 2D example of a solute cavity. In blue the region treated as dielectric media replacing the explicit solvent Ω_{out} and characterized by a relative permittivity ϵ_{out}, in white the solute region where atoms are treated explicitly Ω_{in} with the vacuum relative permittivity set to $\epsilon_{in} = 1$, and in yellow the solute-solvent interface Γ.

A common choice in QM calculation is to use the relative permittivity of vacuum for the inner region Ω_{in}, thus setting $\epsilon_{in} = 1$, while ϵ_{out} for the outer region Ω_{out} is set to the relative permittivity of the solvent.

Then to ensure the uniqueness of the solution of Poisson's equation and a continuous electric displacement $\mathbf{D}(\boldsymbol{r})$ across the surface [91] the following dielectric jump conditions [92, 93] have to applied at the sharp interface Γ:

$$\begin{cases} \phi_{in}(\boldsymbol{r}) = \phi_{out}(\boldsymbol{r}) \\ \epsilon_{in}\mathbf{n} \cdot \nabla\phi(\boldsymbol{r})\Big|_{in} = \epsilon_{out}\mathbf{n} \cdot \nabla\phi(\boldsymbol{r})\Big|_{out} \end{cases} \tag{2.7}$$

where $\mathbf{n}$ is the outer normal direction to Γ and the potentials $\phi(\boldsymbol{r})$, and derivatives of the potential $\nabla\phi(\boldsymbol{r})$, are taken at neighboring points immediately inside and outside the surface Γ [23].

Under these conditons the Poisson equation can be solved for $\phi(\boldsymbol{r})$ throughout space in both the inner and outer regions Ω previously defined, with the potential $\phi(\boldsymbol{r})$ given as sum of two potentials:

$$\phi(\boldsymbol{r}) = \phi_{\rho}(\boldsymbol{r}) + \phi_{RF}(\boldsymbol{r}) \tag{2.8}$$

where $\phi_{\rho}(\boldsymbol{r})$ is the potential of the charge density of the solute $\rho(\boldsymbol{r})$, originally responsible for the polarization of the dielectric media and the appearance of an additional charge density term $\rho_{pol}(\boldsymbol{r})$ acting as source term of the reaction-field potential $\phi_{RF}(\boldsymbol{r})$.

The associated reaction-field energy E_{RF} to $\phi_{RF}(\boldsymbol{r})$ corresponds to the work to place the solute within the polarized media and is given by half the Coulombic interaction between the solute potential $\phi_{\rho}(\boldsymbol{r})$ and the polarized charge density $\rho_{pol}(\boldsymbol{r})$ [94, 95]:

$$E_{RF} = \frac{1}{2}\int \phi_{RF}(\boldsymbol{r})\rho(\boldsymbol{r})d\boldsymbol{r} = \frac{1}{2}\int \phi_{\rho}(\boldsymbol{r})\rho_{pol}(\boldsymbol{r})d\boldsymbol{r} \tag{2.9}$$

where the 1/2 factor accounts for the non-reversible work required to polarize an isotropic linear dielectric [25].

In a QM calculation the electrostatic free energy change ΔG_{el} contributing to the solvation energy is then given by the following equation [64, 96, 25]:

$$\Delta G_{el} = \langle\psi_s| \, H_g + V_{int} \, |\psi_s\rangle - \left(\langle\psi_g| \, H_g \, |\psi_g\rangle + E_{\mathrm{RF}} \right) \tag{2.10}$$

in which the first term $\langle\psi_s| \, H_g + V_{int} \, |\psi_s\rangle$ refers to the solute-solvent system energy with the solvated wave function ψ_s while V_{int} is the interaction potential which is related to the reaction-field energy by:

$$\langle\psi_s| \, V_{int} \, |\psi_s\rangle = 2E_{RF}. \tag{2.11}$$

The second term is the sum of the gas phase energy $\langle\psi_g| \, H_g \, |\psi_g\rangle$ and of the reaction field energy E_{RF}. As previously said, this last component is also equal to the non-reversible work necessary to polarize the dielectric media and has to be removed to obtain a free energy [25].

At this point the dielectric media has been polarized by the solute charge density, but to ensure the mutual solute-solvent polarization and a realistic representation of electrostatic effects the iterative Self-Consistent Reaction Field (SCRF) [79] is considered. The naming of which comes from the central role of the dielectric media reaction-field (RF) [72] in the solvation process.

2.2.1.1 Self-Consistent Reaction Field Formalism

The Self-Consistent Reaction Field (SCRF) procedure requires the introduction of an effective Hamiltonian H_{sol} for the solute-solvent system [25, 23, 79, 97]. This is done by augmenting the gas-phase Hamiltonian H_g with the interaction potential V_{int} obtained from solving the generalized Poisson equation:

$$H_{sol} = H_g + V_{int}. \tag{2.12}$$

The inclusion of the polarized charge density $\rho_{pol}(\boldsymbol{r})$ through the interaction potential V_{int} within the system Hamiltonian is what allows for the mutual solute-solvent polarization within QM computations. A flowchart of a SCRF procedure is reported in Fig. 2.3.

The SCRF procedure starts with the computation of the gas-phase wave function ψ_g and the corresponding charge density of the solute $\rho_{gas}(\boldsymbol{r})$. The density $\rho_{gas}(\boldsymbol{r})$ is then used as source term for the generalized Poisson equation which solution is the solute potential $\phi_\rho(\boldsymbol{r})$ which causes the initial polarization of dielectric media $\rho_{pol}(\boldsymbol{r})$. With the potential and the polarized charge density stems the interaction potential V_{int} used to define a new Hamiltonian H_{sol}, which will now include the effects of the solvent. A condensed-phase calculation is then performed to obtain a new solute charge density $\rho_{sol}(\boldsymbol{r})$ perturbed by the polarized dielectric media. The new solute charge density $\rho_{sol}(\boldsymbol{r})$ will then be used together with the necessary computed quantities to obtain the electrostatic solvation energy ΔG_{el}. The difference between electrostatic solvation energies ΔG_{el} calculated in the latest and previous SCRF cycles of the procedure is used to determine if to stop or continue the SCRF procedure. In the latter case, the generalized Poisson equations is solved once again with the latest computed solute charge density $\rho_{sol}(\boldsymbol{r})$, and the cycle continues until convergence of the electrostatic solvation energy ΔG_{el}.

As mentioned in Sec. 2.1 on page 23 the SCRF is what allows to computationally reproduce the physics of mutual solute-solvent polarization, and is the physical foundation of many of the most advanced and accurate implicit solvation models to date [80, 98, 37, 99, 81].

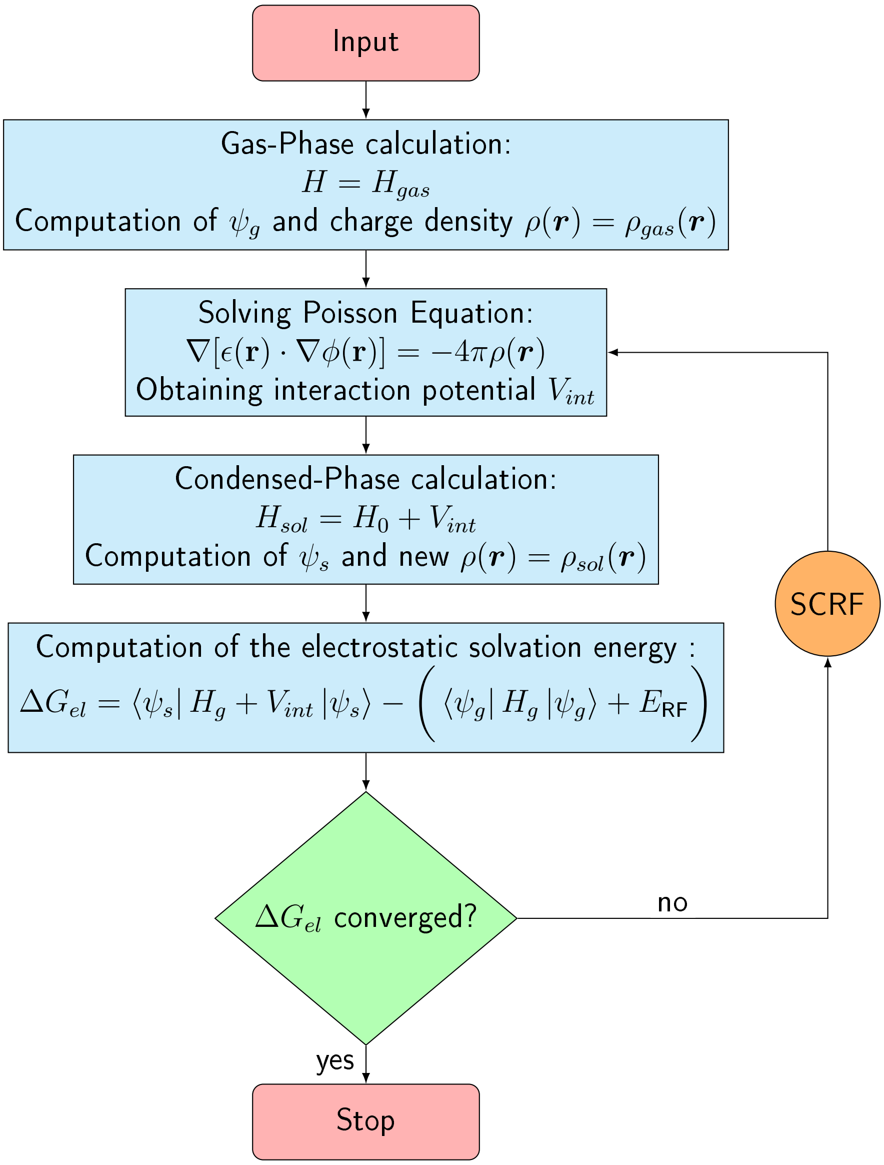

Input
Gas-Phase calculation:
$H = H_{gas}$
Computation of ψ_g and charge density $\rho(\boldsymbol{r}) = \rho_{gas}(\boldsymbol{r})$
Solving Poisson Equation:
$\nabla[\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r})] = -4\pi\rho(\boldsymbol{r})$
Obtaining interaction potential V_{int}
Condensed-Phase calculation:
$H_{sol} = H_0 + V_{int}$
Computation of ψ_s and new $\rho(\boldsymbol{r}) = \rho_{sol}(\boldsymbol{r})$
Computation of the electrostatic solvation energy :
$\Delta G_{el} = \langle\psi_s| H_g + V_{int} |\psi_s\rangle - \Big(\langle\psi_g| H_g |\psi_g\rangle + E_{\mathrm{RF}} \Big)$
ΔG_{el} converged?
no
yes
SCRF
Stop

Figure 2.3: Flowchart of a generic SCRF procedure.

2.2.1.2 Poisson-Boltzmann

The Poisson equation can be extended to consider the dielectric media as an overall neutral electrolyte with uniformly distributed ions within, as shown in Fig. 2.4. For example, one can think of modeling a molecule of biological interest dissolved in a saline solution. In this case the equation becomes non-linear and is known as the Poisson-Boltzmann equation [76, 100, 101, 102, 103], and the solute charge density $\rho(\boldsymbol{r})$ is augmented by a further charge density $\rho_{ions}(\boldsymbol{r})$ to consider the ions:

$$\nabla[\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r})] = -4\pi[\rho(\boldsymbol{r}) + \rho_{ions}(\boldsymbol{r})]. \tag{2.13}$$

This charge density $\rho_{ions}(\boldsymbol{r})$ is treated statistically via mean-field assumption [104, 105], and each ionic specie i on the whole contributes:

$$\rho_{ions}(\boldsymbol{r}) = \lambda(\boldsymbol{r}) \sum_i^N q_i c_i \tag{2.14}$$

where N is the number of ionic species (e.g. N=2 for NaCl), q_i is the total charge of a specie, $\lambda(\boldsymbol{r})$ is a step function to exclude ions from within the solute cavity, and c_i is the local ion density given by Boltzmann equation [106]:

$$c_i = c_i^0 exp\left(\frac{-q_i\phi(\boldsymbol{r})}{k_bT}\right) \tag{2.15}$$

with c_i^0 the ion bulk concentration, k_B the Boltzmann constant, T the temperature, and $-q_i\phi(\boldsymbol{r})$ is the work to move the charge from infinity. The expression for the ion charge density $\rho_{ions}(\boldsymbol{r})$ can be simplified in the case of monovalent salts (e.g. NaCl), reducing to:

$$\rho_{ions}(\boldsymbol{r}) = -2c\lambda(\boldsymbol{r})sinh\left(\frac{-q_i\phi(\boldsymbol{r})}{k_bT}\right), \tag{2.16}$$

and in the limit of a low potential solution typical of physiological solutions ($-q_i\phi(\boldsymbol{r}) << k_bT$) the hyperbolic sine function can be linearized giving what is called the linearized Poisson-Boltzmann equation [107, 108]:

$$\nabla[\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r})] = -4\pi\rho(\boldsymbol{r}) + \lambda(\boldsymbol{r})l_D^2\phi(\boldsymbol{r}), \tag{2.17}$$

where l_D is defined as:

$$l_D = \left(\frac{8\pi e^2 c}{k_bT}\right)^{\frac{1}{2}} \tag{2.18}$$

where the inverse of l_D^{-1} is the Debye screening length [109] measuring the charge electrostatic effect in a solution and how far its electrostatic effect persists. The linearized Poisson-Boltzmann equation greatly decreases the computational cost of including salt effects compared to the non-linear Poisson-Boltzmann equation, and without introducing significant errors [107, 26] under the above mentioned conditions.

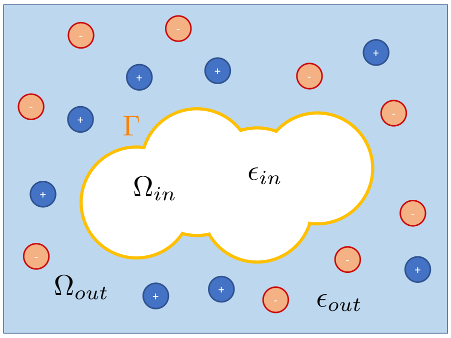


Figure 2.4: Ions distributed within the dielectric media. In blue and red negatively and positively charged species, respectively.

Overall, the Poisson-Boltzmann equation has been widely used in the study of molecules of biological interest due to the possibility of considering conditions similar to that of a physiological solutions, and thus closer to the experiment, achieving accuracies on par with explicit solvents [110, 107]. The Poisson-Boltzmann equation is also of great interest in electrochemical process, as the equation is able to describe the electrochemical potential of ions in the diffuse layer. In fact, the equation is able to describe both the Gouy-Chapman [111], and Stern layer models [112]. The Gouy-Chapman model describes the interaction between a charged surface and an electrolyte solution and the formation of a double layer with diffusive counter-ions, while the Stern Layer model also includes the finite ion size.

2.2.1.3 Atomic Point Charge Approximation

Other than the charge density of the solute, a classical or semi-classical approximation can be used to solve Poisson Equation, in which the charge density of the solute is replaced by atomic point charges [113, 114]:

$$\rho(\boldsymbol{r}) = \sum_{i}^{atoms} q_i \delta(\boldsymbol{r} - \boldsymbol{r}_i). \tag{2.19}$$

The use of point charge models which repartition the charge density of the solute from QM calculations to atomic contributions (see Sec. 3.1.1.1 on page 62) still allow for a SCRF treatment of solvation effects. Furthermore, point charges easily overcome some issues in Apparent Surface Charges (ASC) implicit solvation models [25, 26], which will be treated in Sec. 2.2.6 on page 40. This is one of the most used family of implicit models to day, which considers exclusively polarization of the dielectric media at the solute-solvent interface to reduce the computational cost. The initial assumption of ASC models is not physically valid when considering a charge density obtained from QM calculations, this is due to the asymptotic behavior of the wave function [115] which extends well beyond any reasonably realistic solute cavity, and causing not only polarization at the solute-solvent interface, but also volume polarization within the dielectric media. A problem which can be overcome when atomic point charges are used.

2.2.2 Overview of Numerical Methods

The generalized Poisson equation (Eq. 2.5) can be solved analytically only for simple cavity geometries such as spheres and ellipsoids which in fact were typically employed in earlier solvation models [70, 72, 71]. Nevertheless, Poisson's equation can be successfully solved numerically when considering complex solute cavities which better model the separation between the solute and the dielectric media. Some of the most used methods in implicit solvation are: Finite Element (FEM) [116, 117, 118, 119, 120, 121], Boundary Element (BE) [73, 122, 123, 124, 125], and Finite Difference (FD) [126, 127, 128, 129, 130, 131, 132, 96] methods.

In general, the Finite Element Method (FEM) divides the physical system governed by a partial differential equation, in this case the generalized Poisson equation, into smaller discrete elements. Each element representing a simple approximation of the overall solution of the problem, as shown in Fig 2.5, and requiring integral relations. The global solution throughout space is then obtained by considering the contribution of each discrete element [133]. Instead than discretization on the whole volume as in FEM, similar results can be obtained by the use of BEM, while expensive integration can be overcome by the use of FDM.

2.2.2.1 Boundary Element Method

The idea behind the BEM is that the solution of a linear partial differential equation can be obtained by a linear system of equations whose unknowns are confined on the boundary of the problem. As such BEM has found wide use in implicit solvation models [98, 92, 135, 136, 137, 138] based on the Apparent Surface Charge (ASC) formalism (see Sec. 2.2.6 on page 40). This formalism recasts the volume polarization of the dielectric

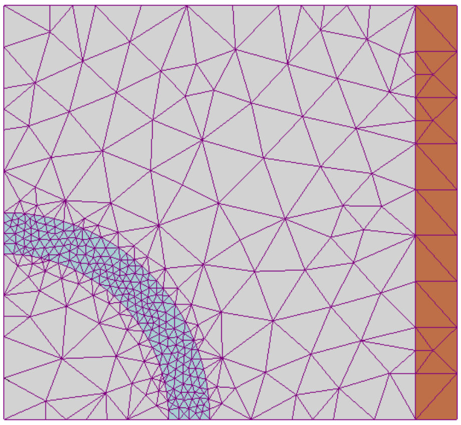


Figure 2.5: Example of a two-dimensional FEM grid around a circular boundary [134].

media to a surface polarization at the solute-solvent interface, making BEM discretization techniques appealing.

As an example of BEM formulation, a two-dimensional non-generalized Poisson equation ($\epsilon(\boldsymbol{r}) \rightarrow \epsilon$) defined on a region of space R will be used [139] for simplicity. The non-generalized Poisson's equations is given by:

$$\nabla^2 \phi(x, y) = -\frac{4\pi}{\epsilon} \rho(x, y), \tag{2.20}$$

where $\rho(\boldsymbol{r})$ is the usual solute charge density, and appropriate Dirichlet and Neumann type boundary conditions defined on a Boundary Γ have to be defined, as in Sec 2.2.1 on page 25. The first step of the BEM formulation is to take the product of $\nabla\phi(x, y)$ with an arbitrary weight function $w(\eta, \xi)$ defined in the same region R, and integrating over the dummy variables η and ξ:

$$\int_\eta \int_\xi w(\eta, \xi) \nabla^2 \phi(x, y)\, d\eta d\xi, \tag{2.21}$$

and by using Gauss' theorem (the variables are omitted to simplify the formulation), one obtains:

$$\int_\eta \int_\xi w \nabla^2 \phi\, d\eta d\xi = \int_\eta \int_\xi \phi \nabla^2 w\, d\eta d\xi + \int_\Gamma \left(w \frac{\partial \phi}{\partial n} + \phi \frac{\partial w}{\partial n} \right) ds \tag{2.22}$$

where the derivatives of ϕ and w are taken in respect to the normal n to the surface Γ. The second integral in the above equation can be reduced to $\phi(x, y)$ if a weight function w can be determined, such that its square gradient is a delta function:

$$\nabla^2 w = \delta(\eta - x, \xi - y), \tag{2.23}$$

with the following property:

$$\int_\eta \int_\xi f(\eta, \xi) \delta(\eta - x, \xi - y)\, d\eta d\xi = f(x, y). \tag{2.24}$$

If this is true, Eq. 2.22 can be rewritten to show the analytical relation between values of $\phi(x, y)$ (within or on the boundary Γ) to a line integral on the boundary Γ:

$$\phi(x, y) = \int_\eta \int_\xi w f\, d\eta d\xi + \int_\Gamma \left(w \frac{\partial \phi}{\partial n} + \phi \frac{\partial w}{\partial n} \right) ds. \tag{2.25}$$

The problem is then discretized by dividing the boundary into elements, and by expressing the unknown function ϕ as a linear combination of a suitable basis function N:

$$\phi = \sum_i \phi_i N_i. \tag{2.26}$$

The problem is then transformed in a matrix-vector equation of the type:

$$\mathbf{Ax} = \mathbf{b} \tag{2.27}$$

where $\mathbf{x}$ is the vector of the unknowns ϕ_i, $\mathbf{b}$ is the known vector for the system physical properties (charge density) of the same size, and $\mathbf{A}$ is a square matrix.

As the electrostatic problem has to be solved only on the boundary, a smaller number of linear equations involving the unknown function are needed [140] compared to other methods. This makes BEM particularly suitable to solve boundary problems like the ones in ASC implicit models, which only consider polarization at the the solute-solvent interface Γ. Nevertheless, this advantage comes at the cost of introducing complicated integral relations, and is lost when treating non-linear partial differential equations, such as the Poisson-Boltzmann equation, as expensive volume integrals have to be reintroduced [141, 140].

2.2.2.2 Finite-Difference Method

The Finite-Difference Method (FDM) is one of the most widespread numerical technique to solve both ordinary and partial differential equations and has a long history in implicit solvation [142, 143, 144], due to the computational efficiency and ease of implementation of the method compared to FEM and BEM [145, 140]. In fact, the FDM is based on the approximation of derivative of a function with the a difference quotient [146]. If one considers a one-dimensional function $f(x)$ as a simple example, the analytical first derivative will be the usual:

$$f'(x) = \lim_{x \to 0} \frac{f(x+h) - f(x)}{h}. \tag{2.28}$$

Finite-Difference considers a fixed size h instead of the limit for x that goes to 0, and by doing so introduces a discretization error which con be quantified through Taylor's theorem [147]:

$$\frac{f(x+h) - f(x)}{h} - f'(x) = O(h). \tag{2.29}$$

This means that the error of approximating the analytical derivative $f'(x)$ with a difference quotient is proportional to h and goes to zero in the limit of $h \to 0$ when using the forward difference of Eq. 2.28. The error for the method can be further reduced if the function f is doubly differentiable. In fact, this allows to consider a difference quotient obtained through a central difference scheme:

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}, \tag{2.30}$$

by doing so the error is reduced to $O(h^2)$, meaning that it is proportional to h^2 and rapidly decreases to zero for $h \to 0$, thus allowing for a larger step size h to be considered for a given accuracy.

In practice, the FDM requires the discretization of the problem on a grid, with a step size h, as shown in Fig. 2.6 in two-dimensional space, and the evaluation of the problem at each grid node characterized by a set of indexes.

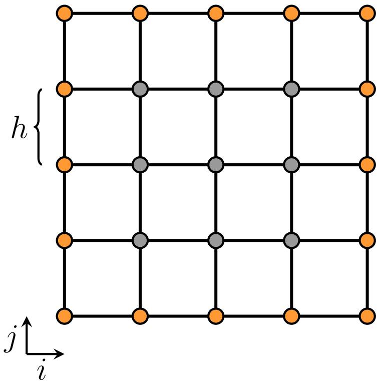


Figure 2.6: 2D example of a finite-difference grid with spacing h. Poisson's equation is solved at the grid nodes. In orange boundary points where to impose boundary conditions.

Using as an example of a FD scheme a generic partial differential equation, such as the non-generalized Poisson equation ($\epsilon(\boldsymbol{r}) \rightarrow \epsilon$) for simplicity:

$$\frac{\partial^2 \phi(x,y)}{\partial x^2} + \frac{\partial^2 \phi(x,y)}{\partial y^2} = -\frac{4\pi}{\epsilon}\rho(x,y) \tag{2.31}$$

the problem at hand is reduced to finding a set of values of ϕ(x,y) through the finite difference approximation of the derivatives at given grid nodes (i,j). Taking the numerical central derivative in regards to x at one such grid nodes, as example, one obtains:

$$\frac{\partial^2 \phi(x,y)}{\partial x^2} \rightarrow \frac{\partial^2 \phi(i,j)}{\partial x^2} \approx \frac{\phi(i-1,j) - 2\phi(i,j) + \phi(i+1,j)}{2h} \tag{2.32}$$

and substituting in Eq. 2.31 together with the derivative of y one is able to obtain a value for $\phi(i,j)$ which is linearly dependent on the adjacent grid nodes. This allows to transform the problem into a matrix-vector equation of the form:

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{2.33}$$

where $\mathbf{x}$ is the vector containing the solutions $\phi(i,j)$ at the grid nodes, and $\mathbf{b}$ is the vector which, in the case of Poisson equation, contains the information of charge densities ρ and boundary conditions on the grid nodes. The most straightforward method to solve the problem for $\mathbf{x}$ is then through matrix inversion:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}. \tag{2.34}$$

Albeit matrix inversion is conceptually straightforward as long as $\mathbf{A}^{-1}$ is obtainable, for larger systems the matrix scales with the power of $\mathbf{x}$ and iterative methods are used is practice such as the Optimal Successive Over Relaxation (O-SOR) [148, 149], or multigrid approaches which lead to increased performances [150, 151].

In general FDM methods have been extensively used in implicit solvation models to solve the generalized Poisson equation, both in classical and QM codes. Some examples of codes which solve the electrostatic problem through a finite difference poisson approach are: UHBD [142], DelPhi [143], MEAD [144]. Furthermore, the non-linear Poisson-Boltzmann can be solved through FDM. In light of the computational efficiency and ease of implementation of FDM the current main issues in implicit solvation is related to forces. This is due to the definition of a the relative permittivity as a step function:

$$\epsilon(\boldsymbol{r}) = \begin{cases} \epsilon_{in} & \text{if } \boldsymbol{r} \in \Omega_{in}, \\ \epsilon_{out} & \text{if } \boldsymbol{r} \in \Omega_{out} \end{cases} \tag{2.35}$$

as the gradient of $\epsilon(\boldsymbol{r})$ is not differentiable on the solute-solvent interface. To solve this issue approximations have to be introduced based on Maxwell stress tensor to calculate forces [152, 153, 154]. Alternatively, a smooth solute cavity can be introduced in which the relative permittivity is a continuous function of the charge density of the solute $\epsilon[\rho]$, as done by Fatteberg and Gygi [99].

2.2.3 Cavities Definition

Up to know we have discussed the solute cavity, acting as an interface between the atomistic solute and dielectric regions, without specifying any shape or size, and instead focusing on its physical significance on the model through Poisson equation. In this section, we will introduce the main solute cavities that have found mainstream use in the literature, focusing in particular on dielectric cavities for which a discontinuity in relative permittivity is present, since this is the cavity type currently implemented in this thesis.

The solute cavities can be roughly subdivided between empirical cavities which stem from empirical Van der Waals radii [155, 156, 157] or based on the molar volume of the solute [72], and non-empirical, which are instead based on an isodensity surface of the electronic density of the solute obtained from a QM calculation [158, 159, 160].

The most straightforward solute cavity type employed by the initial electrostatic models developed by Born [70], Onsager [72], Kirkwood [71] and others is a simple spherical cavity surrounding the solute charge density, and for which the Poisson equation has an analytical solution. The radius for such cavity was suggested [72] to be:

$$R = \left(\frac{3V_m}{4\pi N_a}\right)^{1/3} \tag{2.36}$$

where V_m is the molar volume of the solute and N_a is the Avogadro's constant. Intuitively, this type of spherical solute cavity is better suited for small sized solutes than for complex molecular shapes, especially in case of molecules of biological interest. Nowadays, spherical cavities have been used in cluster-implicit solvation models and as a lightweight replacement for periodic boundary condition in QM/MM calculations [161, 162, 163, 164] where a certain number of explicit solvent molecules surrounding the solute are considered. Noteworthy, are also ellipsoid cavity which better adapt to some molecular shapes (for example bi-atomic molecules), but which overall share the same limitations of spherical cavities [23, 165].

Passing from a spherical, or ellipsoid, cavity to more complex geometries typical of molecules and chemical systems does not allow for analytical solutions of Poisson's equation anymore, and a numerical resolution is needed. The lower accuracy and higher computational time required for numerically solving Poisson equation are nevertheless well justified, as molecular shaped cavities allow for a more physical representation of the target system and a better description of solvation effects and properties [23, 25, 24, 26].

One way to take in account complex molecular geometries is to consider a union of atom centered Van der Waals spheres. These spheres are constructed from empirically obtained radii R_{vdW}, and various sets exist in the literature. The most popular is the one defined by Bondi [156] and validated by a wide number of data from crystal structures. Other commonly used sets are also available from the CRC Handbook of Chemistry and Physics [157], or from the UFF universal forcefield [166].

The van der Waals based solute cavity is also the starting point to define two more surfaces which share the same assumption: reducing the solvent molecule to a sphere with a volume equal to the van der Waals volume of the solvent. These cavities are called: the Solvent-Accessible Surface (SAS) [167, 168], and the Solvent-Excluded Surface (SES) [169, 170].

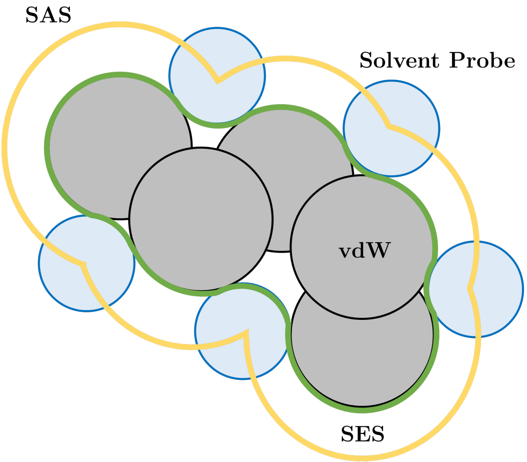


Figure 2.7: 2D example of a SAS in yellow and a SES solute cavity in green, with Solvent Probe in blue, and vdW surfaces and sphere in black and gray, respectively.

- The SAS solute cavity is defined as the locus of points traced by the center of a solvent shaped sphere (solvent probe) rolling on the outer surface of atom centered van der Waals spheres, as shown in Fig. 2.7. Alternatively, it can also be defined as augmented Van der Waals spheres, where each atomic sphere with radius R_{VdW} is expanded by a solvent dependent radius R_p (probe radius), defining a new sphere of radius $R_i = R_{i,VdW} + R_p$, and excluding any point that fall within another sphere domain. Physically, this solute cavity corrects the Van der Waals cavity for the solvent exclusion effect [171, 172], for which the region surrounding the solute should be void of any solvent molecules.
- The SES solute cavity is instead build to avoid high dielectric regions within the solute itself (highlighted in Fig. 2.7), an event less relevant in SAS cavities [173]. The SES uses the same solvent probe as the SAS, but instead of considering the center of the solvent probe to define the surface it uses the probe surface itself. The generated surface can be subdivided in three different patches [174, 175]: i) contact patches are part of the atomic surface that are directly in contact with the probe, ii) toric patches shown in Fig. 2.8a are the collection of curves connecting two contact surfaces when the probe rolls over two contacted spheres, and iii) reentrant patches shown in Fig. 2.8b instead are formed when the probe is in contact with three or more spheres.

Clearly, the use of a solvent probe radius to define the dielectric cavity is an arbitrary choice that only partially resembles the complexity of the physical situation. For example a commonly used solvent probe with radius of $R_p = 1.4$ Å is used for water [176]. This value corresponds to half the distance of the first peak in the radial distribution function between oxygens, but other values have also been used in literature [177], and a common practice

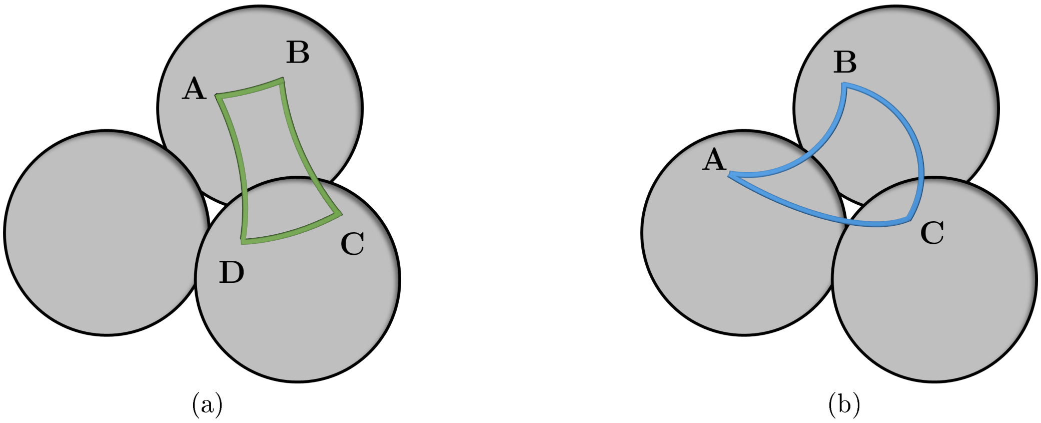


Figure 2.8: Examples of, (a) toric surface patch (ABCD) connecting two spheres, and (b) reentrant patch (ABC) connecting three spheres, for a SES type surface.

in implicit solvation models is to optimize the solvent probe based on the solvation effect or property one is interested in [178]. For example, this is done in the widely used SMx family of solvation models [82, 83, 81, 80], in which solute cavities have been optimized for the prediction for solvation energies of charged molecules and then equally employed for neutral molecules. In general, problems also arise in defining an effective solvent probe for larger solvent molecules, as in the case of complex or long-chain organic molecules, for which the use of a simple spherical probe radius might not be always physically sound.

A more physical definition of a solute cavity can be reached by using an isodensity surface based on the electronic density [158, 159, 160], thus obtaining a realistic representation of the solute-solvent interface. Typical values for the isodensity surface usually vary in a range between 0.0004-0.0010 a.u. [26, 25]. Albeit the physical soundness of such solute cavity, technical aspects limit its applicability, as failure in surface generation are possible. Another issue is related to the definition of the normal surface vector $\mathbf{n}$ being dependent on the gradient of the electronic density [158, 159, 160]. Since this point introduces considerable hurdles in the implementation of electrostatic forces for solute in solvent geometry optimization, van der Waals solute cavities remain widely used, as analytical gradients of the surface are more straightforward to obtain.

In addition, cavities have also been defined based on a smooth permittivity function instead of the discontinuous $\epsilon(\boldsymbol{r})$ as defined in Eq. 3.3. One such cavities, originally developed by Fattebert and Gygi [99, 179, 180], uses a charge density $\rho(\boldsymbol{r})$ dependent relative permittiviy ϵ defined here generically as:

$$\epsilon[\rho] = \begin{cases} \epsilon_{in} & \rho(\boldsymbol{r}) > \rho_{max} \\ f[\rho(\boldsymbol{r})] & \rho_{min} < \rho(\boldsymbol{r}) < \rho_{max} \\ \epsilon_{out} & \rho(\boldsymbol{r}) < \rho_{min} \end{cases} \tag{2.37}$$

where ρ_{min} and ρ_{max} are charge density parameters used in the definition of ϵ which varies between ϵ_{in} and the solvent ϵ_{out} relative permittivity, while the function $f[\rho(\boldsymbol{r})]$ allows for a smooth transition between ϵ_{in} and ϵ_{out}. The main advantages of smooth solute cavities are related to faster convergence of the numerical solution of Poisson equation with a FD scheme and the lack of discontinuities in $\epsilon(\boldsymbol{r})$ allowing for straightforward force calculations [152, 153, 154, 99].

Overall, no unique correct definition of a solute cavity exista, and many solute cavities and variations of the above models have been developed in the literature which try to balance between technical feasibility and realistic molecular shaped cavities [181, 182, 95, 183]. The latters are generally considered more physically sound and also more accurate to compute solvation properties in regards to experimental data [23, 24, 25, 26].

2.2.4 Born and Generalized Born Model

One of the first electrostatic models for solvation is the one developed by Born for charged species [70]. In this model a spherical solute cavity surrounding an ion is considered and the solute charge density $\rho(\boldsymbol{r})$ within it is approximated as a single point charge Q.

Under these assumptions, Poisson equation can be solved analytically and the electrosatic energy is given by:

$$\Delta G_{el} = -\left(\frac{\epsilon - 1}{\epsilon}\right)\frac{Q^2}{2\alpha} \tag{2.38}$$

where ϵ is the solvent relative permittivity outside the solute cavity, and α is the radius of the spherical cavity, also called effective Born radius.

An extension from a single charged sphere to multiple spheres, each with a charge at its center has been developed and is the Generalized Born Approximation (GBA) [86, 87], which allows for molecular shaped cavities to be considered.

Many formulations of the GBA model have been developed [184, 81, 82, 83, 86, 185, 186], both for a classical and quantum-mechanical description of solvation effects. Here we report the expression given by Still et al. [187]:

$$\Delta G_{el} = -\left(\frac{\epsilon - 1}{\epsilon}\right)\sum_{i=1}^{atoms}\sum_{j>i}^{atoms}\frac{q_i q_j}{2 f_{ij}} \tag{2.39}$$

where the index i and j run over the atom centered point charge q in the solute molecule and f_{ij} is called Coulomb operator and is not uniquely defined. A widely used expression also found in the SMx [82, 83, 81] family of solvation models is:

$$f_{ij} = \left(R_{ij}^2 + \alpha_i\alpha_j \cdot \exp\left(\frac{-R_{ij}^2}{d \cdot \alpha_i\alpha_j}\right)\right)^{1/2} \tag{2.40}$$

where R_{ij} is the internuclear distance, and α_i and α_j are the effective Born radius, of atoms i and j, while d is a constant. The exponential term in Eq. 2.40 is a damping factor, which for large distances between atoms ensures that f_{ij} goes to α, leading back to Eq. 2.38 for a mono-atomic system, albeit the effective born radius of an isolated atom should differ from that of the atom within a molecule due to the different chemical surrounding [64].

A problem for GBA is the definition of the effective Born radii α for which many formulations have been proposed [187, 188, 25, 189, 190]. The exact formulations re-introduce the need of solving Poisson's equation for each solute atom, but offers little advantages compared to approximate methods [25] based on optimizable parameters. An example of an approximate method is the one also used in the most recent SMx solvation models (x=6,8,12), such as the SM12 [81] variant.

In the SM12 model the intrinsic Born radius α of an atom i is defined as:

$$\alpha_i = \left(\frac{1}{R_i} + \int_{R_{Z_i}^C}^{R_i}\frac{A_i(r)}{4\pi r^4}dr\right)^{-1} \tag{2.41}$$

where R_i is the radius of an atom centered sphere big enough to include the entire molecular solute cavity, $R^C_{Z_i}$ is an atomic number z dependent parameter to optimize, called the intrinsic Coulomb radius, and A_i is the Solvent-Accessible Surface Area, in brief SASA [167, 168].

Another issue to take into account when developing and using GBA electrostatic models are atomic charges q [64]. In fact, no best and universal procedure exists to obtain them and they have to be carefully chosen to obtain accurate solvation properties [82, 83, 81, 191]. Furthermore, in a quantum-mechanical calculation these charges will be free to evolve together with the polarized electronic wave function of the solute at every step in a self-consistent manner.

The topic of atomic charge models will be treated in more details in Sec. 3.1.1.1 on page 62. Here we will limit ourselves to point out that the choice of the model is central in obtaining accurate solvation energies and properties, and list some of the charge models that have been used within the GBA electrostatic model in quantum-mechanical calculations over the years. Mulliken [192], and Löwdin [193] charge models have been the first to be considered, while the CMx family [194, 195, 196] of models have been used more recently. Although all these models are based on the partition of the electron density between the solute atoms, other strategies such as ChElPG [197], and MK [198], where charges are fitted to quantum-mechanical computed electrostatic potentials, have been used.

Overall, for small to average sized organic molecules the solvation models based on GBA electrostatics have shown similar results to models based fully on Poisson equation when using the same solute cavity definition [86]. Inconsistencies between the two models increasingly occur for larger molecules, such as biological polymers [199]. Larger molecules also introduce some technical issues which arise in the computational protocol of GBA models, which have shown to occasionally assign the solvent relative permittivity to volumes within the solute, thus introducing an unphysical description of the system [200]. This is a key problem when considering solvation of large periodic systems such as surfaces in which the solvent should not penetrate the surface.

2.2.5 Multipolar Expansion Methods

The use of spherical or ellipsoid solute cavities does not limit the representation of the charge distribution $\rho(\boldsymbol{r})$ to atomic charges q in order to obtain analytical solutions of Poisson's equation. In fact, any charge distribution $\rho(\boldsymbol{r})$ can be expressed as a single-center multipole expansion [201, 202, 72, 71, 203, 204], a mathematical series representing a function that depends on angles expressed as a sum of terms with progressively finer angular features called moments [205].

A general formulation for the electrostatic energy using multipole expansion is the one derived by Kirkwood [71] for a spherical cavity, according to:

$$\Delta G_{el} = -\frac{1}{2} \sum_{l=0}^{+\infty} \sum_{m=-l}^{l} \frac{(1+l)(\epsilon-1)}{(1+l)\epsilon+1} \frac{{M_{ml}}^2}{R^{2l+1}} \tag{2.42}$$

where M_{ml} is the m^{th} component of the multipolar moment of order l describing the solute charge distribution calculated at the center of the spherical cavity of radius R, and ϵ is the solvent relative permittivy. Formulations of multipolar expansions have also been developed for ellipsoid cavities [71, 206, 207, 208, 23].

Eq. 2.42 is also referred to generalized Kirkwood model and many solvation models can be seen as a special case of it [26]. For example, using a multipolar moment of order $l = 0$ and atomic charges in place of the solute charge density corresponds to Born's model [70], which was already mentioned in Sec. 2.2.4 on page 37. For higher order multipoles $l = 1$ and using a point dipole μ instead of an atomic charge, one instead obtains Bell's solvation model [209]:

$$\Delta G_{el} = -\frac{(\epsilon - 1)}{(2\epsilon + 1)} \frac{\mu^2}{R^3}. \tag{2.43}$$

In the quantum-mechanical formulation of MPE methods the charge density $\rho(\boldsymbol{r})$ can be expressed as a multipole expansion, and the solute gas phase Hamiltonian can be augmented by the corresponding interaction potential V_{int} [64, 25, 23], allowing for a self-consistent treatment of solvation effects.

The Onsager model was one of the first self-consistent MPE approach [72], and used a spherical cavity. In this model the solute is approximated by a dipole moment μ, and polarizability α obtained self-consistently from a quantum-mechanical calculation, with the electrostatic solvation energy given by:

$$\Delta G_{el} = -\frac{(\epsilon - 1)}{(2\epsilon + 1)} \frac{\mu^2}{R^3} \left[1 - \frac{(\epsilon - 1)}{(2\epsilon + 1)} \frac{2\alpha}{R^3} \right] \tag{2.44}$$

where R is the radius of the spherical cavity, and ϵ the solvent relative permittivity.

The Onsager model has been widely used in the past, but has currently been surpassed by more advanced models. This is due to its non-molecular shaped cavity, clear limitations for solutes with no permanent dipole, and its generation of unphysical solvated charge densities [210, 25].

Mikkelsen's model [211] has tried to improve over Onsager dipole approximation by using quantum-mechanically obtained charges within the spherical solute cavity, but single-center multipole expansions of the charge distributions $\rho(\boldsymbol{r})$ have, overall, shown to have limitations. These limitations are mostly related to problems in convergence and asymptotic behavior of the series, which brought criticisms on the unreliability of MPE codes, especially in case of larger molecules [212, 213, 25, 23], together with the lack of physical molecular cavities.

One of the most complete MPE models is the one developed by Rivail et al [214, 215], which solves the convergence and asymptotic behavior issues of the single-center MPE by using a multiple atom-centered one, also referred to as distributed multipole analysis (DMA). DMA was developed by Stone [88], and earlier proposed by Rein [216]. An example of a strategy from Sokalski and Poirier [217] for QM calculations is where the m^{th} component of the multipole of order l on the center i is defined as:

$$M_{ml}(i) = \sum_{\mu \in i} \sum_{\nu} P_{\mu\nu} \langle \mu | M_{ml} | \nu \rangle \tag{2.45}$$

where $P_{\mu\nu}$ is the population matrix and μ and ν are atomic orbitals. The model by Rivail et al. [215] also replaces the unphysical spherical and ellipsoid cavities of previous MPE models by a molecular shaped solute cavity, thus allowing for a more realistic treatment of solvation effects. Furthermore, as all other MPE methods seen so far, Rivail's model is able to compute fast and reliable forces through the analytical gradients of its reaction-field potential [215].

Result wise, the multicentered MPE method developed by Rivail et al. gives very similar values in both interaction energies and dipole moments to other solvation models [215]. This analytical reaction field potential comes however at the price of a formidable mathematical complexity when considering molecular shaped solute cavities [23]. This complicates significantly its extension to periodic systems, for which other solvation models are more suitable.

2.2.6 Apparent Surface Charge Formalism

Apparent surface charge (ASC) models recast the volume polarization of the dielectric media due to the solute charge density to a surface polarization at the solute-solvent interface, thus simplifying the mathematical formalism and increasing the computational efficiency of the problem at hand [25, 218, 219, 220, 221]. Furthermore, it offers an exact solution of the electrostatic problem when the solute charge density is completely within the solute cavity [25, 222]. For these reasons ASC models have become one of the most widespread class of implicit solvation models nowadays.

Common to all ASC methods is the definition of a reaction-field potential $\phi_{RF}(\boldsymbol{r})$ for the whole volume domain $\boldsymbol{r} \in V$ whose computation requires the charge density $\sigma(\boldsymbol{s})$ on the solute-solvent surface $\boldsymbol{s} \in \Gamma$ [25]:

$$\phi_{RF}(\boldsymbol{r}) = \int_{\boldsymbol{s}\in\Gamma} \frac{\sigma(\boldsymbol{s})}{|\boldsymbol{s}-\boldsymbol{r}|} d\boldsymbol{s}. \tag{2.46}$$

This reaction-field potential $\phi_{RF}(\boldsymbol{r})$ arises from the dielectric jump conditions imposed on the solute-solvent interface Γ to ensure the continuousness of the dielectric displacement throughout the interface and where first seen in Sec. 2.2.1 on page 25 and are here reported again [92, 93]:

$$\begin{cases} \phi_{in}(\boldsymbol{r}) = \phi_{out}(\boldsymbol{r}) \\ \epsilon_{in}\mathbf{n} \cdot \nabla\phi(\boldsymbol{r})\Big|_{in} = \epsilon_{out}\mathbf{n} \cdot \nabla\phi(\boldsymbol{r})\Big|_{out} \end{cases} \tag{2.47}$$

In fact, to satisfy the dielectric jump conditions in the case of a discontinuous surface ($\epsilon_{in} \neq \epsilon_{out}$), polarization proportional to the normal $\mathbf{n}$ electric field at the dielectric interface is required [223, 23]. This polarization is manifested from the appearance of a charge density $\sigma(\boldsymbol{s})$ on the dielectric surface Γ which acts as source term for the potential $\phi_{RF}(\boldsymbol{r})$ in Eq. 2.46.

Albeit the reduction of the source term of the reaction-field potential to a charge distribution $\sigma(\boldsymbol{s})$ limited to the solute-solvent interface Γ, the integration in Eq. 2.46 is still computationally demanding for molecular shaped solute cavities [25]. Thus, it is common to discretize the integral in a finite number of elements N each with an area A_i sampled on the interface Γ:

$$\phi_{RF}(\boldsymbol{r}) \approx \sum_{i}^{N} \frac{\sigma(\boldsymbol{s}_i)A_i}{|\boldsymbol{s}_i - \boldsymbol{r}|} = \sum_{i}^{N} \frac{q(\boldsymbol{s}_i)}{|\boldsymbol{s}_i - \boldsymbol{r}|}, \tag{2.48}$$

where $q(\boldsymbol{s}_i) = \sigma(\boldsymbol{s}_i)A_i$ are the apparent surface charges. The discretization of the surface is usually performed with small flat triangular surface elements called tesserae, one of the most used algorithm for surface discretization being GePol [181], widely used in most PCM variants. This is combined with numerical methods to solve Poisson Equation for the potential and obtain the surface charge densities $\sigma(\boldsymbol{s}_i)$ of a dielectric [98, 122, 116, 127] in order to solve the electrostatic problem.

The inclusion of the charges q into the solute-solvent interaction potential V_{int} within the Hamiltonian allows for a self-consistent reaction field treatment of polarization effects [79], where the local value of the potential from solving Poisson equation at each iteration is used to compute the apparent surface charges, which on the following iteration are used to compute a new value of the potential, and so on until convergence of the electrostatic solvation energy as seen in Sec. 2.2.1.1 on page 27.

The electrostatic solvation ΔG_{el} energy for ASC methods is obtained as in Eq. 2.10 in Sec. 2.2.1 on page 25, and is here reported again:

$$\Delta G_{el} = \langle \psi_s | H_g + V_{int} | \psi_s \rangle - \Big(\langle \psi_g | H_g | \psi_g \rangle + E_{\mathrm{RF}} \Big) \tag{2.49}$$

where ψ_s is the solute solvated wave function, and E_{RF} is the energy associated to the reaction-field potential $\phi_{RF}(\boldsymbol{r})$, which in its discretized form is given as the following Coulomb interaction between the solute potential ϕ_ρ and the i^{th} charge q at its position $\boldsymbol{r}_i$ [64, 96, 25]:

$$E_{RF} = \frac{1}{2} \sum_{i}^{N} q(\boldsymbol{s}_i) \phi_\rho(\boldsymbol{s}_i) = \frac{1}{2} \langle \psi_s | V_{int} | \psi_s \rangle \tag{2.50}$$

and, as mentioned in Sec. 2.2.1 on page 25, is equal to the non-reversible work necessary to polarize a linear isotropic dielectric [25].

We will now briefly cover the first ASC method: the D-PCM developed by Tomasi et al. [73], where D stands for Dielectrics, and its main issue in considering a QM obtained solute charge density while ignoring volume polarization. We will then treat the C-PCM [224, 225] and COSMO [226, 227, 228, 229] formalisms together as they both solve Poisson equation by considering the solvent media as a conductor. Finally, we will treat the latest iteration of the PCM family which is based on the Integral Equation Formalism (IEF-PCM) [98] to solve the electrostatic problem using BEM will be presented. This last model overcomes the problem of neglecting the volume polarization within the QM- ASC formalism, and can be seen as a generalization of both D-PCM and C-PCM.

2.2.6.1 D-PCM

In the original formulation of the D-PCM [73] electrostatic model the charge density $\sigma(\boldsymbol{s})$ in Eq. 2.46 is given by two equal expressions which arise from the dielectric jump conditions:

$$\sigma(\boldsymbol{r}) = \frac{1}{4\pi} \left(\frac{\epsilon_{out} - \epsilon_{in}}{\epsilon_{in}} \right) \mathbf{n} \cdot \nabla \phi(\boldsymbol{r}) \Big|_{in} = \frac{1}{4\pi} \left(\frac{\epsilon_{out} - \epsilon_{in}}{\epsilon_{in}} \right) \mathbf{n} \cdot \nabla \phi(\boldsymbol{r}) \Big|_{out} \tag{2.51}$$

where the potential is defined as sum of the reaction-field and solute potential ($\phi(\boldsymbol{r}) = \phi_\rho(\boldsymbol{r}) + \phi_{RF}(\boldsymbol{r})$) and its derivatives are taken immediately inside and outside the solute-solvent interface Γ. Albeit the outer expression is usually preferred as it avoids self-polarization effects of the dielectric media.

From a physical point of view, the main problem of D-PCM and earlier ASC models is what is usually referred to as the escape charge problem [92, 135, 137, 158]. As mentioned previously, the ASC formalism allows for an exact formulation of the electrostatic problem as long the solute charge density is fully within the solute cavity [136]. This is the case when considering a classical system represented by atomic point charges. Instead in the quantum-mechanical formulation a portion of the solute charge density escapes any reasonable solute cavity definition due to the asymptotic behavior of the wave function, which causes not only surface, but also volume polarization in the dielectric media which is neglected in Eq. 2.46.

The issue of an escaped charge in ASC models was first addressed by Chipmann et al. [136, 158, 230] which corrects the reaction-field potential $\phi_{RF}(\boldsymbol{r})$ by considering it sum of two contributions which arise from surface polarization $\phi_{\Gamma}(\boldsymbol{r})$ and volume polarization $\phi_{V}(\boldsymbol{r})$:

$$\phi_{RF}(\boldsymbol{r}) = \phi_{\Gamma}(\boldsymbol{r}) + \phi_{V}(\boldsymbol{r}). \tag{2.52}$$

This expression forms the base assumption of the Surface and Volume Polarization for Electrostatics (SVPE) model [137]. Other than the SVPE model the problem of escaped charge has been solved in a more recent PCM model based on a Integral Equation Formalism (IEF-PCM) [98]to solve the electrostatic problem, and which is completely equivalent to an approximation of SVPE known as SS(V)PE [158, 137, 230], which stands for Surface and Simulation of Volume Polarization for Electrostatics. Alternatively, the escaped charge problem can be addressed through a renormalization procedure of $\sigma(\boldsymbol{r})$ at each SCRF cycle. This using the relation between theoretical sum of apparent surface charges Q_{asc}^{theory} and the computed sum Q_{asc}, according to [231]:

$$Q_{asc}^{theory} = -\frac{\epsilon_{out} - 1}{\epsilon_{out}} \cdot Q_{ASC} \tag{2.53}$$

The difference between the theoretical and computed sums of ASCs before renormalization, can be used to verify the quality of the ASC formalism.

2.2.6.2 C-PCM and COSMO

The Conductor-like Screening Model (COSMO) [226, 227, 228, 229] and C-PCM (C for conductor) [224, 225] are ASC method which are in sharp contrast to other implicit solvation models as they replace the dielectric media with a conductor. The advantage of this approach is that conductors are characterized by an infinite relative permittivity $\epsilon = +\infty$, and thus no potential exist within the media simplifying the electrostatic problem due to the lack of boundary conditions. A surface charge density $\sigma^{\infty}(\boldsymbol{s})$ can then be computed directly from the electrostatic potential generated by the solute charge density, and is later scaled down by a function of the real relative permittivity of the solvent $f(\epsilon)$ to obtain the proper charge density $\sigma(\boldsymbol{s})$:

$$\sigma(\boldsymbol{s}) = f(\epsilon)\sigma^{\infty}(\boldsymbol{s}). \tag{2.54}$$

The scaling function $f(\epsilon)$ is empirically determined as:

$$f(\epsilon) = \frac{\epsilon - 1}{\epsilon + k} \tag{2.55}$$

with $k = 0.5$ in the original pubblication of COSMO [226, 227, 228], although different values have been given in the literature [232, 25] .

Result wise, the different assumption of treating the dielectric media as a conductor to obtain the apparent surface charges seem to have little impact on the results at least compared to the other ASC methods [232, 233] as the electrostatic energy is not a physical observable [64]. The main advantage of the C-PCM and COSMO models is their computational robustness as well as the lack of potential numerical instabilities of other ASC formalism, along with greater ease of implementation due to the simplified electrostatic formalism [26].

2.2.6.3 IEF-PCM

The IEF-PCM [98] was the first model based on integral operators in quantum-chemistry, and together with the Solvation Model based on solute electron Desnity (SMD) developed by Marenich et al. [80] (with whom it shares the same electrostatic formalism) is one the most used implicit solvation models to date.

As all ASC model the starting point of IEF-PCM is the definiton of the reaction-field potential as in Eq. 2.46:

$$\phi_{RF}(\boldsymbol{r}) = \int_{\boldsymbol{s}\in\Gamma} \frac{\sigma(\boldsymbol{s})}{|\boldsymbol{s}-\boldsymbol{r}|} d\boldsymbol{s} \tag{2.56}$$

together with the decomposition of the potential as sum of the solute potential $\phi_\rho(\boldsymbol{r})$ and the reaction-field potential $\phi_{RF}(\boldsymbol{r})$ (see Sec. 2.2.1 on page 25):

$$\phi(\boldsymbol{r}) = \phi_\rho(\boldsymbol{r}) + \phi_{RF}(\boldsymbol{r}) \tag{2.57}$$

which are then redefined in terms of Green functions [234] $G(\boldsymbol{r},\boldsymbol{r'})$ which represent the potential a point charge at position $\boldsymbol{r'}$ generates at $\boldsymbol{r}$. The relation between the potential in Eq. 2.57 and the Green functions is then the following:

$$\begin{cases} \phi(\boldsymbol{r}) = \int_V G_\rho(\boldsymbol{r},\boldsymbol{r'})\rho(\boldsymbol{r'})d\boldsymbol{r'} \\ \phi_\rho(\boldsymbol{r}) = \int_V G(\boldsymbol{r},\boldsymbol{r'})\rho(\boldsymbol{r'})d\boldsymbol{r'} \\ \phi_{RF}(\boldsymbol{r}) = \int_V G_{RF}(\boldsymbol{r},\boldsymbol{r'})\rho(\boldsymbol{r'})d\boldsymbol{r'} \end{cases} \tag{2.58}$$

with $\rho(\boldsymbol{r'})$ the solute charge density, $G(\boldsymbol{r},\boldsymbol{r'}) = 1/|\boldsymbol{r}-\boldsymbol{r'}|$ the green function associated to the operator $-\nabla^2$, $G_\rho(\boldsymbol{x},\boldsymbol{y})$ the function associated to the operator $-\epsilon\nabla^2$, together with $G_{RF}(\boldsymbol{r},\boldsymbol{r'}) = G_\rho(\boldsymbol{r},\boldsymbol{r'}) + G(\boldsymbol{r},\boldsymbol{r'})$ definitions, and the usual boundary conditions.

Under these conditions the surface charge density $\sigma(\boldsymbol{s})$ of Eq. 2.46 can be recast as an integral equation [235] on the interface Γ and the surface charge density $\sigma(\boldsymbol{s})$ is given by:

$$\left(2\pi\left(\frac{\epsilon+1}{\epsilon-1}\right) - \hat{D}\right)\hat{S}\sigma(\boldsymbol{s}) = (-2\pi + \hat{D})\phi_\rho(\boldsymbol{s}) \tag{2.59}$$

where $\hat{S}$ is the integral operator which acts on the surface charge density $\sigma(\boldsymbol{s})$ to give the single-layer potential [236], and related apparent surface charges:

$$\hat{S}\sigma(\boldsymbol{s}) = \int_{\boldsymbol{s'}\in\Gamma} \frac{\sigma(\boldsymbol{s'})}{|\boldsymbol{s'}-\boldsymbol{s}|} d\boldsymbol{s'} \tag{2.60}$$

while $\hat{D}$ is the integral operator which acts on $\sigma(\boldsymbol{s})$ to give the double layer potential [236]:

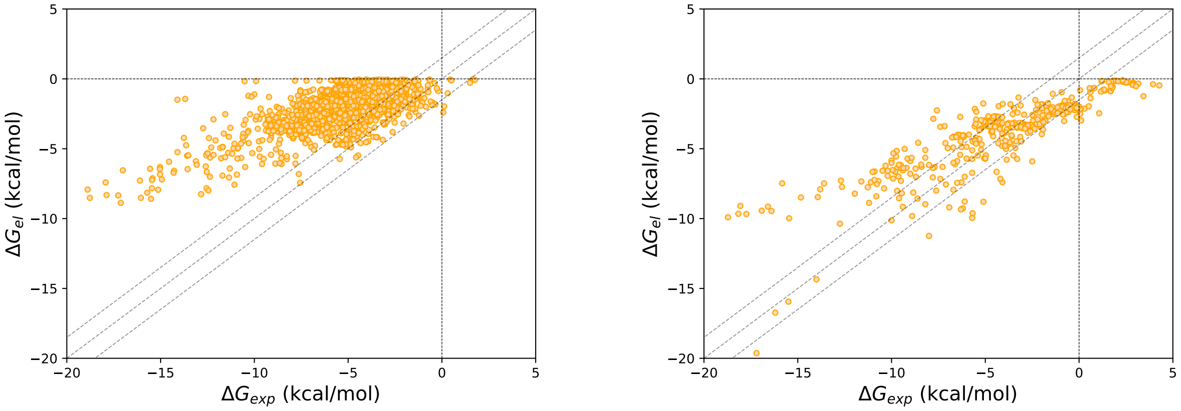
$$\hat{D}\sigma(\boldsymbol{s}) = \int_{\boldsymbol{s'}\in\Gamma} \frac{\partial}{\partial \boldsymbol{n}_{\boldsymbol{s'}}}\left(|\boldsymbol{s'}-\boldsymbol{s}|^{-1}\right)\sigma(\boldsymbol{s'})d\boldsymbol{s'} \tag{2.61}$$

The use of Eq. 2.59 allows to have an exact formulation of the electrostatic problem as the equation also implicitly contains the effects of volume polarization in an approximate way similarly to the SS(V)PE model [158, 137, 230], thus solving the problem of escaped charge typical of earlier ASC models. Another advantage of IEF-PCM over D-PCM is that surface charge density $\sigma(\boldsymbol{s})$ in IEF-PCM is not dependent to the gradient of the potential. In fact, the Poisson equation is numerically solved using a Boundary Element Method (BEM) and the numerical derivative introduce instabilities in the computation of $\sigma(\boldsymbol{s})$ and of forces [26]. The full mathematical treatment can be found in the original publication [98].

2.3 Non-Electrostatic Contribution

Historically, solvation models mostly focused on the electrostatic contribution to the total solvation energy. Such as in the case of the earlier models developed by Born [70], Kirkwood [71], and Onsager [72] which completely ignored the non-electrostatic contribution, with only few attempts being made by Claverie [237] and Rivail [214] focusing in particular on the dispersion effects. The reason for this can be attributed to the fact that most models focused on water due to its importance as a solvent in biology [25]. Water is in fact characterized by a high dipole moment and consequentially a high relative permittivity [157], meaning that the electrostatic term should be the dominant contribution to the total solvation energy [191], which justified the earlier lack of development of non-electrostatic models. This approximation is however rather coarse as a balance exists between the electrostatic and non-electrostatic contributions and solvation models which have included the latter have achieved more accurate results in the prediction of solvation energies and properties [82, 83, 81, 238, 239, 37, 240].

Fig. 2.9 qualitatively shows the performance of the widely used IEF-PCM electrostatic model in predicting solvation energies for both non-aqueous and aqueous solvents for a varied set of solutes widely used both for parametrize and validate solvation models [36, 241] (more in Sec. 6.1.4.1 on page 111).



(a) Non-Aqueous Solvents (b) Aqueous Solvent

Figure 2.9: Calculated electrostatic solvation energies ΔG_{el} as a function of total experimental solvation energies ΔG_{exp} for (a) non-aqueous, and (b) aqueous solvents. The ΔG_{el} values are calculated with IEF-PCM (B3LYP/6-311G**) on 2530 neutral solutes of the MNSOL database [36, 241] in 91 solvents with ϵ_{out} ranging from ≈ 1.00 to 181.56.

For non-aqueous solvents the electrostatic model alone is here unable to predict accurate solvation energies within the experimental uncertainty for neutral molecules, of about 1.00 kcal/mol. On the other hand, IEF-PCM [98, 233] performance in aqueous solvent achieves better results due to the solvent high relative permittivity, which makes the electrostatic term ΔG_{el} the main contribution to the solvation energy, as expected, although this approximation does not hold for larger ΔG_{el}. In fact, on average the error compared to the experimental data is still higher than 1.0 kcal/mol, making it clear that the non-electrostatic contribution cannot be neglected in either cases.

In light of this, more effort has been given in developing non-electrostatic models compared to the pioneering years of implicit solvation. This has been done both by focusing on more physically accurate models for each of the main contributions to the ΔG_{ne} [23, 25, 24, 26], which are:

$$\Delta G_{ne} = \Delta G_{rep} + \Delta G_{dis} + \Delta G_{cav} + ... \tag{2.62}$$

the cavitation ΔG_{cav}, dispersion ΔG_{dis}, and repulsion ΔG_{rep}, and also by considering the ΔG_{ne} related to one or more properties. These latter models usually consider the non-electrostatic contribution to be proportional to the exposed area of the solute, corresponing to the Solvent-Accessibile Surface Area (SASA) [167, 168], previously mentioned in Sec. 2.2.3 on page 34.

In the following, we will first discuss the more physically sound models which treat each contribution independently; starting with cavitation, repulsion, and dispersion, and then we will focus on the models which mostly relate the non-electrostatic energy to the SASA.

2.3.1 Cavitation

The cavitation energy ΔG_{cav} is related to the reversible work needed to form a cavity of suitable size within the solvent. This term is a net positive contribution to the total solvation energy, as work has to be done to break the stabilizing intermolecular interactions between the solvent molecules, and move them apart to form the cavity.

Many methods have been developed to obtain estimates of the cavitation energy [242, 243, 167, 244, 245], but in practice two methods have been actively used. The first are based on the solute-solvent interface surface area [167, 168] (see Sec 2.3.4 on page 47) or cavity volume, while the second are based on statistical mechanics where the solvent, and more in general a fluid, is represented by hard spheres. Albeit, the overall statistical mechanics approach has shown to have low accuracy in estimating properties of real fluids, it has long been used, and is still used, in computational chemistry to compute the cavitation energy [25].

Nevertheless, the cavitation energy for implicit solvation models is based on the Scaled Particle Theory (SPT) [244, 246]. According to the SPT implementation in PCM, the energy for a spherical cavity is given by the following simplified formula which neglects pressure contributions:

$$\Delta G_{cav} = RT\left[-ln(1-y) + \frac{3y}{1-y}\left(\frac{R_s}{R_p}\right) + \left[\frac{3y}{1-y} + \frac{9}{2}\left(\frac{y}{1-y}\right)^2\right]\left(\frac{R_s}{R_s}\right)^2\right] \tag{2.63}$$

where R_s is the radius of the atom approximated as a hard sphere, and R_p is the probe radius (see Sec. 2.2.3 on page 34), R is the gas constant, T the temperature, and y is a function of both the atomic radius R_s and of the number density of the solvent ρ:

$$y = \frac{4\pi}{3}\rho R_p^3. \tag{2.64}$$

As such the model is completely defined through two parameters: R_s, R_p, and one empirical descriptor for the solvent ρ. The extension of SPT to molecular cavities defined as sum of interlocked spheres of Radius $R_i = R_{s,i} + R_p$ is given by the Pierotti-Claverie formula in the C-SPT [245, 247] model:

$$\Delta G_{cav} = \sum_i^N \frac{A_i}{4\pi R_i^2}\Delta G_{cav}(R_i), \tag{2.65}$$

in the which the total cavitation energy ΔG_{cav} is the sum of atomic $\Delta G_{cav}(R_i)$ contributions weighted by the ratio between the exposed surface area of the atom A_i and the squared atomic radius R_i.

We note that the SPT model in computational chemistry has been mostly tested on small solvents such as water with fewer practical uses with larger solvents typical of organic chemistry [25].

2.3.2 Dispersion

To treat the dispersion contribution many QM approaches have been developed over the years [237, 248, 249, 250, 251, 252, 253, 254, 255]. These are mostly based on analytical expressions to be included within the Hamiltonian and which depend explicitly on the solute charge density of the solute. As an example, the dispersion energy ΔG_{dis} contribution of one such methods implemented in PCM and based on dynamic polarizability [256, 257] is:

$$\Delta G_{dis} = -\frac{\beta}{2}\sum_{ijkl}[ij|kl]P_{il}S^{-1}_{jk} + \frac{\beta}{4}\sum_{ijkl}[ij|kl]P_{il}P_{jk} \tag{2.66}$$

with the subscripts $ijkl$ refering to elements of the basis set. The integral $[ij|kl]$ is instead given by the following expression:

$$[ij|kl] = \frac{1}{2}\int_{\Gamma}\left(\phi_{ij}(\boldsymbol{s})E_{kl}(\boldsymbol{s}) + \phi_{kl}(\boldsymbol{s})E_{ij}(\boldsymbol{s})d\boldsymbol{s}\right) \tag{2.67}$$

where $\phi(\boldsymbol{s})$ and $E_{ij}(\boldsymbol{s})$ are respectively the potential and the normal component of the electrical field on the cavity surface at point $\boldsymbol{s}$ when expressed on the chosen basis set elements. P_{ij} and S_{ij} are instead matrix elements of the density (P) and overlap (S) matrix, respectively. The expression for β is instead:

$$\beta = \frac{n_s^2 - 1}{4\pi n_s\left(n_s + \frac{\omega}{I_s}\right)} \tag{2.68}$$

where n_s and I_s are the solute refractive index and first ionization potential, respectively, and ω is the solute average transition energy. Usually approximated by defining an interval around the energy of the highest occupied orbital, as the most significant transitions occur at the frontier orbitals [25]. In general, the computational demands of methods which include dispersion effects within the system Hamiltonian are quite high due to the complex integrals involved [25]. Furthermore, the methods are very sensitive to the choice of basis set, thus requiring the use of rather large sets, which further aggravates the computational cost of including dispersion effects [25]. This makes the dispersion effects treated in such a way excessively expensive for periodic systems, such as extended surfaces.

2.3.3 Repulsion

The Pauli repulsion energy ΔG_{rep} contributing to the non-electrostatic term arises from the overlap between the tails of non-bonded densities of molecules. Compared to the previusly observed dispersion terms this quantity is defined more straightforwardly, and is less computationally demanding. One functional form is the following [256]:

$$\Delta G_{rep} = \alpha\int_{r\notin\Omega_{out}}\rho(\boldsymbol{r})dr \tag{2.69}$$

where $\rho(\boldsymbol{r})$ is the electronic density of the solute, and α is a constant for a given solute, which can be redefined as:

$$\Delta G_{rep} = k\rho_s \frac{n_s}{M_s}, \tag{2.70}$$

where s here refers to quantities that characterize the solvent, which are: the density ρ_s, the molecular weight M_s, the number of valence electrons n_{val}, and k being an empirically determined constant. The integral represents the portion of solute charge lying outside the cavity in Ω_{out}. Another form for the repulsion term is given from the study of hydrogen atoms [258], which is:

$$\Delta G_{rep} = \alpha \int_{r \notin \omega_{out}} |\nabla \rho(\boldsymbol{r})| dr. \tag{2.71}$$

A classical approximation based on empirical determined potentials also exists [25] These expressions have been used in the development of a variant of PCM for the study of systems in extreme pressure environments know as XP-PCM [259, 260] for which intermolecular repulsions plays a critical role.

2.3.4 SASA based models

A computational efficient way to treat within a single term the contributions to the non-electrostatic energy seen so far is based on the SASA [167, 168]. The total SASA A_{tot} is given by the sum of its atomic contributions, according to:

$$A_{tot} = \sum_{i}^{atoms} A_i(R_i^{vdW} + R_{probe}) = \sum_{i}^{atoms} A_i \tag{2.72}$$

where R_i^{vdW} is the radius of the i^{th} atomic vdW sphere, and R_{probe} is the probe radius of a sphere approximating the solvent (see Sec. 2.2.3), albeit some models consider R_{probe} as a solvent independent constant [82, 83, 81, 80]. An example of a SAS is given in Fig. 2.10 for a benzene molecule.

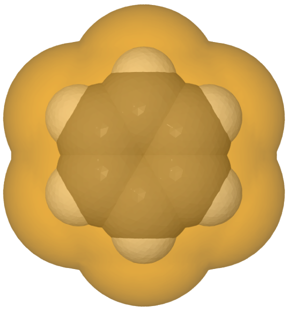


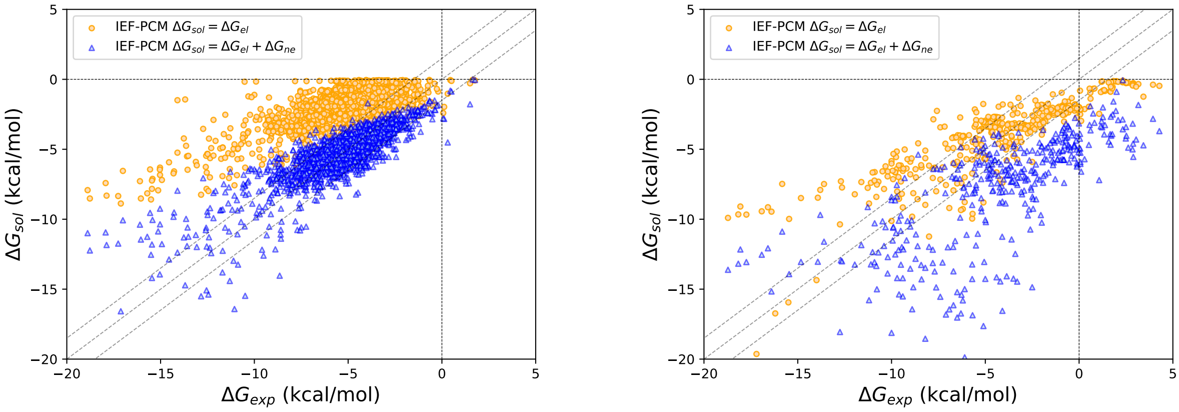
Figure 2.10: In orange the SASA of a benzene molecule, in black and white the van der Waals hard spheres for carbon and hydrogen atoms, respectively.

The physical justification of employing the SASA to account for the non-electrostatic term is that dispersion, repulsion, and other solvation effects not considered in the electrostatic term (e.g. charge transfer, electrostriction, dielectric saturation) mostly affect the first solvation shell of the solute. As such an intuitive approximation is to consider their contribution proportional to the area of the solute-solvent interface. This is based on the high correlation existing between the solvation energy of non-polar organic molecules and the SASA [261, 64]. Furthermore, for relatively small sized solutes with a high surface-area-to-volume ratio, it can also be considered as a good approximation for the cavitation contribution in place of the molecular volume [262, 263, 261, 264]. Due to this, its computational efficiency and relative simplicity, SASA based non-electrostatic models have been widely employed [81, 80, 238, 37, 84].

The most straightforward model for non-electrostatics based on SASA is then to consider the non-electrostatic energy ΔG_{ne} proportional to the total SASA of a solute A_{tot}, according to:

$$\Delta G_{ne} = \alpha \cdot A_{tot} + \beta \tag{2.73}$$

where α and β are the line slope and intercept coefficients, respectively. These two coefficients are the only parameters of this model and are usually obtained from linear regression of experimental data. As an example, Fig. 2.11 shows qualitatively the effectiveness of this simple model and its limitations when coupled with IEF-PCM [98] electrostatics ΔG_{el}, for a varied set of solutes in both aqueous and non-aqueous solvent [36, 241].



(a) Non-Aqueous solvents

(b) Aqueous solvent

Figure 2.11: Calculated ΔG_{sol} and ΔG_{el} as function of total experimental ΔG_{exp} solvation energies for (a) non-aqueous, and (b) aqueous solvent, using ΔG_{el} values from IEF-PCM (B3LYP/6-311G**) on 2530 neutral solutes of the MNSOL database [36, 241] in 91 solvents with ϵ_{out} ranging from $\approx$ 1.00 to 181.56. The ΔG_{ne} are obtained from Eq. 2.73 with coefficients $\alpha = -0.025$ kcal/mol·Å^2 and $\beta = 1.049$ kcal/mol, and $R_{probe} = 0.400$ Å.

From Fig. 2.11 it is clear how this simple SASA based model is effective for non-aqueous solvents, lowering the Mean Unsigned Error (MUE) of the solvation from 3.26 kcal/mol, when considering the electrostatic contribution from IEF-PCM only, to 0.95 kcal/mol. On the other hand, results for aqueous solvent are less encouraging, and an increase in MUE is observed from 1.91 kcal/mol when considering ΔG_{el} alone to 4.09 kcal/mol. This decrease in accuracy is mostly due to water standing out to other solvents due to its high dipole moment and tendency for hydrogen bonding. This partially justifies the choice of earlier solvation models to ignore the non-electrostatic contribution for water and focusing on the electrostatic contribution as previously mentioned, but also makes it clear the importance of the non-electrostatic contribution for solvents other than water. Better results for both aqueous and non-aqueous can nevertheless be achieved with relatively more complex models.

Less coarse models based on SASA which do not stray from the initial idea have been developed which account for the contribution of each solute atom to the total solvation energy, and are usually parametrized for a given solvent [64]. These models consider the non-electrostatic energy proportional to the sum of atomic SASA A_i each weighted with an atomic number dependent Z_i parameter w_{Z_i} (unit wise, a surface tensions), according to:

$$\Delta G_{ne} = \sum_{i}^{atoms} w_{Z_i} A_i. \tag{2.74}$$

The goal of such models is to distinguish between different atoms, as intuitively they will not contribute equally to the non-electrostatic energy. This concept can be further expanded by including atom types to better consider the atomic contribution within a molecular function as is usually done in molecular dynamics [265, 266, 267, 268]. Albeit accurate for a given solvent, these models lack in flexibility, as a different set of parameters should be ideally developed for each solvent, and so requiring experimental data which might not be available, or sufficient, to develop and test these parameters.

2.3.4.1 Cavity Dispersion and Solvent structural effects: CDS

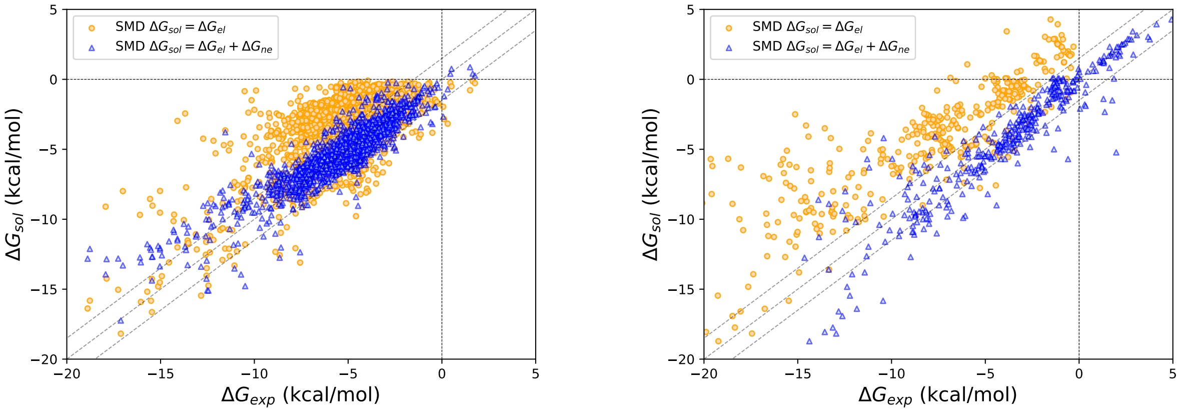
Accounting for different solvents in a generalized fashion is usually done by including solvent-dependent terms, as is done in the SMx family of solvation models [82, 83, 81, 80]. In SMx the non-electrostatic model goes under the name of Cavity, Dispersion, and Solvent structural effects (CDS), and in the latest iteration of the model [81] the non-electrostatic energy is given by:

$$\Delta G_{ne} = \sum_{i}^{atoms} \sigma_i A_i + \sigma_{solv} A_{tot} \tag{2.75}$$

where the first term accounts for specific interaction between the solute's atoms and the solvent, with σ_i being a function of both solvent empirical descriptors and solute atom dependent parameters, while the second term σ_{solv} is instead independent from the solute atoms and is given uniquely by the solvent empirical descriptors. Overall 5 empirical descriptors are required to fully describe the solvent within the CDS model. More details on the descriptors and σ_i, σ_{solv} are given in Sec. 3.2.1 on page 71, where the CDS model has been extended for periodic systems, and new sets of parameters have been developed for Finite-Difference Poisson based electrostatics.

In general, it should be noted that solvation models which employ the CDS model for the non-electrostatic contribution usually predict solvation energies of neutral solutes well below the 1.0 kcal/mol threshold, for both aqueous and non-aqueous solvents [36]. Fig. 2.12 shows the good performance of the SMD electrostatic model (an ASC method based on IEF electrostatics) with and without the non-electrostatic contribution from CDS, for both aqueous and non-aqueous solvents.

Furthermore, the use of empirical solvent descriptors and its parametrization on a wide set of solutes in 91 solvents allows the model to be "universal", meaning that accounting for different solvents requires uniquely empirical descriptors without any modification of the existing parameters, at least for a given electrostatic model.



(a) Non-Aqueous solvents

(b) Aqueous solvent

Figure 2.12: Calculated ΔG_{sol} and ΔG_{el} as function of total experimental ΔG_{exp} solvation energies for (a) non-aqueous, and (b) aqueous solvent, using ΔG_{el} values from SMD (B3LYP/6-311G**) on 2530 neutral solutes of the MSNOL database [36, 241] in 91 solvents with ϵ_{out} ranging from ≈ 1.00 to 181.56. The ΔG_{ne} are obtained from Eq. 3.33 as implemented in Gaussian16 [30].

The CDS parameters, are not fully transferable between electrostatic models, and a new set has to be developed for each one [191]. Two examples of popular solvation models with two different electrostatics which employ CDS successfully are SMD and SM12. The former is an ASC [218] method, while the latter is a GBA [86] one. The number of parameters required by the model is also quite flexible and allows the model to adapt to different electrostatic models. For example, for SMD [80] the CDS model requires only 34 parameters, while in SM12 [81], 72 parameters are required to achieve a similar accuracy in the prediction of solvation energies.

2.3.4.2 SCCS and SCCS-P

Following the CDS model, other non-electrostatic models have been developed which agument the SASA based approach with further terms such as the volume of the cavity V, which correlates not only with the cavitation energy but also with polarizability [269]. An example of such non-electrostatic models is found within the Self-Consistent Continuum Solvation [238, 239] (SCCS) model, where the non-electrostatic energy is given by:

$$\Delta G_{ne} = \Big(\alpha + \gamma\Big) A_{tot} + \beta V_{cav}. \tag{2.76}$$

Here, each solvent is considered through two parameters α and β fitted to experimental data, and γ is the empirical macroscopic surface tensions of the solvent, thus the model is dependent on one empirical descriptor. The accuracy of SCCS is similar to that of SMD and SM12 with CDS, requiring only 2 parameters per solvent, but as previously mentioned for SASA based models this model to lacks in versatility in accounting for different solvents. In fact, the first publication of the model was originally parametrized only for water. To partially solve this issue further parameters for 67 non-aqueous solvents have been recently published, together with a variation of the model, SCCS-P [37], where P stands for polarizability. The main difference between SCCS and SCCS-P is the replacement of the cavity volume V with the isotropic solute polarizability, as they are directly proportional. Result wise, replacing the cavity with the solute polarizability allowed SCCS-P for a overall increase in accuracy, and in particular more accurate results for aromatic compounds.

2.3.4.3 xESE and uESE

A recent non-electrostatic model developed for C-PCM and COSMO like electrostatics (see Sec. 2.2.6.2 on page 42) is the one within the Extended Easy Solvation Estimation model [84] (xESE) and its generalization to solvents other then water uESE [270]. These models are actually loosely based on SASA and can be considered as an example of empirical models which are efficient and accurate, but with many physically arbitrary choices.

xESE was initially developed for the prediction of solvation energies in water alone focusing in particular on charged species. The non-electrostatic contribution is accounted not only through the solute SASA and the cavity volume as in SCCS [238], but considers additional terms accounting for eventual shortcomings in the electrostatic contribution related to ASC values, and is independent of any solvent empirical descriptors. The uESE model is a generalization of the xESE model which extends the non-electrostatic contribution to 57 solvents. This is done by subdividing solvents into four classes: water, polar protic, polar aprotic, and non-polar, with a non-electrostatic term for each:

- A) For **Water** the uESE model is the same as to its predecessor xESE, and the expression for the non-electrostatic term is given by:

  $$\Delta G_{ne} = \sum_i \alpha_i A_i + \sum_i \beta_i q_i^2 + SRC, \tag{2.77}$$

  where α_i, and β_i are element specific parameters, A_i is the atomic SASA, q_i is the average total ASC on atom i, while SRC should account for short range corrections, and is defined as:

  $$SRC = f(C)(A|\sigma_{min}|^3 + B|\sigma_{max}|^3), \tag{2.78}$$

  where σ_{min} and σ_{max} are the minimum and maximum values of Q_i/A_i, and $f(C)$ is a damping factor which limits the value of the ΔG_{ne} based on the value of a parameter C.

- B) For **Polar protic solvents** an additional term is considered to Eq. 2.77 proportional to the volume of the cavity V:

  $$\Delta G_{ne} = \sum_i \alpha_i A_i + \sum_i \beta_i q_i^2 + SRC + \gamma V \tag{2.79}$$

  with the parameter γ being solute and solvent non-specific. The choice of using the cavity volume V is related to its direct proportionality relation with the solute polarizability.

- C) For **Polar aprotic solvents** the SRC term is replaced by a term linearly dependent on q_i weighted by an atomic element dependent parameter δ_i:

  $$\Delta G_{ne} = \sum_i (\alpha_i + \alpha_{solv}) A_i + \sum_i \delta_i q_i + \sum_i \beta_i q_i^2 + \gamma V \tag{2.80}$$

  and furthermore, α_i is shifted by a solvent specific parameter α_{solv}.

- D) The last class of solvents are the **non-polar** one, for which the non-electrostatic term is defined:

  $$\Delta G_{ne} = \sum_i (\alpha_i + \alpha_{solv}) A_i + \sum_i \delta_i |q_i| + \gamma V \tag{2.81}$$

  where the dependency on q_i^2 has been removed due to small values which increased numerical instabilities, and the linear term with Q_q is redefined with its absolute value $|q_i|$.

xESE and uESE are able to achieve accurate solvation energies for neutral solutes on par to other solvation models [84, 270], and are also able to achieve better results for charged solutes with errors below the experimental uncertainty of 3.0 kcal/mol. Albeit, it is clear that the initial intuitive assumption of considering the non-electrostatic term proportional to the SASA is somewhat overstretched.

2.4 Periodic Implicit Solvation

Compared to finite molecular system the modeling of solvation effects for periodic systems through implicit solvation models has seen much less developments. This can be probably attributed to several reasons such as the increased complexity of the equation within the QM and electrostatic formalisms, the already high computational cost for a QM treatment of periodic systems, together with a general lack of experimental solvation energies. For periodic systems, solvation energies of surfaces can however be related to the experimental contact angle θ_c [271] a liquid drop makes with a surface, as shown in Fig. 2.13a. The contact angle is described by Young's equation, which relates the Liquid-Gas γ_{LG}, Solid-Gas γ_{SG}, and Solid-Liquid γ_{SL} surface tensions to the equilibrium contact angle:

$$\gamma_{LG} \cdot \cos\theta_c - \gamma_{SG} + \gamma_{SL} = 0. \tag{2.82}$$

As shown in Fig. 2.13b, three possible cases can be observed: i) when $\gamma_{SG} > \gamma_{SL} + \gamma_{SG}$ no wetting of the surface by the liquid drop occurs, ii) partial wetting of the surface when $\gamma_{SG} < \gamma_{SL} + \gamma_{SG}$, and iii) complete wetting when $\gamma_{SG} = \gamma_{SL} + \gamma_{SG}$.

A relation between the experimental contact angle and the computed solvation energy per unit of area A of a surface slab exists [272], and is given by:

$$\cos\theta_c = \frac{1}{\gamma_{LG}} \frac{\Delta G_{tot}}{2A}. \tag{2.83}$$

This expressions can be used to evaluate the accuracy of periodic solvation models in predicting solvation energies for surfaces, but unfortunately is limited by the experimental uncertainty on θ_c. In fact, experimental contact angle measurements are very sensitive to contamination, and show a low reproducibility between different experimental conditions, reducing the reliability of the already limited experimental data available [273, 274, 275]. In practice, this complicates both the validation and the development of new models, and explains the fewer variety of models available for these type of solutes, which can overall be subdivided in only two main families. A first one which stems from the work of Fattebert and Gygi [99, 179] based on a smooth cavity surface focusing on DFT and plane-waves basis sets, and a second family based on C-PCM developed by Cossi [276] using a discontinuous cavity with a sharp variation of the relative permittivity at the solute cavity.

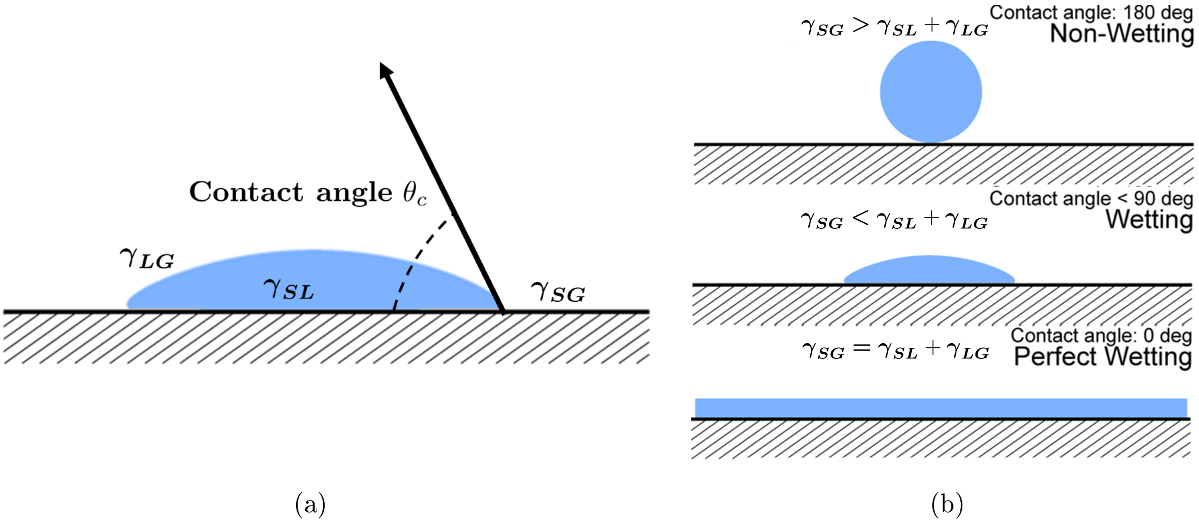

Contact angle θ_c
γ_{LG}
γ_{SL}
γ_{SG}
$\gamma_{SG} > \gamma_{SL} + \gamma_{LG}$
Contact angle: 180 deg
Non-Wetting
$\gamma_{SG} < \gamma_{SL} + \gamma_{LG}$
Contact angle < 90 deg
Wetting
$\gamma_{SG} = \gamma_{SL} + \gamma_{LG}$
Contact angle: 0 deg
Perfect Wetting
(a)
(b)

Figure 2.13: (a) Liquid drop showing the quantities in the Young's equation. (b) Possible wetting scenarios based on the different relationships between surface tensions.

2.4.1 Smooth Cavity Periodic Models

Most codes for the study of the electronic structure of periodic systems are based on the work of Fattebert and Gygi (FG) [99, 179]. Their model was originally developed for finite molecular systems, using Density Functional Theory (DFT), plane-wave basis sets, and a smooth dielectric cavity.

This model extends the standard Hohenberg-Kohn energy functional by including a term for the solute-solvent interaction energy E_s:

$$E[\rho] = T[\rho] + \int v(\boldsymbol{r})\rho(\boldsymbol{r})d\boldsymbol{r} + E_{xc}[\rho] + E_{es}[\rho] \tag{2.84}$$

where the first three terms correspond to: kinetic energy, the interaction energy with an external potential, and the exchange and correlation energies. While the last term E_{es} is the electrostatic energy defined as:

$$E_{es}[\rho] = \int \rho\phi[\rho]d\boldsymbol{r} \tag{2.85}$$

where the potential ϕ is obtained from solving Poisson Equation with a relative permittivity which depends on a smooth function of the solute charge density ρ:

$$-\nabla(\epsilon[\rho]\nabla\phi) = 4\pi\rho. \tag{2.86}$$

When taking the functional derivative of $E[\rho]$ with respect to the solute charge density ρ the E_{es} derivate gives the potential ϕ, plus an additional term:

$$V_\epsilon(\boldsymbol{r}) = -\frac{1}{8\pi}(\nabla\phi(\boldsymbol{r}))^2\frac{\delta\epsilon(\boldsymbol{r})}{\delta\rho} \tag{2.87}$$

to be included in the Kohn-Sham potential. The Kohn-Sham equations are then solved self-consistently to take in account the mutual solute-solvent polarization effects like in other SCRF methods.

The dependence of the relative permittivity on the solute charge density is given by the following surface model:

$$\epsilon(\rho(\boldsymbol{r})) = 1 + \frac{\epsilon_\infty - 1}{2}\left(1 + \frac{1 - (\rho(\boldsymbol{r})/\rho_0)^{2\beta}}{1 + (\rho(\boldsymbol{r})/\rho_0)^{2\beta}}\right) \tag{2.88}$$

where ϵ_∞, β, and ρ_0 are the three parameters needed to define the solute cavity. The first parameter ϵ_∞ control the asymptotic behavour of the function in the limit of $\rho(\boldsymbol{r})$ that goes to zero, and is then equal to the solvent relative permittivity. The second parameter ρ_0 controls the density value in the middle of the solute-solvent interface (it can be seen as the solute cavity size), while the last parameter β controls the width of the interface.

The reason that brought to define a smooth relative permittivity function of the solute charge density is related to the computation of forces which lead to difficulties when considering ASC schemes on the solute-solvent interface. Furthermore, the use of a discontinuous relative permittivity of the solvent complicates force calculation when using regular grids, as in the case of solving Poisson equation using a finite-difference approach [152, 153, 154]. The FG model has been widely adopted and extended for periodic systems in a large variety of quantum-chemical programs which employ plane-wave basis sets. Some examples of periodic codes which based their implicit solvation models on the work of Fattebert and Gygi are: Quantum Espresso [277, 278, 279, 280], VASP[281, 282], ONETEP [283, 284], CP2K[285], and BigDFT[240, 272], JDFTx [286].

2.4.2 Periodic C-PCM

In 2001, Cossi proposed a generalization of the C-PCM formalism to one- and two- dimensional periodic systems. In C-PCM [224, 225] the solute cavity is given by the GEPOL [181] algorithm which discretizes the vdW surface in small patches of area a_i refereed to as tesserae. The solvent is treated as a conductor and the electrostatic equations are hence greatly simplified (see 2.2.6.2 on page 42):

$$\mathbf{S}\mathbf{q} = \boldsymbol{\phi} \tag{2.89}$$

where $\mathbf{q}$ and $\boldsymbol{\phi}$ are vectors containing the apparent surface charges, and the potential at the position of each tesserae. $\mathbf{S}$ is instead a $N \times N$ matrix, where N is the number of tesserae, where elements are defined as:

$$\begin{cases} S_{ii} = 1.0694\left(\dfrac{4\pi}{a_i}\right)^{\frac{1}{2}} \\ S_{ij} = \dfrac{1}{r_{ij}} \end{cases} \tag{2.90}$$

where a_i is the area of the i^{th} tesserae, and r_{ij} is the distance between of them. To obtain the correct charge for the dielectric the vector $\mathbf{q}$ is scaled by a factor $f(\epsilon)$ defined in 2.2.6.2 on page 42, and here reported again:

$$f(\epsilon) = \frac{\epsilon - 1}{\epsilon + k} \tag{2.91}$$

where k is a value which ranges between zero and 0.5, the latter being the original value used.

The extension to periodic systems [276, 287] is done both for one- (1D) and two- (2D) dimensional periodic systems. For 2D systems, a reference cell is replicated along independent primitive displacement vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ which form the lattice vector $\boldsymbol{a} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2$ which indicates uniquely each of the replica of the reference cell through the integers n_1, n_2. This requires the generalization of the C-PCM equations to compute the apparent surface charges in the reference unit cell by taking into account the interaction with the periodic replicas. For the potential, the i^{th} element of the vector $\boldsymbol{\phi}$ for a particular lattice vector $\boldsymbol{a}$ defined by n_1, n_2 is then given by:

$$\phi_i = S_{ii}q_i + \sum_{j\neq i}^{N} S_{ij}q_j + \sum_{\boldsymbol{a}} \phi_i^{\boldsymbol{a}} \tag{2.92}$$

where the last term is the extension to all the direction of the lattice vector $\boldsymbol{a}$, and is given by:

$$\phi_i^{\boldsymbol{a}} = 2q_i \sum_{n=1}^{+\infty} \frac{1}{n|\boldsymbol{a}|} + \sum_{j\neq i}^{N} q_j \sum_{n=1}^{+\infty} \left(\frac{1}{|\boldsymbol{r}_{ij} + n\boldsymbol{a}|} + \frac{1}{|\boldsymbol{r}_{ij} - n\boldsymbol{a}|} \right). \tag{2.93}$$

This expression is recast in terms of Legendre polynomials to improve the stability and convergence of the series for systems where the total charge is equal to zero. For one-dimensional periodic systems with $\boldsymbol{a} = \boldsymbol{a}_1$ the potential of the i^{th} tesserae is:

$$\phi_i^{\boldsymbol{a}} = \sum_{j\neq i}^{N} q_j \left[\frac{2}{|\boldsymbol{a}|} \sum_{m=1}^{+\infty} \left(\frac{\boldsymbol{r}_{ij}}{\boldsymbol{a}} \right)^{2m} \zeta(2m+1)P_{2m} \right], \tag{2.94}$$

and for two-dimensional periodic systems the total potential is given by:

$$\sum_{\boldsymbol{a}} \phi_i^{\boldsymbol{a}} = \sum_{j\neq i}^{N} q_j \left[\sum_{n_1}\sum_{n_2} \frac{2}{|n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2|} \cdot \sum_{m=1}^{+\infty} \left(\frac{\boldsymbol{r}_{ij}}{|n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2|} \right)^{2m} \zeta(2m+1)P_{2m} \right], \tag{2.95}$$

where P_m is the Lengendre Polynomial of order m (in this case $2m$), and $\zeta = \sum_n^{+\infty} n^{-1}$ is the Riemann function. The double summation on n_1 and n_2 requires $n_1 \neq 1$ and couples of n_1, n_2 with greater common divisor of 1.

The periodic generalization of $\mathbf{S}$ instead takes the following form, for one-dimensional periodic systems, of:

$$\begin{cases} S_{ii} = 1.0694\left(\frac{4\pi}{a_i}\right)^{\frac{1}{2}} \\ S_{ij} = \frac{1}{r_{ij}} + \frac{2}{|\boldsymbol{a}|} \sum_{m=1}^{+\infty} \left(\frac{\boldsymbol{r}_{ij}}{\boldsymbol{a}} \right)^{2m} \zeta(2m+1)P_{2m} \end{cases} \tag{2.96}$$

while it is:

$$\begin{cases} S_{ii} = 1.0694\left(\frac{4\pi}{a_i}\right)^{\frac{1}{2}} \\ S_{ij} = \frac{1}{r_{ij}} + \sum_{n_1}\sum_{n_2} \frac{2}{|n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2|} \cdot \sum_{m=1}^{+\infty} \left(\frac{\boldsymbol{r}_{ij}}{|n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2|} \right)^{2m} \zeta(2m+1)P_{2m} \end{cases} \tag{2.97}$$

for two-dimensional systems, where only the non-diagonal elements are modified by the generalization to periodic systems when compared to molecules. The full treatment can be found in the original publication by Cossi [276].

Overall, is clear how the passage from finite to periodic systems increases the complexity of the mathematical formalism for C-PCM, which compared to other ASC methods simplifies the electrostatic problem by considering the solvent as a conductor. Furthermore, to our knowledge, the only application of this model is to an all-trans poly-glycine chain [276] using a classical description of the solute with atomic charges. This was done comparing the computed values obtained with the periodic C-PCM code to other computed solvation energies in the limit of an infinite chain with excellent agreement. As a comparison, modern implicit solvation models are routinely tested on hundred, if not thousands [36, 37, 38] of experimental solvation energies for both neutral and charged finite (non-periodic) solutes.

Chapter 3

Implicit solvation in CRYSTAL

This chapter focuses on the main methods which compose the implicit solvation model implemented in CRYSTAL for both finite molecular and periodic systems, and used in this thesis. In particular, the electrostatic model will be described first, focusing on the Self-Consistent Reaction Field (SCRF) method, the finite-difference (FD) resolution of the generalized Poisson equation, the FD Apparent Surface Charge (ASC) formalism, together with the solute cavity definition. Furthermore, the various atomic charge models used to approximate the solute charge density within the generalized Poisson equation will be presented. In the second part of the chapter the non-electrostatic part of the solvation model will be considered. Specifically, we will discuss the Cavity Dispersion and Solvent structural effects (CDS) model and its gradients.

3.1 Electrostatic Contribution

In CRYSTAL [31, 39], the electrostatic model is based on a Self-Consistent Reaction Field (SCRF) Method [79, 97] using an Apparent Surface Charge (ASC) [218, 219, 220] formalism; where the underlying electrostatic problem is numerically solved through a finite-difference generalized Poisson equation scheme.

For simplicity, the electrostatic model can be divided in three interconnected components, each treating a separate aspect of the model:

- **Finite-Difference generalized Poisson (FDP):** the approach allows to solve Poisson's equation for the electrostatic potential generated by the solute charge density. The solute potential is used to compute the ASC, and to calculate other necessary quantities for the solvation model. The FDP approach is dependent on the choice of **atomic charge model**.

- **Apparent Surface Charges (ASC):** account for the dielectric media surface polarization and are mapped on the SES cavity [169, 170]. The ASC inclusion within the solute Hamiltonian allows for a Quantum-Mechanical (QM) treatment of solvation effects.

- **Solvent-Excluded Surface (SES):** defines the solute cavity, and allows to separate the system in an atomistic region where the solute atoms are treated explicitly, and a dielectric region where the solvent is treated as a structureless dielectric media.

All three component play a fundamental in the implicit solvation model and, with the exception of the SES cavity, within the iterative SCRF procedure.

3.1.1 Finite-Difference Generalized Poisson

In the electrostatic model the potential $\phi(\boldsymbol{r})$, generated by a solute characterized by a charge density $\rho(\boldsymbol{r})$ in a non-homogeneous dielectric media with relative permittiviy $\epsilon(\boldsymbol{r})$, is obtained by solving the generalized Poisson equation:

$$\nabla[\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r})] = -4\pi\rho(\boldsymbol{r}), \tag{3.1}$$

where the solute charge density $\rho(\boldsymbol{r})$ is discretized as atomic point charges (see Sec. 3.1.1.1):

$$\rho(\boldsymbol{r}) = \sum_{i}^{atoms} q_i \delta(\boldsymbol{r} - \boldsymbol{r}_i). \tag{3.2}$$

Using a Solvent-Excluded Surface (SES) [169, 170] solute cavity to separate the atomistic region Ω_{in}, where atoms are treated explicitly, from the dielectric region Ω_{out} representing the solvent, the relative permittivity is set as a step function:

$$\epsilon(\boldsymbol{r}) = \begin{cases} \epsilon_{in} & \text{if } \boldsymbol{r} \in \Omega_{in}, \\ \epsilon_{out} & \text{if } \boldsymbol{r} \in \Omega_{out} \end{cases} \tag{3.3}$$

where ϵ_{in} is equal to 1 in the atomistic region Ω_{in}, and ϵ_{out} is equal to the solvent relative permittivity in the dielectric region Ω_{out}. As the generalized Poisson equation can be solved analytically only for simple cavity geometries, like ellipsoids and spheres, numerical methods have to be invoked for its solution when using molecular shaped cavities. In this thesis a Finite-Difference (FD) method has been used to numerically solve Poisson's equation, where the physical properties of the sytem are mapped onto a grid and numerical derivatives are solved at the grid nodes. The FD method has been mentioned in Sec. 2.2.2.2 on page 32 in the wider contest of implicit solvation models, and below we will focus on the FD implementation in CRYSTAL for both molecular and periodic systems.

Taking as an example a finite molecular system and a cubic FD grid with spacing h and (N_x, N_y, N_z) grid points per direction, where each grid point is characterized by three integers (i, j, k) defining its position in space, the physical properties of the system are initially mapped onto a grid and boundary conditions are assigned along the FD grid edges Γ. An example of cubic a FD grid and property mapping for a finite molecular system is shown in Fig. 3.1.

The dielectric constant is mapped at the facets between the grid points (grid facet points) to improve the dielectric description, and its value is assumed to be ϵ_{in} or ϵ_{out} if the grid facet falls within the atomistic region Ω_{in} or in the dielectric region Ω_{out}, respectively.

The other physical property of the system is the solute charge density $\rho(\boldsymbol{r})$, which is approximated as atomic point charges. These charges are mapped at the grid node (i, j, k), and as each charge q_i position in space (x, y, z) may not fall on a grid node (i, j, k), the atomic charges q_i are distributed on the nearest 27 grid point using a quadratic inverse interpolation algorithm [288]. Furthermore, a special set of points called Boundary Grid Points (BGP) can be defined. These points are located at the boundary between the atomistic Ω_{in} and the dielectric region Ω_{out}, and are points with at least one dielectric neighboring point in a different dielectric media than the other five. These points will be used in the definition of the Apparent Surface Charges and in the construction of the SES in the next sections.

Having mapped the physical properties onto the FD grid, boundary conditions are assigned to the grid edges Γ based on the periodicity of the system as shown in Tab. 3.1.

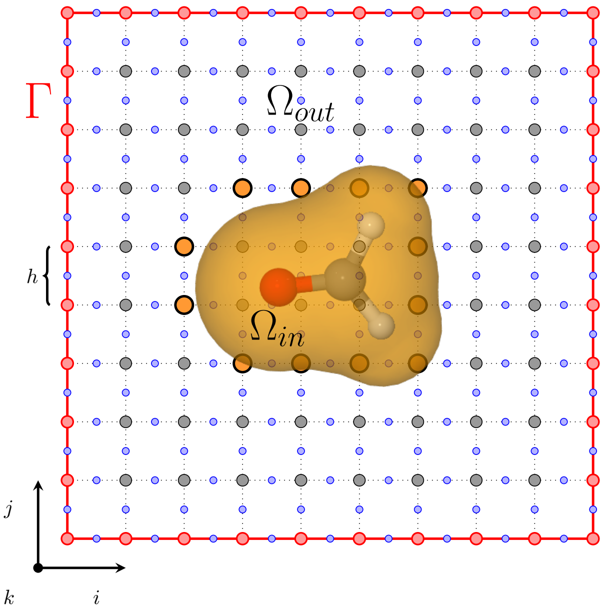


Figure 3.1: 2D example of a finite-difference grid (10 × 10) with spacing h. In gray the grid points where the generalized Poisson equations is solved using an iterative solver. In blue the relative permittivity mapped between the grid nodes. In red edge grid points where Dirichlet boundary conditions or Ewald formulas have to be imposed, and in orange the boundary grid points at the solute/solvent interface. The atomistic region within the solute cavity is indicated by Ω_{in}, and the dielectric region by Ω_{out}.

Periodicity	Dirichlet	Periodic (Ewald)
0	x,y,z	–
1	y,z	x
2	z	x,y

Table 3.1: Boundary conditions for the potential depending on the system periodicity.

For finite molecular systems 0D, and for the non-periodic direction of 1 and 2D systems, Dirichlet boundary conditions are imposed. In particular, the Debye-Hückel potentials $\phi_{ijk} = \phi(i,j,k)$ of the equivalent dipole of the system [143] are computed and assigned at each boundary node on the grid edges Γ, according to:

$$\forall i,j,k \in \Gamma : \phi_{ijk} = q^{+}\frac{\exp(-R^{+}_{ijk}\lambda_D)}{\epsilon_{out} \cdot R^{+}_{ijk}} + q^{-}\frac{\exp(-R^{-}_{ijk}\lambda_D)}{\epsilon_{out} \cdot R^{-}_{ijk}} \tag{3.4}$$

where R^{+}_{ijk} and R^{-}_{ijk} are the distances from the grid node (i,j,k) to the center of positive $Q^{+} = \sum_i q^{+}_i$, and negative $Q^{-} = \sum_i q^{-}_i$ charge, respectively. The term λ_D is the Debye-Hückel length and is set to $\lambda_D = +\infty$ for solvents with no salts effects. The Cartesian coordinates $\boldsymbol{r} = (x,y,z)$ of the centers of charge are, taking the positive center as an example, given by:

$$\boldsymbol{r} = \sum_{i}^{atoms} \frac{q^{+}_i \boldsymbol{r}_i}{Q^{+}}, \tag{3.5}$$

where r_i and q^{+}_i, are the coordinates and charge of the i^{th} positively charged atom, respectively. The same applies for the center of negative charge.

Once the physical properties of the system are mapped onto the FD grid, and the appropriate boundary conditions are imposed, the potential $\phi_{ijk} = \phi(i,j,k)$ can be obtained at each of the grid nodes by solving Poisson's equation derivatives numerically. For a finite molecular system and a FD grid with spacing h along all directions, the potential at each grid node (i,j,k) depends on the charge at the grid node $q_{ijk} = q(i,j,k)$, the grid spacing h, together with the potentials and relative permittivity values at the six adjacent grid points, and is given by [289]:

$$\phi_{ijk}^{0D} = \frac{1}{\sum\limits_{i}^{6} \epsilon_i} \left(\sum_{i}^{6} \epsilon_i \phi_i + 4\pi \frac{q_{ijk}}{h} \right) \tag{3.6}$$

where the summation over i runs on the six adjacent grid nodes, as shown in Fig. 3.2. Furthermore, Eq. 3.22 can be generalized to any parallelepiped grid shape with lengths (h_i, h_j, h_k), and inclination angles (α, β, γ), according to [289]:

$$\phi_0 = \frac{\sin\alpha \frac{h_y h_z}{h_x} \sum_{i=1}^{2} \epsilon_i \phi_i + \sin\beta \frac{h_x h_z}{h_y} \sum_{i=3}^{4} \epsilon_i \phi_i + \sin\gamma \frac{h_x h_y}{h_z} \sum_{i=3}^{6} \epsilon_i \phi_i + 4\pi q_0}{\sin\alpha \dfrac{h_y h_z}{h_x} \sum\limits_{i=1}^{2} \epsilon_i + \sin\beta \dfrac{h_x h_z}{h_y} \sum\limits_{i=3}^{4} \epsilon_i + \sin\gamma \dfrac{h_x h_y}{h_z} \sum\limits_{i=5}^{6} \epsilon_i} \tag{3.7}$$

where ϕ_i and ϵ_i are, respectively, potentials and relative permittivities at grid nodes adjacent to ϕ_0 at grid node (i,j,k), as shown in Fig. 3.2.

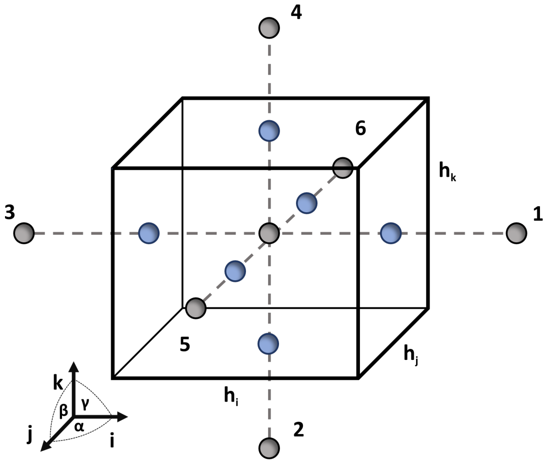


Figure 3.2: Volume element for a 3D FD cubic grid. Each grid point (gray) is neighbored by six grid points, and six relative permittivity values mapped at the facets points (blue). (α, β, γ) are angles along directions (i,j,k) with grid step sizes (h_i, h_j, h_k).

Applying Eq. 3.6 yields a system of $(N_x \cdot N_y \cdot N_z)$ linear equations which are solved iteratively with assigned potentials on the non periodic edge grid nodes Γ, according to Tab. 3.1. The linear system is solved using the Optimal Successive Over-relaxation algorithm [148, 149, 130], where the potential ϕ_{ijk} at each non boundary grid point is given by:

$$\phi_{ijk}^{(n)} = (1-\omega) \cdot \phi_{ijk}^{(n-1)} + \omega \cdot \phi_{ijk}, \tag{3.8}$$

where $\phi_{ijk}^{(n)}$ and $\phi_{ijk}^{(n-1)}$ are the values of the potential at different iterations, and ω is a tunable parameter to optimize the algorithm convergence. The parameter is estimated from the spectral radius of the Gauss-Seidel approach [130], and the number of iterations necessary for convergence scales linearly with the length of the grid.

For periodic systems, the Ewald formulas for systems with reduced periodicity are used instead [290], and Dirichlet conditions are imposed only on the non periodic directions. In particular, for a 1D periodic system in the x direction, such as a polymer composed of N atoms per unit cell, each with charges q_n, the potential at site $\boldsymbol{r}_m$ in space is given by:

$$\begin{aligned}\phi^{1D}(\boldsymbol{r}_m) =& \sum_{\mathbf{R}}\sum_{n=1}^{N'} q_n \frac{\operatorname{erfc}(\xi|\boldsymbol{r}_m - \boldsymbol{r}_n + \mathbf{R}|)}{|\boldsymbol{r}_m - \boldsymbol{r}_n + \mathbf{R}|} \\ &+ \frac{1}{a}\sum_{k_i\neq 0}\sum_{n=1}^{N} q_n e^{-ik_1(\boldsymbol{r}_m - \boldsymbol{r}_n)} K_0(\frac{k_1^2}{4\xi^2}, \rho_{mn}^2\xi^2) \\ &- \frac{1}{a}\sum_{n=1}^{N} q_n \Big(\gamma + \log\big(\rho_{mn}^2\xi^2\big) + E_1(\rho_{mn}^2\xi^2)\Big) - \frac{2\xi}{\sqrt{\pi}} q_m,\end{aligned} \tag{3.9}$$

where the first term is a sum in direct space, and N' indicates that the cases $n = m$ with translation vector $\mathbf{R}$=0 are excluded from the real space sum, and ξ is a parameter affecting the relative decay of the sums. The second and third terms are reciprocal space contributions, where γ is the Euler Mascheroni constant, k_1 forms a discrete set of k points in the reciprocal space, and ρ_{mn} is given by:

$$\rho_{mn} = \Big((z_m - z_n)^2 + (y_m - y_n)^2\Big)^{\frac{1}{2}}. \tag{3.10}$$

The function $K_0(u, v)$ is an incomplete modified Bessel function of the second kind, defined as:

$$K_0(u, v) = \int_1^\infty \frac{1}{t} e^{-(ut - \frac{v}{t})}\, dt, \tag{3.11}$$

and $E_1(v)$ is an exponential integral defined as:

$$E_1(v) = \int_1^\infty \frac{1}{t} e^{-vt}\, dt, \tag{3.12}$$

where v and u where set to: $v = \rho_{mn}^2\xi^2$, and $u = k_1^2/4\xi^2$, respectively. The potential for a 2D periodic system ϕ^{2D} in the direction x and y, such as a slab with unit cell area A, is instead given by:

$$\begin{aligned}\phi^{2D}(\boldsymbol{r}_m) =& \sum_{\mathbf{R}}\sum_{n=1}^{N'} q_n \frac{\operatorname{erfc}(\xi|\boldsymbol{r}_m - \boldsymbol{r}_n + \mathbf{R}|)}{|\boldsymbol{r}_m - \boldsymbol{r}_n + \mathbf{R}|} \\ &+ \sum_{n=1}^{N} q_n \sum_{k\neq 0} e^{-i\mathbf{k}(\boldsymbol{r}_m - \boldsymbol{r}_n)} \cdot \frac{1}{\mathbf{k}} g(\mathbf{k}, z_m - z_n, \xi) \\ &- \frac{2\sqrt{\pi}}{A}\sum_{n=1}^{N}\bigg(\frac{1}{\xi} e^{\xi^2(z_m - z_n)^2} + \sqrt{\pi}\cdot(z_m - z_n)\cdot\operatorname{erf}(\xi(z_m - z_n))\bigg) + \frac{2\xi}{\sqrt{\pi}} q_m,\end{aligned} \tag{3.13}$$

and the function $g(\mathbf{k}, z_{mn}, \xi)$ is given by:

$$g(\mathbf{k}, z_{mn}, \xi) = e^{\mathbf{k}z_{mn}} \operatorname{erfc}\bigg(\frac{\mathbf{k}}{2\xi} + \xi z_{mn}\bigg) + e^{\mathbf{k}z_{mn}} \operatorname{erfc}\bigg(\frac{\mathbf{k}}{2\xi} - \xi z_{mn}\bigg) \tag{3.14}$$

where $z_{mn} = z_m - z_n$ has been used, and $\mathbf{k}$ forms a finite set of vector in the reciprocal space.

The potentials obtained from the FD generalized Poisson scheme, together with the Boundary Grid Points, are fundamental to compute the Apparent Surface Charges (ASC), which allow for the mutual solute-solvent polarization within the SCRF procedure. The ASC will be discussed below in Sec. 2.2.6 on page 40, but first, we will overview the atomic point charge models used by the electrostatic model to approximate the solute charge density.

3.1.1.1 Atomic Charge Models

Atomic charge models play a critical role in the FD generalized Poisson scheme discussed so far. In fact, the solute charge density $\rho(\boldsymbol{r})$, acting as source term in Poisson's equation, is approximated by a discrete sum of atomic point charges q_i in the model:

$$\nabla[\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r})] = -4\pi \sum_{i}^{atoms} q_i \delta(\boldsymbol{r} - \boldsymbol{r}_i). \tag{3.15}$$

Atomic charge models can be divided into different classes [64]: in this thesis class II and class IV charge models have been considered within the SCRF procedure. For class II models, which partition the electron density from a quantum mechanical calculation into atomic contributions, Mulliken [291], Hirshfeld (HPA) [292], and Hirshfeld-I (HPA-I) [293, 294] have been considered. On the other hand, for class IV models, which apply a semi-empirical mapping to a precursor class II or class III charge to reproduce an experimentally-determined or calculated observable, CM5 model [295] has been chosen, together with two modifications of the model to be discussed below. We will now recall the fundamentals of the different charge models considered, highlighting possible advantages and shortcoming in their applications in the electrostatic model.

- **Mulliken**: In this approach, electrons are partitioned amongst the atoms according to the contribution of the different atomic orbitals (AO) μ and ν to the wave function. The Mulliken charge $q_{\mathrm{i}}^{\mathrm{M}}$ of an atom i is given by:

$$q_{\mathrm{i}}^{\mathrm{M}} = Z_i - \sum_{\mu \in \mathrm{i}} \sum_{\nu} \sum_{\mathbf{R}} P_{\nu\mu}^{\mathbf{R}} S_{\nu\mu}^{\mathbf{R}} \tag{3.16}$$

  where Z_i is the nuclear charge, $\mathbf{R}$ is a translation vector in the direct lattice, and P and S are the density and overlap matrices, respectively. Due to their conceptual simplicity and fast calculation time, this is the charge model which has been previously considered for the implicit solvation of both finite and infinite periodic systems in CRYSTAL [289]. However, a number of drawbacks make their use delicate [296, 297, 298, 299, 300, 301, 302, 303], especially due to the equal division of the contribution involving basis functions between different pairs of atoms[302], nonphysical electron partition in case of unbalanced or very large basis sets [303], along with high basis set dependence, making in principle a comparison of partial charges and related properties between different levels of theory difficult. This last point is particularly problematic in case of a SCRF approach in which the source term of the Poisson equation is approximated as a sum of distributed atomic charges, since a high basis set dependence of the solvation energy can be expected especially for finite molecular systems for which diffuse basis functions are routinely used for the calculation of various properties.

- **Hirshfeld and Hirshfeld-I**: In this approach, the total electron density $\rho(\mathbf{r})$ is partitioned into atom-in-molecule densities $\rho_\mathrm{i}(\mathbf{r})$ via a weighting function $w_\mathrm{i}(\mathbf{r})$ according to:

  $$\rho_\mathrm{i}(\mathbf{r}) = \rho(\mathbf{r}) \cdot w_\mathrm{i}(\mathbf{r}) = \rho(\mathbf{r}) \cdot \frac{\rho_i^0(\mathbf{r})}{\sum_\mathrm{j} \rho_\mathrm{j}^0(\mathbf{r})} \tag{3.17}$$

  where $\rho^0(\mathbf{r})$ is the spherical pro-atom density of the isolated and neutral atom i in vacuum and $\sum_\mathrm{j} \rho_\mathrm{j}^0(\mathbf{r})$ is the so-called promolecular density. The corresponding Hirshfeld atomic charge $q_\mathrm{i}^\mathrm{HPA}$ is then obtained by integration of the atomic contributions over the entire space v:

  $$q_\mathrm{i}^\mathrm{HPA} = Z_\mathrm{i} - \int_v \rho_\mathrm{A}(\mathbf{r})\, dv. \tag{3.18}$$

  In particular, for periodic systems the electron density $\rho(\boldsymbol{r})$ is obtained from:

  $$\rho(\boldsymbol{r}) = \sum_{\mu\nu} \sum_{\mathbf{R}} P_{\mu\nu}^{\mathbf{R}} \sum_{\mathbf{h}} \chi_\mu^{(\mathbf{R}+\mathbf{h})}(\boldsymbol{r}) \chi_\nu^{\mathbf{h}}(\boldsymbol{r}), \tag{3.19}$$

  where χ are Atomic Orbitals (AO), $P_{\mu\nu}$ is the density matrix in terms of the AOs, and both $\mathbf{h}$ and $\mathbf{R}$ are lattice vectors indexes. Instead, the spherical pro-atom density $\rho_\mathrm{A}^0(\mathbf{r})$ is given by:

  $$\rho_\mathrm{i}^0(\boldsymbol{r}) = \sum_{\mu\nu\in i} P_{\mu\nu}^0 \sum_{\mathbf{R}} \chi_\mu^{\mathbf{R}}(\boldsymbol{r}) \chi_\nu^{\mathbf{R}}(\boldsymbol{r}), \tag{3.20}$$

  where $\boldsymbol{r}$ is contained in the spatial region of the reference unit cell v of the crystal lattice. Due to the choice of spherically averaged densities of neutral isolated atoms as pro-atom densities, Hirshfeld charges tend to be small in magnitude [293] leading to a possible underestimation of electrostatic interactions [293, 304], which are crucial for calculation of solvation energies in implicit solvation models, especially for charged species. To overcome Hirshfeld charges limitations, other definitions of $w_\mathrm{A}(\mathbf{r})$ can be considered, since Eq. (3.17) does in fact offer significant freedom in choosing it. For example, one possible strategy is the Hirshfeld-I extension [293, 294], where each pro-atom density is obtained as a linear interpolation between the densities of spherically-averaged isolated atoms and/or ions with an integer charge [305]. An iterative procedure is then used to maximize similarity between atom-in-molecule and pro-atom densities. The resulting Hirshfeld-I charges are then larger in magnitude than Hirshfeld charges, sometimes overestimating electrostatic interactions , however [306]. On the other hand, both atomic charge models are well-known to be less basis set dependent than Mulliken charges [307, 308, 309].

- **CM5**: In this model, Hirshfeld charges $q_\mathrm{i}^\mathrm{HPA}$ are corrected by a semi-empirical mapping, according to:

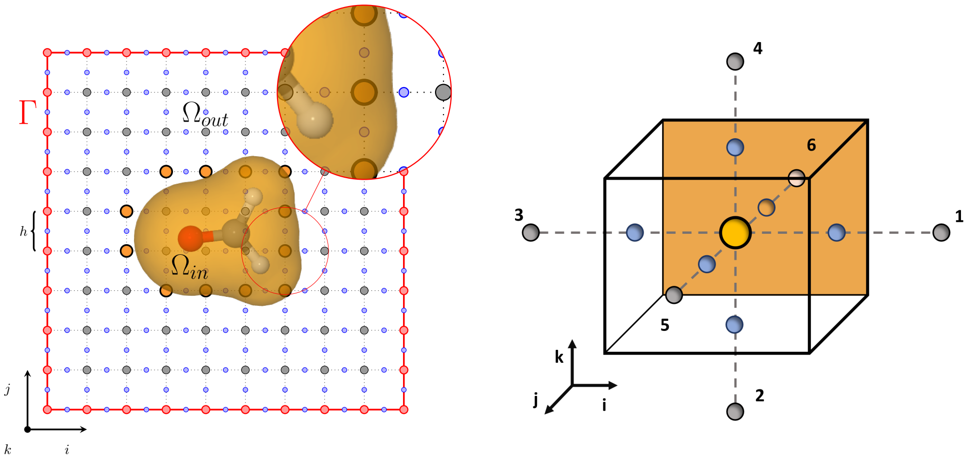
  $$q_i^\mathrm{CM5} = q_i^\mathrm{HPA} + \sum_{\mathbf{R}} \sum_{j\neq i\in v} T_{ij}^{\mathbf{R}} B_{ij}^{\mathbf{R}} = q_i^\mathrm{HPA} + \sum_{\mathbf{R}} \sum_{j\neq i\in v} T_{ij}^{\mathbf{R}} \exp\{-\alpha(R_{ij}^{\mathbf{R}} - \mathbf{r}_i - \mathbf{r}_j^{\mathbf{R}})\} \tag{3.21}$$

  where the α and T_{ij} parameters are fitted to reproduce both experimental and theoretical reference gas phase dipoles, and the B_{ij} term corresponds to the Pauling bond order which approximates the electron density overlap between two atoms i and j at positions $\mathbf{r}_i$ and $\mathbf{r}_j$ respectively, separated by R_{ij}, with radii r_i and r_j. For periodic system **R** is a direct lattice vector, and the summations over i and j are over atoms in the spatial region v of the reference unit cell. The mapping is therefore independent from the molecular electronic structure, and CM5 atomic charges should improve on both Hirshfeld and Hirshfeld-I electrostatics, while still remaining weakly basis set dependent [295]. In addition, CM5 atomic charges have already been found to perform well for solvation energy calculations [81, 310].

In this thesis, two additional modifications to the original CM5 charge model are introduced: (i) the Hirshfeld charge found in Eq.(3.21) is replaced by the Hirshfeld-I version, leading to the iterative CM5 variant (CM5-I) charge model; (ii) a scaling of the CM5 charges is applied, leading to the s·CM5 charge model, where the scaling factor is determined by minimizing the mean unsigned error (MUE) of the computed solvation energies compared to available experimental data of a given test set.

3.1.2 Apparent Surface Charges

Apparent Surface Charges (ASC) [218, 219, 220] $\{q_{asc}\}$ account for the surface polarization at the solute-solvent interface, and are determined from converged electrostatic potentials at Boundary Grid Points (BGP). These are special points of the FD grid characterized by having at least one dielectric neighboring point in a different dielectric media, as shown in Fig. 3.3a and Fig.3.3b.



(a) (b)

Figure 3.3: (a) Focus on a boundary grid point on a 2D cubic finite-difference grid, and (b) elementary volume element containing one boundary grid point at its center in orange. The solute-solvent interface is represented by a orange plane.

For a regular cubic grid with spacing h the ASC q_{asc} at a BGP is obtained from the value of the converged potential ϕ_{BGP} at the BGP grid position (i, j, k), and from the potentials ϕ_n at the neighboring grid nodes ($n \in [1; 6]$), according to [289]:

$$q_{asc} = -q_{BGP} + \frac{3h}{2\pi}\left(\phi_{BGP} - \frac{1}{6}\sum_{n}^{6}\phi_n\right), \tag{3.22}$$

where the solute charge mapped at the BGP during the FD resolution of the generalized Poisson equation q_{BGP} has to be removed from the ASC. Eq. 3.22 can be obtained through Gauss law:

$$\nabla \cdot \mathbf{E}(\boldsymbol{r}) = 4\pi\sigma(\boldsymbol{r}) \xrightarrow[\sigma=q_{asc}/h^3]{\mathbf{E}=-\nabla\phi} -\nabla^2\phi(i, j, k) = 4\pi q_{asc}(i, j, k)/h^3, \tag{3.23}$$

by considering the spatial position $\boldsymbol{r}$ at the BGP coordinates $\boldsymbol{r} = (i, j, k)$, the charge density $\sigma(\boldsymbol{r})$ as a discrete charge q_{asc} per grid volume h^3, and a FD form for the gradient. Furthermore, Eq. 3.22 can be generalized to any parallelepiped grid shape with lengths (h_i, h_j, h_k), and inclination angles (α, β, γ), according to [289]:

$$\begin{cases} V = h_x h_y h_z\Big((1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma) + 2\cos\alpha\cos\beta\cos\gamma\Big)^{\frac{1}{2}} \\ q_{asc} = \frac{V}{4\pi h_x^2 h_y^2 h_z^2}\Big(h_y^2 h_z^2(2\phi_0 - \phi_1 - \phi_2) + h_x^2 h_z^2(2\phi_0 - \phi_3 - \phi_4 + h_y^2 h_z^2(2\phi_0 - \phi_5 - \phi_6)\Big) - q_0 \end{cases} \tag{3.24}$$

where V is the volume of the elementary parallelepiped cell, and the potentials $\{\phi_1, ..., \phi_6\}$ are adjacent to grid node (i, j, k) with potential ϕ_0, as shown in Fig. 3.2.

Since only surface, and not volume, polarization is accounted for in the FD-ASC formalism (see Sec. 2.2.6 on page 40) a renormalization procedure for the ASC can be additionally performed [231]. Positive (q^+) and negative (q^-) ASC are normalized according to the following set of equations:

$$\begin{cases} \overline{q}_i^+ = q_i^+ \left(1 + \dfrac{\Delta Q_{err}}{2\sum_i q_i^+}\right) \\ \overline{q}_i^- = q_i^- \left(1 + \dfrac{\Delta Q_{err}}{2\sum_i q_i^-}\right), \end{cases} \tag{3.25}$$

where ΔQ_{err} is the error between the theoretical sum ΔQ_{asc}^{Th} and the computed sum of ASC Q_{asc} before renormalization:

$$\Delta Q_{err} = Q_{asc}^{Th} - Q_{asc} = Q_{asc}^{Th} - \sum_i^{asc} q_i. \tag{3.26}$$

The theoretical sum ΔQ_{asc}^{Th} is obtained from Gauss theorem and is obtained by considering the total solute charge Q_s and the solute relative permittivity ϵ_{out}, according to:

$$Q_{asc}^{Th} = -\frac{\epsilon_{out} - 1}{\epsilon_{out}} \cdot Q_s. \tag{3.27}$$

This renormalization procedure is applied at each SCRF cycle, and Eq. 3.26 allows to monitor the quality of the ASC formalism throughout the SCRF procedure. In the current implementation typical errors are well lower than 10^{-5} e for ΔQ_{err}.

The ASC are then projected from the FD grid to the SES solute cavity, and their final position in space are used to build the solute-solvent interaction potential within the solute Hamiltonian.

3.1.3 Solvent Excluded Surface

A Solvent-Excluded Surface (SES) [169, 170] cavity has been used throughout this thesis. In particular, a semi analytical grid based approach developed by Rocchia et al. [96] was used to construct the SES.

In general, the SES is formed by three different surface patches which arise from the contact of a solvent spherical probe with the surfaces of atom centered van der Waals (vdW) spheres, as described in Sec. 2.2.3 on page 34. When the solvent probe is in contact with only one vdW sphere a contact patch is generated, instead when the solvent probe is in contact with two or three vdW spheres a reentrant or toric patch is generated, respectively.

In the approach proposed by Rocchia et al. [96] toric patches are neglected, and contact patches are constructed analytically from atom centered vdW spheres of a given radius (in this case Bondi radii [156]), and the reentrant patches are constructed numerically using an iterative procedure. The first step in the construction of the SES is to map the surface onto the grid, and to assign any facet grid point, located at the interface between two grid points (i, j, k), as inside or outside the solute cavity, as shown in Fig. 3.4. In particular, points within the atomistic region Ω_{in} are assigned the relative permittivity ϵ_{in} and a status of "in" points, while points within the dielectric region Ω_{out} are assigned the relative permittivity ϵ_{out} and a status of "out" points. Furthermore, just as in the previous sections BGP are

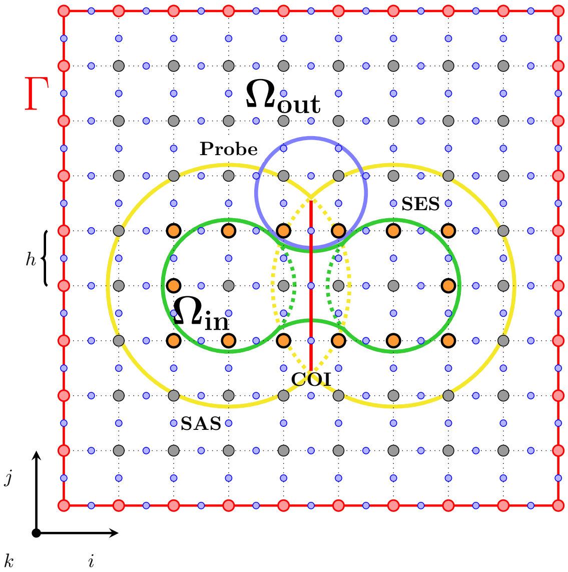


Figure 3.4: Example of solute cavities and surface elements mapped on a 2D FD grid. In yellow the SAS, in green the SES, and in blue the solvent probe radius. Boundary grid point are represented in orange, while grid nodes are in gray. Smaller blue circles represent the relative permittivity mapped at the grid node facets. The grid boundary is in red.

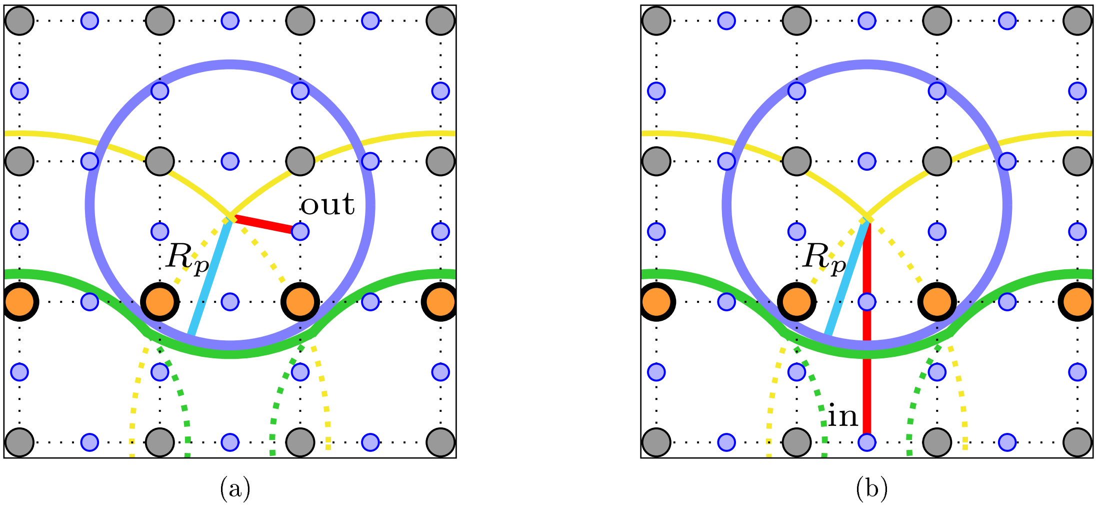
defined as points that have at least one adjacent grid facet point belonging to a different dielectric media.

The SES construction algorithm requires the definition of a Solvent-Accessible Surface (SAS), which is the surface generated by vdW spheres expanded by a solvent probe radius, as seen in 2.2.3 on page 34, and shown again in Fig. 3.4. To build the SES cavity from the SAS the distance from each "out" facet point neighboring a Boundary Grid Point to the SAS has to be calculated. Depending on the distances between this "out" facet point and the SAS two possible cases exist:

- If the point is closer to the SAS than a probe radius the point status is not updated and it will remain as an "out" point, as shown in Fig. 3.5a.
- If the point is further from the SAS than a probe radius it implies that the point cannot be reached by the solvent probe, and is an internal point not accessible to the solvent. In this case the point status is updated to "in", as shown in Fig. 3.5b.

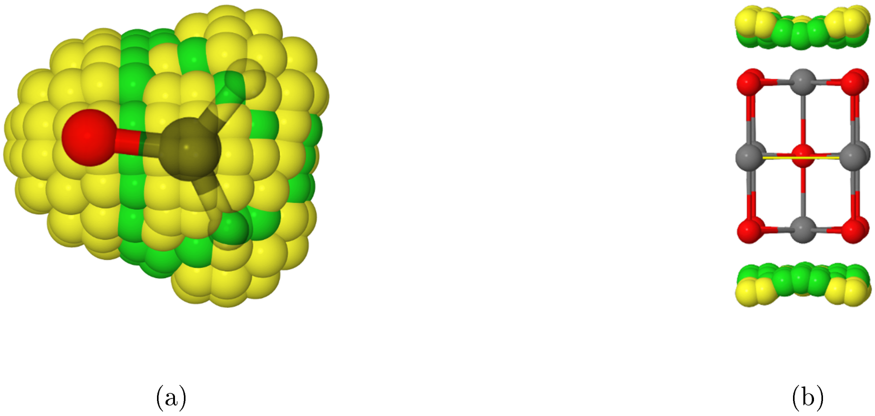
As the contact parts of the solute cavity coincide with the van der Waals surface, only a limited number of points on the SAS are required to calculate the distances to the "out" facet points neighboring a BGP. These SAS points are located on the arcs at the intersections between SAS spheres, called Circle Of Intersection (COI), as shown in Fig. 3.4. As such, the SES contact regions are constructed analytically, while the reentrant regions are constructed numerically using an iterative process to check if the status of any grid point has changed after the assignment of the "in" and "out" status. The final set of BGP are then projected onto the resulting SES, leading to a set of ASC to be included into the solute Hamiltonian.

For periodic systems, supercells are built using 3 and (3×3) unit cells for polymers and slabs, respectively, and only BGP belonging to the reference unit cell are projected onto the SES generating an extended cavity [289]. Examples of SES cavities for periodic and non-periodic systems are shown in Fig. 3.6.



(a) (b)

Figure 3.5: Example of selection process for an "in" and "out" grid facet point. (a) The point status as an external point remains unchanged (b) the point status changes from external to internal. The yellow lines represent the SAS, the green lines the SES, and the blue lines the probe sphere. In light blue the probe radius R_p, and in red the line connecting the closest point on the SAS to a given facet grid point.



(a) (b)

Figure 3.6: Examples of ASC projected on a SES cavity for (a) a finite molecular system (formaldehyde), and (b) a 2D periodic MgO slab. Contact and reentrant ASC are displayed in yellow and green, respectively.

3.1.4 Self-Consistent Reaction Field

A solute-solvent interaction potential V_{int} (see Sec. 2.2.6 on page 40) can be defined from the Apparent Surface Charges (ASC) mapped on the solute Solvent-Excluded Surface (SES). This interaction potential, when included within the gas phase Hamiltonian of the solute H_{gas}:

$$H_{solv} = H_{gas} + V_{int}, \tag{3.28}$$

allows for a quantum mechanical treatment of polarization effects, and mutual solute-solvent polarization effects can further be considered through an iterative process called Self-Consistent Reaction Field (SCRF) [79, 97], first introduced in Sec. 2.2.1.1 on page 27, which forms the heart of the electrostatic solvation model.

The SCRF procedure in Crystal is based on an ASC formalism where the underlying electrostatic problem is governed by the numerically solved generalized Poisson equation using a FD approach, and it can be summarized as follows: for a given solute, a reference gas phase calculation is performed, where the gas phase wave function ψ_g and the solute charge density $\rho_{gas}(\boldsymbol{r})$ are computed. The next step involves setting up the FD grid. In particular, the solute cavity (SES) is defined, the relative permittivity is mapped onto the grid, and boundary grid nodes are found. The previously computed gas phase charge density $\rho_{gas}(\boldsymbol{r})$ is then approximated as atomic point charges (see Sec. 3.1.1.1 on page 62 for the implemented charged models), and the atomic point charges are mapped at the grid nodes using an inverse quadratic interpolation.

Having mapped all the physical properties of the system onto the FD grid, the boundary conditions in the non-periodic directions are set using the Debye-Hüchel potentials of the equivalent dipole of the system (instead, for the periodic direction the Ewald formulas for systems with reduced periodicity are used, see Sec. 3.1.1 on page 58), and the generalized Poisson equation is solved iteratively for $\phi(i, j, k)$ at the FD grid nodes using the O-SOR algorithm. The converged potentials are used to obtain the ASC at the boundary grid nodes, which are projected onto the SES, and all elements for the interaction potential V_{int} are computed. The interaction potential allows for a condensed phase self-consistent field quantum-mechanical calculation, at the Hartree-Fock or DFT level of theory, which gives the last remaining quantities needed to compute the electrostatic solvation energy ΔG_{el}: the solute-solvent system wave function ψ_s, the solvated solute charge density $\rho_{sol}(\boldsymbol{r})$, and the reaction field energy E_{RF} (see below). The ΔG_{el} is then compared to the ΔG_{el} computed at the previous SCRF cycle, and if the difference is below a predetermined threshold the SCRF procedure is concluded successfully. Otherwise, the last computed solvated charge density $\rho_{sol}(\boldsymbol{r})$ is used to obtain new set of atomic point charges, and the SCRF cycle begins anew until convergence of the electrostatic contribution ΔG_{el}. The SCRF procedure is shown as a flowchart in Fig 3.7.

One of the fundamental contributions to the SCRF procedure just described, and to the calculation of the electrostatic contribution ΔG_{el}, is the reaction field energy E_{RF}. In fact, the ΔG_{el} is obtained by the difference between the solute-solvent system energy fully interacting $\langle\psi_s| H_{gas} + V_{int} |\psi_s\rangle$, and the sum of the solute gas phase energy $\langle\psi_g| H_{gas} |\psi_g\rangle$ and the reaction field energy E_{RF}, according to:

$$\Delta G_{el} = \langle\psi_s| H_{gas} + V_{int} |\psi_s\rangle - \Bigg(\langle\psi_g| H_{gas} |\psi_g\rangle + E_{\mathrm{RF}} \Bigg), \tag{3.29}$$

where ψ_s and ψ_g are the system solvated and gas phase wave functions, respectively. The reaction field energy E_{RF} is instead given by a sum of Coulomb interactions between the i^{th} ASC q_i and the solute potential $\phi_s(\boldsymbol{r})$ evaluated at the position of this ASC charges ($\boldsymbol{r}_i^{asc}$), according to:

$$E_{RF} = \frac{1}{2} \langle\psi_s| V_{int} |\psi_s\rangle = \frac{1}{2} \sum_i^{N_{asc}} q_i \cdot \phi(\boldsymbol{r}_i^{asc}). \tag{3.30}$$

From the physical point of view, the reaction field energy E_{RF} corresponds to the work needed to place the ASC $\{q_i\}$, at the positions $\boldsymbol{r}_i^{asc}$, within the field generated by the solute charge distribution, and is also equal to the non-reversible work needed to polarize an isotropic linear dielectric, and has to be removed to obtain the status of a free energy [25, 23]. For periodic system the ASC $\{q_i\}$ are computed in the reference unit cell, and the solute potential ϕ is obtained through the Ewald formulas seen in Sec. 3.1.1 on page 58 at the ASC spatial positions.

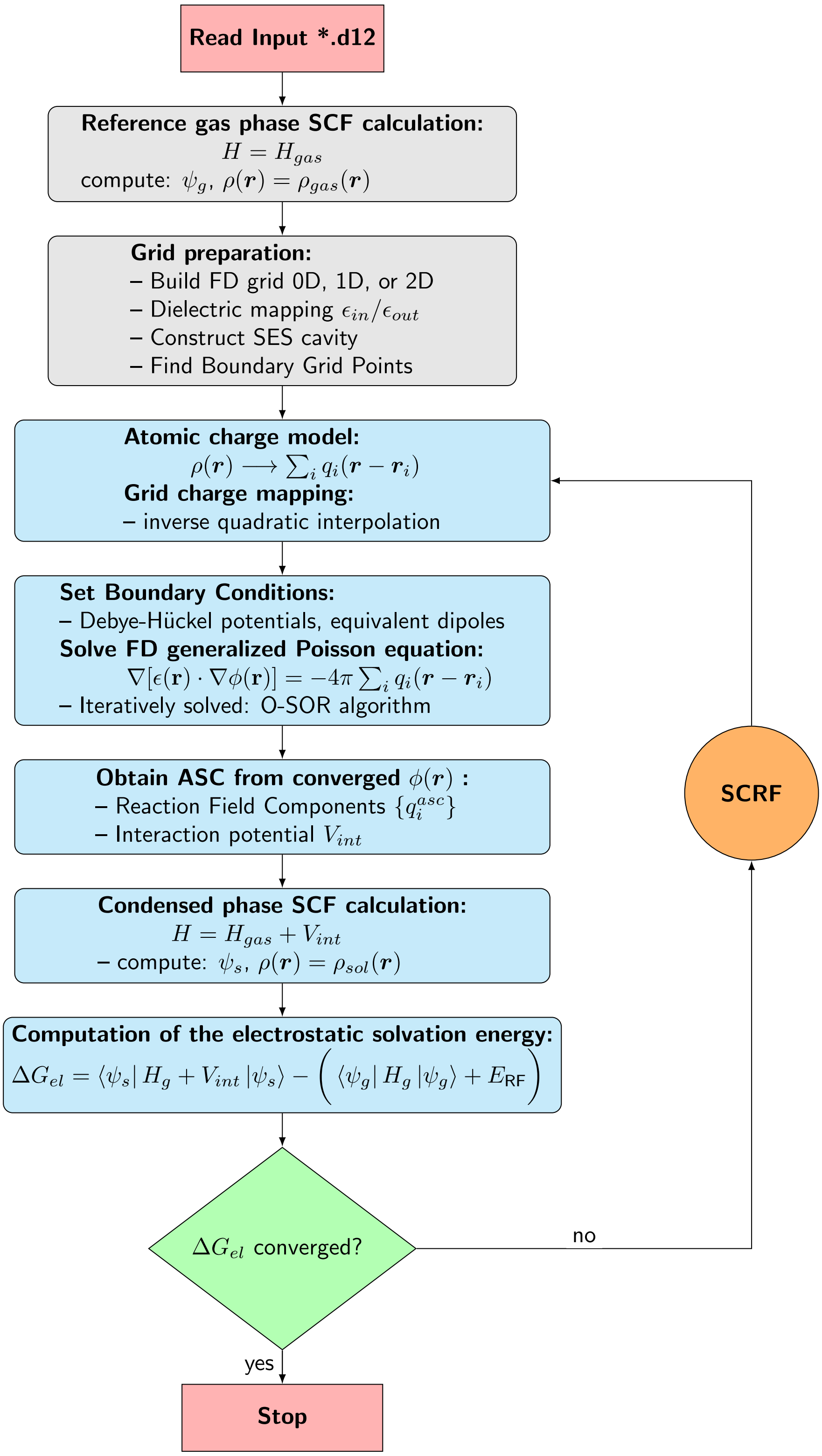

Read Input *.d12
Reference gas phase SCF calculation:
$H = H_{gas}$
compute: ψ_g, $\rho(\boldsymbol{r}) = \rho_{gas}(\boldsymbol{r})$
Grid preparation:
– Build FD grid 0D, 1D, or 2D
– Dielectric mapping $\epsilon_{in}/\epsilon_{out}$
– Construct SES cavity
– Find Boundary Grid Points
Atomic charge model:
$\rho(\boldsymbol{r}) \longrightarrow \sum_i q_i(\boldsymbol{r} - \boldsymbol{r}_i)$
Grid charge mapping:
– inverse quadratic interpolation
Set Boundary Conditions:
– Debye-Hückel potentials, equivalent dipoles
Solve FD generalized Poisson equation:
$\nabla[\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r})] = -4\pi \sum_i q_i(\boldsymbol{r} - \boldsymbol{r}_i)$
– Iteratively solved: O-SOR algorithm
Obtain ASC from converged $\phi(\boldsymbol{r})$:
– Reaction Field Components $\{q_i^{asc}\}$
– Interaction potential V_{int}
Condensed phase SCF calculation:
$H = H_{gas} + V_{int}$
– compute: ψ_s, $\rho(\boldsymbol{r}) = \rho_{sol}(\boldsymbol{r})$
Computation of the electrostatic solvation energy:
$\Delta G_{el} = \langle\psi_s| H_g + V_{int} |\psi_s\rangle - \left(\langle\psi_g| H_g |\psi_g\rangle + E_{\text{RF}}\right)$
ΔG_{el} converged?
no
yes
Stop
SCRF

Figure 3.7: Flow chart of the SCRF procedure in CRYSTAL, in light blue the steps iteratively repeated during the SCRF.

3.2 Non-Electrostatic Contribution

In this section, we will discuss the main non-electrostatic models used throughout this thesis to account for the remaining solvation effects which are not considered by the electrostatic model, such as cavitation, and other effects found mainly in the first solvation shell, such as dispersion and repulsion [23, 25, 24, 26].

These non-electrostatic models are based on the total molecular Solvent-Accessible Surface Area (SASA) of the solute A_m, which is given by the sum of atomic SASA contributions A_i, according to:

$$A(R_i; r_p) = \sum_{i}^{atoms} A_i(R_i + r_p) \tag{3.31}$$

where R_i are the radii of atom centered spheres, and r_p is the radius of a spherical solvent probe, as described in Sec. 2.2.3 on page 34. The atomic SASA contributions are obtained through an analytical stereographic projection approach which will be covered in greater detail in chapter 4 on page 75.

The first non-electrostatic model implemented considered the non-electrostatic contribution simply proportional to the total molecular SASA according to:

$$\Delta G_{ne} = \alpha \cdot A_M + \beta \tag{3.32}$$

where the line coefficients α and β are obtained from a least-square fit of the experimental solvation free energy of hydrocarbons in water against the total molecular SASA [167, 168]. This model, albeit conceptually straightforward, lacks the versatility required to treat more complex systems for which it was not parametrized. In fact, the parameters have to be obtained for each solvent, for which limited or no experimental data might exist. In addition, within this model the ΔG_{ne} is always a positive contribution to the total solvation energy, which might not always be true. As an example, one can think of a benzene solute within a non-polar solvent, where the dominant contribution to the total solvation energy are due to stabilizing non-electrostatic effects. Consequentially, the Cavity, Dispersion, and Solvent structural effects model (CDS), developed by Marenich et al. [81, 80, 83], has been implemented due to its versatility in accounting for different solvents for which the model was not originally parametrized.

3.2.1 Cavity, Dispersion, and Solvent Structural Effects Model

For reliable calculation of solutes in solutions, the non-electrostatic contribution to the total solvation free energy should also account for any inexactness and uncertainties of the electrostatic model. The CDS model takes into account both the non-electrostatics effects and, through its semi-empirical parametrization, for any shortcoming withing the electrostatic model itself.

As previously mentioned, since the non-electrostatic effects are mainly associated with the first solvation shell the CDS model acounts for the work to create the solute cavity, intermolecular dispersion effects, and other explicit solvation effects like hydrogen bonding, exchange repulsion, dielectric saturation and electrostriction effects, through a weighted contribution of both atomic A_i and molecular A_m SASA, according to:

$$\Delta G_{ne}^{\mathrm{CDS}} = \sum_{i}^{atoms} \sigma_i A_i + \sigma_M A_M, \tag{3.33}$$

where atomic σ_i and molecular weights σ_M are surface tensions, unit wise, and are dependent on solvent empirical descriptors and atomic-number-specific parameters, as we will soon see.

The atomic surface tensions σ_i in Eq. 3.33 are given by the following expression:

$$\sigma_i = \left(\tilde{\sigma}^n_{Z_i} n + \tilde{\sigma}^\alpha_{Z_i} \alpha + \tilde{\sigma}^\beta_{Z_i} \beta\right) + \sum_{j \neq i}^{atoms} \left(\tilde{\sigma}^n_{Z_i Z_j} n + \tilde{\sigma}^\alpha_{Z_i Z_j} \alpha + \tilde{\sigma}^\beta_{Z_i Z_j} \beta\right) T(\{Z_j, R_{ij}\}), \tag{3.34}$$

and are function of three solvent empirical descriptors, namely: the refractive index n (at 293K), as well as Abraham's hydrogen acidity α [311, 312, 313] and basicity parameters β [314]. The former descriptor is related to the polarizability of the solvent, while the latter two are introduced in order to describe the tendency of the solvent to act as a hydrogen bond donor or acceptor, respectively. The first term in Eq. 3.34 can be seen as an atomic number dependent term as the parameters $\sigma^n_{Z_i}$, $\sigma^\alpha_{Z_i}$, and $\sigma^\beta_{Z_i}$, multiplying the empirical descriptors, depend uniquely on the atomic number Z_i of atom i. Instead, the second term depends on atomic pair parameters $\sigma^n_{Z_i Z_j}$, $\sigma^\alpha_{Z_i Z_j}$, and $\sigma^\beta_{Z_i Z_j}$, and a cut-off function $T(R_{ij}, r_{Z_i Z_j}, \Delta r_{Z_i Z_j})$, defined as:

$$T(R_{ij}, r_{Z_i Z_j}, \Delta r_{Z_i Z_j}) = \begin{cases} exp\left(\dfrac{\Delta r_{Z_i Z_j}}{R_{ij} + r_{Z_i Z_j} - \Delta r_{Z_i Z_j}}\right) & \text{if } R_{ij} < r_{Z_i Z_j} - \Delta r_{Z_i Z_j} \\ 0 & \text{otherwise,} \end{cases} \tag{3.35}$$

which accounts for atom types based on pair atomic number Z parameters $r_{Z_i Z_j}$, $\Delta r_{Z_i Z_j}$, and the distance R_{ij} between atoms i and j.

On the other hand, the molecular surface tension σ_M is independent from the solute's chemical composition, and is given by the sum of four empirical solvent descriptors and parameters $\tilde{\sigma}$ according to:

$$\sigma_M = \tilde{\sigma}^\gamma \gamma + \tilde{\sigma}^{\phi^2} \phi^2 + \tilde{\sigma}^{\psi^2} \psi^2 + \tilde{\sigma}^{\beta^2} \beta^2 \tag{3.36}$$

where γ is the macroscopic surface tension of the solvent, ϕ^2 represents the square of the fraction of non-hydrogen solvent atoms that are aromatic carbon atoms, ψ^2 is the square of the fraction of non-hydrogen solvent atoms that are F, Cl or Br, and β^2 is the square of Abraham's basicity parameter of the solvent[314].

Overall, the CDS model is highly parametrized, with 4 global parameters within the molecular surface tension σ_M term and up to 3 parameters for each atomic surface tension σ_i. This number that increases to 4 when considering water; as this solvent has a special set of parameters due to its importance as a solvent in biological processes, and its particular behavior compared to other typical organic solvents which makes the prediction of hydration energies particularly complicated. This means that for N atoms, or atom types, a maximum of $(4 \cdot N + 4)$ parameters are needed for this model. These parameters are only partially transferable between different families of electrostatic models as they account for any shortcoming within the electrostatic model itself. As such, a new *ad hoc* set of parameters had to be developed for the FD electrostatic model to achieve maximum accuracy, as we will see in chapter 6 on page 107, where the CDS model has been reparametrized and extended for charged species. Nevertheless, once the parameters have been obtained, the CDS model is considered "universal" as is able to accurately predict solvation energies for solvents and solutes for which it was not originally parametrized, given that the empirical descriptor of the solvent are available.

Furthermore, an advantage of the CDS model are the analytical gradients of the non-electrostatic energy ($\frac{\partial \Delta G_{ne}}{\partial x}$, $\frac{\partial \Delta G_{ne}}{\partial y}$, $\frac{\partial \Delta G_{ne}}{\partial z}$), which allow the computation of fast and accurate non-electrostatic forces for use in solute geometry optimization in solvent. In fact, the non-electrostatic force in the CDS model for an atom j in respect to a generic Cartesian coordinate r is given by:

$$F_j^{ne} = -\sum_i \left(\frac{\partial \sigma_i}{\partial r_j} A_i + (\sigma_i + \sigma_M) \frac{\partial A_i}{\partial r_j} \right), \tag{3.37}$$

and depends uniquely on the gradient of the SASA and of the atomic surface tensions σ_i. Analytical derivatives of the cutt-off function $T(R_{ij}, r_{Z_iZ_j}, \Delta r_{Z_iZ_j})$ within each atomic surface tension σ_i are also required, but readily available [315]:

$$\begin{cases} \dfrac{\partial T_{ij}}{\partial r_i} = -T_{ij} \cdot \left(\dfrac{\Delta r_{Z_iZ_j}}{R_{ij} + r_{Z_iZ_j} - \Delta r_{Z_iZ_j}} \right) \cdot \left(\dfrac{r_i - r_j}{R_{ij}} \right) \\ \dfrac{\partial T_{ij}}{\partial r_j} = -T_{ij} \cdot \left(\dfrac{\Delta r_{Z_iZ_j}}{R_{ij} + r_{Z_iZ_j} - \Delta r_{Z_iZ_j}} \right) \cdot \left(\dfrac{r_j - r_i}{R_{ij}} \right), \end{cases} \tag{3.38}$$

where T_{ij} is given by Eq. 3.35. As such, to obtain accurate non-electrostatic forces it is crucial to have accurate SASA gradients computed analytically, as we will see in the next chapter.

Chapter 4

Analytical calculation of the SASA and its nuclear gradients by stereographic projection

The solute-solvent interactions are traditionally partitioned into two contributions summing to the total solvation energy ΔG_{solv}, according to:

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{ne}. \tag{4.1}$$

Electrostatic interactions (ΔG_{el}) are related to the bulk polarization of the solvent due to the solute charge density and are usually computed by solving the Poisson equation in a self-consistent reaction-field (SCRF) approach[289]. On the other hand, since many effects related to the non-electrostatic contribution occur in the first solvation shell, a computationally efficient treatment of non-electrostatic interactions is to relate the non-electrostatic energy (ΔG_{ne}) to the Solvent-Accessible Surface Area (SASA) of the solute, as described in 2.3.4 on page 47, and illustrated in Fig. 4.1.

A further refinement can be made by considering an additional dependence on the total solvent accessible volume for instance[316]. However, although such formulations only require a limited number of solvent-specific parameters, to our knowledge, these are only available for water, making in principle the application of such models limited to aqueous solution. More general and robust treatments of the non-electrostatic contribution have been proposed[252, 251, 253]. For example, the CDS model[83, 81, 80, 82] which introduces empirical atomic and macroscopic surface tensions parameters combined to total and atomic SASA to re-create some of the lost structural information that comes from the assumption that the liquid is a continuum, has been proposed as a general model applicable to any solute in any solvent. In particular, short-range effects dominating in an explicit treatment of the first solvation shell such as hydrogen bonding, solvent structure breaking, exchange repulsion, solute-solvent charge transfer, dielectric saturation and electrostriction effects[81] are all accounted for in this model. Therefore, apart from solvent-specific parameters, the key to the inclusion of the ΔG_{ne} contribution in an implicit solvation model is the evaluation of both the total and atomic SASA, that is the area of an atom that is not buried by the neighboring atoms[167]. When geometry optimization in solvent is needed, SASA nuclear gradients are thus also required.

Computational methods available to estimate the SASA can be classified into approximate and exact methods. Among the approximate methods, both numerical[317, 318, 319, 320, 321] and analytical[322, 187, 323, 324, 325] approaches have been proposed. Exact

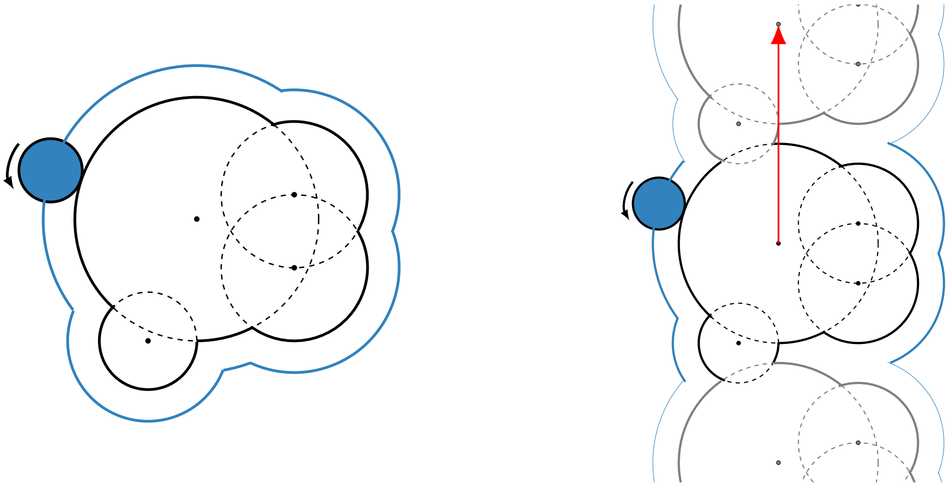

(a) (b)

Figure 4.1: Schematic representation of the van der Waals (in black) and solvent-accessible (in blue) surfaces of a system with four atomic spheres, in the case of (a) a finite system, and (b) a 1D infinite periodic system with a single lattice basis vector (in red). The blue circle is the solvent probe and the black dots represent the centers of the atomic spheres. For the periodic system, the solvent-accessible surface of the central **0**-cell is emphasized with thick lines.

analytical treatments were first published by Connolly[169, 170] and Richmond[172], and subsequent modifications aimed at improving both computational efficiency and robustness of these two strategies.Some of these strategies can also provide nuclear gradients, where the same distinction between approximative [324, 326, 325] and exact [172, 327, 328] analytical approaches applies. Most of the exact analytical strategies mentioned above rely on the general Gauss-Bonnet theorem[329], which requires the definition of the Gauss-Bonnet arcs and related exterior angles together with the full topological analysis of the oriented Gauss-Bonnet paths, and subsequently applies analytical geometry to compute surface areas and, in a limited number of cases, its nuclear gradients. For example, the Analytical Surface Calculation package [330] is one of the reference methods available for computing both the total and atomic SASA (but not its nuclear gradients) based on the work of Connolly[169, 170], with formulas corrected for singularities to make the algorithm robust and efficient.

In this chapter, we present the implementation in the CRYSTAL code of a previously-proposed approach for finite systems[331], which does not rely on the Gauss-Bonnet theorem but instead uses the concept of stereographic projection (SP)[332]. The cases of finite and infinite periodic systems are both considered, and the full derivation of the SASA nuclear gradients is provided since they were not fully-given in the original publication. A simple iterative perturbation scheme of the atomic coordinates is also introduced in order to stabilize the gradients calculation in the case of certain symmetric systems in which the intersection of multiple atomic spheres may lead to gradients instability. The proposed method is therefore generally applicable to any finite or infinite periodic system and only requires the choice of a set of atomic radii along with a solvent probe radius.

In Sec. 4.1, the SP technique for SASA and SASA gradients calculation is presented, along with modifications introduced for periodic systems and details regarding the specific implementation of the algorithm in Crystal. Then, after presenting computational details in Sec. 4.2, results obtained on finite systems of varying sizes as well as on selected examples of infinite periodic systems such as helices and slabs are discussed in Sec. 4.3. Finally, conclusions are drawn in Sec. 4.4.

4.1 Methodology

4.1.1 SASA calculation

We first briefly summarize the main concepts and equations of the SP algorithm for SASA calculation. The key idea is that spherical circles of intersection (COI) resulting from the intersection of atomic spheres in three dimensions are projected as circles on the tangent planes to the south pole of the spheres (see Fig. 4.2). Since SP preserves angles, the surface integrals involved in three-dimensional space for SASA and SASA gradients calculation are transformed into curve integrals on the plane of projection. In the following equations, for periodic systems, index i is used for an atom belonging to the reference **0**-cell, while all other atomic indices refer to the **0**-cell or to a **g**-cell atom, where **g** is a translation vector of the 1D or 2D lattice. We note that, although not presented here, the application of the proposed approach to a 3D lattice is straightforward.

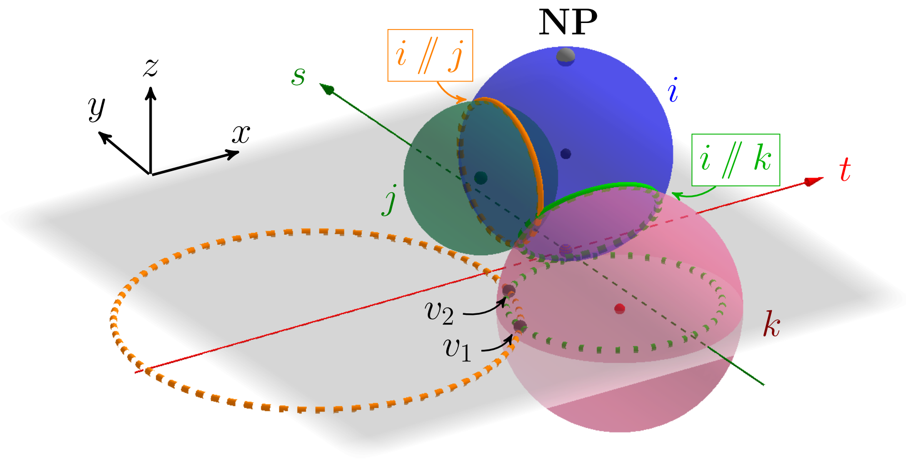


Figure 4.2: Intersection of reference atomic sphere i with neighboring atomic spheres j and k, showing two COI ($i \,/\!\!/\, j$ in solid orange and $i \,/\!\!/\, k$ in solid green), along with their corresponding stereographic projections (dotted line) in the (t, s) tangent projection plane (in grey) to the south pole of the reference sphere i. NP is the North Pole of the reference sphere through which the stereographic projection is performed. The two COI have two intersection points (v_1 and v_2) in the (t, s) plane.

The accessible surface area A_i of sphere i with solvent-expanded radius $r_i = r_{\mathrm{VdW},i} + r_p$, centered at position $\mathbf{r}_i = (x_i, y_i, z_i)$ and intersecting with N_i spheres of indices j can be shown to be[331]:

$$A_i = \chi\left(\Omega_i\right) + \sum_{j \in N_i} \sum_{\lambda=1}^{\Lambda_j^i} I_{j,\lambda}^i \tag{4.2}$$

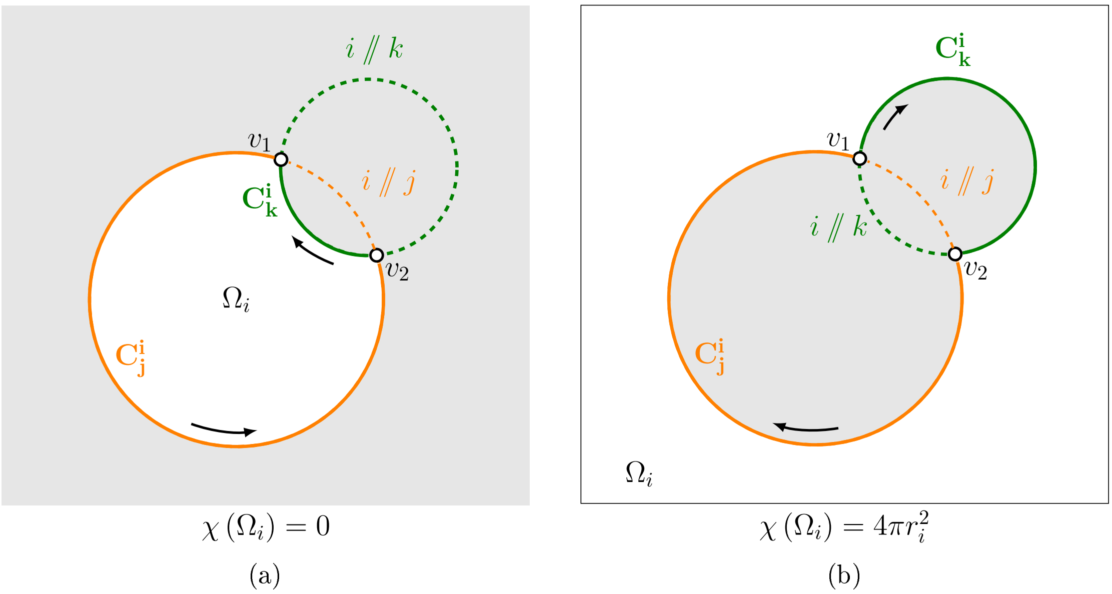


Figure 4.3: Stereographic projections of the two COI ($i \,/\!\!/\, j$ in dashed orange and $i \,/\!\!/\, k$ in dashed green) in the (t, s) tangent plane of the reference sphere i, showing (a) a bounded Ω_i domain made of two circular arcs C^i_j (solid orange, from intersection points v_1 to v_2) and C^i_k (solid green, from v_2 to v_1); (b) an unbounded Ω_i domain with two circular arcs C^i_j (from v_2 to v_1) and C^i_k (from v_1 to v_2). The white section corresponds to the area of the reference sphere i to be computed. The arrows indicate the integration direction along the circular arcs.

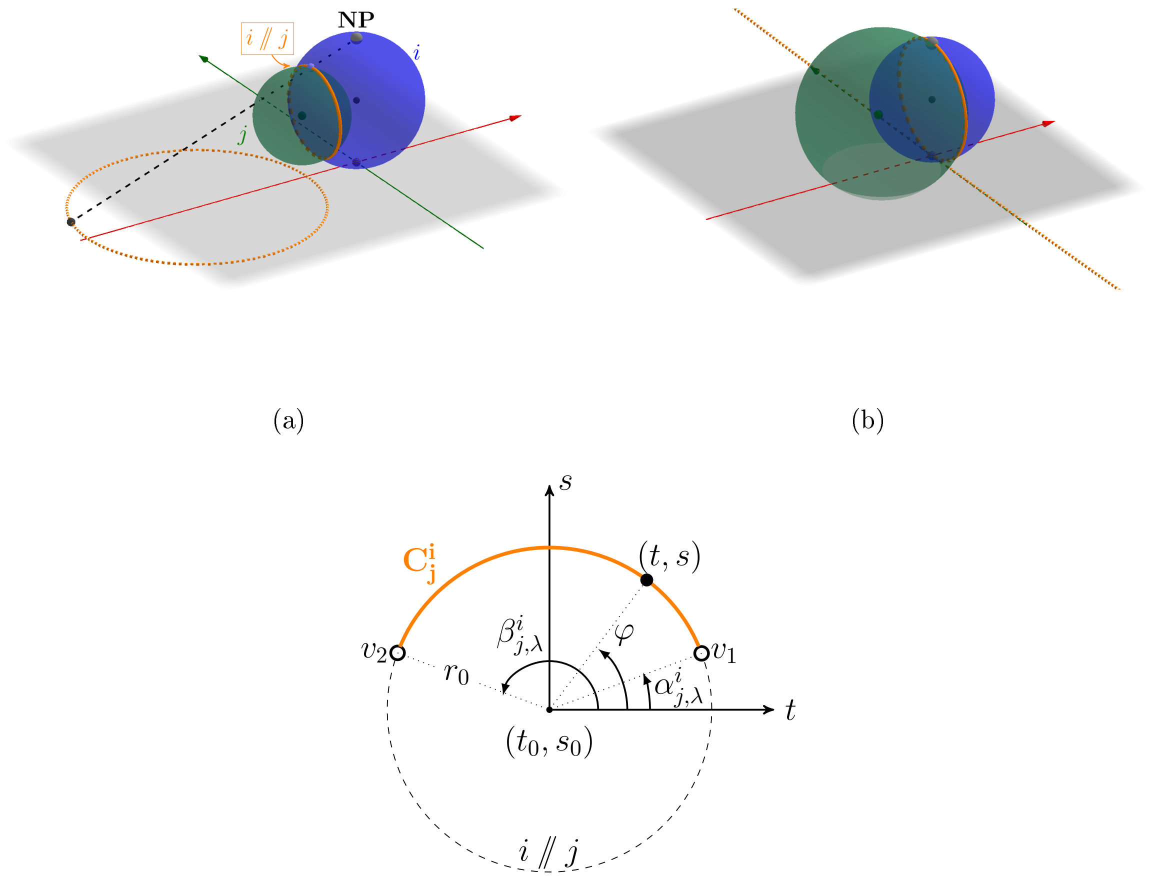
where:

$$\chi(\Omega_i) = \begin{cases} 0 & \text{if } \Omega_i \text{ is bounded} \\ 4\pi r_i^2 & \text{if } \Omega_i \text{ is the whole plane except some regions} \end{cases} \tag{4.3}$$

and Ω_i is the surface region of sphere i whose area is to be computed, N_i is the set of spheres which intersect with the i-th sphere, Λ^i_j is the number of arcs originating from the j-th sphere that participate to the boundary of Ω_i, and $C^i_{j,\lambda}$ is an arc on sphere i originating from sphere j (see Fig. 4.3). The fundamental quantity to compute is then the curve integral $I^i_{j,\lambda}$ whose expression depends on $C^i_{j,\lambda}$ and which can either be positive or negative. It should be noted that, in principle, $C^i_{j,\lambda}$ can be a circular arc or a line, depending on how close the COI is to the north pole (NP) of the reference sphere i (see Fig. 4.4a and 4.4b), leading to different expressions for the curve integral $I^i_{j,\lambda}$.

As previously suggested[331], to avoid multiple cases and possible related numerical instabilities, if at least one sphere surface is found too close to the NP of a reference sphere, a random rotation matrix is applied to the full solute system until a suitable orientation of the solute is found. This allows to reduce the problem to the case of circular arcs $C^i_{j,\lambda}$ only. In the current implementation, this rotation matrix is defined from random values of the two angles φ and γ:

$$R^{\mathrm{rot}} = \begin{pmatrix} \cos\varphi\sin\gamma & -\sin\varphi & \cos\varphi\cos\gamma \\ \sin\varphi\sin\gamma & \cos\varphi & \sin\varphi\cos\gamma \\ -\cos\gamma & 0 & \sin\gamma \end{pmatrix}. \tag{4.4}$$



(a) (b)

(c)

Figure 4.4: Intersection of two spheres i and j results in a COI ($i \mathbin{/\!/} j$, solid orange) which is projected (dotted orange) in the (t, s) tangent plane (in grey) at the south pole of reference sphere i: (a) as a circle when the COI is far enough from the NP of i, (b) as a line when the COI passes through the NP; (c) local polar coordinate system used to define the $\alpha^i_{j,\lambda}$ and $\beta^i_{j,\lambda}$ polar angles with respect to the center (t_0, s_0) of the projected COI $i \mathbin{/\!/} j$ with radius r_0.

In the general case of an arc of circle defined by two vertices with starting and ending polar angles $\alpha^i_{j,\lambda}$ and $\beta^i_{j,\lambda}$, respectively, with respect to the center of the projected COI of i and j (see Fig. 4.4c), it can be shown that:

$$
\begin{aligned}
I^i_{j,\lambda} = r_i^2 \Bigg[\left(\alpha^i_{j,\lambda} - \beta^i_{j,\lambda}\right) \cdot \operatorname{sign}\left(a^i_j\right) + \frac{d^i_j + 4 r_i^2 a^i_j}{V^i_j} \cdot \\
\left(\pi - 2 \arctan \frac{U^i_j}{2 {a^i_j}^2 V^i_j \sin\left(\left(\beta^i_{j,\lambda} - \alpha^i_{j,\lambda}\right)/2\right)} \right) \Bigg]
\end{aligned}
\tag{4.5}
$$

where:

$$\begin{cases} V_j^i = \left[\left(4r_i^2 a_j^i - d_j^i\right)^2 + 4r_i^2\left(b_j^i b_j^i + c_j^i c_j^i\right)\right]^{1/2} \\ a_j^i = (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i + r_i - z_j)^2 - r_j^2 \\ b_j^i = 8r_i^2\,(x_i - x_j) \\ c_j^i = 8r_i^2\,(y_i - y_j) \\ d_j^i = 4r_i^2\left[(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - r_i - z_j)^2 - r_j^2\right] \end{cases} \tag{4.6}$$

and:

$$\begin{aligned} U_j^i = {} & \left|a_j^i\right|\left({b_j^i}^2 + {c_j^i}^2 - 2a_j d_j + 8r_i^2 {a_j^i}^2\right)\cos\left(\left(\beta_{j,\lambda}^i - \alpha_{j,\lambda}^i\right)/2\right) - a_j^i\left({b_j^i}^2 + {c_j^i}^2 - 4a_j^i d_j^i\right)^{1/2} \\ & \cdot\left(b_j^i\cos\left(\left(\beta_{j,\lambda}^i + \alpha_{j,\lambda}^i\right)/2\right) + c_j^i\sin\left(\left(\beta_{j,\lambda}^i + \alpha_{j,\lambda}^i\right)/2\right)\right) \end{aligned} \tag{4.7}$$

When $C_{j,\lambda}^i$ is a full circle with no intersection with other projected COI, this expression reduces to:

$$I_{j,\lambda}^i = 2r_i^2\pi \cdot \left(-\operatorname{sign}\left(a_j^i\right) + \frac{d_j^i + 4r_i^2 a_j^i}{V_j^i}\right). \tag{4.8}$$

While the evaluation of $I_{j,\lambda}^i$ can therefore be directly made from Eq. (4.8) and (4.6) when $C_{j,\lambda}^i$ is a full circle, the case of an arc of circle requires the additional determination of the polar angles $\alpha_{j,\lambda}^i$ and $\beta_{j,\lambda}^i$ of the two arc vertices v_1 and v_2 with coordinates $(t_{j,k,v_1}^i, s_{j,k,v_1}^i)$ and $(t_{j,k,v_2}^i, s_{j,k,v_2}^i)$ in the (t, s) tangent plane. If these vertices are formed from the intersection of the projected $i \mathbin{/\!\!/} j$ and $i \mathbin{/\!\!/} k$ COI, then:

$$\begin{cases} t_{j,k,v_1} = \dfrac{-2\left(b_j^* - b_k^*\right)\left(d_j^* - d_k^*\right) - \left(b_k^* c_j^* - b_j^* c_k^*\right)\left(c_j^* - c_k^*\right) + \left(c_j^* - c_k^*\right) D_{j,k}^{1/2}}{2\left(\left(b_j^* - b_k^*\right)^2 + \left(c_j^* - c_k^*\right)^2\right)} \\ s_{j,k,v_1} = \dfrac{-2\left(c_j^* - c_k^*\right)\left(d_j^* - d_k^*\right) - \left(c_k^* b_j^* - c_j^* b_k^*\right)\left(b_j^* - b_k^*\right) - \left(b_j^* - b_k^*\right) D_{j,k}^{1/2}}{2\left(\left(b_j^* - b_k^*\right)^2 + \left(c_j^* - c_k^*\right)^2\right)} \\ t_{j,k,v_2} = \dfrac{-2\left(b_j^* - b_k^*\right)\left(d_j^* - d_k^*\right) - \left(b_k^* c_j^* - b_j^* c_k^*\right)\left(c_j^* - c_k^*\right) - \left(c_j^* - c_k^*\right) D_{j,k}^{1/2}}{2\left(\left(b_j^* - b_k^*\right)^2 + \left(c_j^* - c_k^*\right)^2\right)} \\ s_{j,k,v_2} = \dfrac{-2\left(c_j^* - c_k^*\right)\left(d_j^* - d_k^*\right) - \left(c_k^* b_j^* - c_j^* b_k^*\right)\left(b_j^* - b_k^*\right) + \left(b_j^* - b_k^*\right) D_{j,k}^{1/2}}{2\left(\left(b_j^* - b_k^*\right)^2 + \left(c_j^* - c_k^*\right)^2\right)} \end{cases} \tag{4.9}$$

where:

$$D_{j,k} = 4\left(b_k^* d_j^* - b_j^* d_k^*\right)\left(b_j^* - b_k^*\right) + 4\left(c_k^* d_j^* - c_j^* d_k^*\right)\left(c_j^* - c_k^*\right) - 4\left(d_j^* - d_k^*\right)^2 + \left(c_i^* b_j^* - c_j^* b_k^*\right)^2 \tag{4.10}$$

and: $b_j^* = b_j^i/a_j^i, c_j^* = c_j^i/a_j^i, d_j^* = d_j^i/a_j^i, b_k^* = b_k^i/a_k^i, c_k^* = c_k^i/a_k^i,\ d_k^* = d_k^i/a_k^i$.
The polar angles $\alpha_{j,\lambda}^i$ and $\beta_{j,\lambda}^i$ can finally be obtained from the following relations:

$$\begin{cases} \alpha_{j,\lambda}^i = \operatorname{atan2}\left(\dfrac{2a_j^i \cdot s_{j,k,v_1} + c_j^i}{2a_j^i \cdot t_{j,k,v_1} + b_j^i}\right) \\ \beta_{j,\lambda}^i = \operatorname{atan2}\left(\dfrac{2a_j^i \cdot s_{j,k,v_2} + c_j^i}{2a_j^i \cdot t_{j,k,v_2} + b_j^i}\right) \end{cases} \tag{4.11}$$

making the application of Eq. (4.5) possible.

4.1.2 SASA nuclear gradients calculation

By application of the chain rule, the derivatives of the SASA of sphere i with respect to the cartesian coordinates $\mathbf{r}_m$ of an arbitrary sphere m can be written as:

$$\begin{aligned}\frac{\partial A_i}{\partial \mathbf{r}_m} &= \frac{\partial A_i}{\partial\left(a_j^i, b_j^i, c_j^i, d_j^i\right)} \cdot \frac{\partial\left(a_j^i, b_j^i, c_j^i, d_j^i\right)}{\partial \mathbf{r}_m} \\ &= \frac{\partial A_i}{\partial a_j^i} \cdot \frac{\partial a_j^i}{\partial \mathbf{r}_m} + \frac{\partial A_i}{\partial b_j^i} \cdot \frac{\partial b_j^i}{\partial \mathbf{r}_m} + \frac{\partial A_i}{\partial c_j^i} \cdot \frac{\partial c_j^i}{\partial \mathbf{r}_m} + \frac{\partial A_i}{\partial d_j^i} \cdot \frac{\partial d_j^i}{\partial \mathbf{r}_m}\end{aligned} \tag{4.12}$$

where both i and m belong to the **0**-cell.

When C_j^i is a full circle without any intersection with other COI ($\lambda = 1$):

$$\frac{\partial A_i}{\partial\left(a_j^i, b_j^i, c_j^i, d_j^i\right)} = \frac{\partial I_{j,\lambda}^i}{\partial\left(a_j^i, b_j^i, c_j^i, d_j^i\right)} \tag{4.13}$$

leading to:

$$\frac{\partial A_i}{\partial \mathbf{r}_m} = \frac{\partial I_{j,\lambda}^i}{\partial\left(a_j^i, b_j^i, c_j^i, d_j^i\right)} \cdot \frac{\partial\left(a_j^i, b_j^i, c_j^i, d_j^i\right)}{\partial \mathbf{r}_m} \tag{4.14}$$

On the other hand, when the j-th circle contributes to Λ_j^i arcs on the border of Ω_i then, from Ref. [331]:

$$\frac{\partial A_i}{\partial\left(a_j^i, b_j^i, c_j^i, d_j^i\right)} = \sum_{\lambda=1}^{\Lambda_j^i}\left(\frac{\partial I_{j,\lambda}^i}{\partial\left(a_j^i, b_j^i, c_j^i, d_j^i\right)} + \frac{\partial I_{j,\lambda}^i}{\partial\left(a_k^i, b_k^i, c_k^i, d_k^i\right)} + \frac{\partial I_{j,\lambda}^i}{\partial\left(a_l^i, b_l^i, c_l^i, d_l^i\right)}\right) \tag{4.15}$$

with k and l two intersecting circles, from which one gets:

$$\begin{aligned}\frac{\partial A_i}{\partial \mathbf{r}_m} = \sum_{\lambda=1}^{\Lambda_j^i}\Bigg(&\frac{\partial I_{j,\lambda}^i}{\partial\left(a_j^i, b_j^i, c_j^i, d_j^i\right)} \cdot \frac{\partial\left(a_j^i, b_j^i, c_j^i, d_j^i\right)}{\partial \mathbf{r}_m} + \frac{\partial I_{j,\lambda}^i}{\partial\left(a_k^i, b_k^i, c_k^i, d_k^i\right)} \cdot \frac{\partial\left(a_k^i, b_k^i, c_k^i, d_k^i\right)}{\partial \mathbf{r}_m} + \\ &\frac{\partial I_{j,\lambda}^i}{\partial\left(a_l^i, b_l^i, c_l^i, d_l^i\right)} \cdot \frac{\partial\left(a_l^i, b_l^i, c_l^i, d_l^i\right)}{\partial \mathbf{r}_m}\Bigg)\end{aligned} \tag{4.16}$$

indicating that the evaluation of the SASA gradients requires the evaluation of twelve additional partial derivatives of curve integrals, whose expressions are given in Appendix A on page 145.

Finally, when $m = i$, the "self-term" $\frac{\partial A_i}{\partial \mathbf{r}_i}$ is obtained from:

$$\frac{\partial A_i}{\partial \mathbf{r}_i} = -\sum_{j \in N_i} \frac{\partial A_i}{\partial \mathbf{r}_j} \tag{4.17}$$

The above expressions also require the derivatives of the a, b, c and d parameters with respect to $\mathbf{r}_m$ which can be obtained from Eq.(4.6) as:

$$\frac{\partial\left(a_j^i, b_j^i, c_j^i, d_j^i\right)}{\partial \mathbf{r}_m} = \begin{pmatrix} 2\left(x_j - x_i\right) & 2\left(y_j - y_i\right) & 2\left(z_j - z_i - r_i\right) \\ -8r_i^2 & 0 & 0 \\ 0 & -8r_i^2 & 0 \\ 8r_i^2\left(x_j - x_i\right) & 8r_i^2\left(y_j - y_i\right) & 8r_i^2\left(z_j - z_i + r_i\right) \end{pmatrix} \quad \text{if} \quad \mathrm{m} = \mathrm{j} \tag{4.18}$$

We note that although the SASA is rotationally-invariant, the SASA nuclear gradients depend on the orientation of the solute. When a R^{rot} rotation matrix has been applied to the solute to avoid the case of lines therefore (see Sec. 4.1), the SASA gradients can be obtained from:

$$\frac{\partial A_i}{\partial \mathbf{r}_m} = \left(\frac{\partial A_i}{\partial \mathbf{r}_m}\right)^{\mathrm{rot}} \cdot M^{\mathrm{rot,T}} \tag{4.19}$$

where $\left(\frac{\partial A_i}{\partial \mathbf{r}_m}\right)^{\mathrm{rot}}$ and $M^{\mathrm{rot,T}}$ correspond to the SASA gradients of the rotated solute and the transpose of the rotation matrix applied, respectively.

4.1.3 Algorithm structure

The algorithm proceeds through the steps shown in Fig. 4.5. The calculations of the atomic SASA (Eq. (4.2)) and of the SASA nuclear gradients (Eq. (4.12)) are only performed when no COI are found close to any NP. In the current implementation, a threshold of 10^{-4} Å is used to check for NP position with respect to atomic spheres. A random rotation matrix (see Eq. (4.4)) is applied otherwise until a suitable solute orientation is found. For periodic systems, supercells with 3 and 27 unit cells are built for 1D and 2D periodic systems, respectively, to generate the full list of atomic spheres and to calculate the SASA and its nuclear gradients for the atoms belonging to the **0**-cell only.

Prior to application of the equations detailed in Sec. 4.1.1 and 4.1.2, an analysis of the mutual positions of the reference sphere and of the neighboring spheres is performed according to the procedure described in Ref. [331] in order to avoid trivial cases. To determine the neighbor set of the current sphere and to avoid unnecessary checks between distant atom/atom pairs, the whole procedure is accelerated by using a cubing algorithm. More precisely, a cubic grid is built with a cell length chosen such that all neighbors of a sphere are found by examining the sphere's own cell and all 26 neighboring cells. The cell side length is $2(r_{max} + r_p)$, where r_{max} is the maximum atomic sphere radius, ensuring that all possible neighbors are located in a $(3 \times 3 \times 3)$ sub grid[333] (see Fig. 4.6).

When very small arc lengths are obtained, to avoid numerical instabilities, the coordinates of atoms i affected are iteratively perturbated with the following approach:

$$\mathbf{r}_i = \mathbf{r}_i + n \cdot 10^{-\mathrm{T}+\text{ncyc-1}} \tag{4.20}$$

where n is a random integer belonging to $[-9; +9]$, ncyc is an integer corresponding to the index of the cycle in the iterative loop and $\mathrm{T} = 14$. In the current implementation, this procedure is activated when $\beta_j^i - \alpha_j^i < 10^{-16}$ degrees. This was found to be important in the case of certain symmetric systems. The number of cycles required in such cases depends strongly on the degree of symmetry of the solute, ranging from 1 (for acetone) to 5 (for highly-symmetric fullerenes or nanotubes).

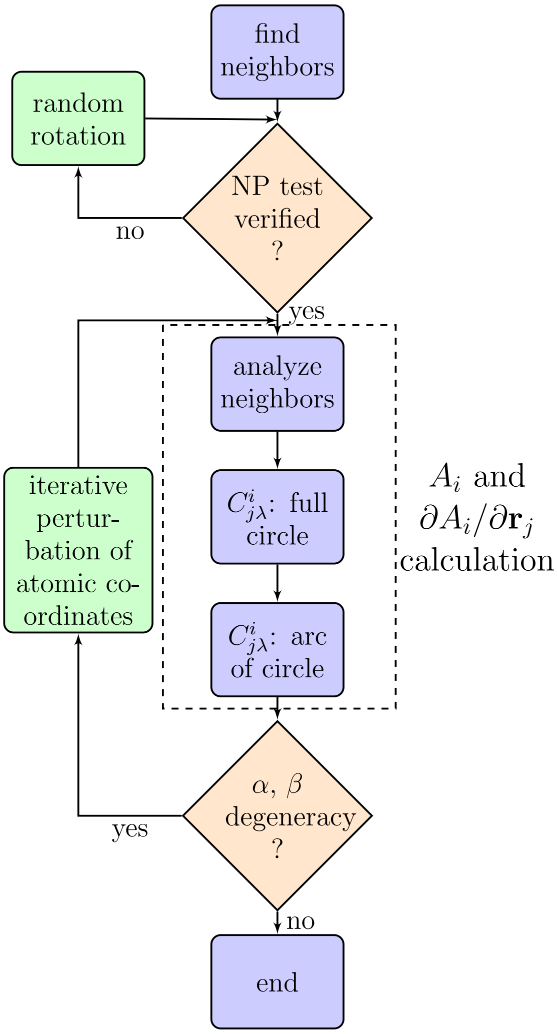

find neighbors
random rotation
NP test verified ?
no
yes
analyze neighbors
$C^i_{j\lambda}$: full circle
$C^i_{j\lambda}$: arc of circle
A_i and $\partial A_i/\partial \mathbf{r}_j$ calculation
iterative perturbation of atomic coordinates
α, β degeneracy ?
yes
no
end

Figure 4.5: Schematic representation of the SP algorithm for the calculation of the SASA and its nuclear gradients.

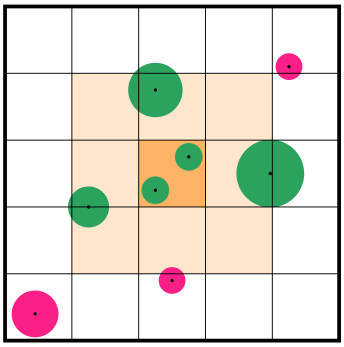


Figure 4.6: 2D view of the neighborhood determination process using a cubing algorithm. For atomic spheres belonging to the dark orange cell, distances are only computed with the green spheres whose centers belong to the dark or light orange grid cells. Red spheres are ignored since they are guaranteed to be out of range.

4.2 Computational details

All calculations have been carried out with a development version of the CRYSTAL code[31, 39], which uses atom centered Gaussian orbitals as a basis set. Unless explicitly stated otherwise, all structures considered have been obtained in gas-phase at the B3LYP/6-311G(d, p) level, considering 75 radial points and 974 angular points for the numerical DFT integration, which ensured an error on the integrated electron density to an accuracy of 10^{-5} e per unit cell or molecule. In this work, SASA calculations have been made using the analytical algorithm presented above, based on the stereographic projection approach originally proposed in Ref. [331], that we refer to as SP. In some cases, the fully-numerical approach of Shrake and Rupley (SR)[317] which has been previously-implemented in the Crystal code[289] and in which N_p equally-spaced points distributed on each solvent-expanded atomic sphere are checked for occlusion by any other atom's accessible surface to determine the atomic and total SASA, as shown in Fig. 4.7, has also been considered. Reference analytical calculations of the total and atomic area have been obtained with analytical surface calculation package [330], which essentially modifies the original Connolly's formulas[169] to improve robustness and stability of the algorithm. Correctness of the analytical gradients has been verified by calculating numerical gradients of the analytical total SASA using a step size $\delta = 10^{-4}$ Å for atomic displacements according to the following equation:

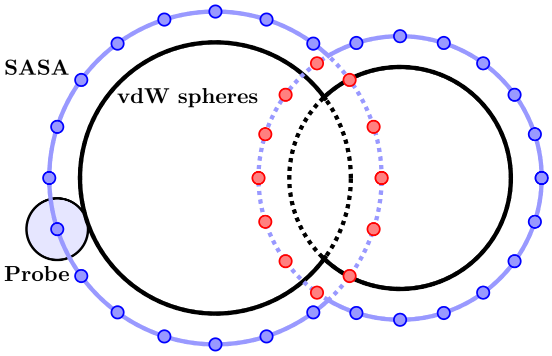


Figure 4.7: Shrake and Rupley algorithm for a system with two van der Waals (vdW) spheres, here represented by black lines. The SASA is represented by blue lines, and blue points on the SASA represent points which do not lie inside any other atom's accessible surface and are therefore considered as accessible to the solvent, while red ones are buried and thus eliminated.

$$\frac{\partial A_i}{\partial \lambda_i} = \frac{A(\lambda_i + \delta) - A(\lambda_i - \delta)}{2\delta} \tag{4.21}$$

where $\lambda_i = x_i, y_i$ or z_i. Bondi atomic radii[156, 157] have been used throughout, along with a solvent probe radius r_p of 0.400 Å corresponding to the solvent probe radius considered in the CDS model[83, 81, 80].

4.3 Results and discussion

We start testing the numerical SR and analytical SP algorithms presented above on the 501 molecules of the Mobley test set[38, 334], with solute sizes ranging from 3 to 34 atoms. We then consider the case of 42 proteins, with sizes reaching several thousands of atoms. Finally, applications to selected periodic systems are presented. When possible, the total and atomic SASA of finite systems obtained are compared to those calculated with analytical surface calculation package[330].

4.3.1 Total and atomic SASA

4.3.1.1 Mobley test set

Fig. 4.8 presents the absolute deviation of the numerically- and analytically-computed total and atomic SASA from the reference analytical values for the 501 molecules of the Mobley test set[38, 334], accounting for 501 total SASA and 8200 atomic SASA values.

It is clear that the analytical algorithm performs much better than the numerical one. For instance, for the total SASA, an average relative unsigned error (RUE) of $7.366e^{-01}$ Å^2 is obtained with the numerical SR algorithm and 500 points per atomic sphere, while a value of $7.583e^{-10}$ Å^2 is obtained with the analytical SP algorithm. When doubling the number of points per atomic sphere, the average RUE of the total SASA is roughly divided by a factor of 2 with the SR algorithm. More precisely, average RUE of $3.935e^{-01}$, $2.350e^{-01}$, $1.582e^{-01}$ and $7.859e^{-02}$ Å^2 are obtained when 1000, 2000, 4000 and 8000 points per atomic sphere are considered, respectively, indicating that only a limited accuracy can be obtained when considering the numerical SR algorithm. The same conclusions hold for the atomic SASA calculation. Although this clearly highlights the correctness of both implementations, it also pinpoints the advantage of using an analytical approach instead of a numerical one when highly-accurate total and atomic SASA are targeted.

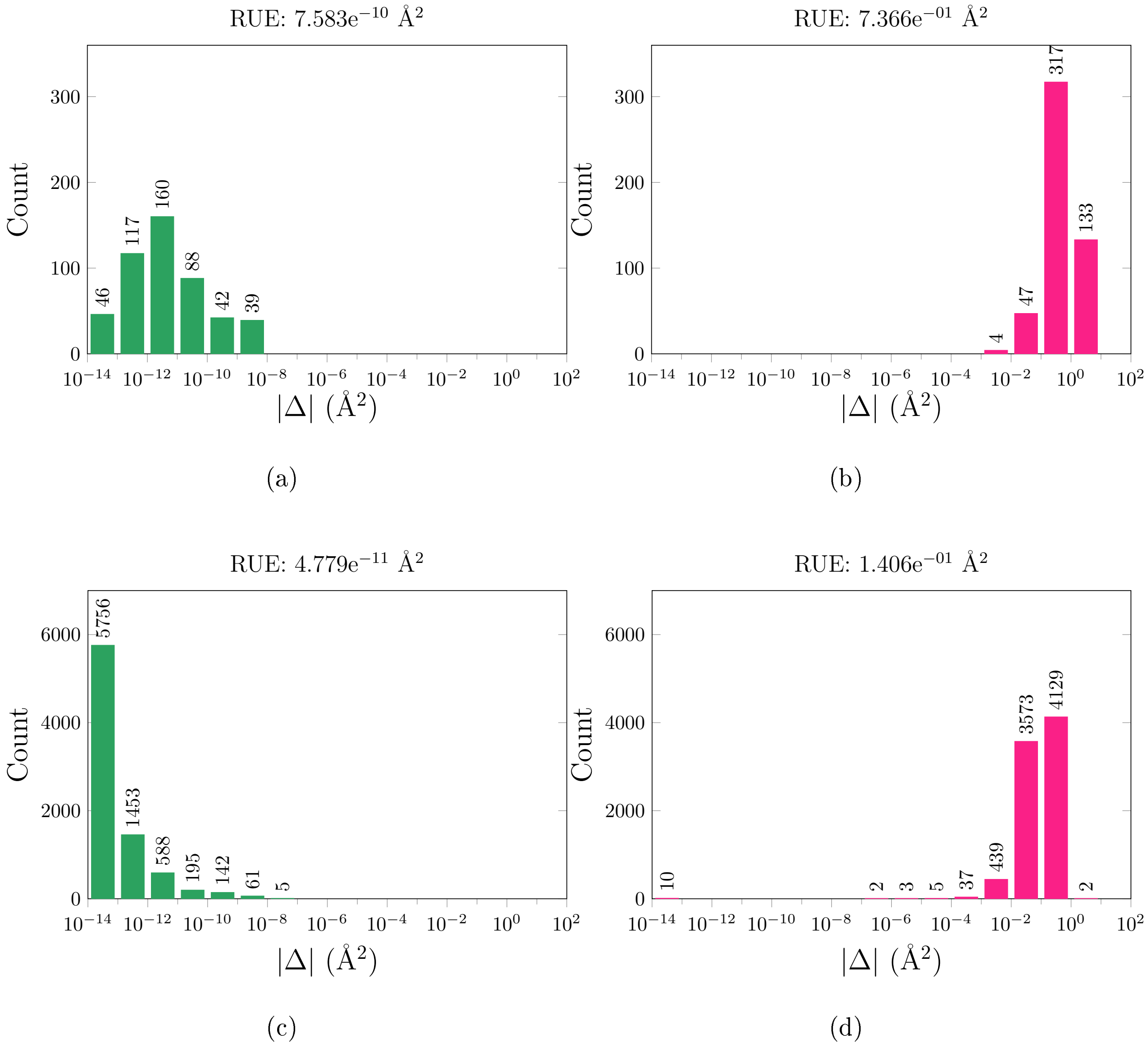
4.3.1.2 42 proteins test set

We now consider the performances of the SP algorithm on larger systems, consisting of a test set of 42 proteins, with sizes ranging from 75 to 11738 atoms. This test set has been considered since it has recently been used to evaluate the TRIFORCE approximate analytical algorithm for SASA and SASA nuclear gradients calculation[335]. All structures have been taken from the RCSB PDB database[336] without further geometry optimization, considering not only “heavy atoms” like previously-done[335], but also hydrogen atoms. Tab. 4.1 presents the average RUE of the atomic SASA for each of the proteins of the test set. An excellent agreement can be seen with the reference values obtained, the average RUE being of $6.264e^{-10}$ Å^2 on the whole test set. Importantly, it is encouraging to notice that computed RUE are very similar on the whole set of proteins tested despite their large size differences, highlighting the robustness of the implementation and the broad applicatibility of the SP algorithm.

4.3.1.3 SASA size-extensivity for periodic systems

When considering periodic systems an important property that should be verified is the size-extensivity of the SASA. More precisely, supercells of increasing sizes can be considered, and the computed SASA can be compared to that of the reference unit cell. For example, for a one dimensional periodic system such as a polymer or an helix, supercells with n unit cells should display a n dependency. Tab. 4.2 presents the ratio of the computed data for different supercells and the reference unit cell of a C_7-13 helix 1D periodic model of glycine, with 49 atoms per unit cell and a lattice parameter of 10.514 Å, a C_7 roto-translational axis and a 13 Å pitch per turn (see Fig. 4.9).

It is clear that the size-extensivity of the SASA is verified, the largest error being lower than 10^{-12} Å^2 among the reported supercells, indicating a precision on the total and atomic SASA for periodic systems largely sufficient for calculating ΔG_{ne}.



(a)

(b)

(c)

(d)

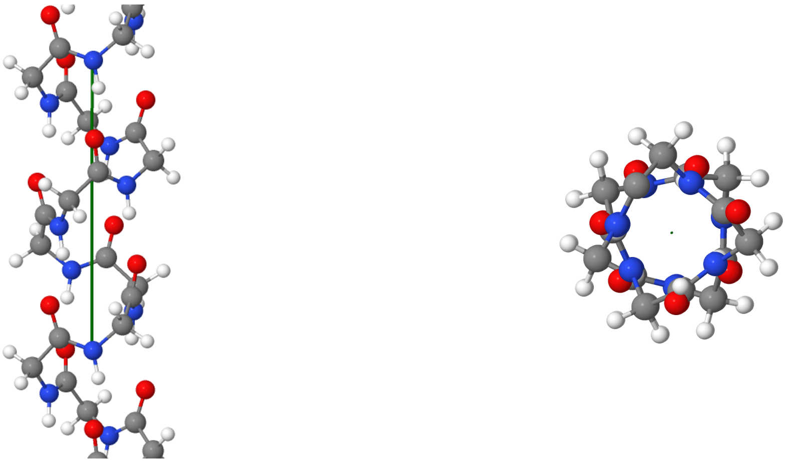
Figure 4.8: Histogram of relative unsigned errors (in Å^2) from reference analytical surface calculation package[330] values of total (top panel) and atomic (bottom panel) SASA, using either the analytical SP (left) or numerical SR (right) algorithms. There are 501 total SASA and 8200 atomic SASA values. For the SR algorithm, 500 points per atomic sphere are used.

4.3.1.4 HCOOH/NiO (100)

Finally, as an example of a 2D slab periodic system, the case of a molecule adsorbed on an oxide surface is now considered. The system consists of a formic acid molecule adsorbed in a bridging bidentate mode on both sides of a (2×3) supercell of the (100)-(2×1) slab unit cell of NiO (100), as a possible anchoring group for dyes in p-type dye-sensitized solar cells applications[337]. Fig. 4.10 shows the optimized PBE0 structure taken from Ref. [338]. As expected only the outermost atoms of the slab contribute to the SASA due to atoms inside the slab having a SASA exactly equal to zero, that is being completely inaccessible to the solvent. In addition, since the top and bottom sides of the slab are symmetrically-equivalent, each side has an identical SASA of 150.804 Å^2, that is half of the total SASA value of 301.608 Å^2 of the slab. The largest contributions to the SASA are given by the non-

Table 4.1: Average RUE (in Å^2) of the atomic SASA calculated with the SP algorithm and compared to the ASC[330] data for the 42 proteins of the TRIFORCE[335] test set. N_{at} is the number of atoms.

PDB code	N_{at}	RUE	PDB code	N_{at}	RUE
1plx	75	$5.202e^{-11}$	6lyz	1102	$3.713e^{-12}$
1cbh	495	$5.716e^{-13}$	1sdf	1124	$7.976e^{-12}$
4pti	514	$1.852e^{-10}$	1lit	1206	$1.045e^{-10}$
1sp2	530	$9.283e^{-11}$	1emr	1231	$1.084e^{-11}$
2cro	537	$7.257e^{-13}$	1mbs	1266	$1.553e^{-11}$
1csp	545	$1.161e^{-11}$	1bj7	1295	$1.527e^{-11}$
1sn3	572	$4.643e^{-12}$	2i1b	1387	$7.285e^{-12}$
1fas	575	$2.492e^{-08}$	1czt	1401	$3.128e^{-11}$
1ubq	660	$4.823e^{-12}$	1j5d	1419	$6.286e^{-12}$
1hip	700	$1.124e^{-11}$	2ptn	1712	$2.905e^{-11}$
1i6f	842	$2.496e^{-11}$	5pad	1715	$6.125e^{-12}$
5rxn	873	$1.284e^{-11}$	1nso	1732	$2.521e^{-12}$
1kte	885	$3.128e^{-12}$	1sur	1739	$9.039e^{-12}$
1opc	890	$2.422e^{-12}$	2hvm	2293	$8.721e^{-12}$
1pht	988	$1.603e^{-11}$	1rhd	2326	$1.400e^{-11}$
1k40	1000	$1.012e^{-11}$	2ts1	2567	$2.916e^{-10}$
2cdv	1020	$6.503e^{-12}$	2cyp	2609	$4.183e^{-11}$
2paz	1021	$5.676e^{-12}$	2tmn	2617	$2.360e^{-11}$
1pdo	1073	$1.933e^{-11}$	1frg	3427	$7.290e^{-11}$
1ooi	1076	$1.112e^{-10}$	1mcp	3544	$1.828e^{-11}$
1fvq	1084	$3.622e^{-12}$	2chn	11738	$8.008e^{-11}$



(a) (b)

Figure 4.9: C$_7$-13 helix model of the poly-glycine 1D periodic system system: (a) side and (b) top views. Red, blue, grey and white balls correspond to O, N, C and H atoms, respectively. The solid green line represents the unit cell.

dissociated H atoms of formic acid on top of the carbon atoms, which are highly-exposed to the solvent.

Table 4.2: Computed total SASA (in Å^2) of the poly-glycine C$_7$-13 helix model with n unit cells and N_{at} atoms per supercell, and RUE when compared to the expected value based on the reference unit cell value. All data obtained with the SP algorithm.

n	N_{at}	SASA	RUE
1	49	328.694	—
2	98	657.388	$6.225e^{-15}$
5	245	1643.472	$2.668e^{-13}$
10	490	3286.944	$6.319e^{-13}$
100	4900	32 869.436	$7.422e^{-14}$

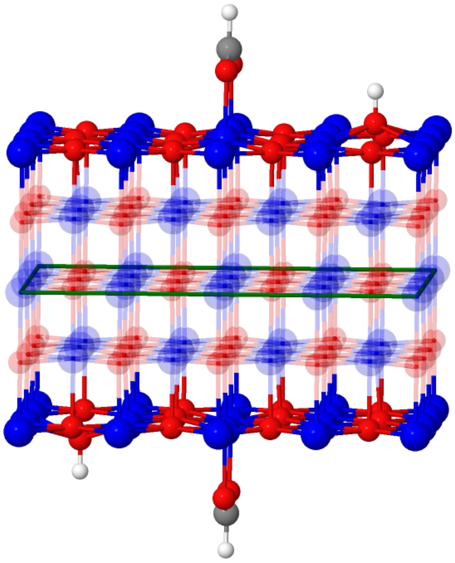


Figure 4.10: HCOOH/NiO (100) 2D slab periodic system. Ni, O, C and H atom as blue, red, grey and white balls, respectively. The solid green line corresponds to the unit cell. The translucent atoms inside the slab have a SASA which is exactly zero.

4.3.2 SASA gradients

4.3.2.1 Mobley test set

Fig. 4.11 presents a comparison between analytically-computed atomic SASA gradients and corresponding numerical gradients, accounting for a total of 24600 gradient components values. With a step size for atomic displacements of 10^{-4} Å an average RUE of $3.250e^{-06}$ Å between analytical and numerical gradients is obtained, outlining the correctness of the analytical gradients calculation.

Additional reliability of the calculated gradients can also be verified by checking their translational invariance: the sums of the derivatives with respect to the same coordinate components should be zero since this corresponds to an overall translation of the solute along the x, y or z axis [325]. Here, for all analytical calculations, the invariance is excellent since these sums are always found lower than $\pm 10^{-14}$ Å. On the other hand, for numerical gradients, values as high as $\pm 10^{-03}$ Å are obtained in some cases, making the comparison between numerical and analytical gradients difficult. Similar conclusions have already been drawn with the popular Gepol approach[325] for instance.

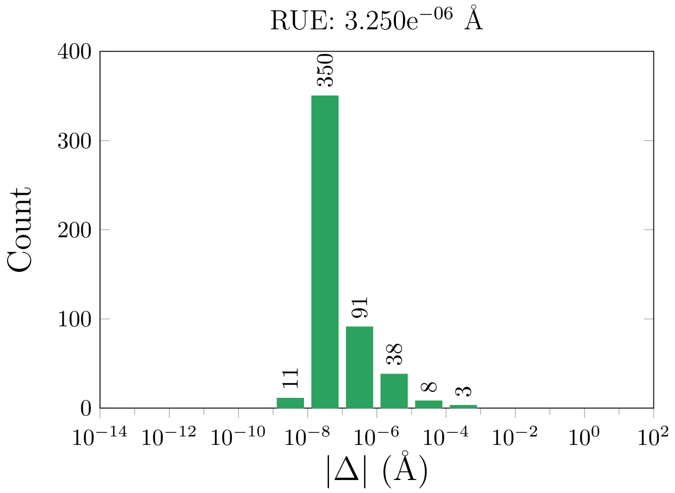
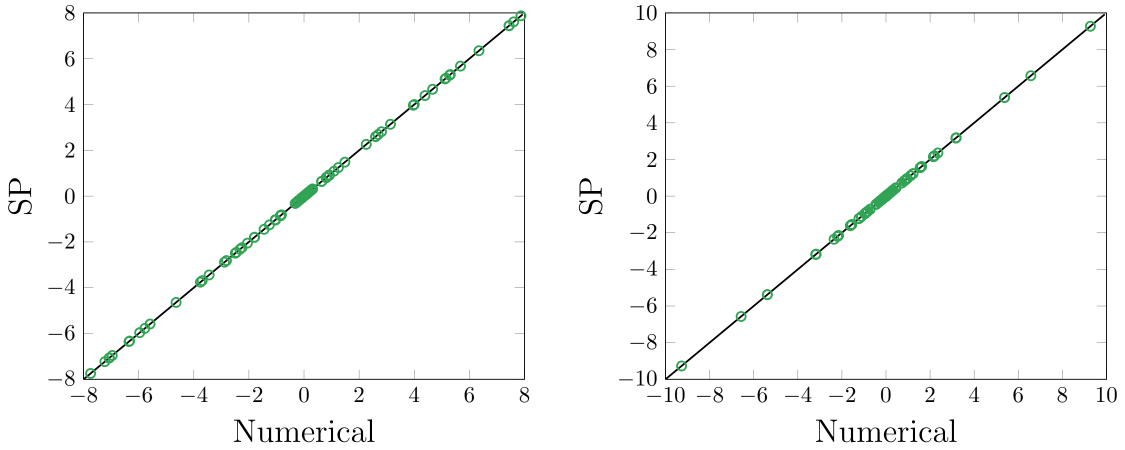


Figure 4.11: Histogram of the relative unsigned error (in Å) per molecule of the analytical nuclear SASA gradients compared to corresponding numerical gradients for the 501 molecules of the Mobley test set.

4.3.2.2 Helix model of poly-glycine and HCOOH/NiO (100)

Fig. 4.12 presents a comparison between analytical and numerical gradients for the periodic poly-glycine C_7-13 helix (Fig. 4.12a) and HCOOH/NiO (100) slab models (Fig. 4.12b) discussed above. Average RUE of $1.074e^{-06}$ and 9.760e^{-09} Å are obtained, respectively.



(a)

(b)

Figure 4.12: Plot of the SASA analytical SP nuclear gradients components as a function of the corresponding numerical gradients for: (a) the 49 atoms of the poly-glycine C_7-13 helix model unit cell; (b) the 130 atoms of the HCOOH/NiO (100) slab unit cell. All data in Å.

The translational invariance of the analytical gradients components is also excellent since the sums of the atomic gradients along both the periodic and the non-periodic directions are found to be lower than $\pm 10^{-11}$ Å in both cases. This clearly indicates that the same numerical precision can be obtained with the SP algorithm for both finite and infinite periodic systems when computing SASA nuclear gradients.

4.4 Conclusions

In this chapter we have presented the implementation of an analytical approach to compute the SASA and its nuclear gradients for both finite and infinite periodic systems in one or two dimensions. The spherical circles of intersection resulting from the intersection of atomic spheres in three dimensions are projected as circles on the projection plane, through stereographic projection. All surface integrals involved in three-dimensional space for SASA and SASA gradients calculation are transformed into curve integrals on the projection plane. An excellent agreement of computed SASA with reference analytical values is found for finite systems, while the SASA size-extensivity is verified for infinite periodic systems. In addition, correctness of the analytical gradients is confirmed by the excellent agreement obtained with numerical gradients and by the translational invariance found, both for finite and infinite periodic systems. Based on the results obtained, the stereographic projection approach appears as a general, simple and efficient technique to compute the basic components needed for the calculation of the non-electrostatic contribution to the solvation energy in an implicit solvation model for both finite and infinite periodic systems.

Chapter 5

Calculating solvation energies with FDPB: effects of the atomic charge and non-electrostatic models

In this chapter we will explore the effects of atomic charge and non-electrostatic models in the prediction of aqueous solvation energies of both neutral and charged molecular solutes.

As we have seen, when solving the generalized Poisson equation the solute charge density can be approximated as atomic point charges, simplifying the electrostatic formalism and avoiding the problem of escaped charge density from the solute cavity, as discussed in Sec 2.2.1.3 on page 30. This is the strategy usually considered in the finite-difference representation of the generalized Poisson equation, especially in the biological field. However, since there is no unique way to obtain atomic charges from a quantum-mechanical calculation, many models have been developed [64, 339] which could potentially be used within an implicit solvation model with various degrees of accuracy. Furthermore, stability with respect to the computational method and the associated basis set are also important criteria to be considered when modeling condensed-phase environments.

In the original implementation [289], the electrostatic contribution contribution to the total solvation energy in Crystal was based on the Class II Mulliken charge model. However, the nonphysical electron partition obtained in case of unbalanced or large basis sets and high dependency on the chosen basis set, have led us to investigate its dependency on the level of theory and accuracy in predicting solvation energies, and to compare the results of five other charge models belonging to the Class II, and cCass IV atomic charge families. More precisely, for Class II charges Hirshfeld [292, 294] and Hirshfeld-I [293, 294] charge models have been considered, while CM5[295] charges alongside two additional modifications have been chosen for Class IV charges. The results are computed both at the HF and B3LYP levels with basis sets of varying quality, on a test set of 501 neutral molecules and 112 charged species in water.

In addition, as the electrostatic contribution is not a direct physical observable, the effect of the non-electrostatic contribution to the total solvation energy cannot be neglected and is also investigated by considering two models of different complexity, both relating the non-electrostatic energy to the SASA [167, 168], computed as discussed in Chapter 4 on page 75. The first model simply relates the non-electrostatic contribution to the total molecular SASA, while the second model is the CDS model, initially proposed in the SMx family of implicit solvation models [82, 83, 81, 80], and based on a weighted contribution of SASA and solvent empirical descriptors, as described in Chapter 3.2.1 on page 71

In Section 5.1, we briefly outline the methods, followed by the computational details in Sec. 5.2, results and discussion in Sec. 5.3, and conclusions are drawn in Sec. 5.4.

5.1 Review of the Methods

In this section, we will briefly outline the methods used to study the effects of atomic charge and non-electrostatic models on the total solvation free energy, or simply solvation energy, of a solute. These methods have been treated in greater detail in Chapter 3 on page 57.

The solvation energy of a solute in a fixed conformation is obtained as a sum of an electrostatic and non-electrostatic contribution, according to:

$$\Delta G_{tot} = \Delta G_{el} + \Delta G_{ne}. \tag{5.1}$$

The non-electrostatic contribution was computed with two different models treated in Sec. 3.2 on page 71, and both based on the SASA of the solute computed using the analytical stereographic projection method discussed previously in Chapter 4 on page 75. The first non-electrostatic model simply considers the ΔG_{ne} contributions proportional to the total molecular SASA A_m, according to:

$$\Delta G_{ne} = \alpha \cdot A_m + \beta, \tag{5.2}$$

where α and β are two parameters obtained from the least-square fit of experimental solvation energies of hydrocarbons in water against the the total molecular SASA (A_m). As this model is rather simple, and parameters are not readily available for every solvent, the more versatile CDS model, discussed in Sec. 3.2.1 on page 71, was here used. Especially in view of further testing with non-aqueous solvents as we will see in Chapter 6 on page 107. Briefly, the CDS model relates the non-electrostatic contribution to a weighted sum of atomic SASA (A_i) and molecular SASA ($A_m = \sum A_i$) of the solute, according to the following set of equations:

$$\begin{cases} \Delta G_{ne}^{\mathrm{CDS}} = \displaystyle\sum_{i}^{atoms} \sigma_i A_i + \sigma_M A_M \\ \sigma_i = \left(\tilde{\sigma}_{z_i}^{n} n + \tilde{\sigma}_{z_i}^{\alpha}\alpha + \tilde{\sigma}_{z_i}^{\beta}\beta\right) + \sum_{j\neq i}^{atoms}\left(\tilde{\sigma}_{z_i z_j}^{n} n + \tilde{\sigma}_{z_i z_j}^{\alpha}\alpha + \tilde{\sigma}_{z_i z_j}^{\beta}\beta\right) T(\{z_j, R_{ij}\}) \\ \sigma_M = \tilde{\sigma}^{\gamma}\gamma + \tilde{\sigma}^{\phi^2}\phi^2 + \tilde{\sigma}^{\psi^2}\psi^2 + \tilde{\sigma}^{\beta^2}\beta^2 \end{cases} \tag{5.3}$$

where σ_i and σ_m are the atomic and molecular surface tensions, unit wise. Both σ_i and σ_m are dependent on a set of solvent specific empirical descriptors $(n, \alpha, \beta, \gamma, \psi^2, \phi^2)$ and parameters. In particular, the atomic surface tensions σ_i depend on solute atomic and atomic pair parameters $\tilde{\sigma}_{Z_i}$ and $\tilde{\sigma}_{Z_i Z_j}$, while the molecular surface tensions σ_m depend on 4 global parameters $\tilde{\sigma}$ independent of the solute. Both empirical descriptors and parameters are treated in Sec. 3.2.1 on page 71. In this chapter the last parametrization for the SM12 implicit solvation model based on the GBA have been used. Since the electrostatic contribution is here described by a PB approach in a FD representation, the transferability of the CDS parameters to another implicit solvation model will also be tested.

On the other hand, the electrostatic contribution ΔG_{el}, accounting for the mutual solute-solvent polarization, is obtained through a SCRF method, as seen in Sec. 3.1.4 on page 68, using an ASC formalism to treat surface polarization effects at the solute-solvent interface.

The ASC are obtained from solving for the potential $\phi(\boldsymbol{r})$ the generalized Poisson equation using a FD scheme, where the solute charge density $\rho(\boldsymbol{r})$, acting as source term, is approximated as atomic point charges q_i:

$$\nabla[\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r})] = -4\pi \sum_{i}^{atoms} q_i \delta(\boldsymbol{r} - \boldsymbol{r}_i). \tag{5.4}$$

The effects of six different charge models belonging to the Class II and IV have then been used, both to validate their accuracy in the prediction of aqueous solvation energies and their dependency on the size of the basis set. In particular, for Class II, which partition the electron density from a quantum-mechanical calculation between atomic contributions, the Mulliken, Hirshfeld (HPA), and Iterative Hirshfeld (HPA-I) have been used. For class IV models, which apply a semi-empirical correction to charge II (or Class III) models to reproduce experimental or calculated observables, the CM5 charge model has been used.

In the molecular formulation of the CM5 charge model, the charges q_i^{CM5} are given by the following expression:

$$q_i^{\mathrm{CM5}} = q_i^{\mathrm{H}} + \sum_{j \neq i} T_{ij} B_{ij} = q_i^{\mathrm{H}} + \sum_{j \neq i} T_{ij} \exp\{-\alpha(R_{ij} - \mathbf{r}_i - \mathbf{r}_j)\} \tag{5.5}$$

where q_i^{HPA} are precursor Class II HPA charges, α and T_{ij} are parameters fitted to reproduces gas phase dipoles, and B_{ij} is related to the Pauling bond border approximating the electron density overlap between two atoms i and j at positions $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$, separated by a distance R_{ij}. Alongside, two additional modification to the CM5 model have also been introduced: (i) HPA charges have been replaced by HPA-I, leading to a variant referred to as CM5-I, and (ii) a scaling of the CM5 charges has been applied leading to a s· CM5 model. The Mulliken, HPA, HPA-I, and CM5 charges been treated in Sec. 3.1.1.1 on page 62 in greater detail.

5.2 Computational Details

All calculations in this chapter have been performed with a modified development version of the Crystal code [31, 39], which uses atom-centered Gaussian orbitals as basis set, in which the CDS non-electrostatic solvation model[80, 81] with the last set of parameters available[81] along with CM5 atomic charges[194] have been implemented. In addition, the original implementation of Hirshfeld and Hirshfeld-I atomic charges for non-periodic and periodic systems [294] has been extended to be used within the SCRF procedure.

Solvation energies have been calculated in water for 501 neutral molecules taken from Mobley's hydration free energy database[38, 334] and 112 charged molecules taken from the MNSOL database, both containing the atomic species H, C, N, O, S, F, P, S, Cl, Br, and I. All solvation energy calculations have been performed with gas-phase geometries obtained at the B3LYP/6-311G(d,p) level for neutral molecules and B3LYP/6-311G++(d,p) for charged ones, and have been verified to be true minima by frequency calculation.

For the electrostatic problem the generalized Poisson equation was solved numerically considering as source term six different charge models: Mulliken, Hirshfeld, Hirshfeld-I, CM5, CM5-I, plus a scaled version of the CM5 charges, s·CM5 that will be discussed below. This was done iteratively with a modified Optimal Successive Over-Relaxation (O-SOR) algorithm [130] and a convergence criteria on the electrostatic potential of 10^{-4} kT at

298.15 K. Dirichlet boundary conditions with Debye Hückel potentials[340] were imposed on the boundaries of the FD grid, using a grid spacing of 0.5 Å, 65 points per grid direction, with the solute filling at most 50% of the grid along each direction. Atomic charges were distributed on the FD grid using an inverse quadratic interpolation algorithm[288].

To reproduce bulk water, the dielectric's and the cavity relative permittivities have been set to 78.3553 and 1.0000, respectively. The SES solute cavity was built from Bondi atomic radii[156, 157] augmented by a solvent probe radius of 1.385 Å. The SCRF convergence criteria was set to $2 \cdot 10^{-5}$ Hartree, corresponding to about 10^{-2} kcal/mol on the solvation energies reported below. Both HF and B3LYP calculations have been carried out, with 6-31G, 6-311G(d,p), and 6-311++G(d,p) basis sets for neutral molecules and only 6-311++G(d,p) for charged molecules. Extra large integration grids with 99 radial points and 1454 angular points have been used for all B3LYP and Hirshfeld-related atomic charges calculations.

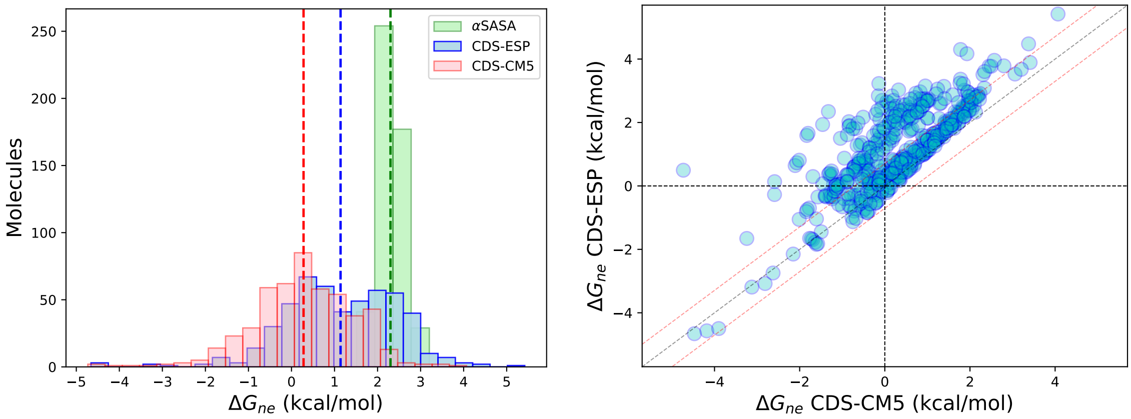
The non-electrostatic contribution to the total solvation energy was calculated through two different approaches described in the Methods section based on analytically-computed total and atomic SASA[341] using Bondi atomic radii[156, 157]: i) the $\propto$SASA model, already used in FD approaches[131, 289], with a solvent probe radius of 1.385 Å, and using as line coefficients the intercept β=0.860 kcal/mol and slope α=0.005 kcal/(mol·Å^2)[131]. ii) the CDS model with a solvent-independent probe radius of 0.400 Å[342], and utilizing two different sets of parameters available from the literature[81], which have been optimized for the electrostatic term of the SM12 solvation model based on the GB approximation[86, 81], both using CM5 and ESP derived atomic charges. Although a FD PB approach is considered here, and although no ESP-derived atomic charges are chosen, for consistency, these parameters have been used as is without further re-optimization and will be referred to as CDS-CM5 and CDS-ESP parameters from here forward.

5.3 Results

In this section, we present the results obtained on a test set of 501 neutral and 112 charged molecules, considering different atomic charge models combined to two different non-electrostatic models, obtained at two levels of theory: B3LYP and HF. In order to highlight the role of each of these points on computed data, we first discuss the effect of the non-electrostatic model on the ΔG_{ne} contribution, we then analyze the impact of atomic charges on ΔG_{el} according to the level of theory and basis set chosen, and both ΔG_{ne} and ΔG_{el} contributions are combined to discuss the effect of all different models and parameters on ΔG_{solv}. Finally, we will discuss the effects of the charge model on the SCRF procedure, and on the computation of the ΔG_{solv} for charged molecules.

5.3.1 ΔG_{ne} contribution

Fig. 5.1a clearly shows the stark differences between the distributions of the ΔG_{ne} values obtained for the test set of 501 neutral molecules in water with the three different non-electrostatic models considered. The $\propto$SASA model results in only positive values, with a distribution sharply centered at 2.30±0.26 kcal/mol, while the CDS-ESP and CDS-CM5 models lead to both positive and negative values, with a distribution loosely centered at 1.14±1.31 kcal/mol and 0.28±1.17 kcal/mol, respectively. In particular, CDS-ESP shows a bimodal distribution compared to CDS-CM5 and $\propto$SASA. This sharp distribution of values with the ∝SASA model can be related to: (i) a single set of α and β parameters used for all solutes with $\alpha \ll \beta$ (see Eq. (2.73)), (ii) the limited sizes of the solutes considered in the chosen test set, with sizes ranging between 3 and 34 atoms, which results in a limited range of total SASA values between 130.5 and 468.5 Å^2. On the other hand, since the CDS model not only accounts for total SASA but also atomic SASA with an additional weighting through geometry-dependent atomic surface tension parameters, a much larger and varied distribution of ΔG_{ne} values is obtained with this model. More precisely, from Fig. 5.1b, noticeable differences between the two sets of parameters available (CDS-ESP and CDS-CM5) in the current SM12 implementation can be evidenced. In particular, the CDS-ESP model generally gives more positive ΔG_{ne} values than the CDS-CM5 one, with differences in magnitude exceeding 1.00 kcal/mol in some cases and a possible sign change in the ΔG_{ne} contribution to the free energy of solvation. When paired with the ΔG_{el} contribution, one can then expect non-negligible differences in the resulting ΔG_{solv}, not only between the ∝SASA and CDS models, but also between the different parametrizations of the latter. This highlights the role of SASA based non-electrostatic models not only as a way to include non-electrostatic effects, but also as a correction of any shortcoming of the electrostatic model itself for which *ad hoc* parameters should be developed.



(a) (b)

Figure 5.1: (a) Comparison of the distributions of ΔG_{ne} values for 501 neutral molecules calculated with the ∝SASA, CDS-CM5 and CDS-ESP models. Dashed lines represent the average values of ΔG_{ne} obtained with the three non-electrostatic solvation models on the whole test set, with values of 2.30±0.26, 1.14±1.31 and 0.28±1.17 kcal/mol for the ∝SASA, CDS-ESP and CDS-CM5 models, respectively; (b) Comparison of computed ΔG_{ne} values obtained with the CDS-CM5 and CDS-ESP models. The dashed red lines represent a ±1 kcal/mol difference.

5.3.2 ΔG_{el} contribution

As mentioned above, several effects can influence the electrostatic contribution to the total free energy of solvation. In this section, we detail the influence of the level of theory and atomic charge model considered. The basis set dependence of the presented data is also quantified.

5.3.2.1 Atomic charge and level of theory effects

Fig. 6.3 first presents the distribution of ΔG_{el} values computed for all 501 neutral molecules with the Mulliken, Hirshfeld, Hirshfeld-I, CM5 and CM5-I atomic charge models, obtained with the 6-311G(d,p) basis set both at the B3LYP and HF levels of theory. We can note that, at the B3LYP level, ΔG_{el} values are generally less negative and less spread out than the ones obtained at the HF level for a given atomic charge model. Indeed, the average of the distributions obtained at the B3LYP level are in kcal/mol: -4.39 ± 3.04 for Mulliken, -3.74 ± 2.48 for Hirshfeld, -5.23 ± 3.21 for Hirshfeld-I, -4.96 ± 2.93 for CM5 and -4.72 ± 3.07 for CM5-I, while the corresponding values at the HF level are: -5.55 ±3.86 for Mulliken, -4.58 ±3.23 for Hirshfeld, -6.47 ±4.02 for Hirshfeld-I, -6.04 ±3.63 for CM5 and -6.04 ±3.85 for CM5-I. We can note that: (i) the average of the Hirshfeld-I distribution is more negatively centered than the Hirshfeld one, qualitatively in line with the overestimation and underestimation of electrostatics with the former and the latter,

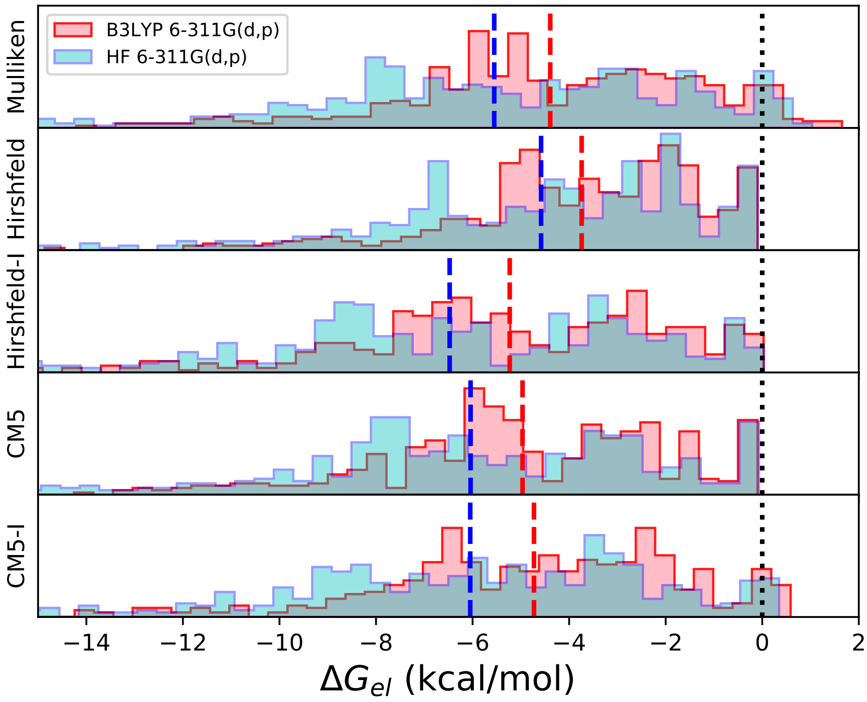


Figure 5.2: Normalized distributions of ΔG_{el} values computed with the Mulliken, Hirshfeld, Hirshfeld-I, CM5 and CM5-I atomic charge models obtained at the B3LYP (in red) and HF (in blue) levels, with the 6-311G(d,p) basis set. Dashed colored lines represent the average of the distributions obtained for the whole 501 molecules test set, while dotted black lines indicate a zero value for ΔG_{el}.

respectively; (ii) the semi-empirical mapping applied to Hirshfeld charges in the CM5 model leads to a description of electrostatics that is in between those obtained with Hirshfeld and Hirshfeld-I. Finally, the overpolarization obtained at the HF level which lead to more negative ΔG_{el} values is in line with the data previously reported on a similar test set of neutral molecules in water[80].

5.3.2.2 Basis set dependence

To assess the stability of ΔG_{el} values with respect to the chosen basis sets, data obtained with the 6-31G, 6-311G(d,p) and 6-311++G(d,p) basis sets are reported in Tab. 5.1, both at the B3LYP and HF levels, as well as in Fig. 5.3. Two criteria have been considered to quantify the basis set dependence of each charge model: (i) the linear regression coefficients of the slope (a), intercept (b), and correlation coefficient (R^2) of the 6-311G(d,p) data with respect to the 6-31G and 6-311++G(d,p) ones, where a fully basis-set independent calculation would then yield: $a = 1.0$, $b = 0.0$, and $R^2 = 1.0$; (ii) the Mean Unsigned Difference (MUD) between the ΔG_{el} data obtained with the 6-311G(d,p) basis set, always with respect to the 6-31G and 6-311++G(d,p) data, defined as:

$$\mathrm{MUD} = \sum_{i}^{N} \frac{|\Delta G_{el,i}^{Basis} - \Delta G_{el,i}^{6-311G^{**}}|}{N} \tag{5.6}$$

where $\Delta G_{el,i}^{Basis}$ is the electrostatic term calculated with the 6-31G or 6-311++G(d,p) basis sets for solute i, and N is the total number of solute molecules in the test set. From the linear regression coefficients and MUD values reported in Tab. 5.1, it is clear that Hirshfeld, Hirshfeld-I, CM5 and CM5-I charges all demonstrate a lower degree of basis set dependence than the Mulliken atomic charge model, with very similar computed data. This conclusion

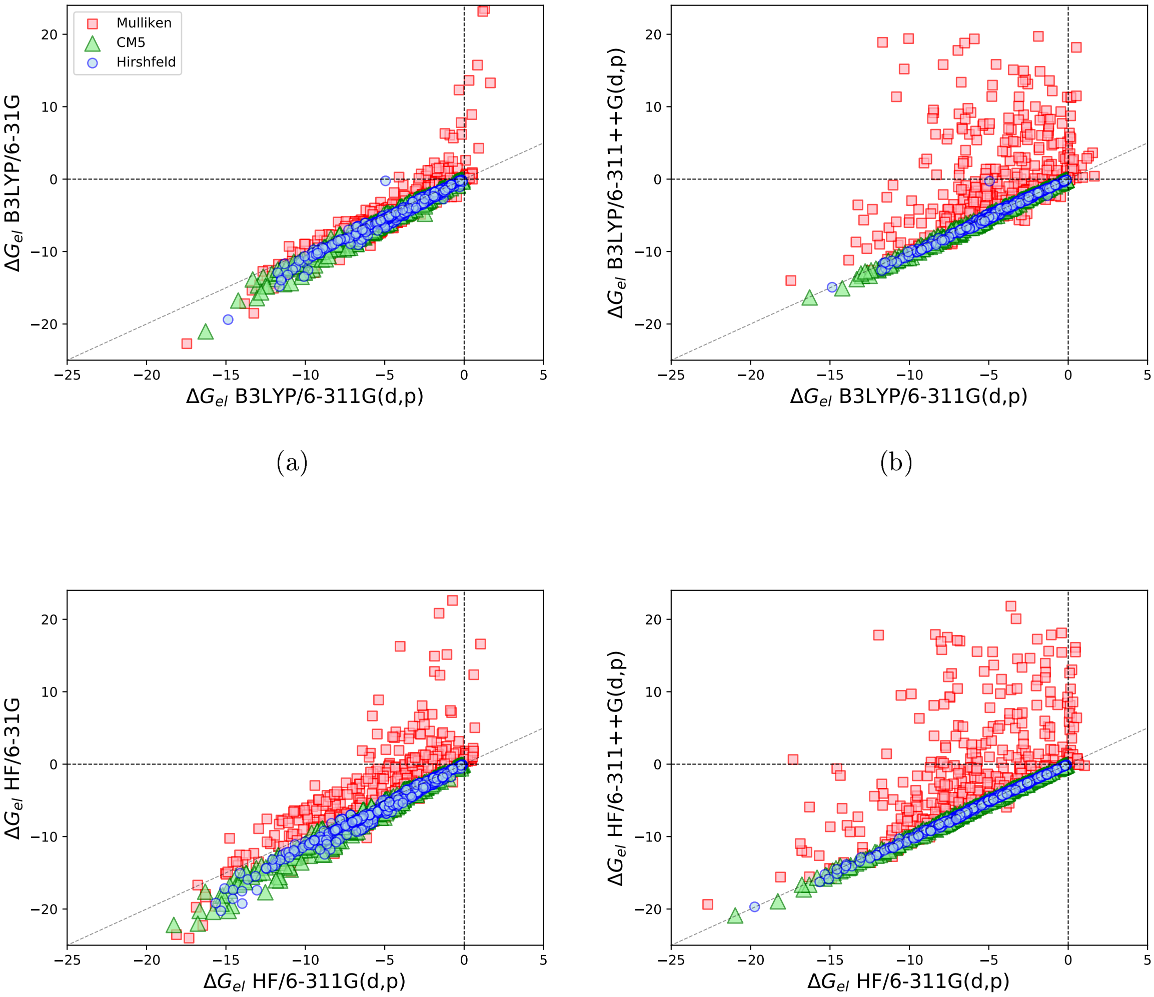
	Mulliken		Hirshfeld		Hirshfeld-I		CM5		CM5-I	
	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF
a	1.343	1.337	1.127	1.184	1.134	1.171	1.175	1.261	1.153	1.201
b	1.633	3.672	-0.017	0.038	0.068	0.089	0.091	0.272	0.111	0.147
R^2	0.864	0.773	0.986	0.990	0.991	0.989	0.987	0.985	0.988	0.987
MUD	1.261	2.535	0.540	0.827	0.702	1.097	0.879	1.467	1.058	1.802

(a) 6-311G(d,p) vs 6-31G data

	Mulliken		Hirshfeld		Hirshfeld-I		CM5		CM5-I	
	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF
a	0.710	0.908	1.049	1.033	1.029	1.052	1.086	1.060	1.039	1.026
b	2.447	3.213	0.047	0.014	0.023	0.005	0.127	0.075	0.040	0.107
R^2	0.710	0.908	0.994	0.999	0.994	0.999	0.995	0.998	0.997	0.998
MUD	3.890	3.922	0.188	0.149	0.216	0.174	0.412	0.351	0.320	0.235

(b) 6-311G(d,p) vs 6-311++G(d,p) data

Table 5.1: Slope coefficient a (kcal/(mol·Å^2)), intercept coefficient b (kcal/mol) and coefficient of determination R^2 obtained from the linear regression of ΔG_{el} values computed with: (a) 6-311G(d,p) and 6-31G; (b): 6-311G(d,p) and 6-311++G(d,p) basis sets. All data obtained on 501 neutral molecules at the both with HF and B3LYP. The corresponding Mean Unsigned Differences (MUD, kcal/mol, see Eq. (6.10)) are also reported.



(a) (b)

(c) (d)

Figure 5.3: Comparison between computed ΔG_{el} values calculated both at the HF and B3LYP levels, with the 6-311G(d,p) basis set and the 6-31G and 6-311++G(d,p) ones. Mulliken (red squares), Hirshfeld (blue circles) and CM5 (green triangles) data are reported.

can be drawn both at the B3LYP and HF levels. As expected, upon going from the smallest 6-31G basis set to the largest 6-311++G(d,p) one the Mulliken atomic charge model leads to unphysical atomic charge values, resulting in a significant increase of the MUD with a value close to 4 kcal/mol. On the other hand, for all other atomic charge models, the MUD decreases to a lower value, ranging between 0.188 and 0.412 kcal/mol at the B3LYP level for instance. This can be clearly evidenced in Fig. 5.3 for three selected atomic charge models. More quantitatively, the degree of basis set dependence evidenced is: Hirshfeld $<$ CM5 $<$ CM5-I $\sim$ Hirshfeld-I $\ll$ Mulliken.

5.3.3 Combined effects on ΔG_{solv}

Tab. 5.2 presents computed MUE values obtained by combining each level of theory (HF and B3LYP) to all atomic charge models considered, along with the three non-electrostatic models, for the full test set of molecules.

		B3LYP				HF			
	ΔG_{ne}	BS1	BS2	BS3	Average	BS1	BS2	BS3	Average
Mulliken	$\propto$SASA	2.02	1.60	4.72	2.72	3.63	1.86	4.51	3.33
Hirshfeld	$\propto$SASA	1.70	1.89	1.85	1.81	2.03	1.91	1.94	1.96
Hirshfeld-I	$\propto$SASA	1.51	1.27	1.43	1.40	2.93	1.81	2.05	2.26
CM5	$\propto$SASA	1.20	1.17	1.22	1.19	2.14	1.48	1.56	1.72
CM5-I	$\propto$SASA	1.57	1.54	1.67	1.59	2.84	1.88	1.97	2.23
Mulliken	CDS-ESP	1.65	**0.97**	3.93	2.18	3.13	1.73	3.99	2.94
Hirshfeld	CDS-ESP	**1.00**	**0.95**	**0.94**	**0.96**	1.73	1.25	1.34	1.44
Hirshfeld-I	CDS-ESP	1.99	1.22	1.54	1.58	3.83	2.42	2.74	2.99
CM5	CDS-ESP	1.58	**1.02**	1.13	1.24	3.04	1.99	2.13	2.36
CM5-I	CDS-ESP	1.73	**1.02**	1.16	1.30	3.75	2.06	2.15	2.65
Mulliken	CDS-CM5	2.24	1.48	3.56	2.42	3.60	2.47	4.03	3.36
Hirshfeld	CDS-CM5	1.36	**1.09**	1.20	1.21	2.36	1.72	1.84	1.97
Hirshfeld-I	CDS-CM5	2.84	2.03	2.36	2.41	4.68	3.26	3.58	3.84
CM5	CDS-CM5	2.40	1.76	1.91	2.02	3.89	2.82	2.97	3.22
CM5-I	CDS-CM5	2.51	1.63	1.82	1.98	4.59	2.86	2.96	3.47

Table 5.2: MUE (in kcal/mol) of the computed total free energy of solvation with respect to the experimental data, both at the HF and B3LYP levels with three basis sets, five atomic charge models and three non-electrostatic models. BS1, BS2 and BS3 correspond to the 6-31G, 6-311G(d,p) and 6-311++G(d,p) basis sets, respectively. Values below the targeted 1.00 kcal/mol threshold are shown in bold.

We can first note that all atomic charge models considered generally better pair with the $\propto$SASA and CDS-ESP non-electrostatic models than with the CDS-CM5 one, especially CM5. This behaviour is mostly due to the more positive distributions of ΔG_{ne} values obtained with the $\propto$SASA and CDS-ESP non-electrostatic models than with the CDS-CM5 one. From Tab. 5.2, by averaging over basis sets, it is noticeable that HF generally tends to give higher MUE compared to B3LYP due to the difference in ΔG_{el} distributions of the two methods, as already discussed in Section 5.3.2. On the other hand, except for the Mulliken atomic charge model, the 6-31G basis set systematically leads to higher errors compared to the 6-311G(d,p) and 6-311++G(d,p) ones, which can be related to the lack of polarization functions in the smallest basis set. The two largest basis sets lead however to very similar results, outlining the high level of basis set independence and the possibility to consider diffuse basis functions with all atomic charge models considered, except for Mulliken as already pointed out before.

Looking at the CM5 charges performances when paired with the CDS-CM5 model, in contrast to previously-reported data obtained with the SM12 solvation model[81], the data obtained is disappointing here, with an average MUE on the three basis sets of 2.02 kcal/mol at the B3LYP level for instance, while it is surprinsingly lower with the CDS-ESP model with a value of 1.24 kcal/mol. This clearly highlights that the assumed transferability of the CDS-ESP and CDS-CM5 parameters from the SM12 solvation model does not hold in the PB framework considered here, due to the difference in the electrostatics treatments of the GB and PB approaches. This conclusion is in line with the reparametrization of the CDS model already proposed in the SMD solvation model [80], when CDS is combined to the electrostatics of the integral equation formalism of PCM[343, 98] for example. In fact, a similar reparametrization of the CDS model should in principle be carried out when this model is combined to electrostatics treatments different from that of the GB approach.

In addition, depending on the non-electrostatic model considered, the use of the iterative variant of Hirshfeld charges might be interesting for implicit solvation modeling. Indeed, upon going from Hirshfeld to Hirshfeld-I charges, a decrease of the MUE is obtained with the $\propto$SASA model while the opposite is true with the two CDS models considered. On the other hand, in all cases, CM5-I charges do not improve upon the CM5 data for condensed-phase modeling.

Finally, we can note that several combinations of level of theory, atomic charge and non-electrostatic models are able to deliver a MUE around the targeted threshold value of 1.00 kcal/mol. Focusing on the B3LYP data for example, Hirshfeld atomic charges are the best performers with the two CDS models considered, while CM5 atomic charges are better paired with the $\propto$SASA model. Since Hirshfeld atomic charges are well-known to underestimate electrostatics as already mentioned previously, this clearly indicates the subtle balance existing between the electrostatic and non-electrostatic contributions to the solvation energy to reach sub 1 kcal/mol accuracy.

5.3.4 Improving CM5 performances: scaling atomic charges

To reach sub 1 kcal/mol accuracy for the test set of neutral molecules without reparametrizing the CDS non-electrostatic model we consider a simpler strategy based on scaling down the CM5 atomic charges, still considering the original ΔG_{ne} contribution from the SM12 parametrization. For each scaling factor s, the ΔG_{el} of the whole set of molecules has been calculated at the B3LYP/6-311G(d,p) level based on CM5 atomic charges, and combined to the three possible ΔG_{ne} values obtained with the $\propto$SASA, CDS-ESP and CDS-CM5 models. The MUE variation as a function of the scaling factor s is shown in Fig. 5.4.

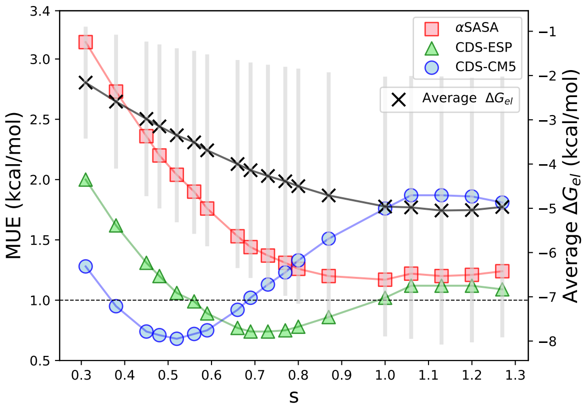


Figure 5.4: MUE of the ΔG_{solv} compared to experimental values and average electrostatic term for 501 neutral molecules against the CM5 charges scaling factor s. The ΔG_{el}, obtained at the B3LYP/ 6-311G level using different scaling factors s, is combined with the ΔG_{ne} obtained with the $\propto$SASA, CDS-ESP and CDS-CM5 models. The dashed line represents the threshold of 1.00 kcal/mol for acceptable MUE. The average ΔG_{el} (solid black line) and the spread of each distribution for each s values (grey vertical lines) are also reported.

From these data, only a shallow minima of 1.17 kcal/mol is recovered around s =1.000 with the ∝SASA model, while two very similar minima of 0.68 and 0.74 kcal/mol are obtained at $s = 0.520$ and $s = 0.730$ with the CDS-CM5 and CDS-ESP models, respectively. Tab. 5.3 compares the unscaled CM5 results with those obtained with the scaled 0.52·CM5 atomic charge model, indicating a significant decrease of the MUE with the scaled charges when combined with the original CDS models, especially with the CDS-CM5 one for all three basis sets considered.

		B3LYP				HF			
	ΔG_{ne}	BS1	BS2	BS3	Average	BS1	BS2	BS3	Average
CM5	∝SASA	1.20	1.17	1.22	1.19	2.14	1.48	1.56	1.72
0.52·CM5	∝SASA	1.73	2.04	1.97	1.91	1.41	1.59	1.55	1.52
CM5	CDS-ESP	1.58	**1.02**	1.13	1.24	3.04	1.99	2.13	2.36
0.52·CM5	CDS-ESP	**0.93**	**1.06**	**1.01**	**1.00**	**0.95**	**0.79**	**0.79**	**0.84**
CM5	CDS-CM5	2.40	1.76	1.91	2.02	3.89	2.82	2.97	3.22
0.52·CM5	CDS-CM5	**0.83**	**0.68**	**0.73**	**0.74**	1.45	**0.98**	**1.04**	1.15

Table 5.3: MUE (in kcal/mol) of the computed total free energy of solvation with respect to the experimental data, both at the HF and B3LYP levels with three basis sets, the CM5 and 0.52·CM5 atomic charge models and three non-electrostatic models. BS1, BS2 and BS3 correspond to the 6-31G, 6-311G(d,p) and 6-311++G(d,p) basis sets, respectively. Values below the targeted 1.00 kcal/mol threshold are shown in bold.

More in details, the behavior of the MUE, and by extension RMSE, can be explained by observing the effects of s on ΔG_{el} and how the different ΔG_{ne} models behave for this set of molecules. Fig. 5.4 shows how increasing the scaling factor in the [0.3;1.0] range lowers the average value of ΔG_{el} and increases the spread of the distribution, while for 1.0<s<1.3 both the average and the distribution of ΔG_{el} plateaus. This indicates that the variation in ΔG_{el} for each molecule decreases by increasing s and, for a given non-electrostatic model, so does ΔG_{solv}. When considering the whole set of molecules this explains the almost constant behavior of the MUE and RMSE for values of s higher than 1.0, instead for values of s lower than 1.0 the variation in MUE can be understood by observing the distributions of ΔG_{ne} values for each model in Fig. 5.1a. Indeed, as already explained in Section 5.3.1, the ∝SASA ΔG_{ne} distribution which is centered at 2.30±0.26 kcal/mol can be in first approximation considered as a constant, while the CDS model parameter sets seem to be constructed to better combine with less negative values of ΔG_{el} than the ones obtained in a PB framework. As a consequence, the CDS-ESP model which leads to more positive contribution of ΔG_{ne} than the CDS-CM5 one has a minimum at higher s value than this latter.

To sum up therefore, the data obtained here indicate that the sub 1 kcal/mol accuracy previously reported in the SM8 and SM12 solvation models[83, 81] for example can be recovered in a PB framework using the original CDS non-electrostatic model by a simple scaling strategy of the atomic charges used in the electrostatics contribution, with computed MUE values in line with those reported previously on similar test sets of neutral molecules[80, 81]. It is also noteworthy that the low basis set dependence pointed out for the CM5 charges in the previous section still holds for the scaled 0.52·CM5 model (see Tab. 5.4).

	6-31G				6-311++G(d,p)			
	CM5		0.52·CM5		CM5		0.52·CM5	
	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF
a	1.175	1.261	1.208	1.301	1.086	1.060	1.064	1.024
b	0.091	0.272	0.014	0.059	0.127	0.075	0.181	0.037
R^2	0.987	0.985	0.976	0.976	0.995	0.998	0.992	0.996
MUD	0.879	1.467	0.410	0.618	0.412	0.351	0.128	0.100

Table 5.4: Slope coefficient a (kcal/(mol·Å^2)), intercept coefficient b (kcal/mol) and coefficient of determination R^2 obtained from the linear regression of ΔG_{el} values computed with the 6-31G and 6-311++G(d,p) basis sets compared to reference 6-311G(d,p) data, for the CM5 and 0.52·CM5 atomic charges. All data obtained on 501 neutral molecules at the HF and B3LYP levels. The corresponding MUD (in kcal/mol, see Eq. (6.10)) are also reported.

5.3.5 Effects of the charge model on the SCRF performance

Fig. 5.5 shows the effects of the charge model on the number of SCRF cycles needed to converge the ΔG_{el} contribution for 501 neutral molecules, both with B3LYP and HF, and considering the 6-31G, 6-311G(d,p) and 6-311++G(d,p) basis sets.

The 0.52·CM5, CM5, Hirshfeld, CM5-I, and Hirshfeld-I charge models perform very similarly between each other at all levels of theory, with the scaled version of the CM5 requiring the lowest number of SCRF cycles for the convergence of the ΔG_{el}. On the other hand, the Mulliken charge model performance is rather low compared to other charge models, showing that the unphysical behavior of the charges not only translates to a low accuracy in predicting solvation energies, but also in overall complications in the convergence procedure, and an increase in computational costs.

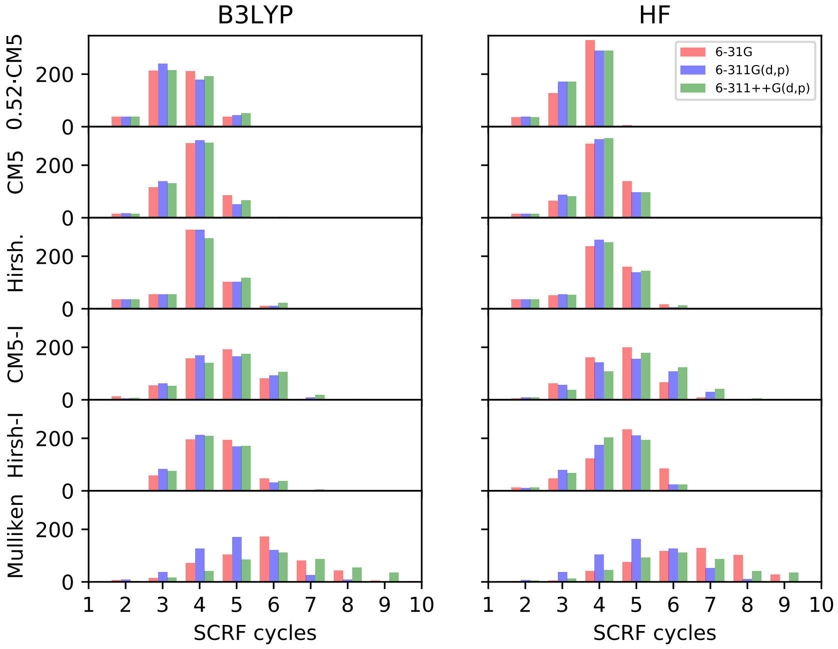


Figure 5.5: Histogram showing the number of SCRF cycles needed for the convergence of the ΔG_{el} with different charge models for 501 neutral molecules in water, at both the B3LYP and HF with the 6-31G, 6-311G(d,p), and 6-311++G(d,p) basis sets.

5.3.6 Effects of the charge model on charged molecules

Tab. 5.5 shows the MUE of 112 charged molecules in water calculated at both the HF and B3LYP levels of theory with the 6-311++G(d,p) basis set for the examined charge models, and the CDS-CM5, CDS-ESP, and αSASA non-electrostatic models. At first glance it is noticeable how the highest MUE are obtained with the Mulliken, and the scaled CM5 charge models for both anions and cations. For the Mulliken charge model this can be attributed to the unphysical behavior already noticed for neutral molecules when using a diffused basis set, while for 0.52·CM5 this can be attributed to the underestimation of the total solute charge due to scaling down atomic charges.

Charge Model	CDS-CM5		CDS-ESP		$\propto$SASA	
	B3LYP	HF	B3LYP	HF	B3LYP	HF
Mulliken	11.13	9.81	12.69	10.78	14.52	12.07
Hirshfeld	6.77	3.46	8.43	4.70	10.39	6.53
CM5	5.02	2.62	6.47	3.24	8.24	4.45
0.52·CM5	18.60	16.05	20.32	17.76	22.34	19.78
Hirshfeld-I	3.70	3.33	4.72	2.77	6.33	3.33
CM5-I	4.19	2.82	5.36	2.67	7.29	3.83

(a) 60 Anions

Charge Model	CDS-CM5		CDS-ESP		$\propto$SASA	
	B3LYP	HF	B3LYP	HF	B3LYP	HF
Mulliken	8.320	9.09	8.912	19.76	9.409	11.34
Hirshfeld	1.893	1.91	2.261	2.29	2.287	2.27
CM5	2.182	2.43	2.121	2.28	1.581	1.59
0.52·CM5	13.941	13.73	14.974	14.76	15.914	15.70
Hirshfeld-I	2.252	2.54	2.160	2.37	1.562	1.55
CM5-I	1.704	2.02	2.268	2.35	2.559	2.26

(b) 52 Cations

Table 5.5: MUE (kcal/mol) of the computed total free energy of solvation with respect to experimental data for 112 charged molecules, both at the HF and B3LYP levels of theory with the 6-311++G(d,p) basis set, with six different atomic charged models and three non-electrostatic models.

Depending on the level of theory the MUE for the remaining charge models is instead closer to the experimental uncertainty of 3.00 kcal/mol for this test set [36], for both anions and cations. In particular, for anions the HF method is able to achieve a lower MUE compared to B3LYP due to overpolarization effects already observed for neutral molecules. The CM5 charge models with CDS-CM5 parameters, and CM5-I with CDS-ESP parameters perform particularly well at the HF level of theory, with a MUE of 2.63 and 2.67 kcal/mol, respectively. These values are on par with recent solvation models especially designed for the prediction of solvation energies of ionic species in water [84]. Instead, at the B3LYP level the lowest MUE observed is 3.70 kcal/mol with Hirshfeld-I and CDS-CM5 parameters, with the remaining charge and non-electrostatic models combinations all achieving MUE above 4.00 kcal/mol. Overall, we can conclude that for anions overpolarization effects and higher charge magnitude achieve better results in contrast to what was seen for neutral systems. Instead, for cations the results, excluding Mulliken and 0.52·CM5 are well within the experimental uncertainty both at the B3LYP and HF levels of theory, with all combinations of charge and non-electrostatic models achieving MUE between 1.50-2.50 kcal/mol.

5.4 Conclusions

In this chapter, we presented the effects of different atomic charge and non-electrostatic models on the solvation free energies of a test set of 501 neutral and 112 charged molecules in water, using an implicit solvation model. The electrostatic contribution to the solvation free energy resulted from a self-consistent reaction field treatment of the bulk electrostatics obtained by solving the Poisson equation in a FD approach where the solute charge density is approximated by atomic charges, while the non-electrostatic contribution accounted interactions between the solute and the solvent dominating in the first solvation shell.

For the electrostatic contribution, five atomic charge models were considered: Mulliken, Hirhsfeld, Hirshfeld-I, CM5, and CM5-I, computed at the HF and B3LYP levels, with the 6-31G, 6-311G(d,p) and 6-311++G(d,p) basis sets. This contribution was combined to two models based on the SASA for the non-electrostatic term: the αSASA model and the CDS one, the latter using two parameter sets obtained from the literature: CDS-CM5, and CDS-ESP.

Based on the atomic charge model and on the levels of theory considered, the distributions of ΔG_{el} values for the test set significantly changes. On average, ΔG_{el} values obtained from the Hirshfeld charge model are lower compared to those obtained with Hirshfeld-I, while Mulliken, CM5 and CM5-I data fall in-between these two charge models. The HF overpolarization globally gives more negative ΔG_{el} distributions when compared to B3LYP, while the basis set dependency of the ΔG_{el} values for each charge model is, from lowest to highest: Hirshfeld $<$ CM5 $<$ CM5-I $<$ Hirshfeld-I $\ll$ Mulliken, the latter giving unphysical solvation energies for larger basis sets including diffuse functions, and issues in the SCRF convergence procedure.

In general, for neutral molecules lower MUE of the total solvation free energy with respect to the experimental data are obtained when the electrostatic contribution of each charge model is paired with the αSASA and CDS-ESP non-electrostatic models than with the CDS-CM5 one. HF overestimates the electrostatics, generally giving higher MUE compared to B3LYP for neutral molecules, and better performances are also obtained with basis sets including polarization functions. Only a limited number of combinations achieved a MUE lower than 1 kcal/mol, most notably the Hirshfeld charge model paired with CDS-ESP non-electrostatic contribution. The poor performances of the CM5 atomic charges obtained here with the chosen approach are in contrast to previously-reported data with the SM12 model, highlighting the low transferability of the CDS parameters from a GB to a PB model in a FD approach without additional reparametrization.

A significant improvement of the performances of the CM5 charges for condensed-phase modeling could however be achieved in the FD strategy chosen by scaling down their values to better balance with the original CDS model with CM5 parametrization. The values of the scaling factor have been obtained by minimization of the MUE obtained at the B3LYP/6-311G(d,p) level on the whole test set of molecules. This has yielded two minima close to the ones reported in the SM8 and SM12 solvation models on a similar test set: 0.68 kcal/mol and 0.74 kcal/mol with scaling factors of 0.52 and 0.73 for the CDS-CM5 and

CDS-ESP models, respectively.

For charged species both Mulliken and the 0.52·CM5 charge models perform poorly for both anions and cations. Instead, the performance of the remaining charge models strongly depends on the level of theory, and if the solute is negatively or positively charged. For anions, the HF overpolarization effects helps to achieve lower MUE compared to B3LYP. In particular the CM5 charge model with CDS-CM5 parameters at the HF level has the lowest MUE of 2.62 kcal/mol, an error within the experimental uncertainty of 3.00 kcal/mol. From the data one can conclude that, both at the HF and B3LYP level of theory, overpolarization effects and charge models with high charge magnitude tend to give better results for negatively charged species. Instead, for cations the MUE is well below 3.00 kcal/mol for all charge and non-electrostatic model combinations, with the exception of Mulliken and 0.52·CM5 charge models.

Overall, this chapter highlights the delicate balance existing between the electrostatic and non-electrostatic contributions to the solvation free energy, and in particular the importance of the development of non-electrostatic models tailored to the chosen electrostatic description for condensed-phase modeling.

Chapter 6

Revised CDS model for FDPB calculations: parametrisation and extension to ionic solutes

In this chapter we mainly focus on the reparametrization of the implicit solvation model implemented in CRYSTAL [289] to both aqueous and non aqueous solvents, and its extension to charged species. In particular, the Cavity, Dispersion, and Solvent structural effects (CDS) model was fully reparametrized taking in account the electrostatic contribution obtained with Hirshfeld (HPA) and CM5 charge models, leading to the development of two separate sets of parameters. The CDS model was then further extended for charged species by using a semi-empirical correction based on Abraham's acidity parameter of the solvent describing the tendency of the solvent to act as a hydrogen bond donor.

The overall performances of the reparametrized CDS model, together with the benefits of the additional corrective term for ions, are tested both on the calculation of solvation energies of neutral and charged species in different solvents, as well as on pKa values in aqueous solution. This last point is particularly challenging for implicit solvation models since it requires to accurately compute solvation energies for both neutral and charged species. Given that experimental solvation energies of the latter are typically one order of magnitude larger than those of the former, this is still very challenging for most implicit solvation models, including widely used models in the quantum chemistry field such as SM12[81] or SMD[80]. We also validate and generalize the reparametrized CDS model by comparison of free energies of solvation, surface energies and band structures of a TiO_2 anatase slab obtained in three different solvents with reference VASPsol [34, 281, 282] data.

This chapter is structured as follows. In Section 6.1, we first present the methods and approaches considered in this work, before introducing the computational details in Section 6.2. Results are then presented and discussed in Section 6.3 and conclusions are finally drawn in Section 6.4.

6.1 Methods

In this section we will present the methods used throughout this chapter. First, briefly outlying the electrostatic ΔG_{el} and non-electrostatic contributions ΔG_{ne} to the total solvation energy:

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{ne}, \tag{6.1}$$

both terms discussed in greater detail in Chapter 4 on page 75, and then discussing the CDS extension for charged species, followed by the reparametrization procedure of the non-electrostatic model.

6.1.1 Electrostatic Contribution

The electrostatic model has been treated in detail in Sec. 3.1 on page 57. Here we remind that the model is based on a SCRF approach where the solute-solvent polarization effects are treated using an ASC formalism, and the underlying electrostatic problem is based on a FD resolution of the generalized Poisson equation (FDPB), with the solute charge density $\rho(\boldsymbol{r})$ approximated by atomic point charges q_i, according to:

$$\nabla[\epsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r})] = -4\pi \sum_{i}^{atoms} q_i \delta(\boldsymbol{r} - \boldsymbol{r}_i). \tag{6.2}$$

As such, based on the conclusions of Chapter 5.4 on page 104, the HPA and CM5 atomic charge models have been chosen to approximate the solute charge density, due to their low basis set dependency and overall acceptable accuracy in predicting solvation energies when coupled with the CDS model using parameters from the literature. These parameters where obtained using GBA electrostatics within the SM12 solvation model [81] using CM5 charges, as well as ChElPG [197] and MK [198] atomic charges fitted to quantum-chemically computed electrostatic potential (ESP). Here, ESP derived charges where not chosen due to complications in their calculation for periodic systems [344], and although strategies have been proposed for such systems [345, 344, 346], they still remain computationally-demanding, especially when applied to a SCRF procedure.

6.1.2 Non-Electrostatic Contribution

The non-electrostatic contribution is based on the CDS model, which has been treated in Sec. 3.2.1 on page 71. Below we will cover the basics of the model which are needed throughout the chapter.

The CDS models relates the non-electrostatic contribution to the total solvation energy ΔG_{ne} as a weighted sum of both atomic SASA (A_i) and molecular SASA ($A_m = \sum A_i$) SASA, according to the following set of equations:

$$\begin{cases} \Delta G_{ne}^{\text{CDS}} = \sum_{i}^{atoms} \sigma_i A_i + \sigma_M A_M \\ \sigma_i = \left(\tilde{\sigma}_{z_i}^{n} n + \tilde{\sigma}_{z_i}^{\alpha} \alpha + \tilde{\sigma}_{z_i}^{\beta} \beta\right) + \sum_{j \neq i}^{atoms} \left(\tilde{\sigma}_{z_i z_j}^{n} n + \tilde{\sigma}_{z_i z_j}^{\alpha} \alpha + \tilde{\sigma}_{z_i z_j}^{\beta} \beta\right) T(\{z_j, R_{ij}\}) \\ \sigma_M = \tilde{\sigma}^{\gamma} \gamma + \tilde{\sigma}^{\phi^2} \phi^2 + \tilde{\sigma}^{\psi^2} \psi^2 + \tilde{\sigma}^{\beta^2} \beta^2 \end{cases} \tag{6.3}$$

where the atomic σ_i and molecular σ_M weights are, unit wise, surface tensions, and both depend on solvent empirical descriptors $(n, \alpha, \beta, \gamma, \psi, \phi)$ which have been described in Sec. 3.2.1 on page 71. In particular, the atomic σ_i surface tension is function of a set of atomic number dependent parameters $\tilde{\sigma}_{z_i}$ as well as atom pair atomic parameters $\tilde{\sigma}_{z_i z_j}$, which are used together with a cutoff function $T(\{z_j, R_{ij}\})$ to defined atom type contributions. The molecular surface tension σ_M is instead independent of the solute and is based on a set of four glabal parameters $\tilde{\sigma}$.

As such, the CDS model requires 4 global parameters $\tilde{\sigma}$ within the molecular surface tensions σ_M term, and up to 3 parameters parameters for each atomic surface tensions σ_i. In practice, this number is increased to 4 as water has its own set of parameters due to its special status as a solvent in the biological field and due to its tendency for hydrogen bonding, which makes the prediction of hydration energies complicated.

As mentioned above, all atomic and molecular surface tension parameters have originally been optimized for the SM12 solvation model [81], which is based on the GBA [86], together with CM5 and ESP-derived atomic charges[197, 198] for neutral species. Since a FD resolution of the generalized Poisson equation is here considered for the electrostatic term, these parameters need to be re-optimized. Here, we limit the parameter optimization of the CDS model to neutral molecules for three main reasons: (i) solvation energies of ions are typically one order of magnitude larger than those of neutral solutes, (ii) the errors in experimental solvation energies measurements are typically of about ±0.1-1.0 kcal/mol for neutrals, while they are of about ±3.0 kcal/mol for ions[36], (iii) there are much fewer available experimental solvation energies for ions than for neutrals, for example, in the MNSOL database[36], almost 90% of solvation energies concern neutral solutes. Thus when optimizing the parameters simultaneously for both types of systems, a non-negligible error is introduced for neutral species. To better describe solvation of charged species, we therefore introduce an additional dedicated corrective term to the CDS model in the next section.

6.1.3 Solvation Energy Correction for Charged Species

The MNSOL database[36] contains the experimental solvation energies of 332 singly-charged species (208 anions and 124 cations) in four different solvents: dimethyl sulfoxide, acetonitrile, methanol, and water. The corresponding averaged experimental solvation energies of these 332 ionic species are reported in Fig. 6.1, as a function of Abraham's hydrogen bond acidity parameter α[311, 312, 313] of these four solvents.

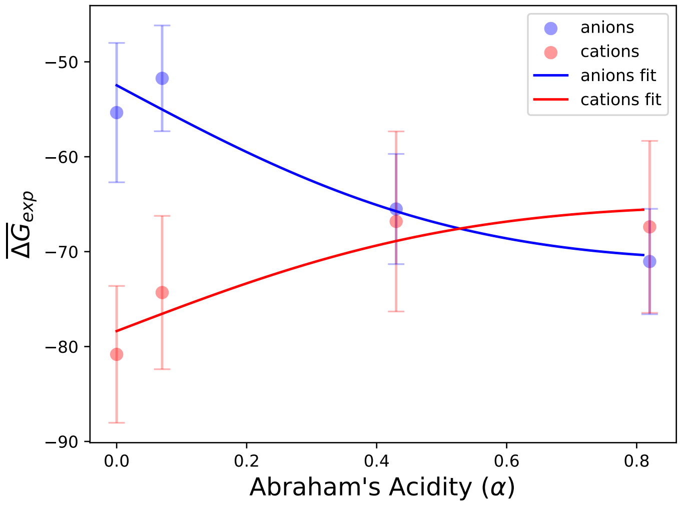


Figure 6.1: Averaged experimental solvation energies $\overline{\Delta G}_{exp}$ (in kcal/mol, colored dots) of 332 solutes taken from the MNSOL database (see Sec. 6.1.4.1) and standard deviations (vertical lines) as a function of Abraham's hydrogen bond acidity parameter α of the solvent. Four solvents have been considered: dimethyl sulfoxide, acetonitrile, methanol and water, in order of increasing α. The solid lines represent the fitted data obtained with the function $f(\alpha) = \dfrac{A\alpha}{(1+\alpha^2)} + B$, as found in Eq. (6.4).

To describe the corrective term for charged species, several functional forms of ΔG_{ion} have been tested to calculate solvation free energies of the ionic solutes mentioned above in four different solvents. Among the tested functions, based on the quasi-linear behaviour observed in Fig. 6.1, a correction is proposed with the following form:

$$\Delta G_{ne}^{ion} = I(Q, C, k) \cdot \left(\frac{A\alpha}{(1+\alpha^2)} + B \right) A_M \tag{6.4}$$

where A,B and C are three parameters to be optimized (see below, Sec. 6.1.4.1 for details), A_M is the total SASA of the solute, α is Abraham's hydrogen bond acidity of the solvent and $I(Q, C, k)$ is a function chosen to easily switch between neutral, positively-, and negatively-charged solutes with no discontinuities, defined as:

$$I(Q, C, k) = \frac{|Q|}{1 + exp(k \cdot Q) + C}. \tag{6.5}$$

This logistic switching function depends on the total solute charge $Q = \sum_i q_i$ and thus modulates the impact of the correction depending on the different kinds of solutes considered. For neutral species ($Q = 0$), the correction is not applied. For anions ($Q < 0$), this function is always positive, while for cations ($Q > 0$) this function decreases very rapidly to zero as a function of α. As a consequence, the proposed corrective term affects almost exclusively anions. This function also depends on k, a constant which controls the steepness of the curve (set to $20/e$ in this work to ensure high steepness of I), and the optimizable parameter C which controls the function asymptote for $Q < 0$. We note that all quantities involved in Eq. (6.4) do not require any additional calculations once the ΔG_{el} and ΔG_{ne} terms described in the previous sections have been computed, resulting in a corrective term obtained at no additional cost.

6.1.4 Training Set and Parameters Optimization

6.1.4.1 Training set

As training set for the optimization of the CDS parameters, and those of the corrective term for charged species, the MNSOL database [36] with solutes composed of H, C, N, O, F, Si, P, S and I atoms was used. In particular, the CDS model was parametrized on 541 neutral solutes in 91 solvents for a total of 2140 nonaqueous and 390 aqueous solvation energies, together with 144 transfer free energies ΔG_{te} between organic solvents and water, defined as:

$$\Delta G_{te} = \Delta G_{solv}^{organic} - \Delta G_{solv}^{water} \tag{6.6}$$

Instead, the corrective term for charged species was parametrized using the database of 213 singly-charged solutes in four solvents: acetonitrile, dimethyl sulfoxide, methanol, and water for a total of 332 solvation energies repartitioned between 208 anions and 124 cations. In particular, among the 213 charged solutes 112 are specific to water and a subset of 31 contains a single water molecule as described in the MNSOL manual [36]. The full list of solvents used together with the number of neutral solutes per solvent is available in Tab. 6.4 with the exception of methanol which was used exclusively for charged species. Furthermore, all solvent empirical descriptors were taken from the Minnesota solvent descriptor database [241].

6.1.4.2 CDS parameter optimization

Solute- ($\tilde{\sigma}_{Z_i}^{n}, \tilde{\sigma}_{Z_i}^{\alpha}, \tilde{\sigma}_{Z_i}^{\beta}, \tilde{\sigma}_{Z_iZ_j}^{n}, \tilde{\sigma}_{Z_iZ_j}^{\alpha}, \tilde{\sigma}_{Z_iZ_j}^{\beta}$) and solvent-dependent parameters ($\tilde{\sigma}^{\gamma}, \tilde{\sigma}^{\psi^2}, \tilde{\sigma}^{\phi^2}, \tilde{\sigma}^{\beta^2}$) of the CDS model have been obtained by minimization of the following cost function:

$$\chi = \sum_{i=1}^{N} \left(\Delta G_i^{exp} - \left(\frac{1}{M} \sum_{j=1}^{M} \Delta G_{ij}^{el} + \Delta G_i^{ne}(\tilde{\sigma}) \right) \right)^2 \tag{6.7}$$

where χ is given by the sum of the squared estimates of errors between the experimental reference solvation energies (ΔG_i^{exp}), and the calculated solvation energies defined as the sum of the parameter-dependent non-electrostatic energy ($\Delta G_i^{ne}(\tilde{\sigma})$) and the average electrostatic energies computed from $M = 10$ different levels of theory. Here, the levels of theory considered have been obtained by combination of five different functionals (B3LYP, B3PW, mPW1PW, PBE0, and PBE) with two basis sets (6-31G*, and 6-311G**). A total of $N = 2674$ solvation energies for neutral solutes have been considered: 2140 nonaqueous solvation energies, 390 aqueous solvation energies and 144 transfer energies. In addition, since the electrostatic term is charge model dependent, two separate optimizations have been performed: one with CM5 and one with HPA charges, resulting in two different sets of parameters for the revised CDS model for FDPB calculations.

Tab. 6.1 lists the resulting 68 solute-dependent parameters and 4 solvent-dependent parameters, for both CM5- and HPA-based electrostatics.

	CM5				HPA			
i	$\tilde{\sigma}_i^n(H_2O)$	$\tilde{\sigma}_i^n$	$\tilde{\sigma}_i^\alpha$	$\tilde{\sigma}_i^\beta$	$\tilde{\sigma}_i^n(H_2O)$	$\tilde{\sigma}_i^n$	$\tilde{\sigma}_i^\alpha$	$\tilde{\sigma}_i^\beta$
H	19.522	-4.074			26.480	-7.697		
C	19.768	52.417	1.558	46.728	-6.239	34.829	14.090	53.066
HC	-72.039	-31.108	15.133	2.571	-68.000	-19.653	14.862	-6.597
CC	-31.556	-61.057	3.483	-20.146	-8.935	-50.762	-14.980	-29.350
N	-5.200	32.153	21.461	71.035	14.273	30.167	-2.042	36.238
HN	-28.656	12.186		-5.411	-81.896	-11.055		-51.519
CN	-58.109	-78.305	129.200	-96.034	-54.835	-59.511	171.165	-67.035
NC	-24.967	-7.320	-23.327		-16.445	-9.898	-55.922	
NC2	-105.672	-159.641			-53.757	-138.080		
O	-61.404	-4.155	23.224	15.492	-63.205	-11.808	3.553	3.528
HO	4.346	53.421		-139.133	-62.775	22.781		-227.585
OC	7.982	-3.409			60.154	18.846		
CO2	92.759	-34.176			71.510	-28.543		
ON	53.822	-7.234	-32.680	2.737	68.347	21.590	31.615	26.200
OO	10.949	-27.982	34.601		44.382	-17.394	-3.933	
F	0.820	-2.985			5.639	-2.440		
Cl	-27.603	-29.302			-20.586	-26.428		
Br	-29.697	-34.731			-22.506	-32.526		
I	-37.382	-38.945			-27.674	-35.646		
Si	-203.089	-211.766			-89.961	-143.804		
P	-68.050				-65.707			
OP	36.492				73.988			
S	-39.233	-37.082			-35.197	-38.315		
OS	-113.993		548.792		-116.654		553.871	
SP	30.083	7.395			34.346	25.482		
X	-19.282	-17.074			-15.995	-18.215		

(a) Solute-dependent parameters

CM5				HPA			
$\tilde{\sigma}^\gamma$	$\tilde{\sigma}^{\phi^2}$	$\tilde{\sigma}^{\psi^2}$	$\tilde{\sigma}^{\beta^2}$	$\tilde{\sigma}^\gamma$	$\tilde{\sigma}^{\phi^2}$	$\tilde{\sigma}^{\psi^2}$	$\tilde{\sigma}^{\beta^2}$
0.355	-2.168	-3.201	-0.354	0.254	-2.118	-4.279	5.207

(b) Solvent-dependent parameters

Table 6.1: Revised CDS parameters for FDPB calculations. The index i refers to the element or atom type. For i=X the index refers to any element other than H, C, N, O, F, Si, P, S, Cl, Br, and I. For the solute-dependent parameters, any parameter not present in this table is set equal to zero. The $\tilde{\sigma}_i^n(H_2O)$ columns correspond to water-specific parameters, while the others correspond to nonaqueous solvents. All data given in cal·mol^{-1}·Å^{-2}.

As in the original implementation of the CDS model, 26 out of the 68 solute-dependent parameters are for water due to its particular importance as a solvent, and the parameter labeled X is for non-defined atom types in order to generalize the non-electrostatic model to any chemical element. Following the same procedure as in the SM12 solvation model [81], optimization of this parameter required to set all other solute-dependent parameters to zero and considering the model as atom-type independent. Optimization led to Root-Mean

Square Errors (RMSE), computed as RMSE $= \left(\frac{\chi}{N}\right)^{\frac{1}{2}}$, of 1.67 and 1.57 kcal/mol for CM5 and HPA electrostatics, respectively. This is in line with the 1.48 and 1.61 kcal/mol values reported for the SM12 model[81] when using CM5 and ESP electrostatics, respectively. We note that, in some cases, significant differences between the FDPB-CM5, FDPB-HPA, and original SM12-CM5 CDS parameters are obtained[81]: for example, values of 19.768, -6.239 and 121.00 cal·mol^{-1}·Å^{-2} are obtained for $\tilde{\sigma}^{n}_{\mathrm{C}}(\mathrm{H_2O})$ with the FDPB-CM5, FDPB-HPA and SM12-CM5 models, respectively. This further highlights the requirement of reoptimization of the parameters when different electrostatic models are considered, and the low direct transferability of the CDS model.

6.1.4.3 Charged species correction and parameter optimization

A similar strategy as the one discussed above has been applied for the minimization of the cost function given in Eq. (6.7) to determine the optimal A, B and C parameters of the proposed corrective term for charged species. In this case, the non-electrostatic term is now defined as:

$$\Delta G_i^{ne} = \Delta G_i^{\mathrm{CDS}}(\tilde{\sigma}) + \Delta G_i^{ion}(A, B, C) \tag{6.8}$$

where $\Delta G_i^{\mathrm{CDS}}(\tilde{\sigma})$ is the non-electrostatic energy obtained from the optimized CDS model for neutral molecules, which is now a constant for a given solute and $\Delta G_i^{ion}(A, B, C)$ is the corrective term for which the parameters A, B, and C minimize χ. In this case, the optimization is carried out by considering the $N = 332$ singly-charged solute/solvent systems from the MNSOL database, with $M = 5$ functionals (B3LYP, B3PW, mPW1PW, PBE0 and PBE) combined with the 6-311++G** basis set. The resulting parameters are listed in Tab. 6.2 for CM5 and HPA atomic charges with similar sign and magnitude for both charge models.

CM5			HPA		
A	B	C	A	B	C
-269.217	78.522	-0.238	-282.079	70.378	-0.276

Table 6.2: A, B and C parameters of the proposed corrective term for charged species (see Eq (6.4) and Sec. 6.1.3 for details). A and B are given in cal·mol^{-1}·Å^{-2}, and C is given in e.

6.2 Computational Details

All calculations have been performed with a modified development version of the CRYSTAL code[31], in which implicit solvation effects are included with the formalism described previously.

The solvation and transfer energies of 2674 neutral molecules, and 332 charged species, composed of H, C, N, O, F, Si, P, S, Cl, Br and I atoms taken from the MNSOL database [36, 241] have been calculated. All solute structures have been optimized in gas phase at the B3LYP/6-311G** and B3LYP/6-311++G** levels for neutral and charged species, respectively, and verified to be true minima by frequency calculations. For each solute/solvent combination in the MNSOL database, the electrostatic contribution to the solvation energies have been obtained with the B3LYP, B3PW, mPW1PW, PBE0 and PBE functionals together with the 6-31G*, and 6-311G** basis sets for neutral molecules, while only the

6-311++G** basis set was used for charged species. In both cases, both CM5 and HPA atomic charges have been used as a source term for the Poisson equation. On the other hand, for pK_a calculations, only B3LYP/Def2-TZVP calculations have been performed. Gas-phase free energies of isolated species have been obtained by numerical frequency calculations, using a step size of 0.003 Å, considering the standard rigid-rotor harmonic approximation[347, 348]. An extra large integration grid with 99 and 1454 radial and angular points, respectively, has been used for the DFT grid integration step, both in the SCF and atomic charge calculations.

Solvation energies of a TiO_2 anatase slab have been obtained in three solvents, namely water, acetonitrile and toluene and compared to VASPsol[34, 281, 282] values obtained with default parameters. Only the PBE exchange-correlation functional has been considered in this case due to the high computational cost of hybrid functionals in plane wave calculations, considering gas-phase optimized geometries. PAW potentials[349, 67] have been considered, with $2s$ and $2p$ electrons of O as well as $3s$, $3p$, $3d$ and $4s$ electrons of Ti treated explicitly. A (12×12×1) Monkhorst-Pack **k**-points grid has been used to sample the Brillouin zone. Only CM5 atomic charges have been considered for the FDPB calculations in this case. In all multislab calculations performed with VASPsol, a 15 Å gap of vacuum has been considered between image slabs along the normal to the surface plane considered.

The SCRF convergence criteria was set to $2 \cdot 10^{-5}$ Hartree (about 10^{-2} kcal/mol), and the Poisson equation was iteratively solved using the Optimal Successive Over-Relaxation algorithm [148] with a convergence criteria on the electrostatic potential of 10^{-5} kT/e at 298.15K, with atomic charges distributed on the FD grid using an inverse quadratic algorithm [288]. The FD grid spacing was set to 0.50 Å, with 65 points per non-periodic grid direction, with the solute filling at most 50% of the grid along each direction. To solve Poisson equation, Dirichlet boundary conditions have been applied on the grid edges, using Debye-Hückel potentials of the equivalent dipole of the system[149] along non-periodic directions, and Ewald potentials along periodic ones[79].

A Solvent Excluded Surface (SES) [169, 170] was used for the solute cavity, built from Bondi atomic radii [156] augmented by a solvent probe radius. A relative permittivity of 1.000 was used inside the cavity, while the static bulk dielectric constants of the different solvents have been applied outside.

The atomic and total SASA of all non-periodic and periodic solutes have been analytically computed with a stereographic projection technique[331, 341] presented in Chapter 4 on page 75, using Bondi's atomic radii and a probe radius of 0.400 Å for all solvents, as originally done in the CDS model. All parameters of the revised CDS model for use in FDPB calculations are collected in Tab. 6.1 and 6.2.

Parameters' optimization of the CDS model along with the additional corrective term has been performed using a Powell's conjugate direction method [350] implemented in a Python code utilizing Scipy [351] and Numpy [352], together with GNU parallel [353].

6.3 Results

In this section, we first discuss results obtained on free energies of solvation of neutral finite and infinite periodic surface systems. In addition, for the periodic system considered, we also investigate surface energies and band structure changes upon solvation as these are key criteria to understand and quantify, especially for photovoltaics and photocatalysis applications. Then, solvation of ionic species is considered, investigating how the proposed corrective term to the CDS model impacts computed free energies of solvation and aqueous pK_a values.

When possible, to evaluate the performances of the revised CDS model for FDPB electrostatics, the difference between calculated and experimental values has been quantified using the Mean Unsigned Error (MUE) according to:

$$\mathrm{MUE} = \sum_{i}^{N} \frac{|y_i - x_i|}{N} \tag{6.9}$$

where N runs on the number of systems considered, and y_i and x_i are the experimental and calculated values, respectively.

6.3.1 Neutral solutes

6.3.1.1 Finite molecular solutes

The overall MUE of computed solvation energies of 2530 neutral solute/solvent combinations in 91 solvents and of 144 transfer energies between water and 14 organic solvent are summarized in Tab. 6.3. The global MUE of computed solvation energies averaged over the ten levels of theory considered are of 0.71 and 0.64 kcal/mol with CM5 and HPA electrostatics, respectively. For transfer energies, the corresponding values are: 0.71 and 0.92 kcal/mol. We also note that the reported MUE values are very similar for all levels of theory tested, both for solvation and transfer energies, with only slight variations of about

	B3LYP		B3PW		mPW1PW		PBE		PBE0		All
	BS1	BS2	BS1	BS2	BS1	BS2	BS1	BS2	BS1	BS2	
CM5	0.70	0.71	0.73	0.71	0.73	0.71	0.72	0.72	0.72	0.71	0.71
HPA	0.63	0.63	0.65	0.64	0.65	0.64	0.65	0.65	0.65	0.64	0.64

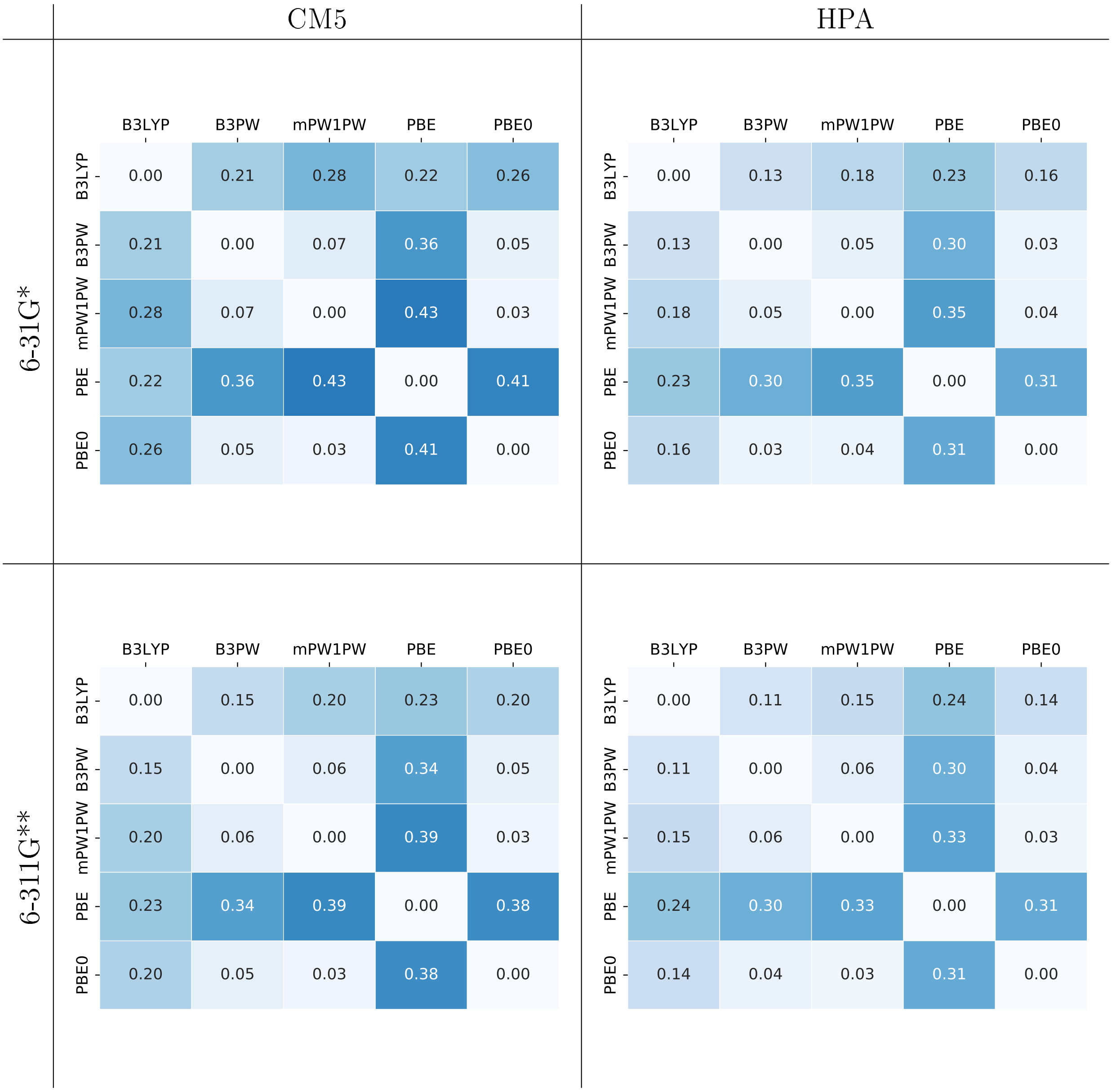
(a) ΔG_{solv}

	B3LYP		B3PW		mPW1PW		PBE		PBE0		All
	BS1	BS2	BS1	BS2	BS1	BS2	BS1	BS2	BS1	BS2	
CM5	0.73	0.73	0.70	0.71	0.69	0.70	0.74	0.75	0.69	0.70	0.71
HPA	0.91	0.92	0.91	0.92	0.91	0.91	0.93	0.95	0.91	0.91	0.92

(b) ΔG_{te}

Table 6.3: MUE of calculated (a) solvation energies (ΔG_{solv}) of 2530 neutral solute/solvent combinations in 91 solvents, and (b) 144 transfer energies (ΔG_{te}) between organic solvents and water, with respect to experimental values, both with CM5 and HPA atomic charges. BS1 and BS2 refer to the 6-31G* and 6-311G** basis sets, respectively. All data in kcal/mol.

0.01 – 0.02 kcal/mol between different levels of theory when considering the global MUE. For reference, the SM12 model [81] reports MUE values of 0.62 and 0.60 kcal/mol at the B3LYP/6-31G* level with CM5 atomic charges for the whole set of 2530 solvation energies and 144 transfer energies, respectively. The data reported here with CM5 charges for solvation and transfer energies of neutral solutes is therefore consistent with the SM12 solvation model data, with a difference in MUE of only about 0.10 kcal/mol. By comparison, when considering the original CDS-CM5 parameters developed for GBA, a MUE of 1.33 kcal/mol averaged on the ten levels of theory is obtained. This value is almost twice the one obtained with the new parametrization of the CDS model for FDPB electrostatics, and is well above the 1.00 kcal/mol threshold targeted for neutral solutes. Furthermore, with HPA, the MUE difference is of about 0.02 kcal/mol when compared to the SM12 model with CM5 charges on solvation energies, and of about 0.26 kcal/mol for transfer energies.



6-31G*, CM5

	B3LYP	B3PW	mPW1PW	PBE	PBE0
B3LYP	0.00	0.21	0.28	0.22	0.26
B3PW	0.21	0.00	0.07	0.36	0.05
mPW1PW	0.28	0.07	0.00	0.43	0.03
PBE	0.22	0.36	0.43	0.00	0.41
PBE0	0.26	0.05	0.03	0.41	0.00

6-31G*, HPA

	B3LYP	B3PW	mPW1PW	PBE	PBE0
B3LYP	0.00	0.13	0.18	0.23	0.16
B3PW	0.13	0.00	0.05	0.30	0.03
mPW1PW	0.18	0.05	0.00	0.35	0.04
PBE	0.23	0.30	0.35	0.00	0.31
PBE0	0.16	0.03	0.04	0.31	0.00

6-311G**, CM5

	B3LYP	B3PW	mPW1PW	PBE	PBE0
B3LYP	0.00	0.15	0.20	0.23	0.20
B3PW	0.15	0.00	0.06	0.34	0.05
mPW1PW	0.20	0.06	0.00	0.39	0.03
PBE	0.23	0.34	0.39	0.00	0.38
PBE0	0.20	0.05	0.03	0.38	0.00

6-311G**, HPA

	B3LYP	B3PW	mPW1PW	PBE	PBE0
B3LYP	0.00	0.11	0.15	0.24	0.14
B3PW	0.11	0.00	0.06	0.30	0.04
mPW1PW	0.15	0.06	0.00	0.33	0.03
PBE	0.24	0.30	0.33	0.00	0.31
PBE0	0.14	0.04	0.03	0.31	0.00

Figure 6.2: Heatmaps of the MUE of calculated solvation energies ΔG_{solv} between pairs of functionals for 2530 neutral solute/solvent combinations, considering different charge models and basis sets combinations. All data in kcal/mol.

Fig. 6.2 presents a more detailed description of the effect of the functionals, basis sets and charge models on computing solvation energies based on the Mean Unsigned Difference (MUD) between solvation energies obtained with different levels of theory according to the following equation:

$$\mathrm{MUD} = \sum_{i}^{N} \frac{|\Delta G_{solv}^{xc_1} - \Delta G_{solv}^{xc_2}|}{N} \tag{6.10}$$

where N runs on the 2530 systems considered and $\Delta G_{solv}^{xc_1}$ and $\Delta G_{solv}^{xc_2}$ are the solvation energies calculated for different pairs of functionals, for a given basis set and charge model combination.

The maximum difference in solvation energies calculated between different functional pairs is 0.45 kcal/mol for PBE when coupled with the mPW1PW functional, considering a 6-31G* basis set and CM5 electrostatics. In general, the highest MUD values are obtained with PBE coupled to any other functional (B3LYP, B3PW, mPW1P1 or PBE0), most probably since PBE is the only non hybrid functional in the chosen set. The remaining hybrid functionals all show lower MUD between each other, with the second highest MUD value being of 0.28 kcal/mol between B3LYP and mPW1PW considering the 6-31G* basis set and CM5 electrostatics, while the lowest value of MUD is 0.03 kcal/mol for multiple pairs of functionals, basis sets and charge models. A general trend can be observed from Fig. 6.2 which shows how the solvation energies differences between functionals tend to slightly decrease by increasing the basis set size from 6-31G* to 6-311G**, and by considering HPA instead of CM5 electrostatics. In the following therefore, due to the low differences in computed MUD and MUE between basis sets and functionals combinations (with the exception of PBE), only data averaged on the different levels of theory will be considered.

Table 6.4: Average MUE of calculated solvation and transfer energies with respect to experimental values for all solvents considered, using CM5 and HPA electrostatics. The average is computed on the 10 levels of theory considered. The number of solute molecules in each solvent (N) along with the corresponding solvent dielectric constant (ϵ) are also reported. All data in kcal/mol.

2530 Solvation energies				
Solvent	ϵ	N	CM5	HPA
1,2,4-trimethylbenzene	2.37	11	0.58 ± 0.11	0.42 ± 0.09
1,2-dibromoethane	4.93	10	0.60 ± 0.09	0.52 ± 0.05
1,2-dichloethane	10.12	39	0.91 ± 0.13	0.63 ± 0.11
1-fluoro-noctane	3.89	6	0.94 ± 0.17	0.77 ± 0.15
2,6-dimethylpyridine	7.17	6	0.53 ± 0.07	0.39 ± 0.05
2-methylpyridine	9.95	6	0.63 ± 0.15	0.42 ± 0.12
4-methyl-2-pentanone	12.89	13	0.87 ± 0.15	0.61 ± 0.08
acetic acid	6.25	7	0.84 ± 0.19	0.80 ± 0.09
acetonitrile	35.69	7	1.07 ± 0.18	0.78 ± 0.18
acetophenone	17.44	9	0.93 ± 0.20	0.75 ± 0.19
aniline	6.89	10	0.77 ± 0.15	0.58 ± 0.13
anisole	4.22	8	0.42 ± 0.05	0.37 ± 0.04
benzene	2.27	75	0.94 ± 0.09	0.74 ± 0.06
benzonitrile	25.59	7	1.13 ± 0.20	0.91 ± 0.18
benzyl alcohol	12.46	10	0.64 ± 0.05	0.53 ± 0.04
bromobenzene	5.39	27	0.79 ± 0.16	0.61 ± 0.11
bromoethane	9.01	7	1.22 ± 0.18	0.94 ± 0.16
bromoform	4.25	12	0.81 ± 0.15	0.52 ± 0.09

Continued on next page

Solvent	ϵ	N	CM5	HPA
bromooctane	5.02	5	1.39 ± 0.19	1.27 ± 0.17
butanone	18.25	13	1.31 ± 0.24	1.05 ± 0.22
butyl acetate	4.99	22	0.67 ± 0.04	0.72 ± 0.04
carbon disulfide	2.61	15	0.49 ± 0.03	0.44 ± 0.03
carbon tetrachloride	2.23	79	0.45 ± 0.04	0.43 ± 0.02
chlorobenzene	5.70	38	0.87 ± 0.15	0.59 ± 0.11
chloroform	4.71	109	0.75 ± 0.13	0.74 ± 0.12
chlorohexane	5.95	11	1.39 ± 0.20	1.20 ± 0.18
cyclohexane	2.02	92	0.46 ± 0.03	0.45 ± 0.02
cyclohexanone	15.62	10	0.69 ± 0.07	0.60 ± 0.09
decalin	2.20	27	1.26 ± 0.02	1.05 ± 0.02
decane	1.99	39	0.42 ± 0.01	0.44 ± 0.04
decanol	7.53	11	0.92 ± 0.15	0.78 ± 0.10
dibutyl ether	3.05	15	0.59 ± 0.08	0.46 ± 0.06
diethyl ether	4.24	72	0.76 ± 0.06	0.69 ± 0.04
diisopropyl ether	3.38	22	0.76 ± 0.01	0.67 ± 0.01
dimethyl sulfoxide	46.83	7	0.77 ± 0.10	0.70 ± 0.06
dodecane	2.01	8	0.28 ± 0.01	0.32 ± 0.03
ethanol	24.85	8	1.78 ± 0.22	1.50 ± 0.19
ethoxybenzene	4.18	7	0.38 ± 0.02	0.34 ± 0.03
ethyl acetate	5.99	24	0.57 ± 0.07	0.53 ± 0.07
ethylbenzene	2.43	29	0.47 ± 0.06	0.43 ± 0.03
fluorobenzene	5.42	7	0.94 ± 0.17	0.75 ± 0.14
heptane	1.91	69	0.46 ± 0.00	0.44 ± 0.01
heptanol	11.32	12	0.72 ± 0.10	0.76 ± 0.08
hexadecane	2.04	198	0.47 ± 0.01	0.47 ± 0.01
hexadecyl iodide	3.53	9	1.28 ± 0.07	1.20 ± 0.09
hexane	1.88	59	0.55 ± 0.02	0.53 ± 0.03
hexanol	12.51	14	0.66 ± 0.08	0.74 ± 0.08
iodobenzene	4.55	20	0.49 ± 0.12	0.48 ± 0.09
isobutanol	16.78	17	0.51 ± 0.02	0.49 ± 0.02
isooctane	1.94	32	0.47 ± 0.07	0.42 ± 0.06
isopropanol	19.27	7	1.33 ± 0.20	1.02 ± 0.17
isopropylbenzene	2.37	19	0.36 ± 0.07	0.31 ± 0.04
m-cresol	12.44	7	0.87 ± 0.11	0.79 ± 0.11
mesitylene	2.27	7	0.49 ± 0.10	0.35 ± 0.07
methoxyethanol	17.20	6	0.59 ± 0.05	0.50 ± 0.05
methylene chloride	8.93	11	1.14 ± 0.18	0.85 ± 0.21
N-butanol	17.33	21	0.65 ± 0.05	0.68 ± 0.06
N-butylbenzene	2.36	10	0.56 ± 0.11	0.43 ± 0.09
N-methylformamide	181.56	7	0.65 ± 0.10	0.56 ± 0.09
nitrobenzene	34.81	15	1.49 ± 0.21	1.08 ± 0.15
nitroethane	28.29	7	0.95 ± 0.17	0.74 ± 0.21
nitromethane	36.56	7	1.10 ± 0.19	0.94 ± 0.15
N,N-dimethylacetamide	37.78	7	0.77 ± 0.12	0.56 ± 0.08
N,N-dimethylformamide	37.22	7	0.70 ± 0.15	0.50 ± 0.13
nonane	1.96	26	0.31 ± 0.01	0.30 ± 0.01
nonanol	8.60	10	0.81 ± 0.07	0.77 ± 0.06
octane	1.94	38	0.35 ± 0.01	0.32 ± 0.01
octanol	9.86	247	0.91 ± 0.03	0.83 ± 0.02
o-dichlorobenzene	9.99	11	1.51 ± 0.17	1.01 ± 0.13
o-nitrotoluene	25.67	6	1.33 ± 0.16	1.02 ± 0.08
pentadecane	2.03	9	0.44 ± 0.01	0.42 ± 0.02
pentane	1.84	26	0.38 ± 0.04	0.31 ± 0.02
pentanol	15.13	22	0.73 ± 0.07	0.74 ± 0.09
perfluorobenzene	2.03	15	0.61 ± 0.10	0.64 ± 0.10
phenyl ether	3.73	6	1.04 ± 0.14	1.00 ± 0.11

Continued on next page

Solvent	ϵ	N	CM5	HPA
p-isopropyltoluene	2.23	6	0.61 ± 0.10	0.43 ± 0.09
propanol	20.52	7	1.17 ± 0.20	0.89 ± 0.17
pyridine	12.98	7	0.67 ± 0.14	0.46 ± 0.12
sec-butanol	15.94	9	0.55 ± 0.06	0.59 ± 0.08
sec-butylbenzene	2.35	5	0.27 ± 0.08	0.24 ± 0.04
t-butylbenzene	2.35	14	0.36 ± 0.09	0.29 ± 0.07
tetrachloroethene	2.27	10	0.58 ± 0.07	0.45 ± 0.04
tetrahydrofuran	7.43	7	0.48 ± 0.04	0.38 ± 0.05
sulfolane	43.96	7	1.27 ± 0.18	0.92 ± 0.13
tetralin	2.77	9	0.94 ± 0.08	1.12 ± 0.09
toluene	2.37	51	0.67 ± 0.07	0.55 ± 0.05
tributylphosphate	8.18	16	0.77 ± 0.06	0.77 ± 0.04
triethylamine	2.38	7	0.84 ± 0.10	0.82 ± 0.07
undecane	1.99	13	0.43 ± 0.02	0.39 ± 0.02
water	78.36	390	0.87 ± 0.04	0.84 ± 0.05
m-, o-, p-xylene mixture	2.39	48	0.61 ± 0.09	0.45 ± 0.05
All		2530	0.71 ± 0.11	0.64 ± 0.09
144 Transfer energies				
1,2-dibromoethane/water		1	0.29 ± 0.14	0.66 ± 0.23
1,2-dichloethane/water		3	1.13 ± 0.07	2.28 ± 0.05
benzene/water		4	1.66 ± 0.26	3.95 ± 0.29
carbon tetrachloride/water		2	1.32 ± 0.26	0.58 ± 0.11
chlorobenzene/water		1	0.99 ± 0.13	1.61 ± 0.23
chloroform/water		7	0.72 ± 0.09	2.18 ± 0.05
cyclohexane/water		5	0.95 ± 0.23	1.00 ± 0.10
dibutyl ether/water		1	1.35 ± 0.24	3.64 ± 0.26
diethyl ether/water		8	0.99 ± 0.10	0.95 ± 0.03
ethyl acetate/water		1	1.02 ± 0.13	0.35 ± 0.15
hpetane/water		6	0.97 ± 0.12	1.09 ± 0.08
hexane/water		1	0.54 ± 0.29	5.60 ± 0.63
nitrobenzene/water		1	1.97 ± 0.05	1.67 ± 0.13
octanol/water		103	0.60 ± 0.01	0.60 ± 0.01
All		144	0.71 ± 0.17	0.92 ± 0.22

More in details, Tab. 6.4 lists the average MUE computed for 2530 solvation energies in 91 solvents and 144 transfer energies between water and 14 organic solvents, while Fig. 6.3 further presents the average MUE as a function of the relative permittivity of the solvent. From these data, it is clear that most MUE values are below the target error threshold of 1 kcal/mol for neutrals. The largest errors are obtained in ethanol ($\varepsilon = 24.85$), with MUE values of 1.78 ± 0.22 and 1.50 ± 0.19 kcal/mol with CM5 and HPA electrostatics, respectively. This MUE might be attributed to the lack of specific parameters developed for ethanol to capture potential hydrogen bond effects, as done for another protic polar solvent such as water for which a specific parametrization with 26 dedicated parameters is used to obtain a MUE of 0.87 ± 0.04 and 0.84 ± 0.05 with CM5 and HPA electrostatics, respectively. In addition, the error tends to increase upon increasing the relative permittivity of the solvent, data obtained with CM5 atomic charges being generally more affected than the one obtained with HPA electrostatics. In general, this behavior can be in part attributed to the larger magnitude of the CM5 charges compared to HPA ones, resulting in a higher solute/solvent polarization effect with the formers than with the latters, which rapidly increases with the value of the solvent relative permittivity.

Transfer energies reported in Tab. 6.4 mainly concern octanol/water data for which low errors are obtained, with a MUE of 0.60 ± 0.01 kcal/mol for both CM5 and HPA electrostatics. On the other hand, larger errors are generally obtained for the remaining organic solvent/water transfer energies. In particular, CM5 charges tend to lead to more stable MUE with acceptable values obtained in all cases, while HPA-based data lead to unacceptable errors in some cases, such as in the hexane/water case with a MUE of 5.60 ± 0.63 kcal/mol or of benzene/water with 3.95 ± 0.29 kcal/mol, indicating that this latter atomic charge model is less preferable in the calculation of transfer energies.

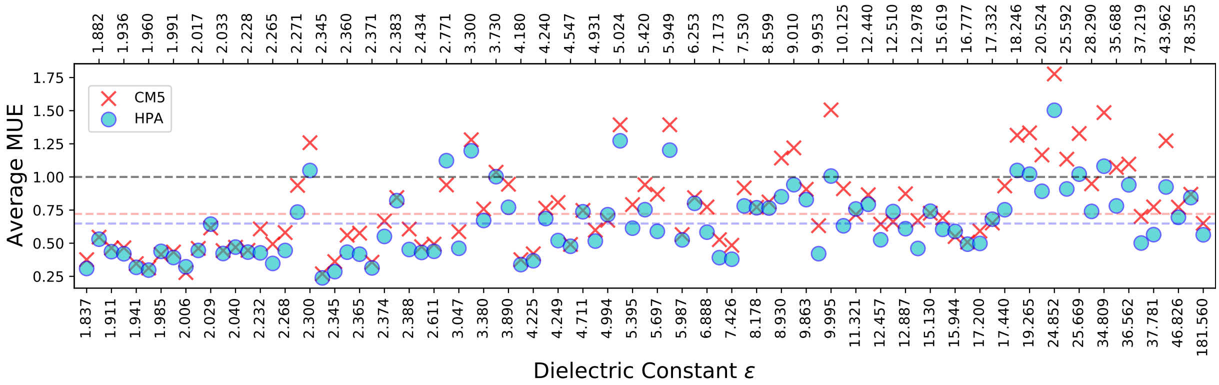


Figure 6.3: Averaged MUE (in kcal/mol) of solvation energies for each solvent given as a function of the dielectric constant of the solvent (ϵ), using both CM5 (red crosses) and HPA (blue dots) electrostatics. Colored dotted lines indicate the average MUE of each charge model: 0.71 ± 0.11 and 0.64 ± 0.09 kcal/mol for CM5 and HPA charges, respectively. The black dashed line highlights the 1.00 kcal/mol threshold targeted for neutral solutes.

To sum up therefore, for finite neutral solutes, slightly more accurate results in the calculation of solvation energies have been obtained with HPA than with CM5 atomic charges, while the opposite has been found for transfer energies. The MUE difference between the two charge models is in the range of 0.10 kcal/mol for both solvation and transfer energies data, with results obtained with both atomic charge models being highly independent from the level of theory chosen.

6.3.1.2 Wetting of TiO_2 anatase (101)

As another example of the general applicability of the reparametrized CDS model, we present solvation of a TiO_2 anatase (110) surface in three different solvents, namely water, acetonitrile and toluene, which have been chosen as examples of polar protic, polar aprotic and apolar solvent media, respectively. Although rutile is the most stable phase of TiO_2 under ambient conditions, TiO_2 nanoparticles are preferentially of the anatase variant, with important applications in photocatalysis and photovoltaics [354, 355], where the anatase (101) surface orientation is commonly found in sensitized solar cells [356, 357] for example due to its high photocatalytic properties[358, 355] especially in aqueous medium [359]. Since no experimental solvation energies are available for surfaces, data obtained with the FDPB approach are here compared to the ones obtained with the reference VASPsol continuum solvation mode l[281, 360], although formalisms of these two models largely differ. A similar strategy has recently been applied to various PbS surfaces in water [361].

A slab made with 10 Ti layers leading to converged surface energies[362] has been considered for all calculations, each Ti layer being composed of three atomic planes with a O-Ti-O sequence. Fig. 6.4 presents the corresponding two-dimensional periodic slab model used, together with the ASC distributed on of the cavity of the solute's unit cell in the FDPB calculations, while Tab. 6.5 presents computed data in the three solvents considered.

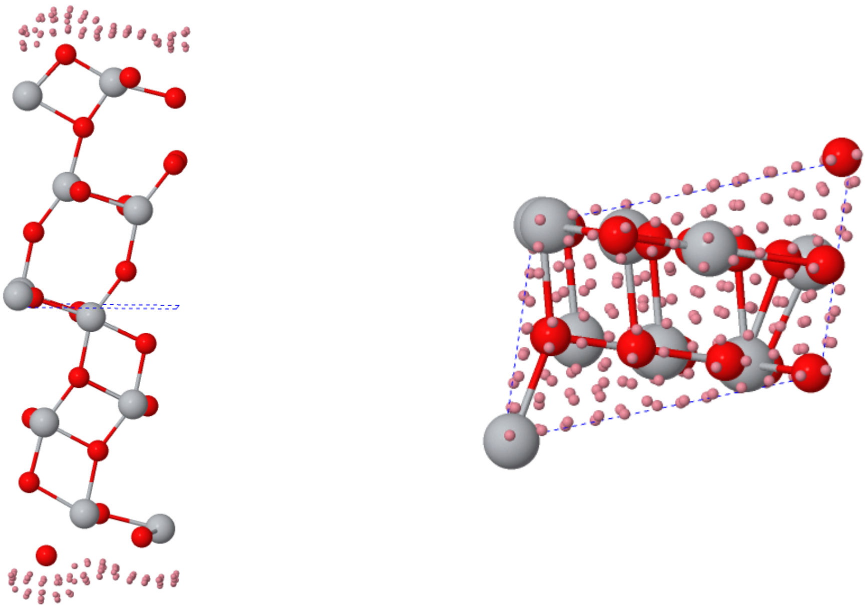

(a) (b)

Figure 6.4: (a) Side and (b) top views of a TiO_2 anatase (101) slab model with 10 Ti layers. Grey and red balls represent Ti and O atoms, respectively. The pink dots correspond to 176 ASC distributed on both sides of the solute cavity exposed to the implicit solvent. The slab unit cell is shown as dotted blue lines.

	Water		Acetonitrile		Toluene	
	FDPB	VASPsol	FDPB	VASPsol	FDPB	VASPsol
ΔG_{el}	-10.37	-10.94	-8.84	-8.32	-3.27	-1.17
ΔG_{ne}	-0.69	$+0.56$	$+0.52$	$+0.56$	$+0.22$	$+0.54$
ΔG_{tot}	-11.06	-10.38	-8.32	-7.76	-3.05	-0.63
$\Delta\Delta G_{el}^{w/x}$	–	–	$+1.53$	$+2.62$	$+7.10$	$+9.77$

Table 6.5: Computed PBE total free energies of solvation (ΔG_{tot}), along with the electrostatic (ΔG_{el}) and non=electrostatic (ΔG_{ne}) components, of a 10 Ti layer TiO_2 anatase (101) slab model solvated in implicit water, acetonitrile and toluene with the FDPB and VASPsol continuum solvation models. The $\Delta\Delta G_{el}^{w/x}$ line corresponds to the difference between the ΔG_{el} data computed in water and the other solvents. All data in kcal/mol. CM5 atomic charges have been used for FDPB calculations.

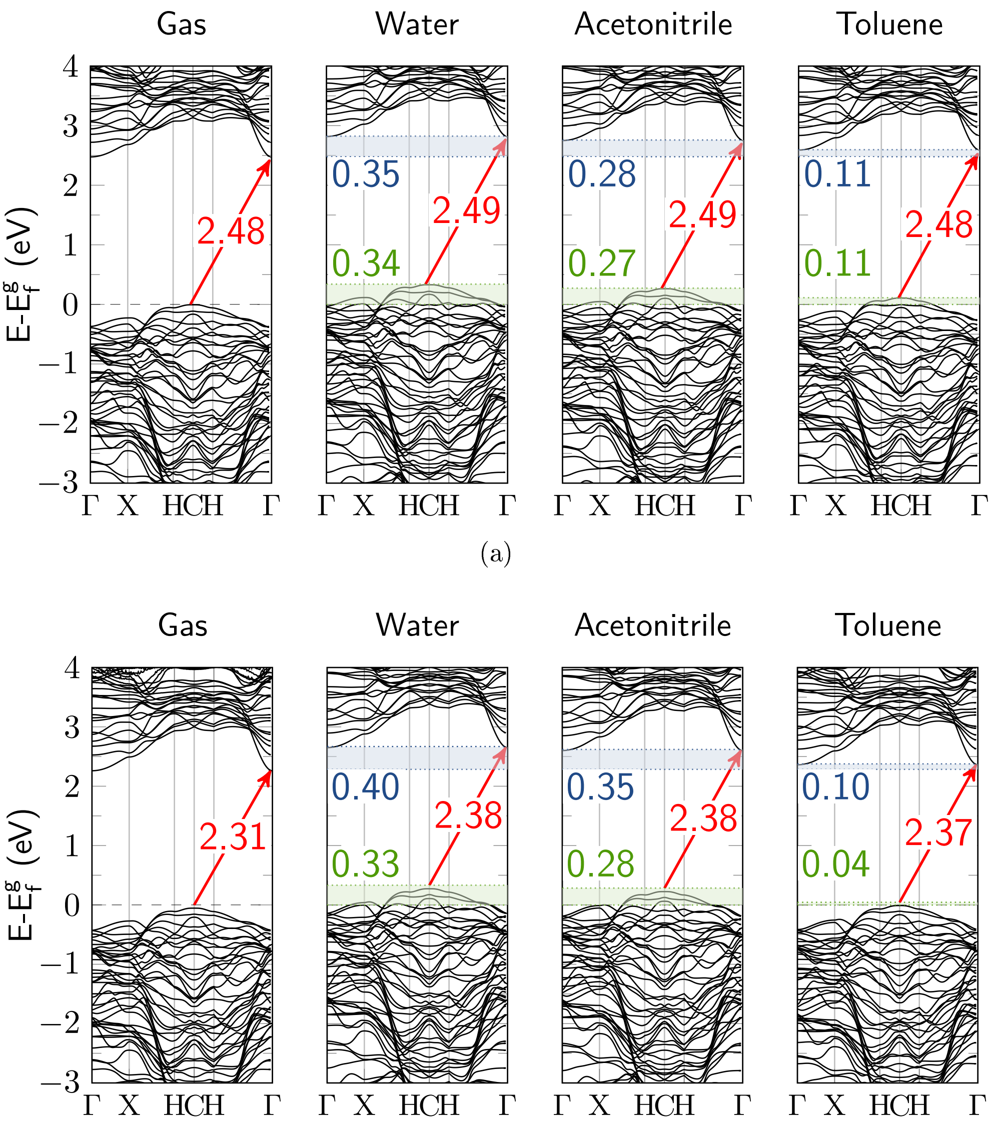
From Tab. 6.5, it is clear that although data obtained for ΔG_{tot} are different between the two continuum solvation models, a general good agreement is however obtained, with trends obtained when going from water to acetonitrile and toluene being very similar and corresponding to a decrease of the free energy of solvation. This is mainly related to a decrease of the electrostatic contribution ΔG_{el} when going from a polar protic solvent to a polar aprotic one and then to an apolar one, which is further quantified by the $\Delta\Delta G_{el}^{w/x}$ data reported, in good agreement between the two solvation models, showing a larger value in the water/toluene case than in the water/acetonitrile one with both models. In addition, for the non-electrostatic contribution ΔG_{ne}, we note that significant differences are obtained between the CDS model used in the FDPB approach and the cavity contribution used in VASPsol [281, 360]. In particular, due to the lack of parameters available for nonaqueous solvents in VASPsol, a very similar positive ΔG_{ne} contribution is obtained for all three solvents considered. On the other hand, a larger range of values is obtained in the FDPB calculations, with either negative (in water) or positive (in acetonitrile and toluene) contributions to the free energy of solvation.

Tab. 6.6 further reports the computed surface energies of the slab both in gas-phase and in solvent. Both continuum solvation models agree very well, the largest discrepancies being obtained for toluene as noted above. In particular, the stabilization of the solvated slab when compared to the gas-phase one are well captured with both continuum solvation models.

	Gas		Water		Acetonitrile		Toluene	
	FDPB	VASPsol	FDPB	VASPsol	FDPB	VASPsol	FDPB	VASPsol
E_s	+0.485	+0.496	+0.293	+0.317	+0.341	+0.362	+0.433	+0.487
$\Delta E_s^{g/x}$	–	–	−0.192	−0.176	−0.144	−0.131	−0.052	−0.009

Table 6.6: Computed PBE surface energies (E_s) of a 10 Ti layer TiO_2 anatase (101) slab model in gas-phase and solvated in implicit water, acetonitrile and toluene with the FDPB and VASPsol continuum solvation models. The variation of surface energy ($\Delta E_s^{g/x}$) when going from gas-phase to solvent x is also reported. All data in J/m^2.

Finally, we also examined the changes in the band structure of the TiO_2 slab upon solvation, which are presented in Fig. 6.5. Overall, the shapes of the band structures are only slightly affected by all three solvents considered. Both a band gap increase and a global shift of the bands to higher energies can be evidenced. As could be expected, these effects are more important when considering water and acetonitrile than toluene since the former two solvents lead to a larger polarization of the slab than the latter one, in line with the computed free energies of solvation discussed above. The band gap increase is more important with the VASPsol continuum solvation model than with the FDPB one, mainly due to a larger upshift of the conduction bands with the former than with the latter. The upshift of the valence bands are however in remarkable agreement between the two models, except for toluene where this effect is larger with the FDPB model than with the VASPsol one. As noted above, this overall agreement between the two continuum solvation models is encouraging for the application of the FDPB model to complex solvated interfaces found in photocatalysis or in the photovoltaic field where band edge displacements are fundamental to better understand.



(a)

(b)

Figure 6.5: Band structures of a 10 Ti layer TiO_2 anatase (101) slab model computed at the PBE level in gas-phase and in implicit water, acetonitrile and toluene with the (a) FDPB and (b) VASPsol continuum solvation models. The zero of the energy axis has been set to the valence band maximum of the gas-phase calculation (E_f^g). The green and blue shaded areas highlight the shifts of the top of the valence and bottom of conduction bands, respectively, with the computed shifts indicated with the same colors. The band gaps are also indicated by the red arrow along with the corresponding values obtained. All data reported in eV.

6.3.2 Ionic solutes

6.3.2.1 Free energies of solvation

To further assess the performances of the proposed implicit solvation model, Tab. 6.7 collects the MUE obtained for the free energies of solvation with respect to the experimental data of 208 anions and 124 cations computed in four solvents (acetonitrile, dimethyl sulfoxide, methanol and water), where the correction for charged species described in Sec. 6.1.3

has been applied.

		B3LYP		B3PW		mPW1PW		PBE		PBE0		SM12^a
	N	CM5	HPA	CM5	HPA	CM5	HPA	CM5	HPA	CM5	HPA	
acetonitrile	30	3.16	3.32	3.55	3.96	3.70	4.13	2.69	2.71	3.59	3.97	2.8-4.7
dimethyl sulfoxide	67	3.90	3.59	3.70	3.37	3.66	3.37	4.26	4.02	3.67	3.36	5.9-7.0
methanol	51	2.37	2.64	2.38	2.86	2.44	2.96	2.62	2.51	2.40	2.88	2.3-3.0
water	60	3.08	3.06	2.64	2.67	2.59	2.62	3.78	3.71	2.65	2.67	2.9-3.8
all	208	3.17	3.16	3.04	3.13	3.05	3.16	3.48	3.36	3.04	3.13	

(a) Anions

		B3LYP		B3PW		mPW1PW		PBE		PBE0		SM12^a
	N	CM5	HPA	CM5	HPA	CM5	HPA	CM5	HPA	CM5	HPA	
acetonitrile	39	4.03	4.10	4.22	3.94	4.16	3.94	4.16	4.04	4.13	3.94	7.2-7.7
dimethyl sulfoxide	4	3.92	1.32	4.09	1.11	4.03	1.14	3.94	1.27	3.98	1.18	2.0-3.1
methanol	29	1.97	3.85	1.65	3.44	1.67	3.51	1.75	3.56	1.69	3.54	1.7-2.5
water	52	3.57	2.48	3.79	2.57	3.72	2.55	3.71	2.54	3.69	2.54	2.9-3.5
all	124	3.35	3.28	3.43	3.16	3.39	3.17	3.40	3.22	3.37	3.18	

(b) Cations

Table 6.7: MUE (in kcal/mol) of the solvation energies of (a) 208 anions, and (b) 124 cations obtained in four different solvents. All data reported include the corrective term for charged species (see Eq. (6.4)). ^{a}: minimal and maximal MUE among the eight values reported in Ref. 81.

By averaging on the five levels of theory and the four solvents considered, the MUE obtained for anions are of 3.15 ± 0.16 and 3.18 ± 0.08 kcal/mol with CM5 annd HPA electrostatics, respectively, while they are of 3.38±0.03 and 3.20±0.04 kcal/mol for cations. For reference, SM12 with CM5 charges at the B3LYP/ 6-31G* level of theory achieved MUE of 4.3 and 4.0 kcal/mol for anions and cations, respectively on the same test set, that is about 1.0 kcal/mol higher than the values obtained with the proposed correction for charged species.

In addition, data obtained without the proposed corrective term are presented in Tab. 6.8.

It is clear that the errors obtained are more than doubled for anions, with computed MUE values of 6.15±0.04 and 7.58±0.09 kcal/mol with CM5 and HPA atomic charges. On the other hand, data obtained for cations are only slightly worse, with MUE of 3.63 ± 0.01 and 3.81±0.07 kcal/mol for CM5 and HPA electrostatics, respectively. This is line with the discussion made in Sec.6.1.3, indicating that the proposed corrective term almost exlcusively affects anions due to the effect of the switching function.

Overall therefore, when considering the whole set of solute/solvent combinations the proposed corrective term ΔG_{ne}^{ion} leads to MUEs very close to the experimental uncertainty on solvation energies[36] of 3.00 kcal/mol for charged solutes in the MNSOL database with

		B3LYP		B3PW		mPW1PW		PBE		PBE0	
	N	CM5	HPA	CM5	HPA	CM5	HPA	CM5	HPA	CM5	HPA
acetonitrile	30	10.07	9.08	10.83	9.93	11.03	10.88	6.38	7.72	8.69	9.95
dimethyl sulfoxide	64	7.68	8.58	8.70	9.63	8.87	8.71	6.14	7.63	6.57	9.61
methanol	51	3.37	4.41	2.67	3.86	2.53	2.64	2.08	5.47	4.61	3.85
water	59	7.85	9.20	6.77	7.98	6.62	6.78	2.25	10.32	8.91	7.99
all	204	7.00	7.58	6.95	7.57	6.95	7.57	6.95	7.65	7.07	7.56

(a) Anions

		B3LYP		B3PW		mPW1PW		PBE		PBE0	
	N	CM5	HPA	CM5	HPA	CM5	HPA	CM5	HPA	CM5	HPA
acetonitrile	39	6.21	3.69	6.46	3.73	6.38	3.71	6.38	3.72	6.35	3.70
dimethyl sulfoxide	4	6.12	1.99	6.28	2.34	6.22	2.31	6.14	2.03	6.18	2.28
methanol	29	2.35	5.28	2.03	4.85	2.07	4.93	2.08	4.96	2.09	4.97
water	51	2.18	3.57	2.25	3.24	2.22	3.28	2.25	3.37	2.21	3.30
all	123	3.62	3.96	3.66	3.74	3.64	3.77	3.64	3.81	3.62	3.79

(b) Cations

Table 6.8: MUE of the calculated solvation energies obtained without the proposed corrective term $\Delta G_{ne}^{\text{ion}}$ with respect to experimental values for (a) 204 anions and (b) 123 cations. All data have been obtained with the 6-311++G** basis set and are reported in kcal/mol.

both CM5 and HPA electrostatics, the former performing better with anions, and the latter performing better with cations. Some noticeable exceptions can however be evidenced: for example, for cations in methanol, CM5 outperforms significantly HPA, with corresponding MUE of 1.77 ± 0.11 and 3.58 ± 0.14 kcal/mol, respectively.

6.3.2.2 pK_a evaluation

As a final example of the evaluation of the performances of the reparametrized CDS model and of the proposed corrective term for charged species, pK_a values have been calculated for a test set [363] of 28 carboxylic acids, 10 aliphatic amines and 45 thiols in aqueous solvent. Calculation of pK_a with implicit solvation models still remains very challenging, in particular since they require highly-accurate solvation energies of both neutral and charged species, which might require potential hydrogen bonding effects between the solute and the solvent, especially in water.

Although different methodologies have been proposed to compute pK_a from continuum solvation models[364], here, we relied on a simple direct approach [64] in which the dissociation of an acid species HA into its conjugate base A$^-$ and the proton is considered according to:

$$\mathrm{HA}_{aq} \rightleftharpoons \mathrm{A}^-_{aq} + \mathrm{H}^+_{aq}. \tag{6.11}$$

The pK_a is then related to the free energy difference ΔG°_{aq} in the 1 M standard state, indicated by the superscript °, directly from the free energies of hydration of the acid

$\Delta G^{\circ}_{solv}(\mathrm{HA})$ and the conjugate base $\Delta G^{\circ}_{solv}(\mathrm{A}^-)$ through the following equations:

$$\begin{cases} pK_a = \dfrac{\Delta G^{\circ}_{aq}}{2.303RT} \\ \Delta G^{\circ}_{aq} = -265.9 + 1.89 + \Delta G^{\circ}_{solv}(\mathrm{A}^-) - \Delta G^{\circ}_{solv}(\mathrm{HA}) + \Delta G^{\circ}_{deprot} \end{cases} \tag{6.12}$$

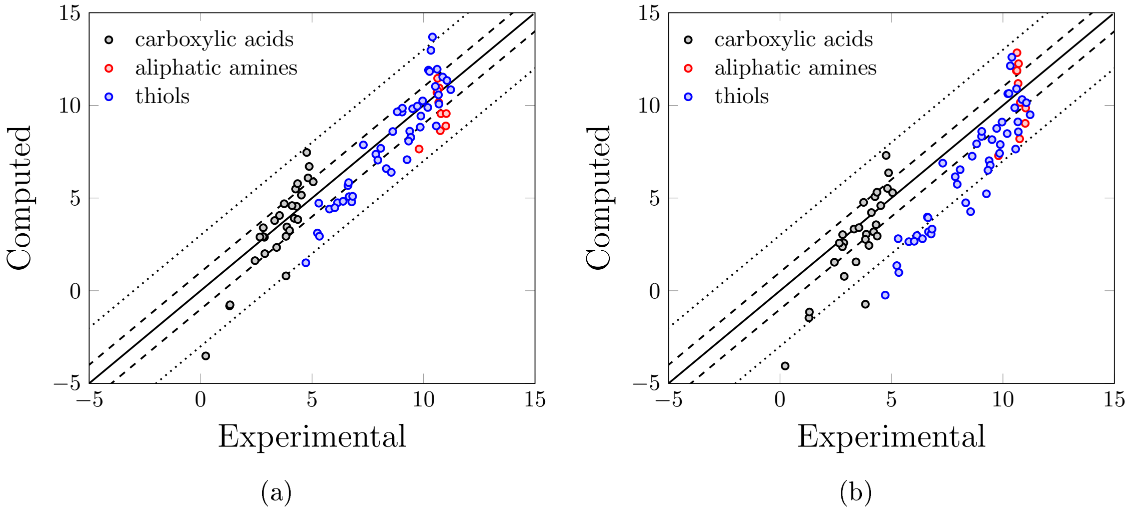
in which R is the ideal gas constant in kcal·K^{-1}·mol^{-1} , T is the temperature 298 K, -265.9 kcal/mol is the experimental solvation free energy of the proton [365, 366, 367, 368], 1.89 kcal/mol is the standard state 1 M correction, and the gas-phase deprotonation standard free energy $\Delta G^{\circ}_{deprot} = G^{\circ}_{g}(\mathrm{A}^-) - G^{\circ}_{g}(\mathrm{HA}) + G^{\circ}_{g}(\mathrm{H}^+)$ is calculated using the experimental gas-phase free energy of the proton $G^{\circ}_{g}(\mathrm{H}^+)$ [365, 366, 367, 368] equal to -6.29 kcal/mol. This approach has already been successfully used for a number of chemical compounds and provided errors lower than 1 pK_a unit (see Ref. 364 for instance and references therein). However, it has also been reported that the precision expected from Eq. (6.12) should be limited due to some systematic errors [369, 370, 364].

All data computed at the B3LYP level are collected in Tab. 6.9. Fig. 6.6 also presents a comparison between the computed and experimental pK_a values. We note that, based on the MUE reported on ΔG_{solv} at the B3LYP level for both neutrals and anions in Tab. 6.3 and 6.7, a potential accumulated error of about 3.8 kcal/mol corresponding to about 3 pK_a units[369, 370, 364] already exists from Eq. (6.12), without considering the uncertainties related to the solvation free energy of the proton[371, 372, 369].

		This work		Ref. 363	
	N	CM5	HPA	SMD	SMD$_{\mathrm{aSAS}}$
Carboxylic acids	28	1.07	1.23	1.38	0.86
Aliphatic amines	10	0.83	2.00	0.68	0.35
Thiols	45	1.33	2.60	7.12	0.48
All	83	1.18	2.07	4.37	0.59

Table 6.9: MUE of computed pK_a in water with respect to experimental data obtained for a set of 83 solutes belonging to the carboxylic acid, aliphatic amine and thiol families, obtained at the B3LYP/Def2-TZVP level with both CM5 and HPA atomic charges. Data obtained at the B3LYP/6-31+G(d,p) level with the standard SMD solvation model [81], as well a modified version considering a scaled solvent-accessible surface approach (SMD$_{\mathrm{aSAS}}$) taken from Ref. 363 are also reported.

On the whole test set, the proposed solvation model presents encouraging results, with MUE values of 1.18 and 2.07 pK_a units with CM5 and HPA electrostatics, respectively. From Fig. 6.6a, it is also clear that all computed CM5 values are within the targeted 3 pK_a units error threshold, while HPA values are slighty above the threshold. As a comparison, the corresponding MUE value of the SMD model is of 4.37 pK_a units on the same test set. Although our model has MUE values on par with the SMD ones for carboxylic acids and aliphatic amines, it outperforms it notably for thiols, where a 7.12 pK_a units MUE is obtained with SMD, while our model reaches values of 1.33 and 2.60 pK_a units with CM5 and HPA atomic charges, respectively. The poor performances of the standard SMD solvation model observed in the pK_a prediction for some classes of organic compounds including thiols when using a direct approach has already been reported[363] and is related to inaccuracies in the solvation free energies of anions. This has led to the development of modified versions of SMD for pK_a calculations, such as the SMD$_{\mathrm{aSAS}}$ approach[363], in which the solute cavity is scaled by considering a scaled solvent-accessible surface approach, leading to an overall improvement of computed pK_a values compared to standard SMD, as shown in Tab. 6.9. In particular, the scaled version of SMD achieves an impressive MUE of 0.59 pK_a units on the whole set of 83 molecules. The application of the scaling in the SMD$_{\mathrm{aSAS}}$ does however increase the MUE of free energies of solvation of the neutral molecules of the considered test set[363], from 0.80 kcal/mol for the default SMD model to 1.70 kcal/mol for the scaled one. The better results obtained with the scaled SMD model might then be due to a larger compensation of errors between the solvation energies of neutral and charged species, as each 1.00 kcal/mol variation in free energy difference for a given deprotonation reaction contributes to about 0.70 pK_a units.



(a)

(b)

Figure 6.6: Comparison between calculated and experimental pK_a values for 28 carboxylic acids (in black), 10 aliphatic amines (in red), and 45 thiols (in blue) in water, obtained with (a) CM5 and (b) HPA atomic charges. The inner and outer dashed lines represent errors of ±1.00 and ±3.00 pK_a units, respectively.

The overall encouraging results in the prediction of pK_a values can be attributed to the correction for charged species based on Abraham's acidity parameter. In fact, by removing this contribution to the total solvation energy of charged species the MUE increases to 4.13 and 4.18 pK_a units for CM5 and HPA electrostatics, respectively. This error is on par with the above mentioned 4.37 pK_a units of the SMD model. Furthermore the corrective term is able to predict accurate pK_a values for deprotonation reactions involving doubly charged species, a class of ions not present in the MNSOL database, and for which no solvation data has been found. This behavior demonstrates a certain degree of robustness of the corrective term and of the overall model, which should be further confirmed by additional calculations.

6.4 Conclusions

In this chapter, we have presented an extension of a generalized FDPB continuum solvation model to non aqueous solvents in CRYSTAL. A reparametrization of the CDS model for neutral species has been presented and a corrective term based on Abraham's acidity parameter of the solvent for singly-charged species has been proposed for the calculation of free energies of solvation of ions. Testing of the reparametrized model has been performed both on finite molecular systems and on an extended periodic TiO_2 surface slab model, by considering free energies of solvation, pK_a, as well as surface energies and band structures calculations. To perform the reparameterization, 10 levels of theory have been considered by combination of 5 functionals (B3LYP, B3PW, mPW1PW, PBE and PBE0) and two basis sets (6-31G* and 6-311G**). In addition, two atomic charge models, namely HPA and CM5, have been considered to solve the Poisson equation, thus resulting in two separate parameterizations.

For finite molecular systems, the MNSOL database has been considered, with calculation of the solvation energies of 2523 neutral species in 91 solvents, 327 charges species in 4 solvents, as well as 144 transfer energies between water and 14 organic solvents. In addition, to test the robustness of the reparametrization and of the proposed corrective term, calculation of 83 pK_a values in water have been considered. For periodic systems, due to the lack of experimental solvation energies, comparison with reference values obtained with the VASPsol continuum solvation model has been performed in three solvents.

For the solvation energies of 2530 neutral solute/solvent combinations in 91 solvents, most computed MUE are below the 1.00 kcal/mol threshold for acceptable results, on par with the reference SM12 solvation model for which the CDS model has been developed, with low dependence on the level of theory considered. Averaging over the ten levels of theory considered, MUE values of 0.71 ± 0.11 and 0.64 ± 0.09 kcal/mol with CM5 and HPA atomic charges, respectively, have been obtained, that is values comparable to the 0.62 kcal/mol one reported with the SM12 solvation model at the B3LYP/6-31G* level with CM5 atomic charges. Similar conclusions have been drawn for transfer energies between water and 14 organic solvents, with MUE values of 0.71 ± 0.17 and 0.92 ± 0.22 kcal/mol obtained with CM5 and HPA atomic charges, respectively, in line with the SM12 corresponding value of 0.60 kcal/mol at the B3LYP/6-31G* level with CM5 atomic charges.

When considering the 327 solvation energies of ions in four solvents, we found that the inclusion of the proposed corrective term significantly improves the computed solvation energies, with MUE of anions being half of those obtained with the original CDS model, while values obtained for cations were only slightly lowered, bringing the error for both cations and anions close to the experimental uncertainty of 3.00 kcal/mol. The benefit of the proposed corrected term was further confirmed by calculation of aqueous pK_a of 83 molecules, with MUE values with respect to the experimental data of 1.18 and 2.07 pK_a units with CM5 and HPA atomic charges, respectively, significantly improving upon the SMD solvation model for which a 4.37 pK_a units MUE value was repported. In particular, when considering CM5 electrostatics for the calculation of pK_a values, the majority of results laid within a 3.00 pK_a units threshold for which results could be considered acceptable.

Finally, an overall very good agreement has been obtained between computed solvation energies, surface energies as well as band structure changes upon solvation of a TiO_2 anatase periodic slab obtained with the FDPB and VASPsol continuum solvation models in three different solvents, namely water, acetonitrile and toluene. In particular, when going from gas-phase to the different solvents, very similar trends have been obtained with the two solvation models for all data, with surface stabilization, slight band gap opening and upshift of computed bands upon solvation. These effects have been found to be more important with polar solvents such as water and acetonitrile than with an apolar one such as toluene.

Overall, these encouraging results demonstrate that the generalized FDPB continuum solvation model can be applied to a broad range of solutes in various solvents, ranging from finite neutral or charged solutes to extended periodic surfaces.

Chapter 7

Towards electrostatic forces in FDPB

To fully account for solvation effects and possible solute conformational changes when passing from a gaseous to a condensed phase, due to solute-solvent interactions, accurate and fast solvation forces are necessary, as they allow for solute geometry optimizations in solvent.

In the same manner as solvation energies the solvation forces $\mathbf{F}_i^{solv}$, acting on an atom i, can be decomposed in the sum of an electrostatic $\mathbf{F}_i^{el}$ and non-electrostatic $\mathbf{F}_i^{ne}$ contributions, according to:

$$\mathbf{F}_i^{solv} = \mathbf{F}_i^{el} + \mathbf{F}_i^{ne}. \tag{7.1}$$

The non-electrostatic forces $\mathbf{F}_i^{ne}$ for the CDS model can be obtained analytically from the gradient of the non-electrostatic energy $(-\nabla \Delta G_{ne})$, as seen in Sec. 3.2.1 on page 71, allowing for computationally inexpensive and accurate non-electrostatic forces.

On the other hand, the assignment of the electrostatic forces $\mathbf{F}_i^{el}$ in FDPB methods is more problematic [373, 374, 375, 376], especially concerning force conservation. This can be mainly attributed to the need of solving the PB equation numerically when considering complex molecular cavities, for which analytical solutions are not available, and thus the introduction of discretization and numerical errors. As such, many methods have been developed to compute electrostatic solvation forces. These methods usually divide the electrostatic forces in the sum of three components [373, 374]:

$$\mathbf{F}_i^{el} = \mathbf{F}_i^{RF} + \mathbf{F}_i^{Os} + \mathbf{F}_i^{DB}, \tag{7.2}$$

where $\mathbf{F}_i^{RF}$ are the reaction field forces, $\mathbf{F}_i^{Os}$ are forces related to the excess osmotic pressure present when electrolytes ions are dissolved within the solvent, and $\mathbf{F}_i^{DB}$ are the dielectric boundary forces.

The reaction field forces $\mathbf{F}_i^{RF}$ act directly on the solute charges (atoms) and are straightforward to compute [152, 153, 154]. In fact, they simply require the atomic charge q_i and the electric field $\mathbf{E}$ generated by the N_{asc} apparent surface charges q_j at the atomic charge q_i position $\boldsymbol{r_i}$, according to:

$$\mathbf{F}_i^{RF} = q_i \mathbf{E} = q_i \cdot \left(\frac{1}{4\pi\epsilon_0} \sum_j^{N_{asc}} \frac{q_j}{|r_j - r_i|^3} \boldsymbol{r}_{i,j} \right), \tag{7.3}$$

where ϵ_0 is the vacuum relative permittivity and $\boldsymbol{r}_{i,j}$ is the unit vector between the atomic charge q_i and ASC charge q_j. The second term $\mathbf{F}_i^{Os}$ is only present when ions are dissolved within the solvent and acts on the Stern layer, and is usually neglected as its magnitude is smaller compared to other solvation forces [152, 153, 154]. The last term, the dielectric boundary forces F_i^{DB}, is physically related to the pressure asserted by the high dielectric media on the low dielectric region and is given by the negative variation of the electrostatic free energy with respect to the change in the boundary location [373, 152, 153, 154], as such these forces act on the solute/solvent interface and have to be distributed to the solute atoms, further introducing numerical errors [175]. The calculation of accurate Dielectric Boundary Forces (DBF) has been a major challenge and source of errors for FDPB models and many formulations have been proposed throughout the years.

In the Following, we will briefly overview various formulations proposed to compute the DBF for FDPB found in the literature in Sec. 7.1. Instead, in Sec. 7.2 we will focus on the current implementation of the formulation by Cai et al. [153, 154] based on the Maxwell stress tensor for discontinuous dielectric cavities in the Crystal code, and we will conclude with initial tests of the implementation on simple model systems in Sec. 7.3.

7.1 Overview of dielectric boundary forces

The computation of DBF has been one of the major challenges to obtain accurate electrostatic forces in FDPB models, as such many formulations for DBF have been proposed, for both smooth and discontinuous dielectric models.

For smooth dielectric models (see Sec. 2.2.3 on page 34) Gilson et al. [127] developed a variational approach for DBF which is given by the following expression:

$$\mathbf{f}^{DB} = -\frac{1}{8\pi}|\mathbf{E}|^2\nabla\epsilon, \tag{7.4}$$

where $\mathbf{f}_{DB}$ is the DBF surface density, which is related to the negative gradient of the pressure P, and the relation to the net forces acting on a differential surface element dS of the solute cavity is given by:

$$d\mathbf{F}^{DB} = \mathbf{f}^{DB}\, dS. \tag{7.5}$$

The use of Eq. 7.4 requires a relative permittivity ϵ which smoothly varies with the position $\mathbf{r}$. In fact, a discontinuous relative permittivity which is a step function of $\mathbf{r}$, as seen in Sec. 3.1.1 on page 58, would yield a non-finite value for $\nabla\epsilon$. As the FDPB approach implemented in Crystal is based on a discontinuous relative permittivy, in this section we will focus on the latter, discontinuous formulations, for the DBF.

A DBF formulation proposed by Davis and McCammon [375] for discontinous dielectric models is based on the integration of the Maxwell stress tensor, and the DBF surface density is given by:

$$\mathbf{f}^{DB} = -\frac{1}{8\pi}(\epsilon_{out} - \epsilon_{in})(\mathbf{E}_{out}\cdot\mathbf{E}_{in})\mathbf{n} \tag{7.6}$$

where the subscripts "in" and "out" refer to the values of the relative permittivity ϵ and electrical field $\mathbf{E}$ immediately within the atomistic region ($\epsilon = 1$) and the solvent region, respectively; while $\mathbf{n}$ is the normal to the solute cavity facing towards the solvent region.

Another DBF formulation for discontinuous dielectric models, developed by Che et al. [377], is instead based on a variational strategy, under the assumption that the normal component to the DBF is predominant compared to the other contributions. Under this assumption, the DBF is given by:

$$\mathbf{f}^{DB} = -\frac{1}{8\pi}\left(\frac{1}{\epsilon_{out}} - \frac{1}{\epsilon_{in}}\right)|\epsilon\nabla\phi|^2\mathbf{n}, \tag{7.7}$$

where $\epsilon\nabla\phi$ corresponds to the normal dielectric displacement vector at the solute/solvent interface. This formulation was further developed by Li et al. [378], giving the following expression:

$$\mathbf{f}^{DB} = -\frac{1}{4\pi}\left[\epsilon_{out}|\nabla\phi_{out}\cdot\mathbf{n}|^2 - \frac{1}{2}\epsilon_{out}|\nabla\phi_{out}|^2 - \epsilon_{in}|\nabla\phi_{in}\cdot\mathbf{n}|^2 + \frac{1}{2}\epsilon_{in}|\nabla\phi_{in}|^2\right], \tag{7.8}$$

which should be consistent with Eq. 7.6.

Instead, the preliminary implementation of the DBF formulation in the CRYSTAL code follows the work of Cai et al. [153, 154] based on an integral approach of the Maxwell stress tensor, as we will see in the next section.

7.2 Methods

The approach of Cai et al. [153, 154] is built on an integral approach of the Maxwell stress tensor, as the differential approach cannot be used with discontinuous dielectric models due to the non-finite numerical values obtained by the gradient of the relative permittivity, as previously mentioned.

The DBF surface density in this formulation is given by the following expression [153, 154]:

$$\mathbf{f}^{DB} = \frac{1}{4\pi}\left[\left(\epsilon_{out}E^2_{out,n} - \frac{1}{2}\epsilon_{out}\mathbf{E}^2_{out}\right) - \left(\epsilon_{in}E^2_{in,n} - \frac{1}{2}\epsilon_{in}\mathbf{E}^2_{in}\right)\right]\mathbf{n}, \tag{7.9}$$

where the electric field $\mathbf{E}$ and the relative permittivity ϵ are evaluated immediately "in" and "out" of the atomistic region. The unit vector $\mathbf{n}$ is the normal to the surface (facing towards the solvent region), and $E_{in,n}$ and $E_{out,n}$ are the normal components of the electric field, again evaluated "in" and "out" the atomistic region. The full mathematical treatment to obtain Eq. 7.9 can be found in the work of Cai et al. [153].

The advantage of this formulation in a FDPB approach, using an Apparent Surface Charge (ASC) formalism, is that it can be recast into a charge-based strategy [153], for which Eq. 7.9 can be related to the surface charge density σ at the solute/solvent interface and approximated using ASC. The DBF surface density for the charge based approach is given by the following expression:

$$\mathbf{f}^{DB} = \frac{1}{2}\sigma\frac{\epsilon_{in}\mathbf{E}_{out}\cdot\mathbf{E}_{in}}{E_{out,n}}\mathbf{n} = \frac{1}{2}\sigma\frac{\mathbf{D}_{in}\cdot\mathbf{D}_{out}}{D_{out,n}}\mathbf{n}, \tag{7.10}$$

which has been expressed more compactly in terms of the dielectric displacement $\mathbf{D}_{in}$ and $\mathbf{D}_{out}$ on the solute and solvent sides of the cavity, respectively.

As mentioned, the surface charge density σ can be approximated by the ratio between the Apparent Surface Charge (ASC), calculated via the FDPB method, and the underlying surface element; by doing so, the discretized DBF surface density $\mathbf{f}^{DB}_i$ can then be written as:

$$\mathbf{f}^{DB}_i = \frac{1}{2}\frac{q^{asc}_i}{a_i}\cdot\frac{\mathbf{D}_{in}\cdot\mathbf{D}_{out}}{D_{in,n}}\mathbf{n}, \tag{7.11}$$

where q^{asc}_i is the i^{th} ASC and a_i is the relative surface element. The total DBF $\mathbf{F}^{DB}$ acting on whole solute/solvent interface can then be obtained by summing over the N ASC, according to [153, 154]:

$$\mathbf{F}^{DB} = \sum_i^N \mathbf{F}^{DB}_i = \sum_i^N \mathbf{f}^{DB}_i a_i = \frac{1}{2}\sum_i^N q^{asc}_i\frac{\mathbf{D}_{in}\cdot\mathbf{D}_{out}}{D_{in,n}}\mathbf{n}, \tag{7.12}$$

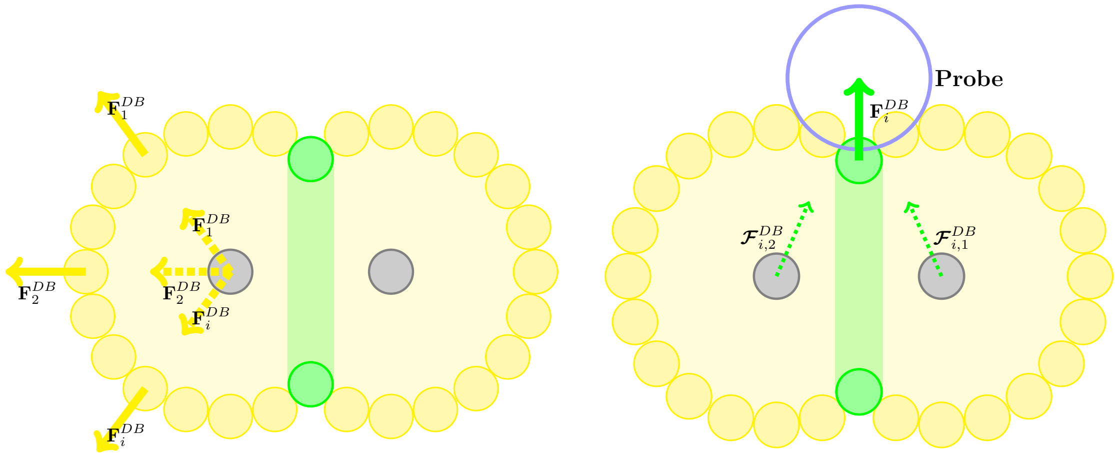
where $\mathbf{F}_i^{DB}$ is the force acting on the i^{th} ASC charge.

Eq. 7.12 can be further simplified in case of solvents with high relative permittivity ϵ, such as water, as the tangential components of the vectors are small in magnitude compared to the normal component ($\approx 1\%$) [153]. Then, in the normal field approximation the total DBF $\mathbf{F}^{DB}$ can be simply expressed by:

$$\mathbf{F}^{DB} = \frac{1}{2}\sum_{i}^{N} q_i^{asc} D_{out,n}\mathbf{n}, \tag{7.13}$$

which should also reduce numerical instabilities and quality of the DBF, especially when considering FDPB calculations with larger grid spacings [153].

As previously mentioned, DBF does not act directly on atoms but on the solute/solvent interface , which in the implicit solvation model is represented by a Solvent Excluded Surface (SES) composed of contact and reentrant segments, see Sec. 2.2.3 on page 34. As such, forces have to be decomposed and assigned to atomic centers [175]. The assignment of DBF from contact surface to atomic center is straightforward, as the force acting on the surface can directly be assigned to the closest atomic center, as shown in Fig. 7.1a. On the other hand, assignment of DBF from reetranct portions of the SES to atomic centers is more problematic, as the forces have to be decomposed and assigned between two different atomic centers.



(a) (b)

Figure 7.1: Two-dimensional section of a SES cavity for a diatomic molecule (atoms in gray). Yellow and green spheres represent ASC assigned to contact and reentrant patches of the SES, respectively. DBF (full lines) are assigned to atomic centers (dotted lines) for (a) contact, and (b) reentrant patches. DBF acting on reentrant patches have to be decomposed and assigned to adjacent atomic centers.

The decomposition of a generic DBF vector $\mathbf{F}_i^{DB}$ acting on the i^{th} reentrant ASC in two vectors, $\boldsymbol{\mathcal{F}}_{i,1}^{DB}$ and $\boldsymbol{\mathcal{F}}_{i,2}^{DB}$ lying on the plane of the ASC and two closest atomic centers, is achieved through the following expressions [175]:

$$\begin{cases} \boldsymbol{\mathcal{F}}_{i,1}^{DB} = \dfrac{\mathbf{F}_i^{DB}\cdot \boldsymbol{f}_{i,1} - \mathbf{F}_i^{DB}\cdot \boldsymbol{f}_{i,2}(\boldsymbol{f}_{i,1}\cdot \boldsymbol{f}_{i,2})}{1-(\boldsymbol{f}_{i,1}\cdot \boldsymbol{f}_{i,2})^2}\boldsymbol{f}_{i,1} \\ \boldsymbol{\mathcal{F}}_{i,2}^{DB} = \dfrac{\mathbf{F}_i^{DB}\cdot \boldsymbol{f}_{i,2} - \mathbf{F}_i^{DB}\cdot \boldsymbol{f}_{i,1}(\boldsymbol{f}_{i,2}\cdot \boldsymbol{f}_{i,1})}{1-(\boldsymbol{f}_{i,2}\cdot \boldsymbol{f}_{i,1})^2}\boldsymbol{f}_{i,2} \end{cases} \tag{7.14}$$

where $\boldsymbol{f}_{i,1}$ and $\boldsymbol{f}_{i,2}$ are unit vectors, and by construction, are directed from the atomic centers to the center of the solvent probe, as shown in Fig. 7.1b.

7.2.1 Electric and Displacement fields calculation

The electric **E** and electric displacement **D** fields required to compute the DBF need the potential ϕ calculated adjacent to the solute/solvent interface. As the potential from the FDPB procedure is calculated only at the FD grid nodes ϕ^{ijk} a first-order least-square interpolation method has been used [153].

The potential at any point in space $\phi(x_0, y_0, z_0)$ is calculated using a potential function Φ of the form:

$$\Phi(x, y, z) = a_0 + a_1(x - x_0) + a_2(y - y_0) + a_3(z - z_0), \tag{7.15}$$

where the coefficients a_0, a_1, a_2, a_3 are obtained from the minimization of the cost function *err*:

$$err = \sum_{i}^{N} |\phi^{ijk}(x_i, y_i, z_i) - \Phi(x_i, y_i, z_i)|^2, \tag{7.16}$$

using $N(\geq 4)$ nearest grid points in the same dielectric region of the selected point (x_0, y_0, z_0). The coefficients are obtained through the Singular Value Decomposition (SVD) algorithm [379] and then related to the potential ϕ and its derivatives with respect to the Cartesian coordinates through the following relations [153]:

$$\begin{cases} \phi(x_0, y_0, z_0) \approx a_0 \\ \dfrac{\partial \phi(x_0, y_0, z_0)}{\partial x} \approx a_1 \\ \dfrac{\partial \phi(x_0, y_0, z_0)}{\partial y} \approx a_2 \\ \dfrac{\partial \phi(x_0, y_0, z_0)}{\partial z} \approx a_3 \end{cases} \tag{7.17}$$

From the potential and its derivatives the electric field **E** and the electric displacement field **D** can be calculated at any given coordinate (x_0, y_0, z_0) using the well known equations [90]:

$$\begin{cases} \mathbf{E}(x_0, y_0, z_0) = -\nabla \phi(x_0, y_0, z_0) \\ \mathbf{D}(x_0, y_0, z_0) = \epsilon \mathbf{E}(x_0, y_0, z_0), \end{cases} \tag{7.18}$$

thus allowing the calculation of the DBF from the values of the fields adjacent to the solute/solvent interface, as seen in Eq. 7.12.

7.3 Model Systems

In this section we validate a first implementation of the electrostatic forces in CRYSTAL's implicit solvation model using two- and three- point charge model systems, a method previously used in the literature [175, 373]. This is done through the comparison of analytical forces, calculated using Eq. 7.2 with DBF from Eq. 7.13, and numerically computed forces obtained from the virtual work principle.

For example, if we consider a system of N atoms, and want to calculate the x component of the forces acting on the i^{th} atom centered at (x_i, y_i, z_i), the force component F_x would be given by the following expression [175]:

$$F_x(\Delta G_{el}) = \frac{\Delta G_{el}(x_i + \delta, y_i, z_i) - \Delta G_{el}(x_i - \delta, y_i, z_i)}{2\delta}, \tag{7.19}$$

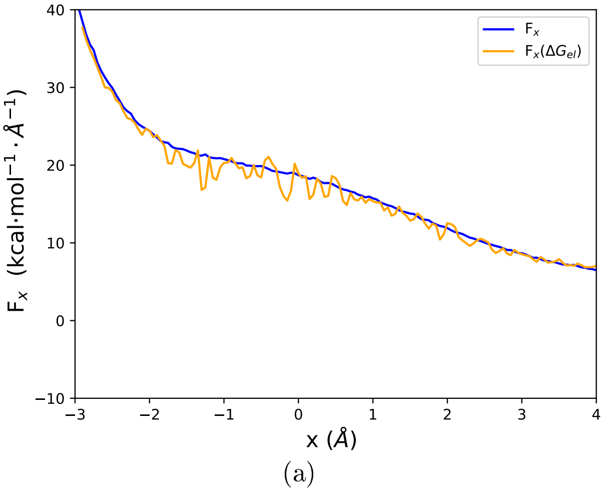
where δ is a small displacement (here set to $\delta = 0.01$Å) and ΔG_{el} is the electrostatic solvation energy of our system of point charges in a given configuration, which depends on the displaced atomic coordinates, as such two calculations are needed for each component. Albeit numerical forces calculated using the virtual work principle are in theory exact, their use is in practice limited by possible numerical noise and their computational cost (as they require $6N$ calculation per degree of freedom). Nevertheless, they allow to qualitatively validate the analytical forces and their current implementation.

For the calculation of the electrostatic solvation energy the default FDPB options, already seen in Sec. 5.2 on page 93, have been used, together with the relative permittivity of water ($\epsilon = 78.36$), and a ionic strength of zero.

7.3.1 Two-atoms System

This system is composed of two positive charged atoms each with a radius of 2.0 Å. The first atom is fixed at $(-3, 0, 0)$ while the second atom is allowed to move from $(-3, 0, 0)$ to $(4, 0, 0)$ with a 0.01 Å displacement. This particular length was chosen as it allows to observe if the SES dissociation, from one cavity to two spherical cavities surrounding the atomic center, will have any effect on the calculation of the forces. The analytical forces F_x and the electrostatic solvation energy ΔG_{el} are computed for every configuration of the model, while numerical forces are calculated using the displaced ΔG_{el}. The goal is to observe if the analytical forces are in line with the numerical ones, and at the same time study the effects of the FD grid spacing on the accuracy of the analytical forces.

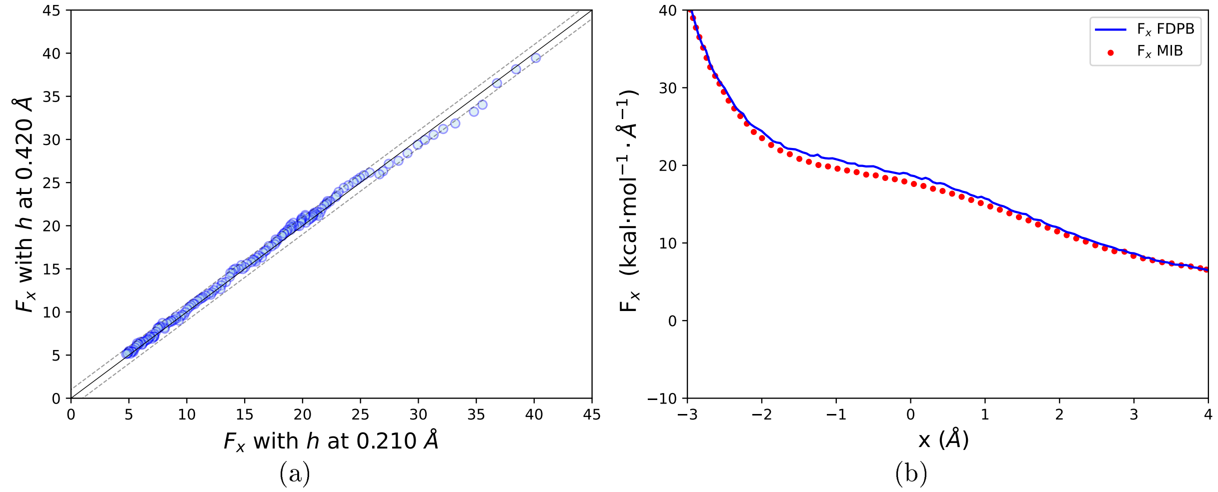
The analytical and numerical forces acting on the moving atom are shown in Fig. 7.2a and Fig. 7.2b with a grid spacing of 0.210 and 0.420 Å, respectively. Clearly, numerical forces show a high degree of noise due to numerical instabilities, especially at higher grid spacings and in the region between (-1,0.0) and (0,0,0). The latter instabilities are due to variations in the reentrant patches of the SES, which are generated following a numerical approach, as seen in Sec. 3.1.3 on page 66. On the other hand, the implemented analytical forces smoothly follow the numerical forces showing a good agreement of the two methods, at least qualitatively. In addition, it is interesting to note that no discontinuities can be observed in the computed analytical forces in the breaking region of the cavity at $x \geq 3.5$.



(a)

(b)

Figure 7.2: Electrostatic forces computed for the moving atom in a two particle model system using a grid spacing of (a) 0.210 Å and (b) 0.410 Å. F_x are the x component of the analytically computed electrostatic forces for the atom moving along the x axis, computed using Eq. 7.19. Instead, $F_x(\Delta G_{el})$ are numerically calculated forces computed using Eq. 7.13.



(a)

(b)

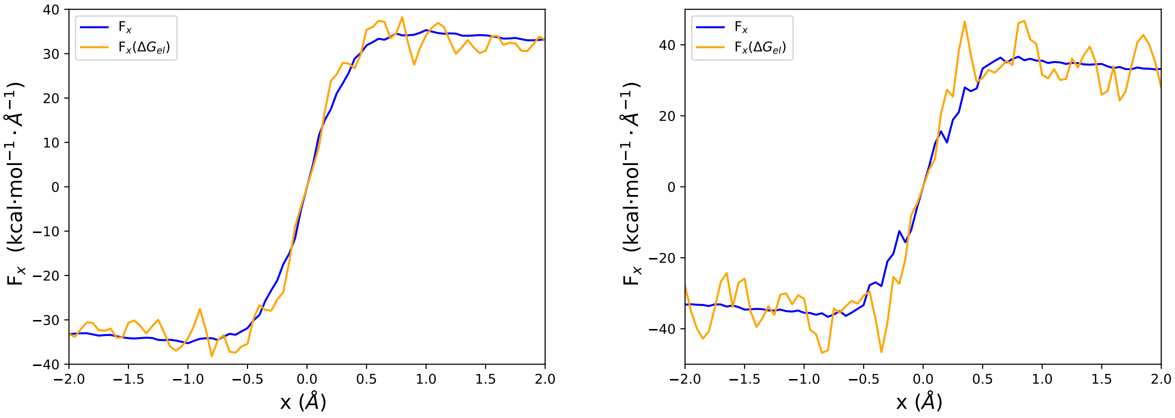
Figure 7.3: (a) Correlation of forces obtained with different grid spacings h for a two point charge model, dotted lines represent a 1 kcal/mol/Å difference. (b) Comparison of electrostatic forces calculated with the FDPB and MIB [175] methods, for a two point charge model system using a grid spacing of 0.210 Å. F_x MIB data has been extracted from Ref. [175].

While the effects of the grid spacing on the solvation energy are minor, and justify the use of larger grid spacings (see Appendix B on page 153), accurate analytical forces require smaller grids to reduce numerical instabilities. This is evident when passing from a larger 0.420 Å to smaller a 0.210 Å grid, as the root-mean square difference between the computed analytical force is of 0.52 kcal/mol/Å, as also shown in 7.3a.

The analytical forces calculated using the FDPB method have also been compared with a more complex mathematical model based on the Matched Interface Boundary (MIB) method [175] using a comparable SES type cavity. Fig.7.3b shows a very good agreement between the two models, with only slight numerical differences in the region between (-2,0,0) and (1,0,0), where only a slight overestimation of the FDPB forces can be observed when compared to the MIB ones.

7.3.2 Three-atoms System

The three point charge systems is composed of three positivity charged atoms each with a radius of 2.0 Å. Two of these atoms are fixed at the coordinate $(0, -1, 0)$ and $(0, 1, 0)$ while the third atom is displaced along the x axis, from (-2,0,0) to (2,0,0) with a 0.01 Å step size. Fig. 7.2a and Fig. 7.2b show the analytical and numerical forces acting on the moving atom calculated using a FD grid spacing of 0.210 and 0.420 Å, respectively.

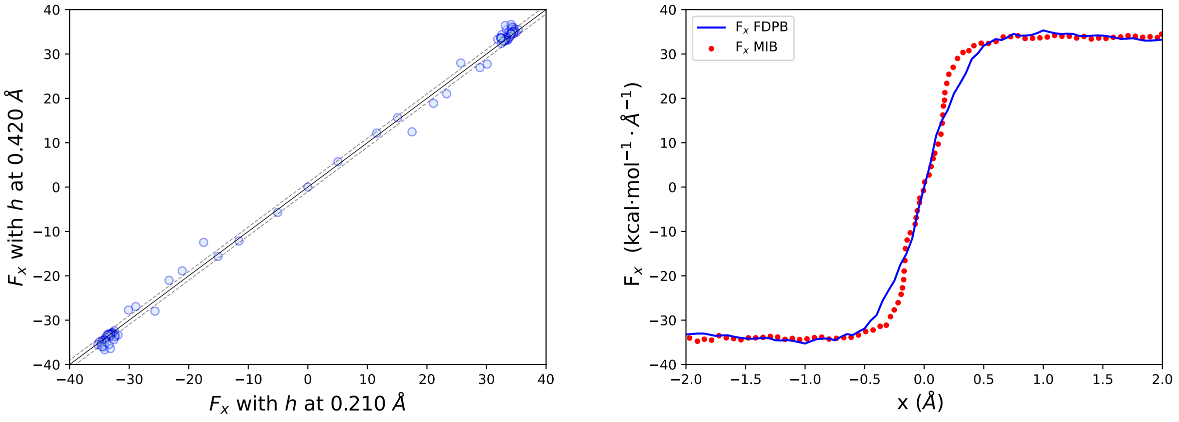


(a) (b)

Figure 7.4: Electrostatic forces computed for the moving atom in a three particle model system using a grid spacing of (a) 0.210 Å and (b) 0.410 Å. F_x are the x component of the analytically computed electrostatic forces for the atom moving along the x axis, computed using Eq. 7.19. Instead, $F_x(\Delta G_{el})$ are numerically calculated forces computed using Eq. 7.13.

As already found in the two point charge model, numerical forces suffer from a high numerical noise, especially at larger grid spacings, and analytical forces qualitatively follow the numerical ones. Albeit, for analytical forces calculated with a grid spacing of 0.420 Å numerical instabilities are present between $(-0.5, 0, 0)$ and $(0.5, 0, 0)$, which can again be attributed to variations of the reentrant patches of the SES. Again, no discontinuities are observed when the cavity breaks. Also, we note that the force profile is symmetric about x=0 Å, as expected from the symmetry of this model system.

Fig. 7.4a shows in greater detail the effects of the FD grid spacing on the calculation of the analytical forces. The root-mean square difference between analytical forces calculated with grid spacing 0.210 Å and 0.420 Å is in fact of 1.38 kcal/mol/A, showing that the effects of the FD grid are also system dependent.



(a) (b)

Figure 7.5: (a) Correlation of forces obtained with different grid spacings h for a three point charge model, dotted lines represent a 1 kcal/mol/Å difference. (b) Comparison of electrostatic forces calculated with the FDPB and MIB [175] methods, for a three point charge model system using a grid spacing of 0.210 Å. F_x MIB data has been extracted from Ref. [175].

As in the two point charge model the analytical forces calculated with FDPB method have been compared to the MIB method [175] using a SES type cavity, as shown in Fig. 7.4b. A good qualitative agreement between MIB and FDPB can be observed, with the major differences between the two methods in the region between $(-0.5, 0, 0)$ and $(0.5, 0, 0)$.

7.4 Molecular Systems

Overall, the computed analytical forces from the FDPB method are in line with the results obtained from both computed numerical forces, and MIB's analytical forces.

As a final example of the application of analytical electrostatic forces, results obtained on the CH_3OH molecule are collected in Tab. 7.1.

		Force components		
n°	atom	x	y	z
1	O	-7.43950	-1.07797	0.00000
2	C	-0.74574	-1.08503	0.00000
3	H	6.68095	-0.30874	-0.00000
4	H	0.16714	1.19917	-0.00000
5	H	0.51042	0.55525	-0.06414
6	H	0.51042	0.55525	0.06414
Components sum:		-0.31631	-0.16206	0.00000

Table 7.1: Electrostatic forces for a CH_3OH molecule, calculated at the B3LYP/6-311G(d,p) level, and a grid spacing of 0.500 Å. Data in kcal/mol/Å.

Although symmetry related atoms (H_5 and H_6 as show in Fig. 7.6) do have symmetry-related values, the sum of the different force component for the system does not equal zero. This can be partially attributed to the numerical instabilities which arise at larger grid spacings, the numerical construction of the reentrant patches of the SES, and DBF decomposition and interpolation of the electrical and displacement fields.

Albeit the results so far are promising, more work is required to both improve and develop more grid independent methods for the calculation of the DBF and of the electrostatic forces in general in the FDPB approach.

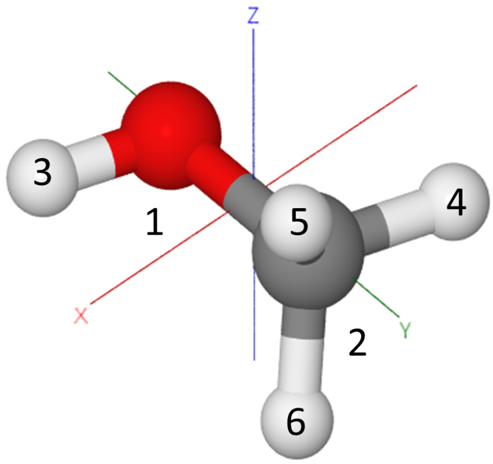


Figure 7.6: Model of the CH_3OH molecule. The geometry was optimized in gas-phase using the B3LYP/6-311G(d,p) level.

Chapter 8

Conclusions and Perspectives

The aim of this thesis was the development, implementation, and improvement of a generalized implicit solvation model in the Crystal code to better model materials by taking into account solvent environmental effects at low computational cost. Both for finite and infinite periodic systems have been considered, at the HF and DFT levels of theory using a linear combination of atom centered Gaussian functions as basis sets.

Throughout the thesis the solvation energy has been used to evaluate the performance of the solvation model, and it has been considered as a sum an electrostatic ΔG_{el} and a non-electrostatic ΔG_{ne} contributions. The electrostatic contribution has been accounted for through a SCRF method using an ASC formalism, where the underlying electrostatic problem is numerically solved through a FD generalized Poisson scheme with the solute charge density approximated as atomic point charges. The non-electrostatic contribution instead accounts for the remaining solvation effects which are not considered by the electrostatic model, and is efficiently computed through a model based on the solute SASA, as the majority of the non-electrostatic effects mainly concert the first solvation shell.

An analytical approach to compute the solute SASA and its nuclear gradients based on a stereographic projection technique, valid for both finite and infinite periodic systems, has been implemented as discussed in chapter 4 on page 75. For finite systems the SASA implementation has been compared to reference analytical values with an excellent agreement, while for infinite periodic systems successful size-extensivity tests have been performed. The correctness of the analytical gradients has instead been confirmed by the excellent agreement obtained by comparison to numerical gradients and the translational invariance found for both finite and infinite periodic systems. The overall results show that the stereographic projection technique is a general, simple, and efficient way to compute the SASA, which is a basic component of many accurate and efficient non-electrostatic solvation models.

As the source term for the generalized Poisson equation is approximated through atomic point charges, the FDPB procedure and the computation of the electrostatic energy is charge model dependent. In chapter 5 on page 91 the effects of different charge and non-electrostatic models on the solvation energy has been investigated on a test set of 501 neutral and 112 charged molecules in water. For the electrostatic contribution five different charge models were considered: Mulliken, Hirshfeld (HPA), Hirshfeld-I (HPA-I), CM5, and CM5-I, calculated using B3LYP and the HF level of theory with the 6-31G, 6-311G(d,p) and 6-311++G(d,p) basis sets. Instead for the non-electrostatic model two approaches based on the solute SASA have been explored: the αSASA and the CDS models. The latter using two separate set of parameters obtained from the literature: CDS-CM5 and CDS-ESP, originally developed for a GBA based solvation model.

The values of electrostatic energy for the test set varies greatly with both the charge models and the level of theory. On average, electrostatic energies obtained with Hirshfeld charge model are the lowest, while values calculated with Hirshfeld-I are the highest, with the remaining charge models falling in-between the two. From the perspective of the level of theory the HF method globally gives more negative electrostatic energies compared to B3LYP due to overpolarization effects, and the least basis set dependent charge model is, from lowest to highest: Hirshfeld<CM5<CM5-I<Hirshfeld-I«Mulliken, with the latter model giving unphysical solvation energies for larger basis sets. Furthermore, Mulliken charge model shows technical issues in the SCRF convergence procedure compared to the other charge models. This increases the total number of SCRF cycles needed for convergence and hinders the model performance.

When pairing the electrostatic contribution computed with the various charge models with the αSASA, CDS-ESP and CDS-CM5 non-electrostatics models for neutral molecules only a limited number of combinations achieved a MUE lower than 1.00 kcal/mol, most notably the Hirshfeld charge model with CDS-ESP. In general, the HF method overestimation of the electrostatic energy translates to a higher MUE compared to B3LYP and better performances are observed when using basis sets which include polarization functions. The high MUE for the CM5 when coupled with the CDS-CM5 also highlights the low transferability of the CDS parameters from a GBA to a FDPB approach.

To improve the performance of the CM5 charges using a FD scheme and the default CDS-CM5 parameters a scaling factor was applied. The value of the scaling has been obtained by minimizing the MUE at the B3LYP/6-311G(d,p) level of theory on the 501 neutral molecules considered as a test set. The process has yielded two minima with MUE of 0.68 and 0.74 kcal/mol with a scaling factor of 0.52 and 0.73 for the CDS-CM5 and CDS-ESP models, respectively.

For charged species, Mulliken and the 0.52·CM5 charge model performed poorly, while the remaining charge models performance is strongly affected by the level of theory. As opposed to neutral molecules, the HF overpolarization effect contributes to a lower MUE compared to B3LYP for anions. Good results are obtained when considering CM5 charge with CDS-CM5 parameters. In fact, this combinations achieves a MUE of only 2.62 kcal/mol, a value within the experimental uncertainly of the test set of 3.00 kcal/mol. In general, we can conclude that overpolarization and charge model with higher charge magnitude are able to achieve lower MUE for negatively charged species and should be preferred. On the other hand the performance for cations is well within the 3.00 kcal/mol for all the combinations of charge and non-electrostatic models, with the exception of Mulliken and 0.52 · CM5 charge models.

Albeit a good accuracy has been reached in the prediction of solvation energies using the 0.52 ·CM5 charges, the results are limited to neutral molecules, as charged species have shown to be extremely sensitive to variations in charge magnitudes and strongly dependent on the level of theory. Furthermore, the scaling has been verified only for aqueous solvent. This highlighted the need of re-optimizing the CDS model parameters and the development of charged specific corrections, while at the same time trying to minimize the dependence of the model on the level of theory.

In chapter 5 on page 91 the FDPB implicit solvation model was extended to non aqueous solvents, and the CDS model was reparametrized for neutral species and extended for charged species by considering an additional corrective term based on Abraham's acidity parameter of the solvent. Testing of the reparametrized and extended CDS model was performed considering solvation energies for finite molecular systems, transfer energies between organic solvents and water, pK_a calculations in aqueous solvent, and band structure calculations on a TiO_2 surface slab model.

The reparametrization of the CDS model is based on 10 levels of theory given by the combination of 5 functionals (B3LYP, B3PW, mPW1PW, PBE and PBE0) and 2 basis sets (6-31G* and 6-311G**). In addition, based on the results of Chapter 4, the HPA and CM5 charge models have been considered to solve the generalized Poisson equation due to their lower dependence on the level of theory, and higher stability, compared to other charge models. This resulted in an independent set of parameters for each of the two charge models considered.

For the calculation of the solvation energies of finite molecular systems the MNSOL database has been used. This database consists of 2523 neutral species, 91 solvents, and 327 charged species in 4 solvents, as well as 144 transfer free energies between water and 14 organic solvents. To further test the robustness of the reparametrized and extended CDS model an additional 83 pK_a calculations in water have been performed. Instead for periodic systems, due to the lack of reliable experimental solvation energies, the results from the FDPB model have been compared with VASPsol's continuum solvation model using 3 different solvents.

Starting with neutral species, very good results have been obtained for 2523 neutral solute/solvent combinations in 91 solvents. In fact, most of the computed MUE fall within the 1.00 kcal/mol threshold for acceptable results, and there is an overall good agreement with the SM12 solvation model used as reference. Furthermore, a very low dependence on the level of theory has been observed for all 10 levels of theory, justifying the choice of using CM5 and HPA charges. Considering an average over the 10 levels of theory, the MUE of the whole test set is 0.71 ± 0.01 and 0.64 ± 0.01 kcal/mol with CM5 and HPA atomic charges, respectively, and are comparable to the 0.62 kcal/mol reported from the SM12 solvation model at the B3LYP/6-31G* level with CM5 atomic charges.

Similar conclusions to neutral molecules can be drawn out when considering the 144 transfer energies between water and 14 organic molecules. The MUE values are of 0.71 ± 0.02 and 0.86 ± 0.01 kcal/mol with CM5 and HPA atomic charges, respectively. These values are again in line with the SM12 model value of 0.60 kcal/mol at the B3LYP/6-31G* level with CM5 atomic charges.

For the 327 charged species in four different solvents, the inclusion of a corrective term based on Abraham's acidity parameter of the solvent significantly improves the computation of solvation energies at the DFT level of theory. In fact, the MUE for anions is halved compared to those obtained with the original CDS model, while the MUE for cations is slightly reduced. For both cations and anions the computed MUE is close to the experimental uncertainty of 3.00 kcal/mol.

The corrective term was further tested by performing pK_a calculations in aquoeus solvent for 83 molecules. The MUE values with respect to the experimental data is of 1.18 and 2.09 pK_a units with CM5 and HPA atomic charges, respectively. These values significantly improves upon the default SMD solvation model using CDS, for which a 4.37 pK_a units MUE value was reported. In particular, when using CM5 charges, the majority of pK_a values fall within a 3.00 pK_a units threshold for which results can be considered acceptable.

Considering the TiO_2 anatase slab, a very good agreement has also been found for computed solvation energies and band structure changes upon solvation between CRYSTAL's FDPB and VASPsol's implicit solvation model. When passing from gas-phase to water, acetonitrile or toluene very similar trends can be observed for both solvation energies, surface stabilization, and slight band gap opening and up shifts of the computed bands upon solvation. All these effects have been found to be more significant with more polar solvents.

These results show that the extended and reparametrized FDPB model is able to achieve accurate solvation energies for a wide range of different solvents and systems, while at the same time exhibiting a low dependence from the level of theory demonstrating the model robustness. Having achieved accurate solvation energies, solvation forces for solute geometry optimization in solvent are needed as a final step to fully complete the FDPB model.

In the current implementation of the solvation forces, described in Chapter 7 on page 129, are decomposed in the sum a non-electrostatic and electrostatic contribution. Accurate and efficient non-electrostatic forces are obtained analytically from the gradient of the CDS non-electrostatic energy, while analytical electrostatic forces are obtained from a charged based formulation developed by Cai et al. [153] using an integral approach of the Maxwell stress tensor.

The analytical electrostatic forces have been tested on two- and three-point charge model systems, and have shown to be qualitatively in line with results obtained from both numerically computed forces and results from the MIB implicit solvation model [**?**], validating the overall implementation of the forces.

Nevertheless, force conservation remains an issue when passing to finite molecular systems. This can be attributed to grid dependent discretization and numerical errors which can be traced back to: (i) the numerical construction of the reentrant patches of the SES, (ii) the decomposition of the dielectric boundary force to the two closest atomic centers, and (iii) the one-sided interpolation of the electric and displacement fields adjacent to the solute/solvent interface. As such, to achieve accurate electrostatic forces more work is still required to improve Cai et al. [153] formulation by decreasing its dependence on the FD grid through the use of more accurate interpolation techniques and by improving the construction of the reetrant patches of the SES.

Overall, the implicit solvation model has shown to be effective and robust, being able to compute solvation energies efficiently for many systems in different solvent environments. Together with the improvement of electrostatic forces, work has also begun on applications of the model to systems of chemical interest, such as the reaction of carbon monoxide and water to form hydrogen and carbon dioxide on a Pt(111) slab (also know as water gas shift reaction) considering an aqueous solvent in order to account for environmental effects and more accurately model the reaction.

Appendices

Appendix A

Partial derivatives of the curve integrals for SASA gradients

This section gives the expressions of all partial derivatives involved in the calculation of Eq. (4.15) presented in Section 4.1.2. Each of the three contributions shown in this equation are successively given below.

A.1 Evaluation of $\frac{\partial I^i_{j,\lambda}}{\partial\left(a^i_j, b^i_j, c^i_j, d^i_j\right)}$

When the j-th circle on the border of Ω_i contains no vertices, by using Eq. (4.8), one obtains:

$$\begin{cases} \dfrac{\partial I^i_{j,\lambda}}{\partial a^i_j} = 8\pi r_i^4 \cdot \dfrac{4r_i^2({b^i_j}^2 + {c^i_j}^2 - 2a^i_j d^i_j) + 2{d^i_j}^2}{{V_j}^{i3}} \\ \dfrac{\partial I^i_{j,\lambda}}{\partial b^i_j} = -8\pi r_i^4 b^i_j \cdot \dfrac{d^i_j + 4r_i^2 a^i_j}{{V_j}^{i3}} \\ \dfrac{\partial I^i_{j,\lambda}}{\partial c^i_j} = -8\pi r_i^4 c^i_j \cdot \dfrac{d^i_j + 4r_i^2 a^i_j}{{V_j}^{i3}} \\ \dfrac{\partial I^i_{j,\lambda}}{\partial d^i_j} = 8\pi r_i^4 \cdot \dfrac{{b^i_j}^2 + {c^i_j}^2 - 2a^i_j d^i_j + 8r_i^2 {a^i_j}^2}{{V_j}^{i3}} \end{cases} \tag{A.1}$$

On the other hand, when the j-th circle on the border of Ω_i contains vertices, by using Eq. (4.5), one obtains:

$$\begin{aligned} \frac{\partial I^i_{j,\lambda}}{\partial\left(a^i_j, b^i_j, c^i_j, d^i_j\right)} = r_i^2 \Bigg[& \left(\frac{\partial \alpha^i_{j,\lambda}}{\partial\left(a^i_j, b^i_j, c^i_j, d^i_j\right)} - \frac{\partial \beta^i_{j,\lambda}}{\partial\left(a^i_j, b^i_j, c^i_j, d^i_j\right)} \right) \cdot \operatorname{sign}\left(a^i_j\right) + \partial\left(\frac{d^i_j + 4r_i^2 a^i_j}{V^i_j} \right) / \partial\left(a^i_j, b^i_j, c^i_j, d^i_j\right) \cdot \\ & \left(\pi - 2\arctan \frac{U^i_j}{2{a^i_j}^2 V^i_j \sin\left((\beta^i_{j,\lambda} - \alpha^i_{j,\lambda})/2\right)} \right) - \frac{d^i_j + 4r_i^2 a^i_j}{V^i_j} \cdot \\ & \partial\left(\frac{U^i_j}{{a^i_j}^2 V^i_j \sin\left((\beta^i_{j,\lambda} - \alpha^i_{j,\lambda})/2\right)} \right) / \partial\left(a^i_j, b^i_j, c^i_j, d^i_j\right) \cdot \frac{1}{1 + \left(\dfrac{U^i_j}{2{a^i_j}^2 V^i_j \sin\left((\beta^i_{j,\lambda} - \alpha^i_{j,\lambda})/2\right)} \right)^2} \Bigg] \end{aligned} \tag{A.2}$$

In the following the four different partial derivatives appearing in Eq. (A.2) are given.

As suggested in Ref. [331], the following expression is used and derivated with respect to a_j^i, b_j^i, c_j^i and d_j^i:

$$\begin{aligned}&\left(a_j^i b_k^i - a_k^i b_j^i\right)^2 + \left(a_j^i c_k^i - a_k^i c_j^i\right)^2 + \left({b_j^i}^2 + {c_j^i}^2 - 4a_j^i d_j^i\right) \cdot {a_k^i}^2 - \left({b_k^i}^2 + {c_k^i}^2 - 4a_k^i d_k^i\right) \cdot {a_j^i}^2 \\ &+ 2({b_j^i}^2 + {c_j^i}^2 - 4a_j^i d_j^i)^{-1/2} \cdot a_k^i \cdot \operatorname{sign} a_j^i \cdot \left(\left(a_j^i b_k^i - a_k^i b_j^i\right) \cos \alpha_{j,\lambda}^i + \left(a_j^i c_k^i - a_k^i c_j^i\right) \sin \alpha_{j,\lambda}^i\right) = 0\end{aligned} \tag{A.3}$$

to get:

$$\left\{\begin{array}{l}
\dfrac{\partial \alpha_{j,\lambda}^i}{\partial a_j^i} = \dfrac{1}{D_1} \cdot \Big[b_k^i (a_j^i b_k^i - a_k^i b_j^i) + c_k^i (a_j^i c_k^i - a_k^i c_j^i) - 2d_j^i {a_k^i}^2 - a_j^i ({b_k^i}^2 + {c_k^i}^2 - 4a_k^i d_k^i) + \\
\qquad (b_k^i \cos \alpha_{j,\lambda}^i + c_k^i \sin \alpha_{j,\lambda}^i) \cdot a_k^i \operatorname{sign} a_j^i ({b_j^i}^2 + {c_j^i}^2 - 4a_j^i d_j^i)^{1/2} \\
\qquad - \dfrac{2d_j^i a_k^i \operatorname{sign} a_j^i}{({b_j^i}^2 + {c_j^i}^2 - 4a_j^i d_j^i)^{1/2}} \cdot \left(\cos \alpha_{j,\lambda}^i (a_j^i b_k^i - a_k^i b_j^i) + \sin \alpha_{j,\lambda}^i (a_j^i c_k^i - a_k^i c_j^i)\right) \Big] \\[2ex]
\dfrac{\partial \alpha_{j,\lambda}^i}{\partial b_j^i} = \dfrac{1}{D_1} \cdot \Big[-a_k^i (a_j^i b_k^i - a_k^i b_j^i) + b_j^i {a_k^i}^2 + \dfrac{b_j^i}{({b_j^i}^2 + {c_j^i}^2 - 4a_j^i d_j^i)^{1/2}} \cdot a_k^i \operatorname{sign} a_j^i \\
\qquad \left((a_j^i b_k^i - a_k^i b_j^i) \cos \alpha_{j,\lambda}^i + (a_j^i c_k^i - a_k^i c_j^i) \sin \alpha_{j,\lambda}^i\right) - ({b_j^i}^2 + {c_j^i}^2 - 4a_j^i d_j^i)^{1/2} {a_k^i}^2 \operatorname{sign} a_j^i \cos \alpha_{j,\lambda}^i \Big] \\[2ex]
\dfrac{\partial \alpha_{j,\lambda}^i}{\partial c_j^i} = \dfrac{1}{D_1} \cdot \Big[-a_k^i (a_j^i c_k^i - a_k^i c_j^i) + c_j^i {a_k^i}^2 + \dfrac{c_j^i}{({b_j^i}^2 + {c_j^i}^2 - 4a_j^i d_j^i)^{1/2}} \cdot a_k^i \operatorname{sign} a_j^i \\
\qquad \left((a_j^i b_k^i - a_k^i b_j^i) \cos \alpha_{j,\lambda}^i + (a_j^i c_k^i - a_k^i c_j^i) \sin \alpha_{j,\lambda}^i\right) - ({b_j^i}^2 + {c_j^i}^2 - 4a_j^i d_j^i)^{1/2} {a_k^i}^2 \operatorname{sign} a_j^i \sin \alpha_{j,\lambda}^i \Big] \\[2ex]
\dfrac{\partial \alpha_{j,\lambda}^i}{\partial d_j^i} = \dfrac{1}{D_1} \cdot \Big[-2a_j^i {a_k^i}^2 - \dfrac{2|a_j^i|}{({b_j^i}^2 + {c_j^i}^2 - 4a_j^i d_j^i)^{1/2}} a_k^i \left((a_j^i b_k^i - a_k^i b_j^i) \cos \alpha_{j,\lambda}^i + (a_j^i c_k^i - a_k^i c_j^i) \sin \alpha_{j,\lambda}^i\right) \Big]
\end{array}\right. \tag{A.4}$$

where: $D_1 = ({b_j^i}^2 + {c_j^i}^2 - 4a_j^i d_j^i)^{1/2} a_k^i \operatorname{sign} a_j^i \cdot \left[(a_j^i b_k^i - a_k^i b_j^i) \sin \alpha_{j,\lambda}^i - (a_j^i c_k^i - a_k^i c_j^i) \cos \alpha_{j,\lambda}^i\right]$.

The expressions of $\dfrac{\partial \beta_{j,\lambda}^i}{\partial a_j^i}$, $\dfrac{\partial \beta_{j,\lambda}^i}{\partial b_j^i}$, $\dfrac{\partial \beta_{j,\lambda}^i}{\partial c_j^i}$ and $\dfrac{\partial \beta_{j,\lambda}^i}{\partial d_j^i}$ can be obtained from Eq. (A.4) by substituting k with l and α with β.

The partial derivatives of $\dfrac{d^i_j + 4r_i^2 a^i_j}{V^i_j}$ are given by:

$$\begin{cases} \partial\left(\dfrac{d^i_j + 4r_i^2 a^i_j}{V^i_j}\right)/\partial a^i_j = \dfrac{4r_i^2}{V^i_j} - (d^i_j + 4r_i^2 a^i_j)\dfrac{\partial V^i_j}{a^i_j}\dfrac{1}{{V^i_j}^2} \\ \partial\left(\dfrac{d^i_j + 4r_i^2 a^i_j}{V^i_j}\right)/\partial b^i_j = -(d^i_j + 4r_i^2 a^i_j)\dfrac{\partial V^i_j}{b^i_j}\dfrac{1}{{V^i_j}^2} \\ \partial\left(\dfrac{d^i_j + 4r_i^2 a^i_j}{V^i_j}\right)/\partial c^i_j = -(d^i_j + 4r_i^2 a^i_j)\dfrac{\partial V^i_j}{c^i_j}\dfrac{1}{{V^i_j}^2} \\ \partial\left(\dfrac{d^i_j + 4r_i^2 a^i_j}{V^i_j}\right)/\partial d^i_j = \dfrac{1}{V^i_j} - (d^i_j + 4r_i^2 a^i_j)\dfrac{\partial V^i_j}{d^i_j}\dfrac{1}{{V^i_j}^2} \end{cases} \tag{A.5}$$

where, from Eq. (4.6):

$$\begin{cases} \dfrac{\partial V^i_j}{\partial a^i_j} = \dfrac{4r_i^2(4r_i^2 - d^i_j)}{V^i_j} \\ \dfrac{\partial V^i_j}{\partial b^i_j} = \dfrac{4r_i^2 b^i_j}{V^i_j} \\ \dfrac{\partial V^i_j}{\partial c^i_j} = \dfrac{4r_i^2 c^i_j}{V^i_j} \\ \dfrac{\partial V^i_j}{\partial d^i_j} = 8\pi r_i^4 \cdot \dfrac{{b^i_j}^2 + {c^i_j}^2 - 2a^i_j d^i_j + 8r_i^2 {a^i_j}^2}{{V^i_j}^3} \end{cases} \tag{A.6}$$

The partial derivatives of $\frac{U_j}{{a^i_j}^2 V^i_j \sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}$ are more complicated and they are thus presented separately below:

$$\begin{aligned} &\partial\left(\frac{U_j}{{a^i_j}^2 V^i_j \sin\left((\beta^i_{j,\lambda} - \alpha^i_{j,\lambda})/2\right)}\right)/\partial a^i_j \\ &= \partial\left(\frac{|a^i_j|({b^i_j}^2 + {c^i_j}^2 - 2a^i_j d^i_j + 8r_i^2 {a^i_j}^2)}{{a^i_j}^2 V^i_j}\right)/\partial a^i_j \cdot \frac{1}{\tan\left((\beta^i_{j,\lambda} - \alpha^i_{j,\lambda})/2\right)} \\ &- \frac{|a^i_j|({b^i_j}^2 + {c^i_j}^2 - 2a^i_j d^i_j + 8r_i^2 {a^i_j}^2)}{2{a^i_j}^2 V^i_j} \cdot \left(\frac{\partial \beta^i_{j,\lambda}}{\partial a^i_j} - \frac{\partial \alpha^i_{j,\lambda}}{\partial a^i_j}\right) \cdot \frac{1}{\sin^2\left((\beta^i_{j,\lambda} - \alpha^i_{j,\lambda})/2\right)} \\ &- \partial\left(\frac{a^i_j({b^i_j}^2 + {c^i_j}^2 - 4a^i_j d^i_j)^{1/2}}{{a^i_j}^2 V^i_j}\right)/\partial a^i_j \cdot \frac{\left(b^i_j \cos\left((\beta^i_{j,\lambda} + \alpha^i_{j,\lambda})/2\right) + c^i_j \sin\left((\beta^i_{j,\lambda} + \alpha^i_{j,\lambda})/2\right)\right)}{\sin\left((\beta^i_{j,\lambda} - \alpha^i_{j,\lambda})/2\right)} \\ &+ \frac{a^i_j({b^i_j}^2 + {c^i_j}^2 - 4a^i_j d^i_j)^{1/2}}{2{a^i_j}^2 V^i_j \sin\left((\beta^i_{j,\lambda} - \alpha^i_{j,\lambda})/2\right)} \cdot \\ &\left[\left(\frac{\partial \beta^i_{j,\lambda}}{\partial a^i_j} - \frac{\partial \alpha^i_{j,\lambda}}{\partial a^i_j}\right) \cdot \frac{\left(b^i_j \cos\left((\beta^i_{j,\lambda} + \alpha^i_{j,\lambda})/2\right) + c^i_j \sin\left((\beta^i_{j,\lambda} + \alpha^i_{j,\lambda})/2\right)\right)}{\tan\left((\beta^i_{j,\lambda} - \alpha^i_{j,\lambda})/2\right)} \cdot \right. \\ &\left. + \left(\frac{\partial \beta^i_{j,\lambda}}{\partial a^i_j} + \frac{\partial \alpha^i_{j,\lambda}}{\partial a^i_j}\right) \cdot \left(b^i_j \sin\left((\beta^i_{j,\lambda} + \alpha^i_{j,\lambda})/2\right) - c^i_j \cos\left((\beta^i_{j,\lambda} + \alpha^i_{j,\lambda})/2\right)\right)\right] \end{aligned} \tag{A.7}$$

$$
\begin{aligned}
&\partial\left(\frac{U_j}{{a_j^i}^2 V_j^i \sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right)/\partial b_j^i\\
&=\partial\left(\frac{|a_j^i|({b_j^i}^2+{c_j^i}^2-2a_j^i d_j^i+8r_i^2{a_j^i}^2)}{{a_j^i}^2V_j^i}\right)/\partial b_j^i\cdot\frac{1}{\tan\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\\
&-\frac{|a_j^i|({b_j^i}^2+{c_j^i}^2-2a_j^i d_j^i+8r_i^2{a_j^i}^2)}{2{a_j^i}^2V_j^i}\cdot\left(\frac{\partial\beta^i_{j,\lambda}}{\partial b_j^i}-\frac{\partial\alpha^i_{j,\lambda}}{\partial b_j^i}\right)\cdot\frac{1}{\sin^2\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\\
&-\partial\left(\frac{a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^i d_j^i)^{1/2}}{{a_j^i}^2V_j^i}\right)/\partial b_j^i\cdot\frac{\left(b_j^i\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)+c_j^i\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)}{\sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\\
&+\frac{a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^i d_j^i)^{1/2}}{2{a_j^i}^2V_j^i\sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\cdot\\
&\Bigg[-2\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)+\left(\frac{\partial\beta^i_{j,\lambda}}{\partial b_j^i}-\frac{\partial\alpha^i_{j,\lambda}}{\partial b_j^i}\right)\cdot\frac{\left(b_j^i\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)+c_j^i\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)}{\tan\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\\
&+\left(\frac{\partial\beta^i_{j,\lambda}}{\partial b_j^i}+\frac{\partial\alpha^i_{j,\lambda}}{\partial b_j^i}\right)\cdot\left(b_j^i\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)-c_j^i\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)\Bigg]
\end{aligned}
\tag{A.8}
$$

$$
\begin{aligned}
&\partial\left(\frac{U_j}{{a_j^i}^2 V_j^i \sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right)/\partial c_j^i\\
&=\partial\left(\frac{|a_j^i|({b_j^i}^2+{c_j^i}^2-2a_j^i d_j^i+8r_i^2{a_j^i}^2)}{{a_j^i}^2V_j^i}\right)/\partial c_j^i\cdot\frac{1}{\tan\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\\
&-\frac{|a_j^i|({b_j^i}^2+{c_j^i}^2-2a_j^i d_j^i+8r_i^2{a_j^i}^2)}{2{a_j^i}^2V_j^i}\cdot\left(\frac{\partial\beta^i_{j,\lambda}}{\partial c_j^i}-\frac{\partial\alpha^i_{j,\lambda}}{\partial c_j^i}\right)\cdot\frac{1}{\sin^2\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\\
&-\partial\left(\frac{a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^i d_j^i)^{1/2}}{{a_j^i}^2V_j^i}\right)/\partial c_j^i\cdot\frac{\left(b_j^i\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)+c_j^i\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)}{\sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\\
&+\frac{a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^i d_j^i)^{1/2}}{2{a_j^i}^2V_j^i\sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\cdot\\
&\Bigg[-2\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)+\left(\frac{\partial\beta^i_{j,\lambda}}{\partial c_j^i}-\frac{\partial\alpha^i_{j,\lambda}}{\partial c_j^i}\right)\cdot\frac{\left(b_j^i\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)+c_j^i\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)}{\tan\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\\
&+\left(\frac{\partial\beta^i_{j,\lambda}}{\partial c_j^i}+\frac{\partial\alpha^i_{j,\lambda}}{\partial c_j^i}\right)\cdot\left(b_j^i\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)-c_j^i\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)\Bigg]
\end{aligned}
\tag{A.9}
$$

$$
\begin{aligned}
&\partial\left(\frac{U_j}{{a^i_j}^2 V^i_j \sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right)/\partial d^i_j \\
&= \partial\left(\frac{|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2)}{{a^i_j}^2V^i_j}\right)/\partial d^i_j\cdot\frac{1}{\tan\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)} \\
&-\frac{|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2)}{2{a^i_j}^2V^i_j}\cdot\left(\frac{\partial\beta^i_{j,\lambda}}{\partial d^i_j}-\frac{\partial\alpha^i_{j,\lambda}}{\partial d^i_j}\right)\cdot\frac{1}{\sin^2\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)} \\
&-\partial\left(\frac{a^i_j({b^i_j}^2+{c^i_j}^2-4a^i_jd^i_j)^{1/2}}{{a^i_j}^2V^i_j}\right)/\partial d^i_j\cdot\frac{\left(b^i_j\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)+c^i_j\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)}{\sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)} \\
&+\frac{a^i_j({b^i_j}^2+{c^i_j}^2-4a^i_jd^i_j)^{1/2}}{2{a^i_j}^2V^i_j\sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\cdot \\
&\left[\left(\frac{\partial\beta^i_{j,\lambda}}{\partial d^i_j}-\frac{\partial\alpha^i_{j,\lambda}}{\partial d^i_j}\right)\cdot\frac{\left(b^i_j\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)+c^i_j\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)}{\tan\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right. \\
&\left.+\left(\frac{\partial\beta^i_{j,\lambda}}{\partial d^i_j}+\frac{\partial\alpha^i_{j,\lambda}}{\partial d^i_j}\right)\cdot\left(b^i_j\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)-c^i_j\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)\right]
\end{aligned}
\tag{A.10}
$$

Evaluation of these last derivatives therefore requires to evaluate the partial derivatives of $\frac{|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2)}{{a^i_j}^2V^i_j}$ and $\frac{a^i_j({b^i_j}^2+{c^i_j}^2-4a^i_jd^i_j)^{1/2}}{{a^i_j}^2V^i_j}$ which are given by:

$$
\left\{
\begin{aligned}
&\partial\left(\frac{|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2)}{{a^i_j}^2V^i_j}\right)/\partial a^i_j=\frac{1}{\left({a^i_j}^2V^i_j\right)^2}\cdot\left({a^i_j}^2V^i_j\left(\frac{a^i_j}{|a^i_j|}({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2)\right.\right. \\
&\qquad\left.\left.+|a^i_j|(-2d^i_j+4\cdot4r_i^2)\right)\right)-|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2)\cdot \\
&\qquad\left(2a^i_jV^i_j+{a^i_j}^2\frac{\partial V^i_j}{\partial a^i_j}\right) \\
&\partial\left(\frac{|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2)}{{a^i_j}^2V^i_j}\right)/\partial b^i_j=\frac{1}{\left({a^i_j}^2V^i_j\right)^2}\cdot\left(2b^i_j|a^i_j|{a^i_j}^2V^i_j\right. \\
&\qquad\left.-|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2){a^i_j}^2\frac{\partial V^i_j}{\partial b^i_j}\right) \\
&\partial\left(\frac{|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2)}{{a^i_j}^2V^i_j}\right)/\partial c^i_j=\frac{1}{\left({a^i_j}^2V^i_j\right)^2}\cdot\left(2c^i_j|a^i_j|{a^i_j}^2V^i_j\right. \\
&\qquad\left.-|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2){a^i_j}^2\frac{\partial V^i_j}{\partial c^i_j}\right) \\
&\partial\left(\frac{|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2)}{{a^i_j}^2V^i_j}\right)/\partial d^i_j=\frac{1}{\left({a^i_j}^2V^i_j\right)^2}\cdot\left(-2a^i_j|a^i_j|{a^i_j}^2V^i_j\right. \\
&\qquad\left.-|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2){a^i_j}^2\frac{\partial V^i_j}{\partial d^i_j}\right)
\end{aligned}
\right.
\tag{A.11}
$$

and:

$$\left\{\begin{aligned}
&\partial\left(\frac{a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}}{{a_j^i}^2V_j^i}\right)/\partial a_j^i = \frac{1}{\left({a_j^i}^2V_j^i\right)^2}\left[\left(({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}-\frac{2a_j^id_j^i}{({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}}\right){a_j^i}^2V_j^i\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\left.-a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}\left(2a_j^iV_j^i+{a_j^i}^2\frac{\partial V_j^i}{\partial a_j^i}\right)\right]\\
&\partial\left(\frac{a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}}{{a_j^i}^2V_j^i}\right)/\partial b_j^i = \frac{1}{\left({a_j^i}^2V_j^i\right)^2}\left[\frac{a_j^ib_j^i}{({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}}{a_j^i}^2V_j^i\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\left.-a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}{a_j^i}^2\frac{\partial V_j^i}{\partial b_j^i}\right]\\
&\partial\left(\frac{a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}}{{a_j^i}^2V_j^i}\right)/\partial c_j^i = \frac{1}{\left({a_j^i}^2V_j^i\right)^2}\left[\frac{a_j^ic_j^i}{({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}}{a_j^i}^2V_j^i\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\left.-a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}{a_j^i}^2\frac{\partial V_j^i}{\partial c_j^i}\right]\\
&\partial\left(\frac{a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}}{{a_j^i}^2V_j^i}\right)/\partial d_j^i = \frac{1}{\left({a_j^i}^2V_j^i\right)^2}\left[\frac{-2{a_j^i}^2}{({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}}{a_j^i}^2V_j^i\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\left.-a_j^i({b_j^i}^2+{c_j^i}^2-4a_j^id_j^i)^{1/2}{a_j^i}^2\frac{\partial V_j^i}{\partial d_j^i}\right]
\end{aligned}\right.\tag{A.12}$$

A.2 Evaluation of $\frac{\partial I^i_{j,\lambda}}{\partial\left(a^i_k,b^i_k,c^i_k,d^i_k\right)}$ and $\frac{\partial I^i_{j,\lambda}}{\partial\left(a^i_l,b^i_l,c^i_l,d^i_l\right)}$

Similarly, it can be shown that:

$$\frac{\partial I^i_{j,\lambda}}{\partial\left(a^i_k,b^i_k,c^i_k,d^i_k\right)} = r_i^2\left[\operatorname{sign} a^i_j \cdot \frac{\partial \alpha^i_{j,\lambda}}{\partial(a^i_k,b^i_k,c^i_k,d^i_k)} - \frac{d^i_j + 4r_i^2 a^i_j}{V^i_j}\cdot\right. \tag{A.13}$$

$$\left.\partial\left(\frac{U^i_j}{{a^i_j}^2 V^i_j \sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right)/\partial(a^i_k,b^i_k,c^i_k,d^i_k)\cdot\frac{1}{1+\left(\frac{U^i_j}{2{a^i_j}^2 V^i_j \sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right)^2}\right] \tag{A.14}$$

where:

$$\begin{aligned}
&\partial\left(\frac{U_j}{{a^i_j}^2 V^i_j \sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right)/\partial\left(a^i_k,b^i_k,c^i_k,d^i_k\right)\\
&=\frac{|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2)}{2{a^i_j}^2V^i_j}\cdot\frac{\partial\alpha^i_{j,\lambda}}{\partial(a^i_k,b^i_k,c^i_k,d^i_k)}\cdot\frac{1}{\sin^2\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\\
&+\frac{a^i_j({b^i_j}^2+{c^i_j}^2-4a^i_jd^i_j)^{1/2}}{2{a^i_j}^2V^i_j\sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\cdot\\
&\left[-\frac{\partial\alpha^i_{j,\lambda}}{\partial(a^i_k,b^i_k,c^i_k,d^i_k)}\cdot\frac{\left(b^i_j\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)+c^i_j\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)}{\tan\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right.\\
&\left.+\frac{\partial\alpha^i_{j,\lambda}}{\partial(a^i_k,b^i_k,c^i_k,d^i_k)}\cdot\left(b^i_j\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)-c^i_j\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)\right]
\end{aligned} \tag{A.15}$$

and:

$$\begin{aligned}
\frac{\partial I^i_{j,\lambda}}{\partial\left(a^i_l,b^i_l,c^i_l,d^i_l\right)} &= r_i^2\left[-\operatorname{sign} a^i_j\cdot\frac{\partial\beta^i_{j,\lambda}}{\partial(a^i_k,b^i_k,c^i_k,d^i_k)}-\frac{d^i_j+4r_i^2a^i_j}{V^i_j}\cdot\right.\\
&\left.\partial\left(\frac{U^i_j}{{a^i_j}^2V^i_j\sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right)/\partial(a^i_l,b^i_l,c^i_l,d^i_l)\cdot\frac{1}{1+\left(\frac{U^i_j}{2{a^i_j}^2V^i_j\sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right)^2}\right]
\end{aligned} \tag{A.16}$$

with:

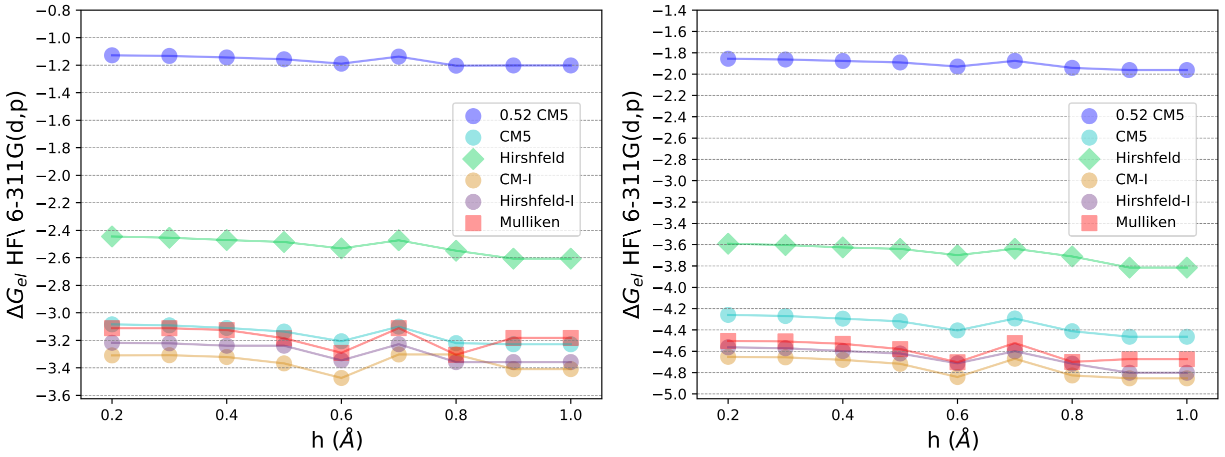
$$\begin{aligned}
&\partial\left(\frac{U_j}{{a^i_j}^2 V^i_j \sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right)/\partial\left(a^i_l,b^i_l,c^i_l,d^i_l\right)\\
&=\frac{-|a^i_j|({b^i_j}^2+{c^i_j}^2-2a^i_jd^i_j+8r_i^2{a^i_j}^2)}{2{a^i_j}^2V^i_j}\cdot\frac{\partial\beta^i_{j,\lambda}}{\partial(a^i_k,b^i_k,c^i_k,d^i_k)}\cdot\frac{1}{\sin^2\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\\
&+\frac{a^i_j({b^i_j}^2+{c^i_j}^2-4a^i_jd^i_j)^{1/2}}{2{a^i_j}^2V^i_j\sin\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\cdot\\
&\left[\frac{\partial\beta^i_{j,\lambda}}{\partial(a^i_k,b^i_k,c^i_k,d^i_k)}\cdot\frac{\left(b^i_j\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)+c^i_j\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)}{\tan\left((\beta^i_{j,\lambda}-\alpha^i_{j,\lambda})/2\right)}\right.\\
&\left.+\frac{\partial\beta^i_{j,\lambda}}{\partial(a^i_k,b^i_k,c^i_k,d^i_k)}\cdot\left(b^i_j\sin\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)-c^i_j\cos\left((\beta^i_{j,\lambda}+\alpha^i_{j,\lambda})/2\right)\right)\right]
\end{aligned} \tag{A.17}$$

Appendix B

Effects of the grid spacing on the solvation energy

In this section the 1-2 dichloroethane, acetic acid (anion), and methylamine (cation) molecules are used to study the effects of the FD grid spacing (h) on the ΔG_{el} contribution, with five charge models: 0.52·CM5, CM5, Hirshfeld, CM5-I, Hirshfeld-I, Mulliken at both HF and B3LYP levels. The 6-311G(d,p) basis set has been used for 1-2 dichloroethane, while 6-311++G(d,p) one has been considered for charged species. A grid spacing ranging from 1.00 to 0.20 Å is considered with a cubic grid. The remaining computational details are shown in Sec. 5.2 on page 93.

B.1 1-2 dichloroethane: grid spacing effects



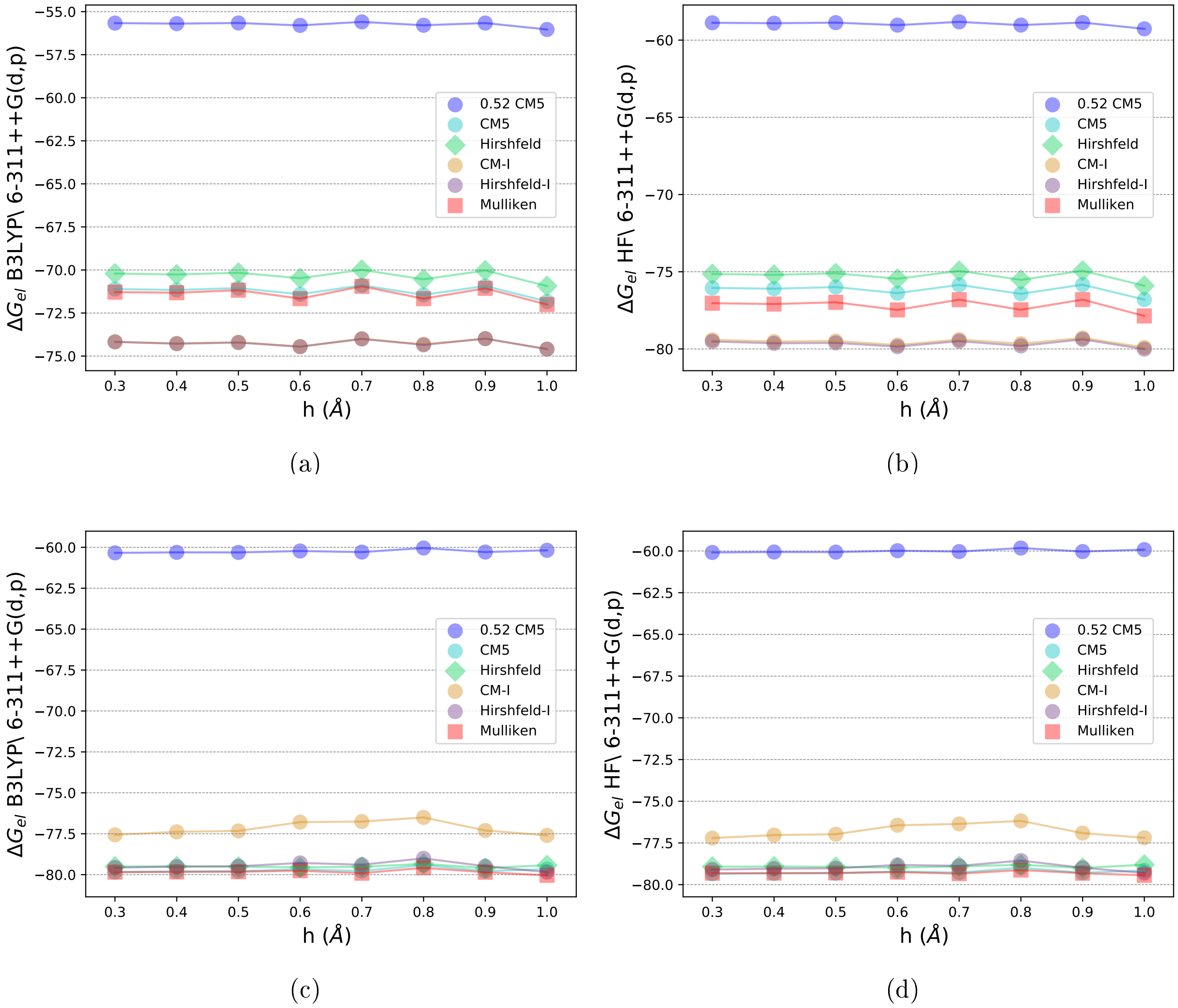
(a) (b)

Figure B.1: FD grid spacing (h) effects on the ΔG_{el} contribution (kcal/mol) for 1,2-dichloroethane in water, at the (a) B3LYP/6-311G(d,p) and (b) HF/6-311G(d,p) levels of theory with various charge models.

	0.52·CM5		CM5		Hirshfeld		CM5-I		Hirhsfeld-I		Mulliken	
h (Å)	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF
0.20	-1.128	-1.856	-3.084	-4.259	-2.445	-3.591	-3.310	-4.652	-3.218	-4.563	-3.112	-4.502
0.30	-1.133	-1.863	-3.091	-4.269	-2.454	-3.603	-3.308	-4.657	-3.222	-4.571	-3.112	-4.508
0.40	-1.144	-1.877	-3.110	-4.294	-2.471	-3.626	-3.321	-4.681	-3.239	-4.598	-3.125	-4.530
0.50	-1.157	-1.890	-3.136	-4.319	-2.485	-3.640	-3.368	-4.718	-3.239	-4.621	-3.185	-4.579
0.60	-1.189	-1.929	-3.207	-4.405	-2.532	-3.698	-3.474	-4.841	-3.346	-4.713	-3.291	-4.703
0.70	-1.138	-1.876	-3.100	-4.293	-2.473	-3.638	-3.303	-4.670	-3.228	-4.599	-3.111	-4.523
0.80	-1.204	-1.942	-3.220	-4.412	-2.549	-3.711	-3.303	-4.827	-3.358	-4.716	-3.305	-4.700
0.90	-1.202	-1.962	-3.229	-4.464	-2.606	-3.815	-3.409	-4.854	-3.358	-4.802	-3.182	-4.673
1.00	-1.202	-1.962	-3.229	-4.464	-2.606	-3.815	-3.409	-4.854	-3.358	-4.802	-3.182	-4.673

Table B.1: FD grid spacing (h) effects on the ΔG_{el} contribution (kcal/mol) for 1,2-dichloroethane in water at the (a) B3LYP/6-311G(d,p) and (b) HF/6-311G(d,p) levels of theory with various charge models.

B.2 Ions: grid spacing effects



(a) (b)

(c) (d)

Figure B.2: FD grid spacing (h) effects on the ΔG_{el} contribution (kcal/mol) for (a) and (b) Acetic Acid (anion) and (c) and (d) methylamine (cation) in water, at the B3LYP and HF levels with the 6-311++G(d,p) basis set with various atomic charge models.

	0.52·CM5		CM5		Hirshfeld		CM5-I		Hirhsfeld-I		Mulliken	
h (Å)	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF
0.30	-55.664	-58.878	-71.109	-76.048	-70.207	-75.142	-74.168	-79.409	-74.179	-79.516	-71.285	-77.037
0.40	-55.696	-58.906	-71.160	-76.098	-70.263	-75.199	-74.267	-79.522	-74.282	-79.635	-71.321	-77.092
0.50	-55.658	-58.869	-71.051	-75.993	-70.162	-75.101	-74.193	-79.488	-74.217	-79.610	-71.174	-76.978
0.60	-55.798	-59.029	-71.412	-76.388	-70.479	-75.452	-74.451	-79.745	-74.453	-79.850	-71.664	-77.472
0.70	-55.594	-58.818	-70.889	-75.850	-69.991	-74.945	-73.992	-79.392	-74.004	-79.495	-70.948	-76.807
0.80	-55.790	-59.028	-71.447	-76.433	-70.541	-75.524	-74.314	-79.662	-74.352	-79.803	-71.666	-77.463
0.90	-55.662	-58.861	-70.929	-75.837	-70.030	-74.934	-73.988	-79.288	-73.987	-79.378	-71.065	-76.804
1.00	-56.034	-59.269	-71.829	-76.804	-70.932	-75.907	-74.596	-79.900	-74.596	-80.013	-72.007	-77.863

(a) Anions

	0.52·CM5		CM5		Hirshfeld		CM5-I		Hirhsfeld-I		Mulliken	
h (Å)	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF	B3LYP	HF
0.30	-60.340	-60.088	-79.835	-79.353	-79.506	-78.925	-77.566	-77.209	-79.568	-79.097	-79.848	-79.313
0.40	-60.309	-60.061	-79.801	-79.326	-79.490	-78.913	-77.382	-77.029	-79.510	-79.043	-79.831	-79.308
0.50	-60.313	-60.066	-79.789	-79.315	-79.507	-78.933	-77.327	-76.978	-79.487	-79.021	-79.819	-79.298
0.60	-60.226	-59.982	-79.674	-79.209	-79.551	-78.976	-76.790	-76.441	-79.283	-78.817	-79.758	-79.249
0.70	-60.292	-60.034	-79.774	-79.278	-79.513	-78.904	-76.756	-76.360	-79.390	-78.873	-79.910	-79.343
0.80	-60.044	-59.821	-79.410	-78.983	-79.343	-78.810	-76.504	-76.179	-79.001	-78.557	-79.580	-79.125
0.90	-60.291	-60.033	-79.782	-79.290	-79.593	-78.999	-77.298	-76.911	-79.484	-78.989	-79.851	-79.316
1.00	-60.182	-59.916	-79.664	-79.150	-79.425	-78.805	-77.600	-77.191	-79.844	-79.300	-80.038	-79.439

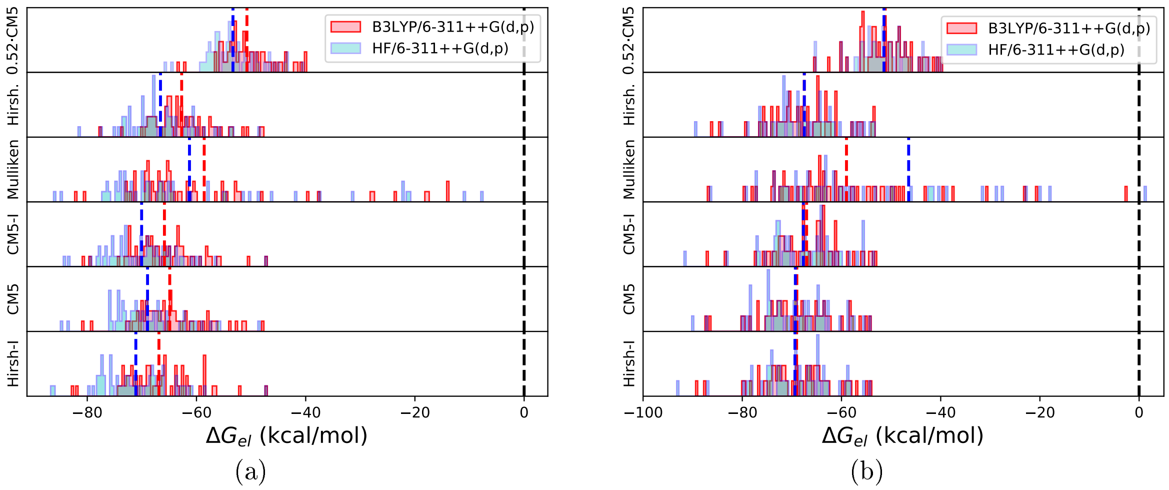
(b) Cations

Table B.2: FD grid spacing (h) effects on the ΔG_{el} contribution (kcal/mol) for (a) Acetic Acid (anion) and (b) methylamine (cation) in water, at the B3LYP and HF levels with the 6-311++G(d,p) basis set with various atomic charge models.

Appendix C

Ions: Effects of the charge and non-electrostatic models

This section focuses on the effects of the charge and non-electrostatic model on 112 charged solutes in water from the MNSOL database [36, 241] considered in Chapter 5 on page 91. In particular, the 0.52·CM5, CM5, Hirshfeld, Hirshfeld-I and Mulliken charge models, at both the B3LYP and HF levels of theory with the 6-311++G(d,p) basis set habe been selected. Computational details are given is Sec. 5.2 on page 93.



(a) (b)

Figure C.1: Distributions of ΔG_{el} contributions for (a) 60 anions, and (b) 52 cations in water at both the B3LYP and HF levels of theory with the 6-311++G(d,p) basis set.

XC	0.52·CM5	CM5	Hirshfeld	CM5-I	Hirshfeld-I	Mulliken
B3LYP	-50.76±4.87	-64.89±6.34	-62.71±5.73	-65.86±6.17	-66.87±6.38	-58.58±16.21
HF	-53.31±4.92	-68.96±6.48	-66.58±5.97	-70.04±6.43	-71.10±6.69	-61.28±18.09

(a) Anions

XC	0.52·CM5	CM5	Hirshfeld	CM5-I	Hirshfeld-I	Mulliken
B3LYP	-51.18±5.47	-69.00±7.26	-67.50±7.25	-67.03±7.03	-69.04±7.33	-58.99±15.69
HF	-51.43±5.56	-69.30±7.43	-67.49±7.41	-67.69±7.23	-69.38±7.51	-46.46±60.04

(b) Cations

Table C.1: Average and standard deviation of the distributions of ΔG_{el} contributions per (a) 60 anions and (b) 52 cations in water at both the B3LYP and HF levels of theory with the 6-311++G(d,p) basis set.

C.1 ΔG_{ne} distributions for ions

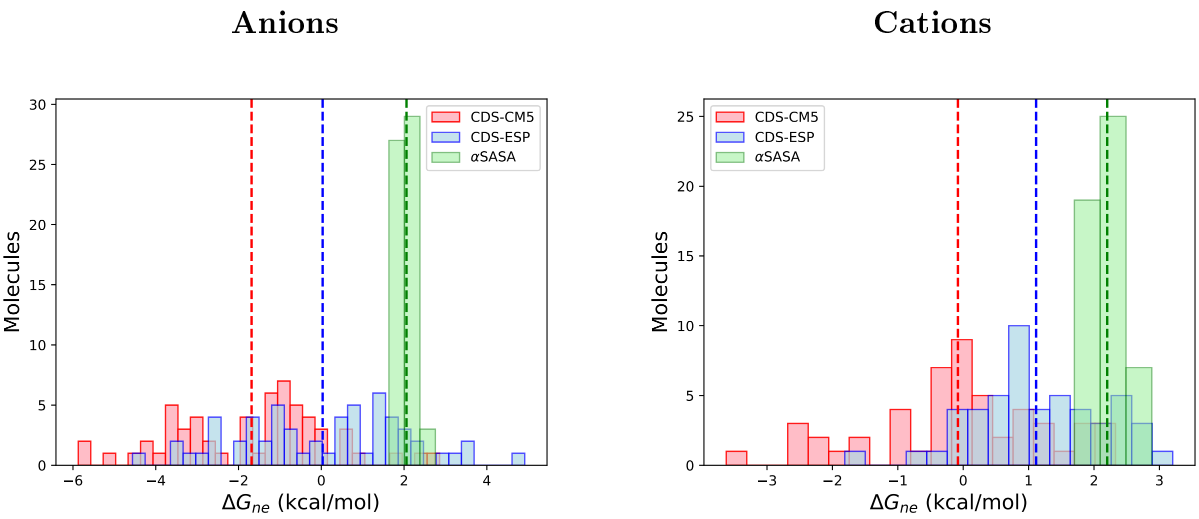


Table C.2: ΔG_{ne} distributions for 60 anions, and 52 cations in water using the αSASA, and CDS-ESP, CDS-CM5 models.

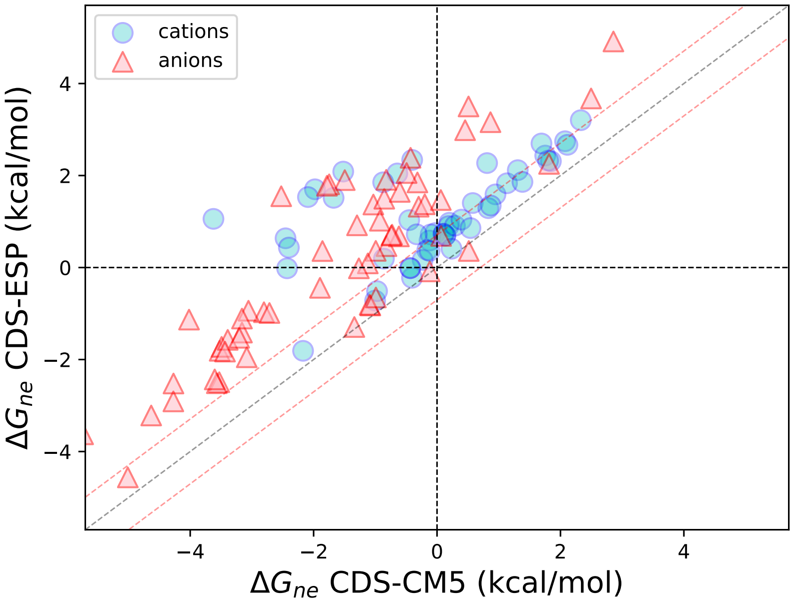


Figure C.2: Scatter plot for 60 anions and 52 cations in water between ΔG_{ne} values obtained using the CDS-ESP and CDS-CM5 model.

C.2 Effects of the charge model on the SCRF performance

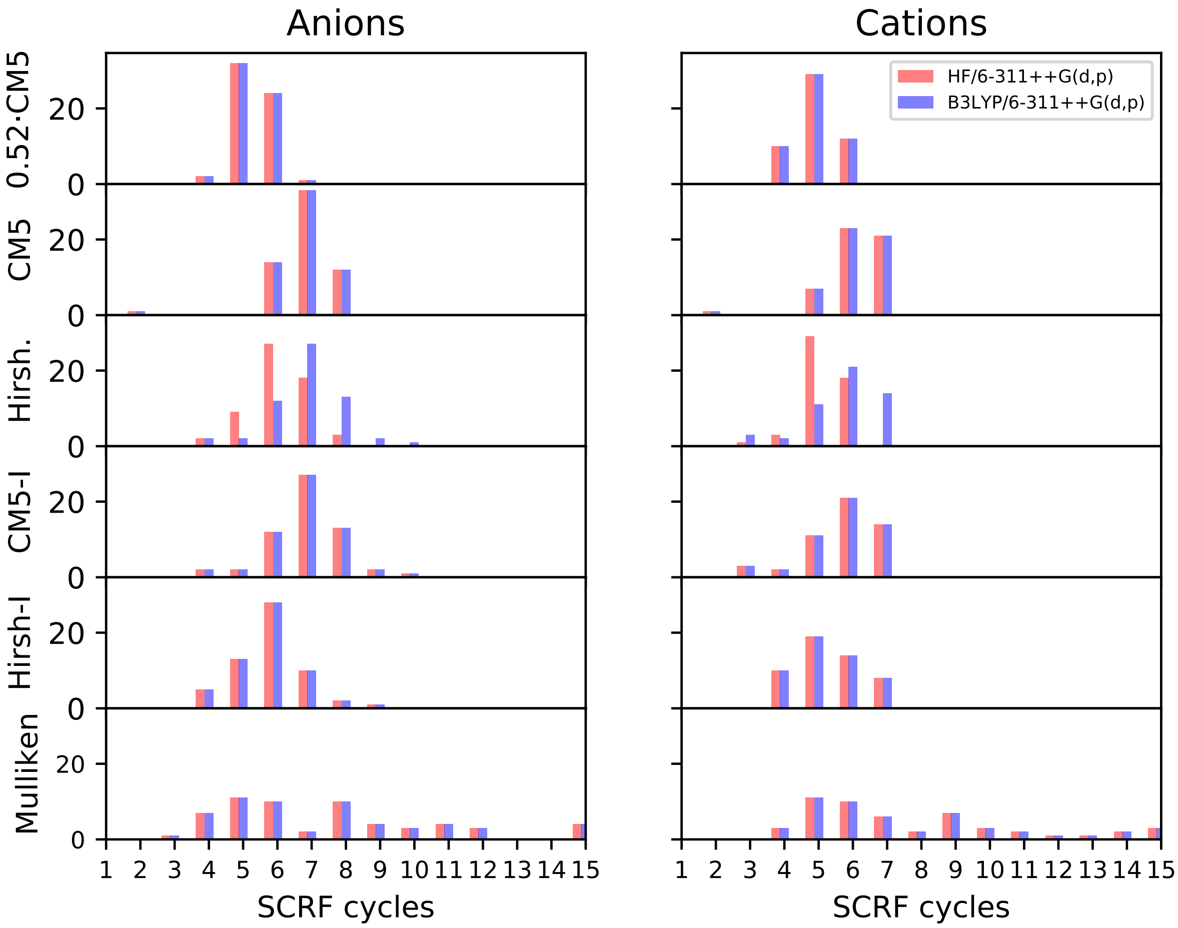


Figure C.3: Distribution of the number of SCRF cycles needed for convergence of 60 anions and 52 cations in water, data obtained at the B3LYP/ 6-311++G(d,p) and HF/ 6-311++G(d,p) levels.

List of Figures

1 Représentations des modèles de solvatation (a) Explicite, (b) cluster-continuum, et (c) Implicite. En bleu clair, la cavité du soluté séparant la région atomistique du soluté et le milieu diélectrique caractérisé par la permittivité relative ϵ. xiv

2 (a) Exemple 2D d'une grille FD avec un espacement h. En gris, les nœuds de grille où l'équation de Poisson généralisée est résolue. En bleu, permittivité relative cartographiée entre les nœuds de la grille. En rouge, les points de la grille aux bords où les conditions aux limites sont imposées, et en orange les points de grille des limites à l'interface soluté-solvant. La région atomistique est indiquée par Ω_{in}, et la région du solvant par Ω_{out}. (b) Élément de volume pour une grille cubique FD. xviii

3 (a) Focus sur un point de grille limite sur une grille de différences finies cubique en 2D, et (b) élément de volume élémentaire contenant un point de grille limite en son centre en orange. L'interface soluté-solvant est représentée par un plan orange. xix

4 Flow chart for the SCRF process. xx

5 En orange, SASA d'une molécule de benzène, en noir et blanc les sphères dures de van der Waals pour les atomes de carbone et d'hydrogène, respectivement. xxi

6 Histogramme des erreurs relatives non signées (en Å^2) des SASA atomiques pour les 501 molécules de l'ensemble de test de Mobley [38], par rapport à une autre approche analytique de référence [330], obtenues soit avec l'algorithme analytique SP (à gauche), soit avec l'algorithme numérique SR (à droite). SASA atomique de la molécule xxiii

7 Comparaison entre les valeurs de ΔG_{el} calculées avec B3LYP/6-311G(d,p) et B3LYP/6-31G et B3LYP/6-311++G(d,p). Les données obtenues avec des charges de Mulliken (carrés rouges), Hirshfeld (cercles bleus) et CM5 (triangles verts) sont représentées. xxiv

8 MUE du ΔG_{solv} comparé aux valeurs expérimentales et terme électrostatique moyen pour 501 molécules neutres en fonction du facteur d'échelle des charges CM5 s. Le ΔG_{el}, obtenu au niveau B3LYP/ 6-311G en utilisant différents facteurs d'échelle s, est combiné avec le ΔG_{ne} obtenu avec les modèles $\propto$SASA, CDS-ESP et CDS-CM5. La ligne en pointillés représente le seuil de 1,00 kcal/mol pour un MUE acceptable. La moyenne des ΔG_{el} (ligne noire pleine) et la dispersion de chaque distribution pour chaque valeur de s (lignes verticales grises) sont également indiquées. xxv

9 MUE moyenne (en kcal/mol) des énergies de solvatation pour chaque solvant donnée en fonction de la constante diélectrique du solvant (ϵ), en utilisant les charges CM5 (croix rouges) et HPA (points bleus). Les lignes en pointillés colorés indiquent le MUE moyen de chaque modèle de charge $0,72 \pm 0,11$ et $0,64 \pm 0,09$ kcal/mol pour les charges CM5 et HPA, respectivement. La ligne pointillée noire souligne le seuil de 1,00 kcal/mol visé pour les solutés neutres. xxvi

10 (a) Energies de solvatation totales calculées à l'aide des charges CM5 au niveau B3LYP/6-311++G(d,p) pour 327 espèces chargées par rapport aux énergies de solvatation expérimentales en utilisant la paramétrisation CDS originale et révisée, et le modèle étendu. (b) Comparaison entre les valeurs pK_a calculées et expérimentales pour 28 acides carboxyliques (en noir), 10 amines aliphatiques (en rouge) et 45 thiols (en bleu) dans l'eau, obtenues avec les charges CM5. Les lignes pointillées intérieures et extérieures représentent des erreurs de ±1,00 et ±3,00 unités pK_a, respectivement. xxvii

11 Representations of (a) Explicit, (b) cluster-continuum, and (c) Implicit solvation models. In light blue the solute cavity separating the solute atomistic region and the dielectric media characterized by the relative permittivity ϵ. 2

1.1 Example of a periodic lattice in two-dimensions. Blue dots represent lattice points, and $\mathbf{a}_1$, $\mathbf{a}_2$ form the basis vectors of the translation vector $\mathbf{R}_n = n_1\mathbf{a}_1 + n_2\mathbf{a}_2$. R_n is here represented with $n_1 = n_2 = 1$ in the direct lattice. In red a unit cell. 16

1.2 Example of the first Brillouin zone for (a) square lattice and (b) hexagonal lattice. Dashed lines represent the path connecting one reciprocal lattice point to its nearest neighbor. In red the area within a closed surface traced by planes orthogonal to the dotted lines midpoints. 17

1.3 Examples of TiO_2 anatase (101) slabs generated with (a) a slab model with 2D periodic boundary conditions in the surface plane, and (b) Multi slab model with 3D periodic boundary conditions. 20

2.1 Thermodynamic cycle depicting the solvation free energy decomposition into contributions [85]. W_{el} stands for the electrostatic work needed to polarize the dielectric media. The cavitation contribution ΔG_{cav} has been considered separate from ΔG_{ne} (gray) to point out the cavity (yellow) formation within the dielectric media (blue). 24

2.2 2D example of a solute cavity. In blue the region treated as dielectric media replacing the explicit solvent Ω_{out} and characterized by a relative permittivity ϵ_{out}, in white the solute region where atoms are treated explicitly Ω_{in} with the vacuum relative permittivity set to $\epsilon_{in} = 1$, and in yellow the solute-solvent interface Γ. 26

2.3 Flowchart of a generic SCRF procedure. 28

2.4 Ions distributed within the dielectric media. In blue and red negatively and positively charged species, respectively. 29

2.5 Example of a two-dimensional FEM grid around a circular boundary [134]. . 31

2.6 2D example of a finite-difference grid with spacing h. Poisson's equation is solved at the grid nodes. In orange boundary points where to impose boundary conditions. 33

2.7 2D example of a SAS in yellow and a SES solute cavity in green, with Solvent Probe in blue, and vdW surfaces and sphere in black and gray, respectively. . 35

2.8 Examples of, (a) toric surface patch (ABCD) connecting two spheres, and (b) reentrant patch (ABC) connecting three spheres, for a SES type surface. 36

2.9 Calculated electrostatic solvation energies ΔG_{el} as a function of total experimental solvation energies ΔG_{exp} for (a) non-aqueous, and (b) aqueous solvents. The ΔG_{el} values are calculated with IEF-PCM (B3LYP/6-311G**) on 2530 neutral solutes of the MNSOL database [36, 241] in 91 solvents with ϵ_{out} ranging from $\approx$ 1.00 to 181.56. 44

2.10 In orange the SASA of a benzene molecule, in black and white the van der Waals hard spheres for carbon and hydrogen atoms, respectively. 47

2.11 Calculated ΔG_{sol} and ΔG_{el} as function of total experimental ΔG_{exp} solvation energies for (a) non-aqueous, and (b) aqueous solvent, using ΔG_{el} values from IEF-PCM (B3LYP/6-311G**) on 2530 neutral solutes of the MNSOL database [36, 241] in 91 solvents with ϵ_{out} ranging from $\approx$ 1.00 to 181.56. The ΔG_{ne} are obtained from Eq. 2.73 with coefficients $\alpha = -0.025$ kcal/mol·Å^2 and $\beta = 1.049$ kcal/mol, and $R_{probe} = 0.400$ Å. 48

2.12 Calculated ΔG_{sol} and ΔG_{el} as function of total experimental ΔG_{exp} solvation energies for (a) non-aqueous, and (b) aqueous solvent, using ΔG_{el} values from SMD (B3LYP/6-311G**) on 2530 neutral solutes of the MSNOL database [36, 241] in 91 solvents with ϵ_{out} ranging from $\approx$ 1.00 to 181.56. The ΔG_{ne} are obtained from Eq. 3.33 as implemented in Gaussian16 [30]. 50

2.13 (a) Liquid drop showing the quantities in the Young's equation. (b) Possible wetting scenarios based on the different relationships between surface tensions. 53

3.1 2D example of a finite-difference grid (10 × 10) with spacing h. In gray the grid points where the generalized Poisson equations is solved using an iterative solver. In blue the relative permittivity mapped between the grid nodes. In red edge grid points where Dirichlet boundary conditions or Ewald formulas have to be imposed, and in orange the boundary grid points at the solute/solvent interface. The atomistic region within the solute cavity is indicated by Ω_{in}, and the dielectric region by Ω_{out}. 59

3.2 Volume element for a 3D FD cubic grid. Each grid point (gray) is neighbored by six grid points, and six relative permittivity values mapped at the facets points (blue). (α, β, γ) are angles along directions (i, j, k) with grid step sizes (h_i, h_j, h_k). 60

3.3 (a) Focus on a boundary grid point on a 2D cubic finite-difference grid, and (b) elementary volume element containing one boundary grid point at its center in orange. The solute-solvent interface is represented by a orange plane. 65

3.4 Example of solute cavities and surface elements mapped on a 2D FD grid. In yellow the SAS, in green the SES, and in blue the solvent probe radius. Boundary grid point are represented in orange, while grid nodes are in gray. Smaller blue circles represent the relative permittivity mapped at the grid node facets. The grid boundary is in red. 67

3.5 Example of selection process for an "in" and "out" grid facet point. (a) The point status as an external point remains unchanged (b) the point status changes from external to internal. The yellow lines represent the SAS, the green lines the SES, and the blue lines the probe sphere. In light blue the probe radius R_p, and in red the line connecting the closest point on the SAS to a given facet grid point. 68

3.6 Examples of ASC projected on a SES cavity for (a) a finite molecular system (formaldehyde), and (b) a 2D periodic MgO slab. Contact and reentrant ASC are displayed in yellow and green, respectively. 68

3.7 Flow chart of the SCRF procedure in CRYSTAL, in light blue the steps iteratively repeated during the SCRF. 70

4.1 Schematic representation of the van der Waals (in black) and solvent-accessible (in blue) surfaces of a system with four atomic spheres, in the case of (a) a finite system, and (b) a 1D infinite periodic system with a single lattice basis vector (in red). The blue circle is the solvent probe and the black dots represent the centers of the atomic spheres. For the periodic system, the solvent-accessible surface of the central **0**-cell is emphasized with thick lines. 76

4.2 Intersection of reference atomic sphere i with neighboring atomic spheres j and k, showing two COI ($i \mathbin{/\!/} j$ in solid orange and $i \mathbin{/\!/} k$ in solid green), along with their corresponding stereographic projections (dotted line) in the (t, s) tangent projection plane (in grey) to the south pole of the reference sphere i. NP is the North Pole of the reference sphere through which the stereographic projection is performed. The two COI have two intersection points (v_1 and v_2) in the (t, s) plane. 77

4.3 Stereographic projections of the two COI ($i \mathbin{/\!/} j$ in dashed orange and $i \mathbin{/\!/} k$ in dashed green) in the (t, s) tangent plane of the reference sphere i, showing (a) a bounded Ω_i domain made of two circular arcs C^i_j (solid orange, from intersection points v_1 to v_2) and C^i_k (solid green, from v_2 to v_1); (b) an unbounded Ω_i domain with two circular arcs C^i_j (from v_2 to v_1) and C^i_k (from v_1 to v_2). The white section corresponds to the area of the reference sphere i to be computed. The arrows indicate the integration direction along the circular arcs. 78

4.4 Intersection of two spheres i and j results in a COI ($i \mathbin{/\!/} j$, solid orange) which is projected (dotted orange) in the (t, s) tangent plane (in grey) at the south pole of reference sphere i: (a) as a circle when the COI is far enough from the NP of i, (b) as a line when the COI passes through the NP; (c) local polar coordinate system used to define the $\alpha^i_{j,\lambda}$ and $\beta^i_{j,\lambda}$ polar angles with respect to the center (t_0, s_0) of the projected COI $i \mathbin{/\!/} j$ with radius r_0. 79

4.5 Schematic representation of the SP algorithm for the calculation of the SASA and its nuclear gradients. 83

4.6 2D view of the neighborhood determination process using a cubing algorithm. For atomic spheres belonging to the dark orange cell, distances are only computed with the green spheres whose centers belong to the dark or light orange grid cells. Red spheres are ignored since they are guaranteed to be out of range. 83

4.7 Shrake and Rupley algorithm for a system with two van der Waals (vdW) spheres, here represented by black lines. The SASA is represented by blue lines, and blue points on the SASA represent points which do not lie inside any other atom's accessible surface and are therefore considered as accessible to the solvent, while red ones are buried and thus eliminated. 84

4.8 Histogram of relative unsigned errors (in Å^2) from reference analytical surface calculation package[330] values of total (top panel) and atomic (bottom panel) SASA, using either the analytical SP (left) or numerical SR (right) algorithms. There are 501 total SASA and 8200 atomic SASA values. For the SR algorithm, 500 points per atomic sphere are used. 86

4.9 C_7-13 helix model of the poly-glycine 1D periodic system system: (a) side and (b) top views. Red, blue, grey and white balls correspond to O, N, C and H atoms, respectively. The solid green line represents the unit cell. . . . 87

4.10 HCOOH/NiO (100) 2D slab periodic system. Ni, O, C and H atom as blue, red, grey and white balls, respectively. The solid green line corresponds to the unit cell. The translucent atoms inside the slab have a SASA which is exactly zero. 88

4.11 Histogram of the relative unsigned error (in Å) per molecule of the analytical nuclear SASA gradients compared to corresponding numerical gradients for the 501 molecules of the Mobley test set. 89

4.12 Plot of the SASA analytical SP nuclear gradients components as a function of the corresponding numerical gradients for: (a) the 49 atoms of the poly-glycine C_7-13 helix model unit cell; (b) the 130 atoms of the HCOOH/NiO (100) slab unit cell. All data in Å. 89

5.1 (a) Comparison of the distributions of ΔG_{ne} values for 501 neutral molecules calculated with the $\propto$SASA, CDS-CM5 and CDS-ESP models. Dashed lines represent the average values of ΔG_{ne} obtained with the three non-electrostatic solvation models on the whole test set, with values of 2.30$\pm$0.26, 1.14$\pm$1.31 and 0.28$\pm$1.17 kcal/mol for the $\propto$SASA, CDS-ESP and CDS-CM5 models, respectively; (b) Comparison of computed ΔG_{ne} values obtained with the CDS-CM5 and CDS-ESP models. The dashed red lines represent a $\pm$1 kcal/mol difference. 95

5.2 Normalized distributions of ΔG_{el} values computed with the Mulliken, Hirshfeld, Hirshfeld-I, CM5 and CM5-I atomic charge models obtained at the B3LYP (in red) and HF (in blue) levels, with the 6-311G(d,p) basis set. Dashed colored lines represent the average of the distributions obtained for the whole 501 molecules test set, while dotted black lines indicate a zero value for ΔG_{el}. 96

5.3 Comparison between computed ΔG_{el} values calculated both at the HF and B3LYP levels, with the 6-311G(d,p) basis set and the 6-31G and 6-311++G(d,p) ones. Mulliken (red squares), Hirshfeld (blue circles) and CM5 (green triangles) data are reported. 98

5.4 MUE of the ΔG_{solv} compared to experimental values and average electrostatic term for 501 neutral molecules against the CM5 charges scaling factor s. The ΔG_{el}, obtained at the B3LYP/ 6-311G level using different scaling factors s, is combined with the ΔG_{ne} obtained with the $\propto$SASA, CDS-ESP and CDS-CM5 models. The dashed line represents the threshold of 1.00 kcal/mol for acceptable MUE. The average ΔG_{el} (solid black line) and the spread of each distribution for each s values (grey vertical lines) are also reported. 100

5.5 Histogram showing the number of SCRF cycles needed for the convergence of the ΔG_{el} with different charge models for 501 neutral molecules in water, at both the B3LYP and HF with the 6-31G, 6-311G(d,p), and 6-311++G(d,p) basis sets. 102

6.1 Averaged experimental solvation energies $\overline{\Delta G_{exp}}$ (in kcal/mol, colored dots) of 332 solutes taken from the MNSOL database (see Sec. 6.1.4.1) and standard deviations (vertical lines) as a function of Abraham's hydrogen bond acidity parameter α of the solvent. Four solvents have been considered: dimethyl sulfoxide, acetonitrile, methanol and water, in order of increasing α. The solid lines represent the fitted data obtained with the function $f(\alpha) = \frac{A\alpha}{(1+\alpha^2)} + B$, as found in Eq. (6.4). 110

6.2 Heatmaps of the MUE of calculated solvation energies ΔG_{solv} between pairs of functionals for 2530 neutral solute/solvent combinations, considering different charge models and basis sets combinations. All data in kcal/mol. . . . 116

6.3 Averaged MUE (in kcal/mol) of solvation energies for each solvent given as a function of the dielectric constant of the solvent (ϵ), using both CM5 (red crosses) and HPA (blue dots) electrostatics. Colored dotted lines indicate the average MUE of each charge model: 0.71 ± 0.11 and 0.64 ± 0.09 kcal/mol for CM5 and HPA charges, respectively. The black dashed line highlights the 1.00 kcal/mol threshold targeted for neutral solutes. 120

6.4 (a) Side and (b) top views of a TiO_2 anatase (101) slab model with 10 Ti layers. Grey and red balls represent Ti and O atoms, respectively. The pink dots correspond to 176 ASC distributed on both sides of the solute cavity exposed to the implicit solvent. The slab unit cell is shown as dotted blue lines. 121

6.5 Band structures of a 10 Ti layer TiO_2 anatase (101) slab model computed at the PBE level in gas-phase and in implicit water, acetonitrile and toluene with the (a) FDPB and (b) VASPsol continuum solvation models. The zero of the energy axis has been set to the valence band maximum of the gas-phase calculation (E_f^g). The green and blue shaded areas highlight the shifts of the top of the valence and bottom of conduction bands, respectively, with the computed shifts indicated with the same colors. The band gaps are also indicated by the red arrow along with the corresponding values obtained. All data reported in eV. 123

6.6 Comparison between calculated and experimental pK_a values for 28 carboxylic acids (in black), 10 aliphatic amines (in red), and 45 thiols (in blue) in water, obtained with (a) CM5 and (b) HPA atomic charges. The inner and outer dashed lines represent errors of $\pm$1.00 and $\pm$3.00 pK_a units, respectively. 127

7.1 Two-dimensional section of a SES cavity for a diatomic molecule (atoms in gray). Yellow and green spheres represent ASC assigned to contact and reentrant patches of the SES, respectively. DBF (full lines) are assigned to atomic centers (dotted lines) for (a) contact, and (b) reentrant patches. DBF acting on reentrant patches have to be decomposed and assigned to adjacent atomic centers. 132
7.2 Electrostatic forces computed for the moving atom in a two particle model system using a grid spacing of (a) 0.210 Å and (b) 0.410 Å. F_x are the x component of the analytically computed electrostatic forces for the atom moving along the x axis, computed using Eq. 7.19. Instead, $F_x(\Delta G_{el})$ are numerically calculated forces computed using Eq. 7.13. 135
7.3 (a) Correlation of forces obtained with different grid spacings h for a two point charge model, dotted lines represent a 1 kcal/mol/Å difference. (b) Comparison of electrostatic forces calculated with the FDPB and MIB [175] methods, for a two point charge model system using a grid spacing of 0.210 Å. F_x MIB data has been extracted from Ref. [175]. 135
7.4 Electrostatic forces computed for the moving atom in a three particle model system using a grid spacing of (a) 0.210 Å and (b) 0.410 Å. F_x are the x component of the analytically computed electrostatic forces for the atom moving along the x axis, computed using Eq. 7.19. Instead, $F_x(\Delta G_{el})$ are numerically calculated forces computed using Eq. 7.13. 136
7.5 (a) Correlation of forces obtained with different grid spacings h for a three point charge model, dotted lines represent a 1 kcal/mol/Å difference. (b) Comparison of electrostatic forces calculated with the FDPB and MIB [175] methods, for a three point charge model system using a grid spacing of 0.210 Å. F_x MIB data has been extracted from Ref. [175]. 137
7.6 Model of the CH_3OH molecule. The geometry was optimized in gas-phase using the B3LYP/6-311G(d,p) level. 138

B.1 FD grid spacing (h) effects on the ΔG_{el} contribution (kcal/mol) for 1,2-dichloroethane in water, at the (a) B3LYP/6-311G(d,p) and (b) HF/6-311G(d,p) levels of theory with various charge models. 153
B.2 FD grid spacing (h) effects on the ΔG_{el} contribution (kcal/mol) for (a) and (b) Acetic Acid (anion) and (c) and (d) methylamine (cation) in water, at the B3LYP and HF levels with the 6-311++G(d,p) basis set with various atomic charge models. 154

C.1 Distributions of ΔG_{el} contributions for (a) 60 anions, and (b) 52 cations in water at both the B3LYP and HF levels of theory with the 6-311++G(d,p) basis set. 157
C.2 Scatter plot for 60 anions and 52 cations in water between ΔG_{ne} values obtained using the CDS-ESP and CDS-CM5 model. 158
C.3 Distribution of the number of SCRF cycles needed for convergence of 60 anions and 52 cations in water, data obtained at the B3LYP/ 6-311++G(d,p) and HF/ 6-311++G(d,p) levels. 159

List of Tables

1 Energies libres totales de solvatation calculées (ΔG_{tot}), composantes électrostatiques (ΔG_{el}) et non électrostatiques (ΔG_{ne}), d'un modèle slab d'une surafce de TiO_2 anatase (101) à 10 couches de Ti, solvaté dans de l'eau, de l'acétonitrile et du toluène implicites avec les modèles de solvatation FDPB et VASPsol. $\Delta\Delta G_{el}^{w/x}$ est la différence entre les données ΔG_{el} calculées dans l'eau et dans les autres solvants. Toutes les données sont en kcal/mol. Les charges atomiques CM5 ont été utilisées. xxviii

3.1 Boundary conditions for the potential depending on the system periodicity. . 59

4.1 Average RUE (in Å^2) of the atomic SASA calculated with the SP algorithm and compared to the ASC[330] data for the 42 proteins of the TRIFORCE[335] test set. N_{at} is the number of atoms. 87

4.2 Computed total SASA (in Å^2) of the poly-glycine C_7-13 helix model with n unit cells and N_{at} atoms per supercell, and RUE when compared to the expected value based on the reference unit cell value. All data obtained with the SP algorithm. 88

5.1 Slope coefficient a (kcal/(mol·Å^2)), intercept coefficient b (kcal/mol) and coefficient of determination R^2 obtained from the linear regression of ΔG_{el} values computed with: (a) 6-311G(d,p) and 6-31G; (b): 6-311G(d,p) and 6-311++G(d,p) basis sets. All data obtained on 501 neutral molecules at the both with HF and B3LYP. The corresponding Mean Unsigned Differences (MUD, kcal/mol, see Eq. (6.10)) are also reported. 97

5.2 MUE (in kcal/mol) of the computed total free energy of solvation with respect to the experimental data, both at the HF and B3LYP levels with three basis sets, five atomic charge models and three non-electrostatic models. BS1, BS2 and BS3 correspond to the 6-31G, 6-311G(d,p) and 6-311++G(d,p) basis sets, respectively. Values below the targeted 1.00 kcal/mol threshold are shown in bold. 99

5.3 MUE (in kcal/mol) of the computed total free energy of solvation with respect to the experimental data, both at the HF and B3LYP levels with three basis sets, the CM5 and 0.52·CM5 atomic charge models and three non-electrostatic models. BS1, BS2 and BS3 correspond to the 6-31G, 6-311G(d,p) and 6-311++G(d,p) basis sets, respectively. Values below the targeted 1.00 kcal/mol threshold are shown in bold. 101

5.4 Slope coefficient a (kcal/(mol·Å^2)), intercept coefficient b (kcal/mol) and coefficient of determination R^2 obtained from the linear regression of ΔG_{el} values computed with the 6-31G and 6-311++G(d,p) basis sets compared to reference 6-311G(d,p) data, for the CM5 and 0.52·CM5 atomic charges. All data obtained on 501 neutral molecules at the HF and B3LYP levels. The corresponding MUD (in kcal/mol, see Eq. (6.10)) are also reported. 102

5.5 MUE (kcal/mol) of the computed total free energy of solvation with respect to experimental data for 112 charged molecules, both at the HF and B3LYP levels of theory with the 6-311++G(d,p) basis set, with six different atomic charged models and three non-electrostatic models. 103

6.1 Revised CDS parameters for FDPB calculations. The index i refers to the element or atom type. For i=X the index refers to any element other than H, C, N, O, F, Si, P, S, Cl, Br, and I. For the solute-dependent parameters, any parameter not present in this table is set equal to zero. The $\tilde{\sigma}_i^n(H_2O)$ columns correspond to water-specific parameters, while the others correspond to nonaqueous solvents. All data given in cal·mol^{-1}·Å^{-2}. 112

6.2 A, B and C parameters of the proposed corrective term for charged species (see Eq (6.4) and Sec. 6.1.3 for details). A and B are given in cal·mol^{-1}·Å^{-2}, and C is given in e. 113

6.3 MUE of calculated (a) solvation energies (ΔG_{solv}) of 2530 neutral solute/solvent combinations in 91 solvents, and (b) 144 transfer energies (ΔG_{te}) between organic solvents and water, with respect to experimental values, both with CM5 and HPA atomic charges. BS1 and BS2 refer to the 6-31G* and 6-311G** basis sets, respectively. All data in kcal/mol. 115

6.4 Average MUE of calculated solvation and transfer energies with respect to experimental values for all solvents considered, using CM5 and HPA electrostatics. The average is computed on the 10 levels of theory considered. The number of solute molecules in each solvent (N) along with the corresponding solvent dielectric constant (ϵ) are also reported. All data in kcal/mol. 117

6.5 Computed PBE total free energies of solvation (ΔG_{tot}), along with the electrostatic (ΔG_{el}) and non=electrostatic (ΔG_{ne}) components, of a 10 Ti layer TiO$_2$ anatase (101) slab model solvated in implicit water, acetonitrile and toluene with the FDPB and VASPsol continuum solvation models. The $\Delta\Delta G_{el}^{w/x}$ line corresponds to the difference between the ΔG_{el} data computed in water and the other solvents. All data in kcal/mol. CM5 atomic charges have been used for FDPB calculations. 121

6.6 Computed PBE surface energies (E_s) of a 10 Ti layer TiO$_2$ anatase (101) slab model in gas-phase and solvated in implicit water, acetonitrile and toluene with the FDPB and VASPsol continuum solvation models. The variation of surface energy ($\Delta E_s^{g/x}$) when going from gas-phase to solvent x is also reported. All data in J/m^2. 122

6.7 MUE (in kcal/mol) of the solvation energies of (a) 208 anions, and (b) 124 cations obtained in four different solvents. All data reported include the corrective term for charged species (see Eq. (6.4)). ^{a}: minimal and maximal MUE among the eight values reported in Ref. 81. 124

6.8 MUE of the calculated solvation energies obtained without the proposed corrective term $\Delta G_{ne}^{\text{ion}}$ with respect to experimental values for (a) 204 anions and (b) 123 cations. All data have been obtained with the 6-311++G** basis set and are reported in kcal/mol. 125
6.9 MUE of computed pK_a in water with respect to experimental data obtained for a set of 83 solutes belonging to the carboxylic acid, aliphatic amine and thiol families, obtained at the B3LYP/Def2-TZVP level with both CM5 and HPA atomic charges. Data obtained at the B3LYP/6-31+G(d,p) level with the standard SMD solvation model [81], as well a modified version considering a scaled solvent-accessible surface approach (SMD_{aSAS}) taken from Ref. 363 are also reported. 126

7.1 Electrostatic forces for a CH_3OH molecule, calculated at the B3LYP/6-311G(d,p) level, and a grid spacing of 0.500 Å. Data in kcal/mol/Å. 137

B.1 FD grid spacing (h) effects on the ΔG_{el} contribution (kcal/mol) for 1,2-dichloroethane in water at the (a) B3LYP/6-311G(d,p) and (b) HF/6-311G(d,p) levels of theory with various charge models. 154
B.2 FD grid spacing (h) effects on the ΔG_{el} contribution (kcal/mol) for (a) Acetic Acid (anion) and (b) methylamine (cation) in water, at the B3LYP and HF levels with the 6-311++G(d,p) basis set with various atomic charge models. 155

C.1 Average and standard deviation of the distributions of ΔG_{el} contributions per (a) 60 anions and (b) 52 cations in water at both the B3LYP and HF levels of theory with the 6-311++G(d,p) basis set. 157
C.2 ΔG_{ne} distributions for 60 anions, and 52 cations in water using the αSASA, and CDS-ESP, CDS-CM5 models. 158

Bibliography

[1] F. W. Fowler, A. R. Katritzky, and R. J. D. Rutherford. The correlation of solvent effects on physical and chemical properties. *Journal of the Chemical Society B: Physical Organic*, page 460, 1971.

[2] Sia Nemat-Nasser, Shahram Zamani, and Yitzhak Tor. Effect of solvents on the chemical and physical properties of ionic polymer-metal composites. *Journal of Applied Physics*, 99(10):104902, May 2006.

[3] C. Gardner Swain, Marguerite S. Swain, Arnet L. Powell, and Sergio Alunni. Solvent effects on chemical reactivity. Evaluation of anion- and cation-solvation components. *Journal of the American Chemical Society*, 105(3):502–513, February 1983.

[4] E. Buncel and R. A. Stairs. *Solvent effects in chemistry.* Wiley, Hoboken, New Jersey, second edition edition, 2016.

[5] Orlando Tapia and J Bertrán. *Solvent effects and chemical reactivity.* Kluwer Academic Publishers, New York, 2002. OCLC: 50321338.

[6] Marcelo Takara and Amando Siuiti Ito. General and Specific Solvent Effects in Optical Spectra of ortho-Aminobenzoic Acid. *Journal of Fluorescence*, 15(2):171–177, March 2005.

[7] Pw Alexander and Rj Sleet. Solvent effects on the ultraviolet absorption spectra of o-, m-, and p-Hydroxybenzylideneimines. *Australian Journal of Chemistry*, 23(6):1183, 1970.

[8] M. S. Zakerhamidi, A. Ghanadzadeh, and M. Moghadam. Solvent Effects on the UV/ Visible Absorption Spectra of Some Aminoazobenzene Dyes. *Chemical Science Transactions*, 1(1):1–8, May 2012.

[9] I. A. Koppel and V. A. Palm. The Influence of the Solvent on Organic Reactivity. In N. B. Chapman and J. Shorter, editors, *Advances in Linear Free Energy Relationships*, pages 203–280. Springer US, Boston, MA, 1972.

[10] Wolfgang Linert. Thermodynamic implications of substituent and solvent effects on reactivity. *Journal of Chemical Information and Computer Sciences*, 32(3):221–226, May 1992.

[11] Jin Zhang, Haiyang Zhang, Tao Wu, Qi Wang, and David van der Spoel. Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents. *Journal of Chemical Theory and Computation*, 13(3):1034–1043, March 2017.

[12] Ronald M. Levy and Emilio Gallicchio. COMPUTER SIMULATIONS WITH EXPLICIT SOLVENT: Recent Progress in the Thermodynamic Decomposition of Free Energies and in Modeling Electrostatic Effects. *Annual Review of Physical Chemistry*, 49(1):531–567, October 1998.

[13] Alexey V. Onufriev and Saeed Izadi. Water models for biomolecular simulations: Water models for biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(2):e1347, March 2018.

[14] Ramu Anandakrishnan, Aleksander Drozdetski, Ross C. Walker, and Alexey V. Onufriev. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophysical Journal*, 108(5):1153–1164, March 2015.

[15] Norbert Schuch and Frank Verstraete. Computational complexity of interacting electrons and fundamental limitations of density functional theory. *Nature Physics*, 5(10):732–735, October 2009.

[16] Dario Vassetti, Marco Pagliai, and Piero Procacci. Assessment of GAFF2 and OPLS-AA General Force Fields in Combination with the Water Models TIP3P, SPCE, and OPC3 for the Solvation Free Energy of Druglike Organic Molecules. *Journal of Chemical Theory and Computation*, 15(3):1983–1995, March 2019.

[17] Jung Mee Park, Alessandro Laio, Marcella Iannuzzi, and Michele Parrinello. Dissociation Mechanism of Acetic Acid in Water. *Journal of the American Chemical Society*, 128(35):11318–11319, September 2006.

[18] Serena Donnini, Florian Tegeler, Gerrit Groenhof, and Helmut Grubmüller. Constant pH Molecular Dynamics in Explicit Solvent with lambda-Dynamics. *Journal of Chemical Theory and Computation*, 7(6):1962–1978, June 2011.

[19] David B. Kony, Wolfgang Damm, Serge Stoll, Wilfred F. van Gunsteren, and Philippe H. Hünenberger. Explicit-Solvent Molecular Dynamics Simulations of the Polysaccharide Schizophyllan in Water. *Biophysical Journal*, 93(2):442–455, July 2007.

[20] Jelle M. Boereboom, Paul Fleurat-Lessard, and Rosa E. Bulo. Explicit Solvation Matters: Performance of QM/MM Solvation Models in Nucleophilic Addition. *Journal of Chemical Theory and Computation*, 14(4):1841–1852, April 2018.

[21] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227–249, May 1976.

[22] Martin J. Field, Paul A. Bash, and Martin Karplus. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry*, 11(6):700–733, July 1990.

[23] Jacopo Tomasi and Maurizio Persico. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chemical Reviews*, 94(7):2027–2094, November 1994.

[24] Christopher J. Cramer and Donald G. Truhlar. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chemical Reviews*, 99(8):2161–2200, August 1999.

[25] Jacopo Tomasi, Benedetta Mennucci, and Roberto Cammi. Quantum Mechanical Continuum Solvation Models. *Chemical Reviews*, 105(8):2999–3094, August 2005.

[26] John M. Herbert. Dielectric continuum methods for quantum chemistry. *WIREs Computational Molecular Science*, 11(4), July 2021.

[27] Josefredo R. Pliego and José M. Riveros. The Cluster-Continuum Model for the Calculation of the Solvation Free Energy of Ionic Species. *The Journal of Physical Chemistry A*, 105(30):7241–7247, August 2001.

[28] Eirik F. da Silva, Hallvard F. Svendsen, and Kenneth M. Merz. Explicitly Representing the Solvation Shell in Continuum Solvent Calculations. *The Journal of Physical Chemistry A*, 113(22):6404–6409, June 2009.

[29] Vyacheslav S. Bryantsev, Mamadou S. Diallo, and William A. Goddard III. Calculation of Solvation Free Energies of Charged Solutes Using Mixed Cluster/Continuum Models. *The Journal of Physical Chemistry B*, 112(32):9709–9719, August 2008.

[30] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian˜16 Revision C.01, 2016.

[31] Roberto Dovesi, Alessandro Erba, Roberto Orlando, Claudio M. Zicovich-Wilson, Bartolomeo Civalleri, Lorenzo Maschio, Michel Rérat, Silvia Casassa, Jacopo Baima, Simone Salustro, and Bernard Kirtman. Quantum-mechanical condensed matter simulations with CRYSTAL. *WIREs Computational Molecular Science*, 8(4), July 2018.

[32] Filipp Furche, Reinhart Ahlrichs, Christof Hättig, Wim Klopper, Marek Sierka, and Florian Weigend. Turbomole. *WIREs Computational Molecular Science*, 4(2):91–100, March 2014.

[33] E. Aprà, E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. J. van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. O. Deegan, T. H. Dunning, M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtl,

L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H. Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J. Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. Martin del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols, J. Nieplocha, A. Otero-de-la Roza, B. Palmer, A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. A. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, Á. Vázquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. D. Vogiatzis, D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao, and R. J. Harrison. NWChem: Past, present, and future. *The Journal of Chemical Physics*, 152(18):184102, May 2020.

[34] J. Hafner and G. Kresse. The Vienna AB-Initio Simulation Program VASP: An Efficient and Versatile Tool for Studying the Structural, Dynamic, and Electronic Properties of Materials. In Antonios Gonis, Annemarie Meike, and Patrice E. A. Turchi, editors, *Properties of Complex Inorganic Solids*, pages 69–82. Springer US, Boston, MA, 1997.

[35] Paolo Giannozzi, Oscar Baseggio, Pietro Bonfà, Davide Brunato, Roberto Car, Ivan Carnimeo, Carlo Cavazzoni, Stefano de Gironcoli, Pietro Delugas, Fabrizio Ferrari Ruffino, Andrea Ferretti, Nicola Marzari, Iurii Timrov, Andrea Urru, and Stefano Baroni. Quantum ESPRESSO toward the exascale. *The Journal of Chemical Physics*, 152(15):154105, April 2020.

[36] A. V. Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer, and D. G. Truhlar. Minnesota solvation database-version 2012, 2012. Publication title: University of minnesota, minneapolis.

[37] Christoph Hille, Stefan Ringe, Martin Deimel, Christian Kunkel, William E. Acree, Karsten Reuter, and Harald Oberhofer. Generalized molecular solvation in non-aqueous solutions by a single parameter implicit solvation scheme. *The Journal of Chemical Physics*, 150(4):041710, January 2019.

[38] David L. Mobley and J. Peter Guthrie. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *Journal of Computer-Aided Molecular Design*, 28(7):711–720, July 2014.

[39] R Dovesi, V R Saunders, C Roetti, R Orlando, C M Zicovich-Wilson, F Pascale, B Civalleri, K Doll, N M Harrison, I J Bush, P D'Arco, M Llunell, M Causa, L Maschio Y. Noel, A Erba, M Rerat, and S Casassa. CRYSTAL17 user's manual (university of torino, torino, 2017). Publication title: CRYSTAL17 user's manual (university of torino, torino, 2017).

[40] Attila Szabo and Neil S. Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory.* Dover Publications, Mineola, N.Y, 1996.

[41] I.N. Levine. *Quantum chemistry.* Pearson Education, 2013.

[42] D. R. Hartree. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(1):111–132, January 1928.

[43] J. C. Slater. The Self Consistent Field and the Structure of Atoms. *Physical Review*, 32(3):339–348, September 1928.

[44] Robert G. Parr. Density Functional Theory of Atoms and Molecules. In Kenichi Fukui and Bernard Pullman, editors, *Horizons of Quantum Chemistry*, pages 5–15. Springer Netherlands, Dordrecht, 1980.

[45] David S Sholl and Janice A Steckel. *Density functional theory a practical introduction*. Wiley, Hoboken, N.J., 2009. OCLC: 1162598133.

[46] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140(4A):A1133–A1138, November 1965.

[47] P. A. M. Dirac. Note on Exchange Phenomena in the Thomas Atom. *Mathematical Proceedings of the Cambridge Philosophical Society*, 26(3):376–385, July 1930.

[48] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Physical Review Letters*, 77(18):3865–3868, October 1996.

[49] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Physical Review*, 136(3B):B864–B871, November 1964.

[50] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58(8):1200–1211, August 1980.

[51] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B*, 23(10):5048–5079, May 1981.

[52] Lee A. Cole and J. P. Perdew. Calculated electron affinities of the elements. *Physical Review A*, 25(3):1265–1271, March 1982.

[53] G. L. Oliver and J. P. Perdew. Spin-density gradient expansion for the kinetic energy. *Physical Review A*, 20(2):397–403, August 1979.

[54] Kieron Burke, John P. Perdew, and Matthias Ernzerhof. Why the generalized gradient approximation works and how to go beyond it. *International Journal of Quantum Chemistry*, 61(2):287–293, 1997.

[55] Kieron Burke, John P. Perdew, and Yue Wang. Derivation of a Generalized Gradient Approximation: The PW91 Density Functional. In John F. Dobson, Giovanni Vignale, and Mukunda P. Das, editors, *Electronic Density Functional Theory*, pages 81–111. Springer US, Boston, MA, 1998.

[56] Matthias Ernzerhof and Gustavo E. Scuseria. Assessment of the Perdew–Burke–Ernzerhof exchange-correlation functional. *The Journal of Chemical Physics*, 110(11):5029–5036, March 1999.

[57] Chengteh Lee and Carlos Sosa. Local density component of the Lee–Yang–Parr correlation energy functional. *The Journal of Chemical Physics*, 100(12):9018–9024, June 1994.

[58] Axel D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, April 1993.

[59] Axel D. Becke. A new mixing of Hartree–Fock and local density-functional theories. *The Journal of Chemical Physics*, 98(2):1372–1377, January 1993.

[60] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics*, 110(13):6158–6170, April 1999.

[61] Benjamin G. Janesko, Thomas M. Henderson, and Gustavo E. Scuseria. Screened hybrid density functionals for solid-state chemistry and physics. *Phys. Chem. Chem. Phys.*, 11(3):443–454, 2009.

[62] Roberto Dovesi, Bartolomeo Civalleri, Carla Roetti, Victor R. Saunders, and Roberto Orlando. Ab Initio Quantum Simulation in Solid State Chemistry. In Kenny B. Lipkowitz, Raima Larter, and Thomas R. Cundari, editors, *Reviews in Computational Chemistry*, pages 1–125. John Wiley & Sons, Inc., Hoboken, NJ, USA, January 2005.

[63] Charles Kittel. *Introduction to solid state physics.* Wiley, Hoboken, NJ, 8th ed edition, 2005.

[64] Christopher J. Cramer. *Essentials of computational chemistry: theories and models.* Wiley, Chichester, West Sussex, England ; Hoboken, NJ, 2nd ed edition, 2004.

[65] Peter Schwerdtfeger. The Pseudopotential Approximation in Electronic Structure Theory. *ChemPhysChem*, 12(17):3143–3155, December 2011.

[66] David J. Singh and Lars Nordström. *Planewaves, pseudopotentials, and the LAPW method.* Springer, New York, NY, 2nd ed edition, 2006.

[67] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953–17979, December 1994.

[68] Neil W. Ashcroft and N. David Mermin. *Solid state physics.* Holt, Rinehart and Winston, New York, 1976.

[69] CRYSTAL TUTORIAL PROJECT.

[70] M. Born. Volumen und Hydratationswärme der Ionen. *Zeitschrift für Physik*, 1(1):45–48, February 1920.

[71] John G. Kirkwood. The Dielectric Polarization of Polar Liquids. *The Journal of Chemical Physics*, 7(10):911–919, October 1939.

[72] Lars Onsager. Electric Moments of Molecules in Liquids. *Journal of the American Chemical Society*, 58(8):1486–1493, August 1936.

[73] S. Miertuš, E. Scrocco, and J. Tomasi. Electrostatic interaction of a solute with a continuum. A direct utilizaion of AB initio molecular potentials for the prevision of solvent effects. *Chemical Physics*, 55(1):117–129, February 1981.

[74] R. Bonaccorsi, R. Cimiraglia, and J. Tomasi. Ab initio evaluation of absorption and emission transitions for molecular solutes, including separate consideration of orientational and inductive solvent effects. *Journal of Computational Chemistry*, 4(4):567–577, 1983.

[75] J. Tomasi. Thirty years of continuum solvation chemistry: a review, and prospects for the near future. *Theoretical Chemistry Accounts*, 112(4), September 2004.

[76] Paweł Grochowski and Joanna Trylska. Continuum molecular electrostatics, salt effects, and counterion binding—A review of the Poisson–Boltzmann theory and its modifications. *Biopolymers*, 89(2):93–113, February 2008.

[77] Josefredo R. Pliego and Jose M. Riveros. Hybrid discrete-continuum solvation methods. *WIREs Computational Molecular Science*, 10(2), March 2020.

[78] Victor Gold, editor. *The IUPAC Compendium of Chemical Terminology: The Gold Book*. International Union of Pure and Applied Chemistry (IUPAC), Research Triangle Park, NC, 4 edition, 2019.

[79] Jacopo Tomasi and Benedetta Mennucci. Self-consistent Reaction Field Methods. In Paul von Rague Schleyer, Norman L. Allinger, Tim Clark, Johann Gasteiger, Peter A. Kollman, Henry F. Schaefer, and Peter R. Schreiner, editors, *Encyclopedia of Computational Chemistry*, page csa003. John Wiley & Sons, Ltd, Chichester, UK, April 2002.

[80] Aleksandr V. Marenich, Christopher J. Cramer, and Donald G. Truhlar. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *The Journal of Physical Chemistry B*, 113(18):6378–6396, May 2009. Publisher: American Chemical Society.

[81] Aleksandr V. Marenich, Christopher J. Cramer, and Donald G. Truhlar. Generalized Born Solvation Model SM12. *Journal of Chemical Theory and Computation*, 9(1):609–620, January 2013.

[82] Casey P. Kelly, Christopher J. Cramer, and Donald G. Truhlar. SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute-Water Clusters. *Journal of Chemical Theory and Computation*, 1(6):1133–1152, November 2005.

[83] Christopher J. Cramer and Donald G. Truhlar. A Universal Approach to Solvation Modeling. *Accounts of Chemical Research*, 41(6):760–768, June 2008.

[84] Alexander A. Voityuk and Sergei F. Vyboishchikov. Fast and accurate calculation of hydration energies of molecules and ions. *Physical Chemistry Chemical Physics*, 22(26):14591–14598, 2020.

[85] Ronald M. Levy, Linda Y. Zhang, Emilio Gallicchio, and Anthony K. Felts. On the Nonpolar Hydration Free Energy of Proteins: Surface Area and Continuum Solvent Models for the Solute-Solvent Interaction Energy. *Journal of the American Chemical Society*, 125(31):9523–9530, August 2003.

[86] Donald Bashford and David A. Case. Generalized Born Models of Macromolecular Solvation Effects. *Annual Review of Physical Chemistry*, 51(1):129–152, October 2000.

[87] Alexey V. Onufriev and David A. Case. Generalized Born Implicit Solvent Models for Biomolecules. *Annual Review of Biophysics*, 48(1):275–296, May 2019.

[88] Anthony J. Stone. Assessment of multipolar approximations to the induction energy. *Chemical Physics Letters*, 155(1):111–118, February 1989.

[89] Benedetta Mennucci and Roberto Cammi, editors. *Continuum Solvation Models in Chemical Physics.* John Wiley & Sons, Ltd, Chichester, UK, November 2007.

[90] Richard P. Feynman, Robert B. Leighton, Matthew L. Sands, and Richard P. Feynman. *Mainly electromagnetism and matter.* Number Feynman, Leighton, Sands ; volume 2 in The Feynman lectures on physics. Basic Books, New York, new millenium edition edition, 2010. OCLC: 838503554.

[91] C. J. F Böttcher and O. C. van Belle. *Dielectrics in static fields.* Elsevier Scientific Pub. Co., Amsterdam; New York, 1973. OCLC: 842939839.

[92] Daniel M. Chipman. Charge penetration in dielectric models of solvation. *The Journal of Chemical Physics*, 106(24):10194–10206, June 1997.

[93] Jaydeep P Bardhan. Biomolecular electrostatics—I want your solvation (model). *Computational Science & Discovery*, 5(1):013001, November 2012.

[94] Marc P. Coons and John M. Herbert. Quantum chemistry in arbitrary dielectric environments: Theory and implementation of nonequilibrium Poisson boundary conditions and application to compute vertical ionization energies at the air/water interface. *The Journal of Chemical Physics*, 148(22):222834, June 2018.

[95] Adrian W. Lange and John M. Herbert. A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: The switching/Gaussian approach. *The Journal of Chemical Physics*, 133(24):244111, December 2010.

[96] Walter Rocchia, Sundaram Sridharan, Anthony Nicholls, Emil Alexov, Alessandro Chiabrera, and Barry Honig. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *Journal of Computational Chemistry*, 23(1):128–137, January 2002.

[97] David J. Tannor, Bryan Marten, Robert Murphy, Richard A. Friesner, Doree Sitkoff, Anthony Nicholls, Barry Honig, Murco Ringnalda, and William A. Goddard. Accurate First Principles Calculation of Molecular Charge Distributions and Solvation Energies from Ab Initio Quantum Mechanics and Continuum Dielectric Theory. *Journal of the American Chemical Society*, 116(26):11875–11882, December 1994.

[98] J. Tomasi, B. Mennucci, and E. Cancès. The IEF version of the PCM solvation method: an overview of a new method addressed to study molecular solutes at the QM ab initio level. *Journal of Molecular Structure: THEOCHEM*, 464(1-3):211–226, May 1999.

[99] Jean-Luc Fattebert and Francois Gygi. Density functional theory for efficientab initio molecular dynamics simulations in solution. *Journal of Computational Chemistry*, 23(6):662–666, April 2002.

[100] Kim A. Sharp and Barry. Honig. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *The Journal of Physical Chemistry*, 94(19):7684–7692, September 1990.

[101] Markus Deserno and Christian Holm. Cell model and Poisson-Boltzmann theory: A brief introduction. *arXiv:cond-mat/0112096*, December 2001. arXiv: cond-mat/0112096.

[102] Gene Lamm. The Poisson-Boltzmann Equation. In Kenny B. Lipkowitz, Raima Larter, and Thomas R. Cundari, editors, *Reviews in Computational Chemistry*, pages 147–365. John Wiley & Sons, Inc., Hoboken, NJ, USA, September 2003.

[103] Nathan A. Baker. Poisson–Boltzmann Methods for Biomolecular Electrostatics. In *Methods in Enzymology*, volume 383, pages 94–118. Elsevier, 2004.

[104] R.R. Netz and H. Orland. Beyond Poisson-Boltzmann: Fluctuation effects and correlation functions. *The European Physical Journal E*, 1(2):203, 2000.

[105] Phil Attard. *Thermodynamics and statistical mechanics: equilibrium by entropy maximisation*. Academic Press, San Diego, Calif, 2002.

[106] Rita G. Lerner and George L. Trigg, editors. *Encyclopedia of physics*. VCH, New York, 2nd ed edition, 1991.

[107] Federico Fogolari, Pierfrancesco Zuccato, Gennaro Esposito, and Paolo Viglino. Biomolecular Electrostatics with the Linearized Poisson-Boltzmann Equation. *Biophysical Journal*, 76(1):1–16, January 1999.

[108] Huan-Xiang Zhou. Macromolecular electrostatic energy within the nonlinear Poisson–Boltzmann equation. *The Journal of Chemical Physics*, 100(4):3152–3162, February 1994.

[109] Peter Debye and Erich Hückel. Zur theorie der elektrolyte. I. Gefrierpunktserniedrigung und verwandte erscheinungen. *Physikalische Zeitschrift*, 24(185):305, 1923. tex.added-at: 2012-04-11T14:04:10.000+0200 tex.biburl: https://www.bibsonomy.org/bibtex/2a0dcc1272bc510726b9c0ecc930f5bd5/pkilian tex.interhash: ed6329cfbfa05dd4c6d2c3350d316b1c tex.intrahash: a0dcc1272bc510726b9c0ecc930f5bd5 tex.timestamp: 2012-04-11T14:14:28.000+0200.

[110] Changhao Wang, Pengyu Ren, and Ray Luo. Ionic Solution: What Goes Right and Wrong with Continuum Solvation Modeling. *The Journal of Physical Chemistry B*, 121(49):11169–11179, December 2017.

[111] David Leonard Chapman. LI. A contribution to the theory of electrocapillarity. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 25(148):475–481, April 1913.

[112] Otto Stern. ZUR THEORIE DER ELEKTROLYTISCHEN DOPPELSCHICHT. *Zeitschrift für Elektrochemie und angewandte physikalische Chemie*, 30(21-22):508–516, 1924.

[113] Hendrik Heinz and Ulrich W. Suter. Atomic Charges for Classical Simulations of Polar Systems. *The Journal of Physical Chemistry B*, 108(47):18341–18352, November 2004.

[114] K A Sharp and B Honig. Electrostatic Interactions in Macromolecules: Theory and Applications. *Annual Review of Biophysics and Biophysical Chemistry*, 19(1):301–332, June 1990.

[115] J. Katriel and E. R. Davidson. Asymptotic behavior of atomic and molecular wave functions. *Proceedings of the National Academy of Sciences*, 77(8):4403–4406, August 1980.

[116] Christian M. Cortis and Richard A. Friesner. An automatic three-dimensional finite element mesh generation system for the Poisson–Boltzmann equation. *Journal of Computational Chemistry*, 18(13):1570–1590, 1997.

[117] Christian M. Cortis and Richard A. Friesner. Numerical solution of the Poisson–Boltzmann equation using tetrahedral finite-element meshes. *Journal of Computational Chemistry*, 18(13):1591–1608, 1997.

[118] N. Baker, M. Holst, and F. Wang. Adaptive multilevel finite element solution of the Poisson–Boltzmann equation II. Refinement at solvent-accessible surfaces in biomolecular systems. *Journal of Computational Chemistry*, 21(15):1343–1352, 2000.

[119] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18):10037–10041, August 2001.

[120] M. Holst, N. Baker, and F. Wang. Adaptive multilevel finite element solution of the Poisson–Boltzmann equation I. Algorithms and examples. *Journal of Computational Chemistry*, 22(4):475–475, 2001.

[121] P.E. Dyshlovenko. Adaptive numerical method for Poisson–Boltzmann equation and its application. *Computer Physics Communications*, 147(1-2):335–338, August 2002.

[122] R. J. Zauhar and R. S. Morgan. The rigorous computation of the molecular electric potential. *Journal of Computational Chemistry*, 9(2):171–187, March 1988.

[123] AndréH Juffer, Eugen F.F Botta, Bert A.M van Keulen, Auke van der Ploeg, and Herman J.C Berendsen. The electric potential of a macromolecule in a solvent: A fundamental approach. *Journal of Computational Physics*, 97(1):144–171, November 1991.

[124] S.A. Allison and V.T. Tran. Modeling the electrophoresis of rigid polyions: application to lysozyme. *Biophysical Journal*, 68(6):2261–2270, June 1995.

[125] Ranganathan Bharadwaj, Andreas Windemuth, S. Sridharan, Barry Honig, and Anthony Nicholls. The fast multipole boundary element method for molecular electrostatics: An optimal approach for large systems. *Journal of Computational Chemistry*, 16(7):898–913, July 1995.

[126] Isaac Klapper, Ray Hagstrom, Richard Fine, Kim Sharp, and Barry Honig. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins: Structure, Function, and Genetics*, 1(1):47–59, January 1986.

[127] Michael K. Gilson, Kim A. Sharp, and Barry H. Honig. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *Journal of Computational Chemistry*, 9(4):327–335, June 1988.

[128] Michael K. Gilson and Barry Honig. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins: Structure, Function, and Genetics*, 4(1):7–18, 1988.

[129] Anthony Nicholls and Barry Honig. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *Journal of Computational Chemistry*, 12(4):435–445, May 1991.

[130] Anthony Nicholls and Barry Honig. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *Journal of Computational Chemistry*, 12(4):435–445, 1991. Publisher: Wiley Online Library.

[131] Doree Sitkoff, Kim A. Sharp, and Barry Honig. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *The Journal of Physical Chemistry*, 98(7):1978–1988, February 1994.

[132] David J. Tannor, Bryan Marten, Robert Murphy, Richard A. Friesner, Doree Sitkoff, Anthony Nicholls, Barry Honig, Murco Ringnalda, and William A. Goddard. Accurate First Principles Calculation of Molecular Charge Distributions and Solvation Energies from Ab Initio Quantum Mechanics and Continuum Dielectric Theory. *Journal of the American Chemical Society*, 116(26):11875–11882, December 1994.

[133] O. C Zienkiewicz, Robert Leroy Taylor, and J. Z Zhu. *The finite element method its basis and fundamentals, seventh edition.* Butterworth-Heinemann, Kidlington, Oxford, UK; Waltham, MA, 2013. OCLC: 881209508.

[134] Wikipedia contributors. *Wikipedia, The Free Encyclopedia.* 2004.

[135] Daniel M. Chipman. Simulation of volume polarization in reaction field theory. *The Journal of Chemical Physics*, 110(16):8012–8018, April 1999.

[136] Chang-Guo Zhan, John Bentley, and Daniel M. Chipman. Volume polarization in reaction field theory. *The Journal of Chemical Physics*, 108(1):177–192, January 1998.

[137] Daniel M. Chipman. Reaction field treatment of charge penetration. *The Journal of Chemical Physics*, 112(13):5558–5565, April 2000.

[138] Daniel M. Chipman. Energy correction to simulation of volume polarization in reaction field theory. *The Journal of Chemical Physics*, 116(23):10129–10138, June 2002.

[139] G.G. Gu and M.A. Gennert. Boundary element methods for solving Poisson equations in computer vision problems. In *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 546–551, Maui, HI, USA, 1991. IEEE Comput. Sco. Press.

[140] Kok Hwa Yu, A. Halim Kadarman, and Harijono Djojodihardjo. Development and implementation of some BEM variants—A critical review. *Engineering Analysis with Boundary Elements*, 34(10):884–899, October 2010.

[141] John T. Katsikadelis. The BEM for Nonlinear Problems. In *The Boundary Element Method for Engineers and Scientists*, pages 373–395. Elsevier, 2016.

[142] Malcolm E. Davis, Jeffry D. Madura, Brock A. Luty, and J.Andrew McCammon. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Computer Physics Communications*, 62(2-3):187–197, March 1991.

[143] Lin Li, Chuan Li, Subhra Sarkar, Jie Zhang, Shawn Witham, Zhe Zhang, Lin Wang, Nicholas Smith, Marharyta Petukh, and Emil Alexov. DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophysics*, 5(1):9, 2012.

[144] Donald Bashford. An object-oriented programming suite for electrostatic effects in biological molecules An experience report on the MEAD project. In G. Goos, J. Hartmanis, J. van Leeuwen, Yutaka Ishikawa, Rodney R. Oldehoeft, John V. W. Reynders, and Marydell Tholburn, editors, *Scientific Computing in Object-Oriented Parallel Environments*, volume 1343, pages 233–240. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997. Series Title: Lecture Notes in Computer Science.

[145] Christian Grossmann, Hans-Görg Roos, Martin Stynes, and SpringerLink (Online service). *Numerical Treatment of Partial Differential Equations.* 2007. OCLC: 1204025903.

[146] Joe D. Hoffman. *Numerical methods for engineers and scientists.* Marcel Dekker, New York, 2nd ed., rev. and expanded edition, 2001.

[147] James Stewart. *Calculus.* Cengage Learning, Boston, MA, USA, eighth edition edition, 2016. OCLC: ocn892432745.

[148] David M. Young. *Iterative solution of large linear systems.* Computer science and applied mathematics. Academic Press, New York, 1971.

[149] A. Hadjidimos. Successive overrelaxation (SOR) and related methods. *Journal of Computational and Applied Mathematics*, 123(1-2):177–199, November 2000.

[150] Pieter Wesseling. *An introduction to multigrid methods.* Pure and applied mathematics. Wiley, Chichester ; New York, 1992.

[151] Michael T. Heath. *Scientific computing: an introductory survey.* McGraw-Hill, Boston, 2nd ed edition, 2002.

[152] Li Xiao, Qin Cai, Xiang Ye, Jun Wang, and Ray Luo. Electrostatic forces in the Poisson-Boltzmann systems. *The Journal of Chemical Physics*, 139(9):094106, September 2013.

[153] Qin Cai, Xiang Ye, Jun Wang, and Ray Luo. Dielectric boundary force in numerical Poisson–Boltzmann methods: Theory and numerical strategies. *Chemical Physics Letters*, 514(4-6):368–373, October 2011.

[154] Qin Cai, Xiang Ye, and Ray Luo. Dielectric pressure in continuum electrostatic solvation of biomolecules. *Physical Chemistry Chemical Physics*, 14(45):15917, 2012.

[155] Manjeera Mantina, Adam C. Chamberlin, Rosendo Valero, Christopher J. Cramer, and Donald G. Truhlar. Consistent van der Waals Radii for the Whole Main Group. *The Journal of Physical Chemistry A*, 113(19):5806–5812, May 2009.

[156] A. Bondi. van der Waals Volumes and Radii. *The Journal of Physical Chemistry*, 68(3):441–451, March 1964.

[157] David R Lide. *Handbook of chemistry and physics*. CRC Press, Boca Ratón (Florida), 2011. OCLC: 782425882.

[158] Daniel M. Chipman. New formulation and implementation for volume polarization in dielectric continuum theory. *The Journal of Chemical Physics*, 124(22):224111, June 2006.

[159] Daniel M. Chipman and Michel Dupuis. Implementation of solvent reaction fields for electronic structure. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 107(2):90–102, February 2002.

[160] Feiwu Chen and Daniel M. Chipman. Boundary element methods for dielectric cavity construction and integration. *The Journal of Chemical Physics*, 119(19):10289–10297, November 2003.

[161] Patricia Schaefer, Demian Riccardi, and Qiang Cui. Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules. *The Journal of Chemical Physics*, 123(1):014905, July 2005.

[162] Tobias Benighaus and Walter Thiel. Efficiency and Accuracy of the Generalized Solvent Boundary Potential for Hybrid QM/MM Simulations: Implementation for Semiempirical Hamiltonians. *Journal of Chemical Theory and Computation*, 4(10):1600–1609, October 2008.

[163] Tobias Benighaus and Walter Thiel. A General Boundary Potential for Hybrid QM/MM Simulations of Solvated Biomolecular Systems. *Journal of Chemical Theory and Computation*, 5(11):3114–3128, November 2009.

[164] Tobias Benighaus and Walter Thiel. Long-Range Electrostatic Effects in QM/MM Studies of Enzymatic Reactions: Application of the Solvated Macromolecule Boundary Potential. *Journal of Chemical Theory and Computation*, 7(1):238–249, January 2011.

[165] James B. Foresman, Todd A. Keith, Kenneth B. Wiberg, John Snoonian, and Michael J. Frisch. Solvent Effects. 5. Influence of Cavity Shape, Truncation of Electrostatics, and Electron Correlation on ab Initio Reaction Field Calculations. *The Journal of Physical Chemistry*, 100(40):16098–16104, January 1996.

[166] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, 114(25):10024–10035, December 1992.

[167] B. Lee and F.M. Richards. The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology*, 55(3):379–IN4, February 1971.

[168] Robert B. Hermann. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. *The Journal of Physical Chemistry*, 76(19):2754–2759, September 1972.

[169] M. Connolly. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221(4612):709–713, August 1983.

[170] M. L. Connolly. Analytical molecular surface calculation. *Journal of Applied Crystallography*, 16(5):548–558, October 1983.

[171] Zixiang Tang, L. E. Scriven, and H. T. Davis. Effects of solvent exclusion on the force between charged surfaces in electrolyte solution. *The Journal of Chemical Physics*, 100(6):4527–4530, March 1994.

[172] Timothy J. Richmond. Solvent accessible surface area and excluded volume in proteins. *Journal of Molecular Biology*, 178(1):63–89, September 1984.

[173] Adrian W. Lange, John M. Herbert, Benjamin J. Albrecht, and Zhi-Qiang You. Intrinsically smooth discretisation of Connolly's solvent-excluded molecular surface. *Molecular Physics*, 118(6):e1644384, March 2020.

[174] Julius Parulek and Ivan Viola. Implicit representation of molecular surfaces. In *2012 IEEE Pacific Visualization Symposium*, pages 217–224, Songdo, Korea (South), February 2012. IEEE.

[175] Weihua Geng and G.W. Wei. Multiscale molecular dynamics using the matched interface and boundary method. *Journal of Computational Physics*, 230(2):435–457, January 2011.

[176] David H. Brookes and Teresa Head-Gordon. Family of Oxygen–Oxygen Radial Distribution Functions for Water. *The Journal of Physical Chemistry Letters*, 6(15):2938–2943, August 2015.

[177] Alexey V. Onufriev and Boris Aguilar. Accuracy of continuum electrostatic calculations based on three common dielectric boundary definitions. *Journal of Theoretical and Computational Chemistry*, 13(03):1440006, May 2014.

[178] Jessica M. J. Swanson, Stewart A. Adcock, and J. Andrew McCammon. Optimized Radii for Poisson-Boltzmann Calculations with the AMBER Force Field. *Journal of Chemical Theory and Computation*, 1(3):484–493, May 2005.

[179] Jean-Luc Fattebert and François Gygi. First-principles molecular dynamics simulations in a continuum solvent. *International Journal of Quantum Chemistry*, 93(2):139–147, 2003.

[180] Damián A. Scherlis, Jean-Luc Fattebert, François Gygi, Matteo Cococcioni, and Nicola Marzari. A unified electrostatic and cavitation model for first-principles molecular dynamics in solution. *The Journal of Chemical Physics*, 124(7):074103, February 2006.

[181] E. Silla, F. Villar, O. Nilsson, J.L. Pascual-Ahuir, and O. Tapia. Molecular volumes and surfaces of biomacromolecules via GEPOL: A fast and efficient algorithm. *Journal of Molecular Graphics*, 8(3):168–172, September 1990.

[182] Peifeng Su and Hui Li. Continuous and smooth potential energy surface for conductorlike screening solvation model using fixed points with variable areas. *The Journal of Chemical Physics*, 130(7):074109, February 2009.

[183] Adrian W. Lange and John M. Herbert. Polarizable Continuum Reaction-Field Solvation Models Affording Smooth Potential Energy Surfaces. *The Journal of Physical Chemistry Letters*, 1(2):556–561, January 2010.

[184] John Mongan, Carlos Simmerling, J. Andrew McCammon, David A. Case, and Alexey Onufriev. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *Journal of Chemical Theory and Computation*, 3(1):156–169, January 2007.

[185] Alexey N. Romanov, Sergey N. Jabin, Yaroslav B. Martynov, Alexey V. Sulimov, Fedor V. Grigoriev, and Vladimir B. Sulimov. Surface Generalized Born Method: A Simple, Fast, and Precise Implicit Solvent Model beyond the Coulomb Approximation. *The Journal of Physical Chemistry A*, 108(43):9323–9327, October 2004.

[186] Grigori Sigalov, Peter Scheffel, and Alexey Onufriev. Incorporating variable dielectric environments into the generalized Born model. *The Journal of Chemical Physics*, 122(9):094511, March 2005.

[187] W. Clark Still, Anna Tempczyk, Ronald C. Hawley, and Thomas Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society*, 112(16):6127–6129, August 1990.

[188] Alexey Onufriev, David A. Case, and Donald Bashford. Effective Born radii in the generalized Born approximation: The importance of being perfect. *Journal of Computational Chemistry*, 23(14):1297–1304, November 2002.

[189] Gregory D. Hawkins, Christopher J. Cramer, and Donald G. Truhlar. Pairwise solute descreening of solute charges from a dielectric medium. *Chemical Physics Letters*, 246(1-2):122–129, November 1995.

[190] Gregory D. Hawkins, Christopher J. Cramer, and Donald G. Truhlar. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *The Journal of Physical Chemistry*, 100(51):19824–19839, January 1996.

[191] Dario Vassetti and Frédéric Labat. Evaluation of the performances of different atomic charge and nonelectrostatic models in the finite-difference Poisson–Boltzmann approach. *International Journal of Quantum Chemistry*, 121(7), April 2021.

[192] R. S. Mulliken. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *The Journal of Chemical Physics*, 23(10):1833–1840, October 1955.

[193] Per-Olov Löwdin. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *The Journal of Chemical Physics*, 18(3):365–375, March 1950.

[194] Aleksandr V. Marenich, Steven V. Jerome, Christopher J. Cramer, and Donald G. Truhlar. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *Journal of Chemical Theory and Computation*, 8(2):527–541, February 2012.

[195] Ryan M. Olson, Aleksandr V. Marenich, Christopher J. Cramer, and Donald G. Truhlar. Charge Model 4 and Intramolecular Charge Polarization. *Journal of Chemical Theory and Computation*, 3(6):2046–2054, November 2007.

[196] Paul Winget, Jason D. Thompson, James D. Xidos, Christopher J. Cramer, and Donald G. Truhlar. Charge Model 3: A Class IV Charge Model Based on Hybrid Density Functional Theory with Variable Exchange. *The Journal of Physical Chemistry A*, 106(44):10707–10717, November 2002.

[197] U. Chandra Singh and Peter A. Kollman. An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry*, 5(2):129–145, April 1984.

[198] Curt M. Breneman and Kenneth B. Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *Journal of Computational Chemistry*, 11(3):361–373, April 1990.

[199] Saeed Izadi, Robert C. Harris, Marcia O. Fenley, and Alexey V. Onufriev. Accuracy Comparison of Generalized Born Models in the Calculation of Electrostatic Binding Free Energies. *Journal of Chemical Theory and Computation*, 14(3):1656–1670, March 2018.

[200] Jayashree Srinivasan, Megan W. Trevathan, Paul Beroza, and David A. Case. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 101(6):426–434, May 1999.

[201] Ramu Anandakrishnan, Charles Baker, Saeed Izadi, and Alexey V. Onufriev. Point Charges Optimally Placed to Represent the Multipole Expansion of Charge Distributions. *PLoS ONE*, 8(7):e67715, July 2013.

[202] Miroslav Medved', Šimon Budzák, Wojciech Bartkowiak, and Heribert Reis. Solvent Effects on Molecular Electric Properties. In Jerzy Leszczynski, Anna Kaczmarek-Kedziera, Tomasz Puzyn, Manthos G. Papadopoulos, Heribert Reis, and Manoj

K. Shukla, editors, *Handbook of Computational Chemistry*, pages 741–794. Springer International Publishing, Cham, 2017.

[203] Ming Wah Wong, Michael J. Frisch, and Kenneth B. Wiberg. Solvent effects. 1. The mediation of electrostatic effects by solvents. *Journal of the American Chemical Society*, 113(13):4776–4782, June 1991.

[204] Ming Wah Wong, Kenneth B. Wiberg, and Michael J. Frisch. Solvent effects. 2. Medium effect on the structure, energy, charge density, and vibrational frequencies of sulfamic acid. *Journal of the American Chemical Society*, 114(2):523–529, January 1992.

[205] A. R Edmonds. *Angular Momentum in Quantum Mechanics.* 2016. OCLC: 1002061151.

[206] Shirley W. Harrison, Hans J. Nolte, and David L. Beveridge. Free energy of a charge distribution in a spheroidal cavity in a polarizable dielectric continuum. *The Journal of Physical Chemistry*, 80(23):2580–2585, November 1976.

[207] S. Ehrenson. Transmission of substituent effects. Generalization of the ellipsoidal cavity field effect model. *Journal of the American Chemical Society*, 98(24):7510–7514, November 1976.

[208] Clifford E. Felder and Jon Applequist. Energies of solute molecules from an atom charge–dipole interaction model with a surrounding dielectric: Application to Gibbs energies of proton transfer between carboxylic acids in water. *The Journal of Chemical Physics*, 75(5):2390–2398, September 1981.

[209] Ronald P. Bell. The electrostatic energy of dipole molecules in different media. *Transactions of the Faraday Society*, 27:797, 1931.

[210] H. Kirchberg, P. Nalbach, and M. Thorwart. Nonequilibrium quantum solvation with a time-dependent Onsager cavity. *The Journal of Chemical Physics*, 148(16):164301, 2018.

[211] Kurt V. Mikkelsen, Poul Jo/rgensen, and Hans Jo/rgen Aagaard Jensen. A multiconfiguration self-consistent reaction field response method. *The Journal of Chemical Physics*, 100(9):6597–6607, May 1994.

[212] Devarajan Thirumalai, Kunizo Onda, and Donald G. Truhlar. Electron scattering by CO_2 : Elastic scattering, rotational excitation, and excitation of the asymmetric stretch at 10 eV impact energy. *The Journal of Chemical Physics*, 74(12):6792–6805, June 1981.

[213] Ove Christiansen and Kurt V. Mikkelsen. A coupled-cluster solvent reaction field method. *The Journal of Chemical Physics*, 110(3):1365–1375, January 1999.

[214] Daniel Rinaldi and Jean-Louis Rivail. Polarisabilites moleculaires et effet dielectrique de milieu l'etat liquide. etude theorique de la molecule d'eau et de ses dimeres. *Theoretica Chimica Acta*, 32(1):57–70, March 1973.

[215] Daniel Rinaldi, Alain Bouchy, Jean-Louis Rivail, and Valerie Dillet. A self-consistent reaction field model of solvation using distributed multipoles. I. Energy and energy derivatives. *The Journal of Chemical Physics*, 120(5):2343–2350, February 2004.

[216] Robert Rein. On Physical Properties and Interactions of Polyatomic Molecules: With Application to Molecular Recognition in Biology. In *Advances in Quantum Chemistry*, volume 7, pages 335–396. Elsevier, 1973.

[217] W.Andrzej Sokalski and R.A. Poirier. Cumulative atomic multipole representation of the molecular charge distribution and its basis set dependence. *Chemical Physics Letters*, 98(1):86–92, June 1983.

[218] Giovanni Scalmani and Michael J. Frisch. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *The Journal of Chemical Physics*, 132(11):114110, March 2010.

[219] Claudio Amovilli, Vincenzo Barone, Roberto Cammi, Eric Cancès, Maurizio Cossi, Benedetta Mennucci, Christian S. Pomelli, and Jacopo Tomasi. Recent Advances in the Description of Solvent Effects with the Polarizable Continuum Model. In *Advances in Quantum Chemistry*, volume 32, pages 227–261. Elsevier, 1998.

[220] Maurizio Cossi, Giovanni Scalmani, Nadia Rega, and Vincenzo Barone. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *The Journal of Chemical Physics*, 117(1):43–54, July 2002.

[221] Filippo Lipparini, Giovanni Scalmani, Benedetta Mennucci, Eric Cancès, Marco Caricato, and Michael J. Frisch. A variational formulation of the polarizable continuum model. *The Journal of Chemical Physics*, 133(1):014106, July 2010.

[222] Eric Cancès and Benedetta Mennucci. The escaped charge problem in solvation continuum models. *The Journal of Chemical Physics*, 115(13):6130–6135, October 2001.

[223] Roald K. Wangsness. *Electromagnetic fields.* Wiley, New York, 2nd ed edition, 1986.

[224] Vincenzo Barone and Maurizio Cossi. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *The Journal of Physical Chemistry A*, 102(11):1995–2001, March 1998.

[225] Maurizio Cossi, Nadia Rega, Giovanni Scalmani, and Vincenzo Barone. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *Journal of Computational Chemistry*, 24(6):669–681, April 2003.

[226] A. Klamt and G. Schüürmann. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2*, (5):799–805, 1993.

[227] Jan Andzelm, Christoph Kölmel, and Andreas Klamt. Incorporation of solvent effects into density functional calculations of molecular energies and geometries. *The Journal of Chemical Physics*, 103(21):9312–9320, December 1995.

[228] Andreas Klamt and Volker Jonas. Treatment of the outlying charge in continuum solvation models. *The Journal of Chemical Physics*, 105(22):9972–9981, December 1996.

[229] Andreas Klamt. The COSMO and COSMO-RS solvation models. *WIREs Computational Molecular Science*, 1(5):699–709, September 2011.

[230] Daniel M. Chipman. Comparison of solvent reaction field representations. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 107(2):80–89, February 2002.

[231] Maurizio Cossi, Benedetta Mennucci, Jesus Pitarch, and Jacopo Tomasi. Correction of cavity-induced errors in polarization charges of continuum solvation models. *Journal of Computational Chemistry*, 19(8):833–846, 1998.

[232] A. Klamt, C. Moya, and J. Palomar. A Comprehensive Comparison of the IEFPCM and SS(V)PE Continuum Solvation Methods with the COSMO Approach. *Journal of Chemical Theory and Computation*, 11(9):4220–4225, September 2015.

[233] Jennifer J. Guerard and J. Samuel Arey. Critical Evaluation of Implicit Solvent Models for Predicting Aqueous Oxidation Potentials of Neutral Organic Compounds. *Journal of Chemical Theory and Computation*, 9(11):5046–5058, November 2013.

[234] Dean G Duffy and Taylor & Francis. *Green's Functions with Applications.* 2020. OCLC: 1110595931.

[235] A. D Polianin and A. V Manzhirov. *Handbook of integral equations.* Chapman & Hall/CRC, Boca Raton, 2008. OCLC: 244159978.

[236] *Integral Operators in Potential Theory.* Springer, Berlin, 1980. OCLC: 838032334.

[237] Marie Jose Huron and Pierre Claverie. Calculation of the interaction energy of one molecule with its whole surrounding. I. Method and application to pure nonpolar compounds. *The Journal of Physical Chemistry*, 76(15):2123–2133, July 1972.

[238] C. Dupont, O. Andreussi, and N. Marzari. Self-consistent continuum solvation (SCCS): The case of charged systems. *The Journal of Chemical Physics*, 139(21):214110, December 2013.

[239] Oliviero Andreussi, Ismaila Dabo, and Nicola Marzari. Revised self-consistent continuum solvation in electronic-structure calculations. *The Journal of Chemical Physics*, 136(6):064102, February 2012.

[240] Giuseppe Fisicaro, Luigi Genovese, Oliviero Andreussi, Sagarmoy Mandal, Nisanth N. Nair, Nicola Marzari, and Stefan Goedecker. Soft-Sphere Continuum Solvation in Electronic-Structure Calculations. *Journal of Chemical Theory and Computation*, 13(8):3829–3845, August 2017.

[241] Paul Winget, Derek Dolney, David Giesen, Christopher Cramer, and Donald Truhlar. Minnesota solvent descriptor database. 1999.

[242] V. Gogonea and E. Ōsawa. Implementation of solvent effect in molecular mechanics Part 3. The first- and second-order analytical derivatives of excluded volume. *Journal of Molecular Structure: THEOCHEM*, 311:305–324, July 1994.

[243] G. Hummer, S. Garde, A. E. Garcia, A. Pohorille, and L. R. Pratt. An information theory model of hydrophobic interactions. *Proceedings of the National Academy of Sciences*, 93(17):8951–8955, August 1996.

[244] Robert A. Pierotti. A scaled particle theory of aqueous and nonaqueous solutions. *Chemical Reviews*, 76(6):717–726, December 1976.

[245] J. Langlet, P. Claverie, J. Caillet, and A. Pullman. Improvements of the continuum model. 1. Application to the calculation of the vaporization thermodynamic quantities of nonassociated liquids. *The Journal of Physical Chemistry*, 92(6):1617–1631, March 1988.

[246] Robert A. Pierotti. THE SOLUBILITY OF GASES IN LIQUIDS 1. *The Journal of Physical Chemistry*, 67(9):1840–1845, September 1963.

[247] P. Claverie. Intermolecular interactions: From diatomics to biopolymers. In B. Pullman, editor, *Intermolecular Interactions: From Diatomics to Biopolymers*. John wiley & sons edition, 1978. tex.city: New York.

[248] Bernard Theodoor Thole and Petrus Theodorus Duijnen. On the quantum mechanical treatment of solvent effects. *Theoretica Chimica Acta*, 55(4):307–318, 1980.

[249] Daniel Rinaldi, Benedito J. Costa Cabral, and Jean-Louis Rivail. Influence of dispersion forces on the electronic structure of a solvated molecule. *Chemical Physics Letters*, 125(5-6):495–499, April 1986.

[250] M.A. Aguilar and F.J.Olivares del Valle. A computation procedure for the dispersion component of the interaction energy in continuum solute-solvent models. *Chemical Physics*, 138(2-3):327–336, November 1989.

[251] F. Floris and J. Tomasi. Evaluation of the dispersion contribution to the solvation energy. A simple computational model in the continuum approximation. *Journal of Computational Chemistry*, 10(5):616–627, July 1989.

[252] F. M. Floris, J. Tomasi, and J. L. Pascual Ahuir. Dispersion and repulsion contributions to the solvation energy: Refinements to a simple computational model in the continuum approximation. *Journal of Computational Chemistry*, 12(7):784–791, September 1991.

[253] F. M. Floris, M. Selmi, A. Tani, and J. Tomasi. Free energy and entropy for inserting cavities in water: Comparison of Monte Carlo simulation and scaled particle theory results. *The Journal of Chemical Physics*, 107(16):6353–6365, October 1997.

[254] Vladimir Frecher, Stanislav Miertus, and Magdaléna Májeková. Modelling of dispersion and repulsion interactions in liquids. *Journal of Molecular Structure: THEOCHEM*, 227:157–173, March 1991.

[255] Christopher J. Cramer and Donald G. Truhlar. General parameterized SCF model for free energies of solvation in aqueous solution. *Journal of the American Chemical Society*, 113(22):8305–8311, October 1991.

[256] Claudio Amovilli and Benedetta Mennucci. Self-Consistent-Field Calculation of Pauli Repulsion and Dispersion Contributions to the Solvation Free Energy in the Polarizable Continuum Model. *The Journal of Physical Chemistry B*, 101(6):1051–1057, February 1997.

[257] Claudio Amovilli and Roy McWeeny. A matrix partitioning approach to the calculation of intermolecular potentials. General theory and some examples. *Chemical Physics*, 140(3):343–361, February 1990.

[258] Anna Pomogaeva and Daniel M. Chipman. New Implicit Solvation Models for Dispersion and Exchange Energies. *The Journal of Physical Chemistry A*, 117(28):5812–5820, July 2013.

[259] Roberto Cammi, Vincenzo Verdolino, Benedetta Mennucci, and Jacopo Tomasi. Towards the elaboration of a QM method to describe molecular solutes under the effect of a very high pressure. *Chemical Physics*, 344(1-2):135–141, February 2008.

[260] R. Cammi, C. Cappelli, B. Mennucci, and J. Tomasi. Calculation and analysis of the harmonic vibrational frequencies in molecules at extreme pressure: Methodology and diborane as a test case. *The Journal of Chemical Physics*, 137(15):154112, October 2012.

[261] David J. Giesen, Christopher J. Cramer, and Donald G. Truhlar. Entropic Contributions to Free Energies of Solvation. *The Journal of Physical Chemistry*, 98(15):4141–4147, April 1994.

[262] Anton E. Krukowski, Hue Sun Chan, and Ken A. Dill. An exact lattice model of complex solutions: Chemical potentials depend on solute and solvent shape. *The Journal of Chemical Physics*, 103(24):10675–10688, December 1995.

[263] Seishi Shimizu, Mitsunori Ikeguchi, Shugo Nakamura, and Kentaro Shimizu. Size dependence of transfer free energies: A hard-sphere-chain- based formalism. *The Journal of Chemical Physics*, 110(6):2971–2982, February 1999.

[264] Jayashree Srinivasan, Thomas E. Cheatham, Piotr Cieplak, Peter A. Kollman, and David A. Case. Continuum solvent studies of the stability of DNA, RNA, and Phosphoramidate-DNA helices. *Journal of the American Chemical Society*, 120(37):9401–9409, 1998.

[265] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, July 2004.

[266] Michael J. Robertson, Julian Tirado-Rives, and William L. Jorgensen. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *Journal of Chemical Theory and Computation*, 11(7):3499–3509, July 2015.

[267] Dawei Zhang and Raudah Lazim. Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Scientific Reports*, 7(1):44651, April 2017.

[268] Shuai Wei, Charles L. Brooks, and Aaron T. Frank. A rapid solvent accessible surface area estimator for coarse grained molecular simulations. *Journal of Computational Chemistry*, 38(15):1270–1274, June 2017.

[269] Tore Brinck, Jane S. Murray, and Peter Politzer. Polarizability and volume. *The Journal of Chemical Physics*, 98(5):4305–4306, March 1993.

[270] Sergei F. Vyboishchikov and Alexander A. Voityuk. Fast non-iterative calculation of solvation energies for water and non-aqueous solvents. *Journal of Computational Chemistry*, 42(17):1184–1194, June 2021.

[271] Jaroslaw W Drelich, Ludmila Boinovich, Emil Chibowski, Claudio Della Volpe, Lucyna Hołysz, Abraham Marmur, and Stefano Siboni. Contact angles: history of over 200 years of open questions. *Surface Innovations*, 8(1-2):3–27, February 2020.

[272] G. Fisicaro, L. Genovese, O. Andreussi, N. Marzari, and S. Goedecker. A generalized Poisson and Poisson-Boltzmann solver for electrostatic environments. *The Journal of Chemical Physics*, 144(1):014103, January 2016.

[273] Lauren Beger, L. Roberts, Kim K. Degroh, and B. Banks. Use of atomic oxygen for increased water contact angles of various polymers for biomedical applications. 2007.

[274] Keiko Gotoh, Akemi Yasukawa, and Yasuyuki Kobayashi. Wettability characteristics of poly(ethylene terephthalate) films treated by atmospheric pressure plasma and ultraviolet excimer light. *Polymer Journal*, 43(6):545–551, June 2011.

[275] Sangwha Lee, Joon-Seo Park, and T. Randall Lee. The Wettability of Fluoropolymer Surfaces: Influence of Surface Dipoles. *Langmuir*, 24(9):4817–4826, May 2008.

[276] Maurizio Cossi. Continuum solvation model for infinite periodic systems. *Chemical Physics Letters*, 384(1-3):179–184, January 2004.

[277] Verónica M. Sánchez, Mariela Sued, and Damián A. Scherlis. First-principles molecular dynamics simulations at solid-liquid interfaces with a continuum solvent. *The Journal of Chemical Physics*, 131(17):174108, November 2009.

[278] Hui-Fang Wang and Zhi-Pan Liu. Formic Acid Oxidation at Pt/H $_2$ O Interface from Periodic DFT Calculations Integrated with a Continuum Solvation Model. *The Journal of Physical Chemistry C*, 113(40):17502–17508, October 2009.

[279] Damián A. Scherlis, Jean-Luc Fattebert, François Gygi, Matteo Cococcioni, and Nicola Marzari. A unified electrostatic and cavitation model for first-principles molecular dynamics in solution. *The Journal of Chemical Physics*, 124(7):074103, February 2006.

[280] Damián A. Scherlis, Jean-Luc Fattebert, and Nicola Marzari. Stacking of oligo- and polythiophene cations in solution: Surface tension and dielectric saturation. *The Journal of Chemical Physics*, 124(19):194902, May 2006.

[281] Kiran Mathew, Ravishankar Sundararaman, Kendra Letchworth-Weaver, T. A. Arias, and Richard G. Hennig. Implicit solvation model for density-functional study of nanocrystal surfaces and reaction pathways. *The Journal of Chemical Physics*, 140(8):084106, February 2014.

[282] Matthew Fishman, Houlong L. Zhuang, Kiran Mathew, William Dirschka, and Richard G. Hennig. Accuracy of exchange-correlation functionals and effect of solvation on the surface energy of copper. *Physical Review B*, 87(24):245402, June 2013.

[283] J. Dziedzic, H. H. Helal, C.-K. Skylaris, A. A. Mostofi, and M. C. Payne. Minimal parameter implicit solvent model for ab initio electronic-structure calculations. *EPL (Europhysics Letters)*, 95(4):43001, August 2011.

[284] Jacek Dziedzic, Stephen J. Fox, Thomas Fox, Christofer S. Tautermann, and Chris-Kriton Skylaris. Large-scale DFT calculations in implicit solvent-A case study on the T4 lysozyme L99A/M102Q protein. *International Journal of Quantum Chemistry*, 113(6):771–785, March 2013.

[285] Wen-Jin Yin, Matthias Krack, Xibo Li, Li-Zhen Chen, and Li-Min Liu. Periodic continuum solvation model integrated with first-principles calculations for solid surfaces. *Progress in Natural Science: Materials International*, 27(2):283–288, April 2017.

[286] Ravishankar Sundararaman, Kendra Letchworth-Weaver, Kathleen A. Schwarz, Deniz Gunceler, Yalcin Ozhabes, and T.A. Arias. JDFTx: Software for joint density-functional theory. *SoftwareX*, 6:278–284, 2017.

[287] B. Delley. The conductor-like screening model for polymers and surfaces. *Molecular Simulation*, 32(2):117–123, February 2006.

[288] J. A. Grant, B. T. Pickup, and A. Nicholls. A smooth permittivity function for Poisson-Boltzmann solvation methods. *Journal of Computational Chemistry*, 22:608–640, 2001.

[289] Frédéric Labat, Bartolomeo Civalleri, and Roberto Dovesi. Implicit solvation using a generalized finite-difference approach in CRYSTAL: Implementation and results for molecules, polymers, and surfaces. *Journal of Chemical Theory and Computation*, 14(11):5969–5983, November 2018.

[290] Anna-Karin Tornberg. The Ewald sums for singly, doubly and triply periodic electrostatic systems. *arXiv:1404.3534 [math-ph]*, April 2014. arXiv: 1404.3534.

[291] R. S. Mulliken. Electronic population analysis on LCAO-MO molecular wave functions. I. *The Journal of Chemical Physics*, 23(10):1833–1840, October 1955.

[292] F. L. Hirshfeld. Bonded-atom fragments for describing molecular charge densities. *Theoretica Chimica Acta*, 44(2):129–138, 1977.

[293] Patrick Bultinck, Christian Van Alsenoy, Paul W Ayers, and Ramon Carbó-Dorca. Critical analysis and extension of the Hirshfeld atoms in molecules. *The Journal of Chemical Physics*, 126(14):144111, 2007. Publisher: American Institute of Physics.

[294] C. M. Zicovich-Wilson, M. Hô, A. M. Navarrete-López, and S. Casassa. Hirshfeld-I charges in linear combination of atomic orbitals periodic calculations. *Theoretical Chemistry Accounts*, 135(8):188, August 2016.

[295] Aleksandr V. Marenich, Steven V. Jerome, Christopher J. Cramer, and Donald G. Truhlar. Charge model 5: An extension of hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *Journal of Chemical Theory and Computation*, 8(2):527–541, February 2012.

[296] Peter Politzer and Robert S. Mulliken. Comparison of two atomic charge definitions, as applied to the hydrogen fluoride molecule. *The Journal of Chemical Physics*, 55(10):5135–5136, November 1971.

[297] P. Ros and G. C.A. Schuit. Molecular orbital calculations on copper chloride complexes. *Theoretica Chimica Acta*, 4(1):1–12, 1966.

[298] Ernest R. Davidson. Electronic population analysis of molecular wavefunctions. *The Journal of Chemical Physics*, 46(9):3320–3324, May 1967.

[299] Alain Veillard. Electronic structure of lithiumacetylene. *The Journal of Chemical Physics*, 48(5):2012–2016, 1968.

[300] L. Ch Cusachs and P. Politzer. On the problem of defining the charge on an atom in a molecule. *Chem. Phys. Letters*, 1(11):529–531, 1968.

[301] Peter Politzer and Roger R. Harris. Properties of atoms in molecules. I. A proposed definition of the charge on an atom in a molecule. *Journal of the American Chemical Society*, 92(22):6451–6454, November 1970.

[302] Peter Politzer and Robert S. Mulliken. Comparison of two atomic charge definitions, as applied to the hydrogen fluoride molecule. *The Journal of Chemical Physics*, 55(10):5135–5136, 1971.

[303] C. W. Kern and M. Karplus. Analysis of charge distributions: Hydrogen fluoride. *The Journal of Chemical Physics*, 40(5):1374–1389, March 1964.

[304] Paul W Ayers. Atoms in molecules, an axiomatic approach. I. Maximum transferability. *The Journal of Chemical Physics*, 113(24):10886–10898, 2000. Publisher: American Institute of Physics.

[305] Maximiliano Riquelme, Alejandro Lara, David L Mobley, Toon Verstraelen, Adelio R Matamala, and Esteban Vöhringer-Martinez. Hydration free energies in the FreeSolv Database calculated with polarized iterative Hirshfeld charges. *J. Chem. Inf. Model*, 58(9):1779–1797, 2018. Publisher: ACS Publications.

[306] Toon Verstraelen, Sergey V Sukhomlinov, Veronique Van Speybroeck, Michel Waroquier, and Konstantin S Smirnov. Computation of charge distribution and electrostatic potential in silicates with the use of chemical potential equalization models. *Journal of Physical Chemistry C*, 116(1):490–504, 2012. Publisher: ACS Publications.

[307] Ye Mei, Andrew C. Simmonett, Frank C. Pickard, Robert A. DiStasio, Bernard R. Brooks, and Yihan Shao. Numerical study on the partitioning of the molecular polarizability into fluctuating charge and induced atomic dipole contributions. *The J. Phys. Chem. A*, 119(22):5865–5882, June 2015.

[308] Kenneth B. Wiberg and Paul R. Rablen. Comparison of atomic charges derived via different procedures. *Journal of Computational Chemistry*, 14(12):1504–1518, December 1993.

[309] Alexander A. Voityuk, Anton J. Stasyuk, and Sergei F. Vyboishchikov. A simple model for calculating atomic charges in molecules. *Physical Chemistry Chem. Phys.*, 20(36):23328–23337, 2018.

[310] Jonah Z. Vilseck, Julian Tirado-Rives, and William L. Jorgensen. Evaluation of CM5 charges for condensed-phase modeling. *Journal of Chemical Theory and Computation*, 10(7):2802–2812, July 2014.

[311] Michael H. Abraham, Priscilla L. Grellier, David V. Prior, Philip P. Duce, Jeffrey J. Morris, and Peter J. Taylor. Hydrogen bonding. Part 7. A scale of solute hydrogen-bond acidity based on log K values for complexation in tetrachloromethane. *Journal of the Chemical Society, Perkin Transactions 2*, (6):699, 1989. Publisher: Royal Society of Chemistry (RSC) tex.source: Crossref.

[312] Michael H. Abraham. Scales of solute hydrogen-bonding: Their construction and application to physicochemical and biochemical processes. *Chemical Society Reviews*, 22(2):73, 1993. Publisher: Royal Society of Chemistry (RSC) tex.source: Crossref.

[313] Michael H. Abraham. Hydrogen bonding. 31. Construction of a scale of solute effective or summation hydrogen-bond basicity. *Journal of Physical Organic Chemistry*, 6(12):660–684, December 1993. Publisher: Wiley tex.source: Crossref.

[314] H. Winde. Quantitative treatments of Solute/Solvent interactions. *Zeitschrift fur Physikalische Chemie*, 193(Part 1):217–217, January 1996. Publisher: Walter de Gruyter GmbH tex.source: Crossref.

[315] Tianhai Zhu, Jiabo Li, Daniel A. Liotard, Christopher J. Cramer, and Donald G. Truhlar. Analytical energy gradients of a self-consistent reaction-field solvation model based on CM2 atomic charges. *The Journal of Chemical Physics*, 110(12):5503–5513, March 1999.

[316] Chunhu Tan, Yu-Hong Tan, and Ray Luo. Implicit Nonpolar Solvent Models. *The Journal of Physical Chemistry B*, 111(42):12263–12274, October 2007.

[317] A. Shrake and J.A. Rupley. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *Journal of Molecular Biology*, 79(2):351–371, September 1973.

[318] Huajun Wang and Cyrus Levinthal. A vectorized algorithm for calculating the accessible surface area of macromolecules. *Journal of Computational Chemistry*, 12(7):868–871, September 1991.

[319] Scott M. Le Grand and Kenneth M. Merz. Rapid approximation to molecular surface area via the use of Boolean logic and look-up tables. *Journal of Computational Chemistry*, 14(3):349–352, March 1993.

[320] Frank Eisenhaber, Philip Lijnzaad, Patrick Argos, Chris Sander, and Michael Scharf. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*, 16(3):273–284, March 1995.

[321] N. Futamura, S. Aluru, D. Ranjan, and B. Hariharan. Efficient parallel algorithms for solvent accessible surface area of proteins. *IEEE Transactions on Parallel and Distributed Systems*, 13(6):544–555, June 2002.

[322] S. J. Wodak and J. Janin. Analytical approximation to the accessible surface area of proteins. *Proceedings of the National Academy of Sciences*, 77(4):1736–1740, April 1980.

[323] Estanislao Silla, Iñaki Tuñón, and Juan Luis Pascual-Ahuir. GEPOL: An improved description of molecular surfaces II. Computing the molecular area and volume: GEPOL: AN IMPROVED DESCRIPTION OF MOLECULAR SURFACES. *Journal of Computational Chemistry*, 12(9):1077–1088, November 1991.

[324] G. Perrot, B. Cheng, K.D. Gibson, J. Vila, K.A. Palmer, A. Nayeem, B. Maigret, and H.A. Scheraga. MSEED: A program for the rapid analytical determination of accessible surface areas and their derivatives. *Journal of Computational Chemistry*, 13(1):1–11, January 1992.

[325] Maurizio Cossi, Benedetta Mennucci, and Roberto Cammi. Analytical first derivatives of molecular surfaces with respect to nuclear coordinates. *Journal of Computational Chemistry*, 17(1):57–73, 1996.

[326] R. Cammi and J. Tomasi. Analytical derivatives for molecular solutes. II. Hartree–Fock energy first and second derivatives with respect to nuclear coordinates. *The Journal of Chemical Physics*, 101(5):3888–3897, September 1994.

[327] Laura Wesson and David Eisenberg. Atomic solvation parameters applied to molecular dynamics of proteins in solution: Atomic solvation parameters of proteins in solution. *Protein Science*, 1(2):227–235, February 1992.

[328] Berthold von Freyberg, Timothy J. Richmond, and Werner Braun. Surface Area Included in Energy Refinement of Proteins. *Journal of Molecular Biology*, 233(2):275–292, September 1993.

[329] Manfredo Perdigão do Carmo. *Differential geometry of curves and surfaces*. Dover Publications Inc, Mineola New York, revised & updated second edition edition, 2016.

[330] Frank Eisenhaber and Patrick Argos. Improved strategy in analytic surface calculation for molecular systems: Handling of singularities and computational efficiency. *Journal of Computational Chemistry*, 14(11):1272–1280, November 1993.

[331] Shura Hayryan, Chin-Kun Hu, Jaroslav Skřivánek, Edik Hayryane, and Imrich Pokorný. A new analytical method for computing solvent-accessible surface area of macromolecules and its gradients. Method for Computing Solvent-Accessible Surface Area. *Journal of Computational Chemistry*, 26(4):334–343, January 2005. Publisher: Wiley tex.source: Crossref.

[332] Richard J. Lisle and Peter R. Leyshon. *Stereographic Projection Techniques for Geologists and Civil Engineers.* Cambridge University Press, 2 edition, April 2004.

[333] M. Krone, K. Bidmon, and T. Ertl. Interactive Visualization of Molecular Surface Dynamics. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1391–1398, November 2009.

[334] David L. Mobley, Christopher I. Bayly, Matthew D. Cooper, Michael R. Shirts, and Ken A. Dill. Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations. *Journal of Chemical Theory and Computation*, 5(2):350–358, 2009.

[335] Nils J. D. Drechsel, Christopher J. Fennell, Ken A. Dill, and Jordi Villà-Freixa. TRIFORCE: Tessellated Semianalytical Solvent Exposed Surface Areas and Derivatives. *Journal of Chemical Theory and Computation*, 10(9):4121–4132, September 2014.

[336] H. M. Berman. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, January 2000.

[337] Yann Pellegrin, Loïc Le Pleux, Errol Blart, Adèle Renaud, Benoit Chavillon, Nadine Szuwarski, Mohammed Boujtita, Laurent Cario, Stéphane Jobic, Denis Jacquemin, and Fabrice Odobel. Ruthenium polypyridine complexes as sensitizers in NiO based p-type dye-sensitized solar cells: Effects of the anchoring groups. *Journal of Photochemistry and Photobiology A: Chemistry*, 219(2-3):235–242, April 2011.

[338] Michael Wykes, Fabrice Odobel, Carlo Adamo, Ilaria Ciofini, and Frédéric Labat. Anchoring groups for dyes in p-DSSC application: insights from DFT. *Journal of Molecular Modeling*, 22(12):289, December 2016.

[339] István Mayer. *Bond orders and energy components: extracting chemical information from molecular wave functions.* CRC Press, 2016.

[340] A. Hadjidimos. Successive overrelaxation (SOR) and related methods. *Journal of Computational and Applied Mathematics*, 123(1-2):177–199, November 2000.

[341] Dario Vassetti, Bartolomeo Civalleri, and Frédéric Labat. Analytical calculation of the solvent-accessible surface area and its nuclear gradients by stereographic projection: A general approach for molecules, polymers, nanotubes, helices, and surfaces. *Journal of Computational Chemistry*, 41(15):1464–1479, 2020.

[342] Jason D Thompson, Christopher J Cramer, and Donald G Truhlar. New universal solvation model and comparison of the accuracy of the SM5.42R, SM5.43R, C-PCM, D-PCM, and IEF-PCM continuum solvation models for aqueous and organic solvation free energies and for vapor pressures. *The J. Phys. Chem. A*, 108(31):6532–6542, 2004. Publisher: ACS Publications.

[343] E Cances, Benedetta Mennucci, and J Tomasi. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *The Journal of Chemical Physics*, 107(8):3032–3041, 1997. Publisher: American Institute of Physics.

[344] De-Li Chen, Abraham C. Stern, Brian Space, and J. Karl Johnson. Atomic charges derived from electrostatic potentials for molecular and periodic systems. *Journal of Physical Chemistry A*, 114(37):10225–10233, September 2010. Publisher: American Chemical Society (ACS) tex.source: Crossref.

[345] Carlos Campana, Bastien Mussard, and Tom K. Woo. Electrostatic potential derived atomic charges for periodic systems using a modified error functional. *Journal of Chemical Theory and Computation*, 5(10):2866–2878, August 2009. Publisher: American Chemical Society (ACS) tex.source: Crossref.

[346] Andrea Gabrieli, Marco Sant, Pierfranco Demontis, and Giuseppe B. Suffritti. Partial charges in periodic systems: Improving electrostatic potential (ESP) fitting via total dipole fluctuations and multiframe approaches. *Journal of Chemical Theory and Computation*, 11(8):3829–3843, July 2015. Publisher: American Chemical Society (ACS) tex.source: Crossref.

[347] F. Pascale, C.M. Zicovich-Wilson, F. López Gejo, B. Civalleri, R. Orlando, and R. Dovesi. The calculation of the vibrational frequencies of crystalline compounds and its implementation in the CRYSTAL code. Crystalline Compounds and the CRYSTAL Code. *Journal of Computational Chemistry*, 25(6):888–897, March 2004. Publisher: Wiley tex.source: Crossref.

[348] CM Zicovich-Wilson, Fabien Pascale, Carla Roetti, VR Saunders, Roberto Orlando, and Roberto Dovesi. Calculation of the vibration frequencies of alpha-quartz: The effect of Hamiltonian and basis set. *Journal of Computational Chemistry*, 25(15):1873–1881, 2004. Publisher: Wiley Online Library.

[349] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59(3):1758, 1999. Publisher: APS.

[350] M.J.D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal*, 7(2):155–162, February 1964. Publisher: Oxford University Press (OUP) tex.source: Crossref.

[351] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R.J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E.A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, February 2020. Publisher: Springer Science and Business Media LLC tex.source: Crossref.

[352] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with

NumPy. *Nature*, 585(7825):357–362, September 2020. Publisher: Springer Science and Business Media LLC tex.source: Crossref.

[353] Ole Tange. GNU parallel 20200522 ('kraftwerk'), May 2020.

[354] A. Vittadini, A. Selloni, F.P. Rotzinger, and M. Grätzel. Structure and energetics of water adsorbed at TiO2 anatase (101) and (001) surfaces. *Physical Review Letters*, 81(14):2954–2957, October 1998. Publisher: American Physical Society (APS) tex.source: Crossref.

[355] Jenny Schneider, Masaya Matsuoka, Masato Takeuchi, Jinlong Zhang, Yu Horiuchi, Masakazu Anpo, and Detlef W. Bahnemann. Understanding TiO2 photocatalysis: Mechanisms and materials. *Chemical Reviews*, 114(19):9919–9986, September 2014. Publisher: American Chemical Society (ACS) tex.source: Crossref.

[356] Andrea Vittadini, Maurizio Casarin, and Annabella Selloni. Chemistry of and on TiO2-anatase surfaces by DFT calculations: A partial review. *Theor Chem Account*, 117(5-6):663–671, December 2006. Publisher: Springer Science and Business Media LLC tex.source: Crossref.

[357] Frédéric Labat, Tangui Le Bahers, Ilaria Ciofini, and Carlo Adamo. First-principles modeling of dye-sensitized solar cells: Challenges and perspectives. *Accounts of Chemical Research*, 45(8):1268–1277, April 2012. Publisher: American Chemical Society (ACS) tex.source: Crossref.

[358] Ulrike Diebold. The surface science of titanium dioxide. *Surface Science Reports*, 48(5-8):53–229, January 2003. Publisher: Elsevier BV tex.source: Crossref.

[359] Mingzheng Ge, Qingsong Li, Chunyan Cao, Jianying Huang, Shuhui Li, Songnan Zhang, Zhong Chen, Keqin Zhang, Salem S. Al-Deyab, and Yuekun Lai. One-dimensional TiO2 nanotube photocatalysts for solar water splitting. *Advanced Science*, 4(1):1600152, September 2016. Publisher: Wiley tex.source: Crossref.

[360] Kiran Mathew, V.S. Chaitanya Kolluru, Srinidhi Mula, Stephan N. Steinmann, and Richard G. Hennig. Implicit self-consistent electrolyte model in plane-wave density-functional theory. *The Journal of Chemical Physics*, 151(23):234101, December 2019. Publisher: AIP Publishing tex.source: Crossref.

[361] I. Can Oguz, D. Vassetti, and F. Labat. Assessing the performances of different continuum solvation models for the calculation of hydration energies of molecules, polymers and surfaces: a comparison between the SMD, VASPsol and FDPB models. *Theoretical Chemistry Accounts*, page submitted.

[362] Frédéric Labat, Philippe Baranek, and Carlo Adamo. Structural and electronic properties of selected rutile and anatase TiO2 surfaces: An ab initio investigation. *Journal of Chemical Theory and Computation*, 4(2):341–352, January 2008. Publisher: American Chemical Society (ACS) tex.source: Crossref.

[363] Peng Lian, Ryne C. Johnston, Jerry M. Parks, and Jeremy C. Smith. Quantum chemical calculation of pKas of environmentally relevant functional groups: Carboxylic acids, amines, and thiols in aqueous solution. *Journal of Physical Chemistry*

A, 122(17):4366–4374, April 2018. Publisher: American Chemical Society (ACS) tex.source: Crossref.

[364] Rodrigo Casasnovas, Joaquin Ortega-Castro, Juan Frau, Josefa Donoso, and Francisco Muñoz. Theoretical pKacalculations with continuum model solvents, alternative protocols to thermodynamic cycles. *International Journal of Quantum Chemistry*, 114(20):1350–1363, May 2014. Publisher: Wiley tex.source: Crossref.

[365] Bishnu Thapa and H. Bernhard Schlegel. Density functional theory calculation of pKa's of thiols in aqueous solution using explicit water molecules and the polarizable continuum model. *Journal of Physical Chemistry A*, 120(28):5726–5735, June 2016. Publisher: American Chemical Society (ACS) tex.source: Crossref.

[366] Donald M. Camaioni and Christine A. Schwerdtfeger. Comment on "Accurate experimental values for the free energies of hydration of H+, OH-, and H3O+". *Journal of Physical Chemistry A*, 109(47):10795–10797, November 2005. Publisher: American Chemical Society (ACS) tex.source: Crossref.

[367] Abdirisak A. Isse and Armando Gennaro. Absolute potential of the standard hydrogen electrode and the problem of interconversion of potentials in different solvents. *Journal of Physical Chemistry B*, 114(23):7894–7899, May 2010. Publisher: American Chemical Society (ACS) tex.source: Crossref.

[368] Aleksandr V. Marenich, Junming Ho, Michelle L. Coote, Christopher J. Cramer, and Donald G. Truhlar. Computational electrochemistry: Prediction of liquid-phase reduction potentials. *Physical Chemistry Chemical Physics*, 16(29):15068–15106, 2014. Publisher: Royal Society of Chemistry (RSC) tex.source: Crossref.

[369] Junming Ho and Michelle L. Coote. A universal approach for continuum solvent pK a calculations: Are we there yet? *Theoretical Chemistry Accounts*, 125(1-2):3–21, November 2009. Publisher: Springer Science and Business Media LLC tex.source: Crossref.

[370] Junming Ho and Michelle L Coote. First-principles prediction of acidities in the gas and solution phase. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):649–660, 2011. Publisher: Wiley Online Library.

[371] Thomas S. Hofer and Philippe H. Hunenberger. Absolute proton hydration free energy, surface potential of water, and redox potential of the hydrogen electrode from first principles: QM/MM MD free-energy simulations of sodium and potassium hydration. *The Journal of Chemical Physics*, 148(22):222814, June 2018. Publisher: AIP Publishing tex.source: Crossref.

[372] Philippe Hunenberger and Maria Reif. *Single-ion solvation. Experimental and Theoretical Approaches to Elusive Thermodynamic Quantities.* Theoretical and computational chemistry series. Royal Society of Chemistry, 2011. ISSN: 2041-319X Pages: P001–664 tex.source: Crossref.

[373] Wonpil Im, Dmitrii Beglov, and Benoît Roux. Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Computer Physics Communications*, 111(1-3):59–75, June 1998.

[374] R.J. Zauhar and R.S. Morgan. A new method for computing the macromolecular electric potential. *Journal of Molecular Biology*, 186(4):815–820, December 1985.

[375] M. E. Davis and J. A. McCammon. Calculating electrostatic forces from grid-calculated potentials. *Journal of Computational Chemistry*, 11(3):401–409, April 1990.

[376] R. J. Zauhar. The incorporation of hydration forces determined by continuum electrostatics into molecular mechanics simulations. *Journal of Computational Chemistry*, 12(5):575–583, June 1991.

[377] Jianwei Che, Joachim Dzubiella, Bo Li, and J. Andrew McCammon. Electrostatic Free Energy and Its Variations in Implicit Solvent Models. *The Journal of Physical Chemistry B*, 112(10):3058–3069, March 2008.

[378] Bo Li, Xiaoliang Cheng, and Zhengfang Zhang. Dielectric Boundary Force in Molecular Solvation with the Poisson–Boltzmann Free Energy: A Shape Derivative Approach. *SIAM Journal on Applied Mathematics*, 71(6):2093–2111, January 2011.

[379] William H. Press, editor. *Numerical recipes in Fortran 77: the art of scientific computing.* Number 1 in Fortran numerical recipes. Cambridge University Press, Cambridge, 2. ed., repr., corr. to software version 2.10 edition, 2007.

RÉSUMÉ

La prise en compte des effets environnementaux est cruciale pour mieux modéliser les systèmes chimiques, car les propriétés chimiques et physiques de la matière sont fortement influencées par leur interaction avec leur environnement. Pour inclure les effets de solvatation à faible coût de calcul dans les calculs de mécanique quantique (MQ), les modèles de solvatation implicite sont très intéressants car ils tiennent compte de l'environnement du solvant en remplaçant les degrés de liberté du solvant par un diélectrique sans structure caractérisé par la permittivité relative du solvant, en se concentrant principalement sur les interactions électrostatiques entre le soluté et le solvant.

Historiquement, les modèles de solvatation implicite ont surtout été développés pour les systèmes moléculaires finis, tandis que les systèmes périodiques, tels que les polymères et les surfaces, ont reçu moins d'attention, à la fois en raison du manque de données expérimentales comme les énergies de solvatation, et de la complexité accrue des formalismes MQ et électrostatiques, à cause de la nature périodique des équations mises en jeu. Dans cette thèse, nous nous concentrerons sur le développement, l'implémentation et la validation d'un modèle de solvatation implicite généralisé dans le code Crystal, applicable aux systèmes finis et périodiques, aux niveaux Hartree-Fock (HF) et théorie de la fonctionnelle de la densité (DFT).

Les effets électrostatiques et la polarisation mutuelle soluté-solvant, sont traités par une approche de champ réactionnel auto-cohérent (Self-Consistent Reaction Field, SCRF), utilisant des charges apparentes de surface (Apparent Surface Charges, ASC) pour représenter la polarisation du solvant, tandis que le problème électrostatique sous-jacent est résolu numériquement par une Équation de Poisson en différences finies (FDP) utilisant des charges ponctuelles atomiques, dont les modèles ont été généralisés pour fonctionner dans le cadre du SCRF. Les effets non électrostatiques sont inclus grâce à l'implémentation, la reparamétrisation et l'extension aux systèmes périodiques du modèle « Cavity, Dispersion, and Solvent structural effects » (CDS). Le modèle CDS a de plus été modifié afin de mieux décrire les espèces chargées, grâce à l'introduction d'un terme supplémentaire basé sur le paramètre d'acidité d'Abraham du solvant. En outre, comme le modèle non électrostatique dépend de la surface accessible au solvant (SASA) du soluté, une technique de projection stéréographique a été mise en œuvre et étendue aux systèmes périodiques pour le calcul analytique de la SASA et de ses gradients atomiques.

Pour les systèmes moléculaires finis, le calcul des énergies de solvatation dans divers solvants et de pKa dans l'eau ont été utilisés pour valider le modèle. En revanche, pour les systèmes périodiques tels que les surfaces, la comparaison avec les données de référence obtenues avec VASPsol a été utilisée, en raison du manque général de données expérimentales pour ces solutés.

Dans l'ensemble, le modèle de solvatation implicite généralisé étudié, implémenté et mis en œuvre dans cette thèse permet une description plus réaliste d'une large gamme de solutés allant de systèmes finis à infinis périodiques, permettant une modélisation plus précise de ces systèmes.

MOTS CLÉS

Continuum, Solvatation, Solvation, Surfaces, CRYSTAL

ABSTRACT

Taking in account environmental effects is crucial to better model chemical systems, since both chemical and physical properties of matter are heavily influenced by their interaction with their surroundings. To include solvation effects at low computational cost in quantum-mechanical (QM) calculations, implicit solvation models are very appealing since they account for the solvent environment by replacing the solvent degrees of freedom with a structureless dielectric characterized by the relative permittivity of the solvent, focusing mainly on the electrostatic interactions between the solute and the solvent.

Historically, implicit solvation models have mostly been developed for finite molecular systems, while periodic systems, such as polymers and surfaces, have received less attention, both for the lack of experimental data, such as solvation energies, and increased complexities in the QM and electrostatic formalisms due to the periodic nature of the equations involved. In this thesis, we will focus on the development, implementation and validation of a generalized implicit solvation model in the Crystal code, applicable to both finite and periodic systems, at the Hartree-Fock (HF) and Density Functional Theory (DFT) levels. The electrostatic effects, and solute-solvent mutual polarization, are treated through a Self-Consistent Reaction Field approach (SCRF), using Apparent Surface Charges (ASC) to represent solvent polarization, while the underlying electrostatic problem is numerically solved through a Finite-Difference generalized Poisson scheme (FDP) using atomic point charges, which models have been extended to work within the SCRF framework. The non-electrostatic effects are included thanks to the implementation, re-parametrization and extension to periodic systems of the Cavity, Dispersion, and Solvent structural effects model (CDS). The CDS model has been further revised to better describe charged species through an additional term based on Abraham's acidity parameter of the solvent. Furthermore, as the non-electrostatic model is dependent on the Solvent Accessible Surface Area (SASA) of the solute, a stereographic projection technique has been implemented and extended to periodic systems for the analytical calculation of SASA and its atomic gradients.

For finite molecular systems, the calculation of solvation energies in various solvents and pKa's in water have been used to validate the model. Instead for periodic systems, such as surfaces, comparison with reference data obtained with VASPsol has been used, due to the general lack of experimental data for such solutes.

Overall, the generalized implicit solvation model investigated and implemented in this thesis allows for a more realistic description of a broad range of solutes ranging from finite to infinite periodic systems, allowing for a more accurate modeling of such systems.

KEYWORDS

Continuum, Solvation, Surfaces, CRYSTAL