Implementation of the SM12 Solvation Model into ADF and ADF-BAND

By

Craig A. Peeples

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Chemistry

University of Manitoba

Winnipeg

Table of Contents

Table of contents……. 2

List of figures……..…….…5

List of Tables……...…7

List of Schemes ……...……9

Abstract……...……10

Acknowledgement……..…11

List of Abbreviations ……12

Chapter 1: Introduction……..13

1 General Introduction……13

1.1 Solvation……..14

1.1.1 Implicit Solvation……..17

1.2 Secular equation and the Slater Determinant……19

1.3 Density Functional Theory……...21

1.3.1 Approximate Functionals……..25

1.3.2 Slater Type Orbitals and Gaussian Type Orbitals……..28

1.4 Solid-State Considerations……...30

1.4.1 Band theory……...…31

1.4.2 ADF-BAND Basis Sets……32

1.5 Organization of this Thesis……...……34

1.6 References ……...…35

Chapter 2: SM12 Theory……...42

2.1 Implicit Solvation Model 12, Polarization……42

2.1.1 SM12 Solvent Parameters…….……49

2.1.2 Overall Solvation Energy……..……53

2.2 References……54

Preface to Chapter 3……...……59

Chapter 3: Implementation of the SM12 Solvation Model into ADF and Comparison with COSMO…… 60

Preface to Chapter 4……...……95

Chapter 4: Periodic Solvation in ADF-BAND……...……...……97

Chapter 5: Summary and Future developments and studies of SM12 in both ADF and ADF-BAND…….135

5.1 Summary……135

5.2 Future Work……...…137

5.3 References……...…140

Appendix 1……...…143

Appendix 2……...…192

List of Figures

Figure 1.1. Schematic representation of three common ways to solvate a system. The blue mesh describes the continuum smeared over the molecule

Figure 2.1. Schematic illustration of the intersecting atomic spheres underlying the ASA. a) Cross section of where the two spheres meet; the radius from the centre to this intersection is defined by r. b) Full intersection.

Figure 3.1. Comparison of CM5 in ADF (TZ2P STOs) to CM5 literature results obtained with Gaussian 09 (MG3S GTOs). The data set comprises 721 molecular compounds (5880 charges), with varying charges and multiplicities. Results can be seen in Appendix 1.2

Figure 3.2. Comparison of HPA in ADF (TZ2P STOs) to HPA in Gaussian 09 (MG3S GTOs). The data set comprises with 721 molecular compounds (5880 charges), with varying charges and multiplicities. Results can be seen in Appendix 1.2.

Figure 3.3. Comparison of SM12 (blue squares) and COSMO (violet circles) water solvation free energies with the GGA PBE using DZP.

Figure 3.4. SM12 comparison to experiment of 376 compounds in a water solvent, HF DZP calculations

Figure 3.5. SM12 comparison to experiment of 272 compounds in a water solvent, M06-2X TZP calculations.

Figure 3.6. SM12 comparison to experiment of 2141 single point calculations of various compounds in the 97 available solvents, B3LYP TZP calculations

Figure 4.1. Schematic representation of the split of the surface area between each image, in order to account for SASA edge effects.

Figure 4.2. Reaction cycle of solvating the adsorption of CO onto a Cu bilayer.

List of Tables

Table 2.1. A list of the elemental types parameterized for equations 2.18 & 2.19. *X is generically set up for the rest of the periodic table.

Table 3.1. Parameters used for describing the solvent (See Text). Each parameter is customizable. *All 92 solvents are listed in Appendix 1.1 in the Supporting Information.

Table 3.2. MUE of ADF based calculations and experiment of 381 molecules. Solvation energies are all calculated in water solvent. Single point calculations were done on optimized gas-phase geometries from the Minnesota Solvation database. All energies are reported in kcal/mol, and the solvation free energy calculations of Hartree-Fock using DZP are in Appendix 1.4. *All Q-Chem calculations are from the Minnesota Solvation database.

Table 3.3. MUE of ADF based calculations and experiment of 272 molecules. Solvation energies are all calculated in water solvent. Calculations were done on optimized gas-phase geometries from the Minnesota Solvent database, and their implementation into Gaussian 09. All energies are reported in kcal/mol, and can be seen in Appendix 1.3.

Table 3.4. MUE of ADF based calculations and experiment of the 92 solvents available within SM12, and varying compounds. Calculations were done on optimized geometries from the Minnesota Solvent database, and their implementation into Gaussian 09. All energies are reported in kcal/mol. 2141 single point calculations were performed.*Marenich et al. had reported 2129 calculations

Table 3.5. MUE of ADF based calculations and experiment of 109 varying compounds in three different solvents. Calculations were done on optimized geometries from the Minnesota Solvent database, and their implementation into Gaussian 09. All energies are reported in kcal/mol, and in Appendix 1.5.

Table 3.6. MUE of ADF based calculations and experiment of 60 anions and MUE of ADF based calculations and experiment of 52 cations. Solvation energies are all calculated in water solvent. Calculations were done with optimizations from the Minnesota Solvent database, and their implementation into Gaussian 09. All energies are reported in kcal/mol; these results can also be seen in Appendix 1.6.

Table 4.1. Periodic zero-dimensional solvation electrostatics: representative examples of the energy of solvation (kcal/mol). Calculations were done on optimized gas-phase geometries from the Minnesota Solvent database, and their implementation into Gaussian 09

Table 4.2. Total bonding energy and energy of solvation. All energies are reported in kcal/mol. These calculations were based on gas phase optimizations, done within ADF-BAND at the LDA/TZP level of theory, of a bilayer of Cu with a CO layer adsorbed to the top.

Table 4.3. Energy of adsorption and energy of solvation for adsorption. All energies are reported in kcal/mol. These calculations were based on gas-phase optimizations, done within ADF-BAND at the LDA/TZP level of theory, of a bilayer of CO on Cu.

List of Schemes

Scheme 2.1. Breakdown of associated interactions with hydrogen and its nearest neighbors within the COT region.

Scheme 2.2. Breakdown of associated interactions with carbon and its nearest neighbors within the COT region.

Abstract

Modeling systems in liquid is imperative to chemistry, as many reactions take place in liquid, and nearly all of biochemistry is in the liquid state. Being able to appropriately account for the energy associated with reactions in liquids is required for full study of these systems. A general overview of solvation and quantum chemistry is presented in Chapter 1. Solvation Model 12 (SM12) is the newest Generalized Born Approximation iteration of a series of solvation models from Minnesota, it shows great promise for accurate, description of solutions, the theory behind SM12 is detailed in Chapter 2. Before the work presented in Chapter 3, there has been no testing or version of SM12 on a Slater Type Orbital code, and it had only been tested on Gaussian Type Orbitals. Shown is the full implementation of SM12 in to the pure Slater Type Orbital code, the Amsterdam Density Functional (ADF) package in particular. The model performs as well as its Gaussian Type Orbital counterpart, and outperforms the native solvation model in ADF, COSMO, while performing similarly to the COSMO-RS code within the ADF package. The model has been extended to account for periodic boundary conditions, as presented by the ADF-BAND code, and is shown to perform exactly as it does in the molecular code. The extension to infinite boundaries creates interesting edge effects that need to be taken into consideration, and are accounted for through cut off approximations and a screening function to ensure the potential is well-behaved. This extension is detailed in Chapter 4, since there are no experimental data for solvation energies of periodic systems, presented is an argument for the implementation and a case study of adsorption solvation energies. Chapter 5 summarizes this thesis, and presents several areas for improvement, and future work.

Acknowledgments

I would like to thank Dr. Georg Schreckenbach, my supervisor, for allowing me to work on my thesis within his group. His leadership, patience, help, understanding, ideas, and critical thought have always been a guide and inspiration for my studies, and continued learning. I would also like to thank the members of my advisory committee, Dr. Jennifer van Wijngaarden, and Dr. Jesko Sirker.

I also would like to thank the past, and current group members of the Schreckenbach group at the University of Manitoba. Grigory Shamov, whom was extremely helpful for my undergraduate studies, as well as help with working within the local clusters. Jeffery Perkins for always being helpful for talking through ideas with, Xiaobin Zhang for extra discussions.

I would also like to thank the people and group members of Software for Chemistry & Materials for travel and in-kind support. Dr. Stan van Gisbergen, for leading the research, Dr. Pier Philipsen for being more than helpful with working within ADF-BAND, Erik van Lenthe for his input and help within ADF.

I would like to thank my family. Without the help and reassurance of my significant other, Amber Williams, my mother father and brother, (Corinne, William, Robert Peeples), my grandmother, Annie Fulton, and my aunt and uncle, Candice and Rick Sherbo, this task would have never been completed.

Abbreviations

ASC	Apparent Surface Charge
ADF	Amsterdam Density Functional
ASA	Analytical Surface Area
AUG	Augmented
CDS	Cavitation, Dispersion, Solvent
CM5	Charge Model 5
COSMO	Conductor like Screening Model
COT	Cut-off Tan
DFT	Density Functional Theory
DMSO	Dimethyl sulfoxide
DSSC	Dye Sensitized Solar Cells
ENP	Electronic, Nuclear, Polarization
FF	Force Field
GBA	General Born Approximation
GGA	Generalized Gradient Approximation
GTO	Gaussian Type Orbital
HPA	Hirshfeld Population Analysis
IEF	Integral Equation Formalism
LCAO	Linear Combinations of Atomic Orbitals
LDA	Local Density Approximation
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
MO	Molecular Orbital theory
MUE	Mean unsigned Error
PBC	Periodic Boundary Conditions
PCM	Polarizable Continuum Model
QM	Quantum Mechanics
R2	Correlation Constant
SASA	Solvent Accessible Surface Area
SCF	Self consistent field
SCM	Scientific Computing and Modeling
SCRF	Self Consistent Reaction Field
SM12	Solvation Model 12
STO	Slater type Orbital
SVPE	Surface and Volume Polarization for Electrostatic

Chapter 1: Introduction

1 General Introduction

Solvation, in its simplest description, defines moving a particle from a gas-phase to a liquid-phase. Solvation is describing the process of attraction when one inserts a particle into a condensed packing of molecules confined to a uniform composition. Doing so requires to take into account the re-arrangement of the condensed medium due to the charge distribution of the inserted particle. This, in turn, has an effect on the inserted particle which constitutes a change in internal energy. The rearrangement of the liquid creates a cavity around the particle which creates an energy change as well. This process has been studied thoroughly through many different models and approximations.[1,2,3] Solvation can be used to calculate many properties that otherwise would not be possible to obtain without the inclusion of the solvation continuum, including but not limited to transition state geometries, electrochemical properties and electronic dynamics.

The inclusion of solvation into theoretical calculations can have a profound effect on the systems of interest. Solvation, in general, can help stabilize compounds that are unstable in gas-phase. The structure of compounds can be highly dependent on whether they are within a stabilizing polarized environment. For instance, glycine in gas-phase shows a very different structure than glycine in solution, requiring the energy of solvation to properly differentiate between the two phases.[4,5,6,7] The Menshutkin reaction does not have a transition state in the gas-phase.[8] For this reaction, the potential energy surface between reactants and products is uphill, and can only be found using solvation.[9]

Theoretical modeling of electrochemistry requires the use of solutions as well. Electrochemistry is used to describe the chemical reactions that take place at an electrode within an electrically conducting solution, these reactions dictate the movement of electrons from the electrode to and from an electrolyte. This is of great interest for use in batteries,[10] fuel cells[11] and solar cells.[12,13] Required is a way to calculate the energy of removing an electron in solution. This can be studied through the use of solvation on the metal-electrolyte interface.[14]

The study of electronic dynamics in solution allows for close characterization of the charge transfer between solute and solvent. This can be used to further understand the charge interactions when inserting a charged particle into a solution.[15] When considering solvation and its effects on the solute there needs to be great attention regarding how the charge is distributed, as outlaying charge may become an issue.[16]

Solvation can indeed be a useful tool in quantum chemistry, and comes in many different formulations. Some approaches require a computationally expensive method while others save computational cost at the expense of properties. All of this is done in an on-going effort to continue to explain how the world works.

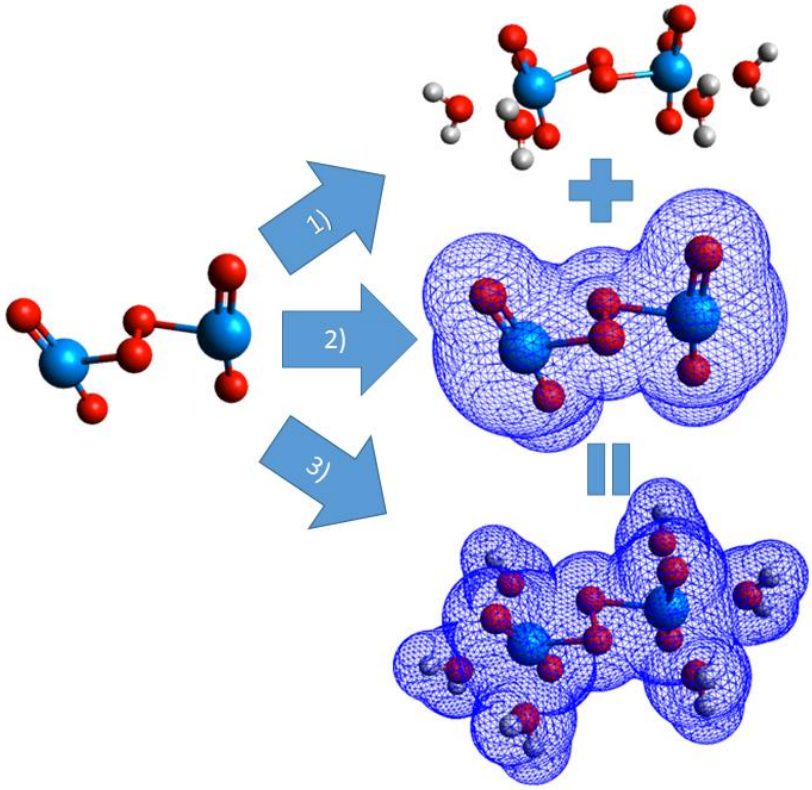
1.1 Solvation

Solvation models can come in several different approaches, all with the overall goal to take quantum mechanical (QM) gas-phase calculations and appropriately model their condensed-phased counterparts. Earlier, Schreckenbach and Shamov proposed that there are three perpendicular axes of approximation to modeling chemistry: level of theory used

for modeling chemistry, relativistic effects, and the choice of modeling complex systems, which may also incorporate solvation methods.[17] There are three basic camps of solvation methods ascribed to in chemistry, those being (Figure 1.1):

1) Explicit Solvation
2) Implicit Solvation
3) Hybrid Solvation

Figure 1.1: *Schematic representation of three common ways to solvate a system. The blue mesh describes the continuum smeared over the molecule.*



Explicit solvation is the most general form of solvation. In order to solvate molecules in this way one needs to physically place solvent molecules around the solute of interest. This allows us to utilize the true structural complex one would expect in solution,

allowing us access to the molecular degrees of freedom associated with placing a molecule into a solvent. Explicit solvation models usually are handled in a few separate ways through molecular mechanics (MM),[18] molecular dynamics (MD), [19,20] Monte Carlo (MC) simulations,[18] and as well as QM calculations of clusters.[18] MM treats the system in a classical mechanical way, where the atoms are treated as balls on springs. One can make use of molecular dynamical methods to model extremely large systems. Such large systems of study are for instance biomolecules where thousands of atoms create much too large an array for quantum mechanical methods. These MD systems then make use of MM methods to calculate the solvent potentials. These calculations can be done over picoseconds of time, and the addition of water will then increase the complexity of the model.

On the other hand, MC simulations are applied over wide ranges of chemical space with random sampling, to get numerical results. MC simulations perform well when large numbers of degrees of freedom are required, such as would be needed for the inclusion of thousands of water or other solvent molecules. Then, pure QM treatment of a system leads to large computational time and costs. As with density functional theory (DFT), detailed in section 1.2.1, calculations on the system scales at N^3, where N is the number of basis functions. Therefore, for even a modest system our computational cost can be quite large. This is but only one of the problems, as there can be huge problems with finding the minimum energy associated with the geometry of the atoms; having more atomic coordinates creates finding the minimum overall coordinates, with respect to energy, a difficult challenge.[2]

Implicit solvation, or otherwise known as continuum solvation, expresses the solvent molecules in a continuous distribution function.[1] This can be thought of as approximating the entirety of the solvent as a smeared function that describes the potential of the solvent, at the cost of structural details. In expressing the solvent in this way we reduce computational effort; this comes at a cost of losing information about the degrees of freedom that the explicit methods will generate. This approach is represented by the blue mesh *(#2)* in *figure 1*.

The third way to handle solvation comprises the so-called hybrid solvation methods. These methods take the best of both of the other models and combine them. There are a few ways to handle this. One such method of solvation is denoted QM/MM, where we treat the solute and near surroundings as purely quantum mechanically then wrap the QM region with an aforementioned explicit method, utilizing MM again. This combines the strength of both methods, with the accuracy of the QM core and the speed of the MM outside. It allows for more manageable solvation of larger species, while not losing the accuracy at the site of interest. There are still several complications within this, such as how does one properly define the region where we cut off the QM and then what are consequences of the transition between the two regions. Another way to combine the two methods would be to include, explicitly, a first and/or second shell of solvent molecules, then shroud the entire molecule in a continuum solvation. This is expressed in *figure 1* as #3.[18]

1.1.1 Implicit Solvation

Continuum solvation models take the solvent molecules and smear them as a charge distribution, the challenge is then how to properly express the electrostatics of inserting a charged particle into a liquid medium. This is not a new problem and has been thoroughly studied through Poisson's equation.[1]

$$-\vec{\nabla}\left[\epsilon(\vec{r})\,\nabla\,V(\vec{r})\right] = 4\pi\rho_M(\vec{r}) \quad (1.1)$$

Solving Poisson's equation will give us information about the electric potential, $V(\vec{r})$, for a given charge distribution. Poisson's equation can be simplified into separate forms for describing a particle, 1.2, within a cavity, and outside the cavity 1.3.

$$-\nabla^2 V(\vec{r}) = 4\pi\rho_M(\vec{r}) \quad (1.2)$$

$$-\epsilon\nabla^2 V(\vec{r}) = 0 \quad (1.3)$$

Equation 1.2 represents the potential, $V(\vec{r})$, as a function of the molecular density, $\rho_M(\vec{r})$, inside the cavity, while equation 1.3 express the potential outside the cavity, with relation to the permittivity of the medium, ϵ. This type of calculation is currently handled by several continuum solvation models such as the Polarizable Continuum Model (PCM) [21], Integral Equation Formalism (IEF) [22], Surface and Volume Polarization for Electrostatic (SVPE and SS(V)PE) [23], the Conductor-like Screening Model (COSMO) of Klamt and Schüürmann [24], as well as others. [1] Collectively, these methods are called the Apparent Surface Charge (ASC) methods.[1] The term ASC is given to these types of models because these methods take the charge and spread it out over the surface

area of the molecule resting in the solvent; making use of tesserae small enough to contain a constant charge over each tesserae element. This essentially equates to doing a separate charge calculation for every grid point on the surface in the molecule at every step of the iterative process, or the self-consistent reaction field, SCRF.[3]

Calculation of the SCRF requires the use of basis sets, and the Hamiltonian and how they are related to the energy of the system. The charge itself is a function of the density of the system. We require an addition to the Hamiltonian in the form of the potential, this is produced by the polarization created with the insertion of the particle, equations 1.1-1.3. Doing so makes the Schrödinger wave equation then doubly dependent on the choice of basis sets, effectively making the Hamiltonian non-linear. This is due to the charge being dependent on the density as well as the Hamiltonian. The basics of how the density may be handled is detailed below, in section 1.2.

1.2 Secular equation and the Slater Determinant

There are only a few cases in which we can get an exact analytical solution of the Schrödinger Wave equation, 1.4,[25] mostly when considering a one-electron system; with such a system there is no need to worry about how the electrons interact with each other. However, through the Variational Principle for the Hamiltonian operator, 1.5, an iterative process can be set up in which the wavefunction is approximated. With the use of an initial guess, continuous approximations may lead to decrease of the energy of the system, if the original guess is good enough. Which is explained by equation 1.5, where E_0 is the exact ground state energy.

$$\widehat{H}\psi = E\psi \text{ (1.4)}$$

$$E = \frac{\int \psi^* H\psi \, \mathrm{dr}}{\int \psi^* \psi \, \mathrm{dr}} \geq E_0 \text{ (1.5)}$$

One-electron molecular orbitals that can be used to approximate the wavefunction, ψ, are built up as a linear combination of atomic orbitals (LCAO), ϕ, described by equation 1.6. We use the variational principle to find the optimal coefficients, a_i, for each basis set, χ_i. In order to achieve that, we want to investigate the minimization of the energy in order to get closer to E_0; this is done through substitution of 1.6 into 1.5, which produces equation 1.7 and simplifies to 1.8. In order to minimize the energy, we take the partial derivatives of the energy with respect to the coefficients and set them to zero, $\partial E/\partial a_k$ for all electrons (N), equation (1.9), to obtain the secular equation, 1.10.[18]

$$\phi = \sum_{i=1}^{N} a_i \chi_i \text{ (1.6)}$$

$$E = \frac{\int \phi^* H\phi \, \mathrm{dr}}{\int \phi^* \phi \mathrm{dr}} \geq E_0 \text{(1.7)}$$

$$E = \frac{\sum_{ij} a_i^* a_j \, H_{ij}}{\sum_{ij} a_i^* a_j S_{ij}} \geq E_0 \text{ (1.8)}$$

$$\sum_{i=1}^{n} a_i (H_{ik} - ES_{ik}) = 0 \text{ (1.9)}$$

$$\begin{vmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} & \cdots & H_{1N} - ES_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{1N} - ES_{1N} & H_{N2} - ES_{N2} & \cdots & H_{NN} - ES_{NN} \end{vmatrix} = 0 \text{ (1.10)}$$

Here we then obtain a polynomial of the N^{th} order with N roots, creating an associated E_j and set of a_{ij} for every root. Where i denotes the basis functions, and j indicates the molecular orbitals that are associated with E_j. Using equation 1.10 we can then find an optimized wavefunction with the basis functions χ_i and the orbital coefficients a_{ij}, 1.11.

$$\phi_j = \sum_{i=1}^{N} a_{ij}\chi_i \quad (1.11)$$

Now, using a similar notation we can move onto the many-electron wavefunctions described in 1.5, ψ. Many-electron wavefunctions are anti-symmetrized representations of the occupied one-electron orbitals, for HartreeFock and DFT (1.2.1). The anti-symmetry is due to electrons being fermions and the wavefunction must be anti-symmetric upon exchange of any two electrons. The interchanging of these two electrons then also consequently satisfies the Pauli Principle, ensuring no two electrons of the same spin occupy the same space, equation 1.12.[26] Slater formulated a determinant which satisfies this exchange principle. It is expressed through orthonormal spin-orbitals, each defining a spin and probability distribution of each electron for the many electron wavefunction, and is known as a Slater determinant, 1.13.[26]

$$\psi(r_1, r_2, r_3, \ldots r_N) = -\psi(r_2, r_1, r_3, \ldots r_N) \quad (1.12)$$

$$\psi(1,2,3\ldots N) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_N(1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \cdots & \phi_N(N) \end{vmatrix} \quad (1.13)$$

1.3 Density Functional Theory

Expanding upon the Schrödinger equation further, the Hamiltonian can be split up into its general contributions, from equation 1.4.

$$\hat{H}\psi = E\psi \quad (1.4)$$

$$\hat{H} = T + V + U_{ee} + U_{Ne} \quad (1.14)$$

Where T is the kinetic energy operator, V is the interaction with an external field, such as solvation, U_{ee} is the mutual electron interaction energy, and U_{Ne} is the nuclear electron interaction energy. In order to solve equation 1.4, through 1.14, we require a new formulation of QM through the use of the density of the system that allows us to construct the Hamiltonian, $\hat{H}$. Thus, DFT is based on two theorems by Hohenberg and Kohn, the first being that the potential $V(r)$ is a functional of the electron density, $\rho(r)$.[27] From the external potential the Hamiltonian is fixed, from which the wavefunction may be extracted. This then leads to second theorem that states that the ground state energy, E_0, can be obtained variationally which is proven via *reduction ad absurdum*.[27]

Using the expressions from section 1.2 in combination with the DFT formulation, we investigate the system with the density, as opposed to the wavefunction, and therefore we search for an optimized density. When the ground state energy is found, through the variational principle, it can be used to properly describe the ground state properties. Using Hohenberg and Kohn's theorems we can now determine an approximated density that describes the full ground state through the Hamiltonian, and the energy can thusly be extracted from equation 1.15, where $T[\rho_0]$, $E_{ee}[\rho_0]$, $E_{Ne}[\rho_0]$ are the kinetic energy, electron-electron repulsion, and nuclear-electron repulsion, respectively. Therefore, if we choose different densities, the ones that provide the lower energies are closer to the ground state density configuration, resembling the wavefunction approach seen in section 1.2.

$$E_0[\rho_0] = T[\rho_0] + E_{ee}[\rho_0] + E_{Ne}[\rho_0] \quad (1.15)$$

This leads us to Kohn-Sham Density Functional Theory.[28] This theory uses a fictitious density system of non-interacting electrons that has the exact same density as the system of interest. The real and virtual (fictitious) systems both contain the same atoms and positions. We can then express the energy as a summation of its components, as in equation 1.16. Where $T_{virtual}$, and $V_{virtual}$ are the kinetic energy and nuclear-electron interactions of the virtual system, respectively.[29] ΔT is the correction to the kinetic energy, V_{ee} is the classical electron-electron repulsion, and ΔV_{ee} is then the non-classical part of the electron-electron repulsion energy.

$$E[\rho(r)] = T_{virtual}[\rho(r)] + V_{virtual}[\rho(r)]$$

$$+V_{ee}[\rho(r)] + \Delta T[\rho(r)] + \Delta V_{ee}[\rho(r)] \text{ (1.16)}$$

Going further with relation to the wave function equations, we can express the energy of the density in orbital form, using ϕ_i as the Kohn-Sham orbitals, Z_k the nuclear charge of atom k, and E_{XC} is the exchange-correlation energy. Using E_{XC}, equation 1.17, to encapsulate the difficult terms $\Delta T[\rho(r)]$ and $\Delta V_{ee}[\rho(r)]$ which provide the difference between the fictitious and actual system,[29] we can then write equation 1.18, with r_i, as the electron location and R_k as the location of the nucleus.

$$E_{XC}[\rho] \equiv (T[\rho] - T_S[\rho]) + (E_{ee}[\rho] - J[\rho]) \text{ (1.17)}$$

$$E[\rho(r)] = \sum_i^N \left(\left\langle \phi_i \middle| -\frac{1}{2}\nabla_i^2 \middle| \phi_i \right\rangle - \left\langle \phi_i \middle| \frac{Z_k}{|r_i - R_k|} \middle| \phi_i \right\rangle \right)$$

$$+ \sum_i^N \left\langle \phi_i \middle| \frac{1}{2} \int \frac{\rho(r_j)}{|r_i - r_j|} dr_j \middle| \phi_i \right\rangle + E_{XC}[\rho(r)] \text{ (1.18)}$$

Using

$$\rho(r) = \sum_{i=1}^N |\phi_i(r)|^2 \text{ (1.19)}$$

Therefore, we have a density, 1.19, that is described by the orbitals that require the density itself for their determination. This type of calculation then has to be done self-consistently, or the SCF, as we use the trial density to obtain a better and better density as described by the secular equation, using $\hat{h}_i^{KS}$, 1.21, as our new KS operator.

$$\hat{h}_i^{KS}\phi_i = e_i\phi_i \text{ (1.20)}$$

$$\hat{h}_i^{KS} = -\frac{1}{2}\nabla_i^2 - \sum_k^{nuclei} \frac{Z_k}{|r_i - R_k|} + \int \frac{\rho(r_j)}{|r_i - r_j|} dr_j + V_{XC} \quad (1.21)$$

Where,

$$V_{XC} \equiv \frac{\delta E_{XC}}{\delta \rho} \quad (1.22)$$

V_{XC}, 1.22, is then a functional derivative of E_{XC}. Therefore, if we know the exact form of both E_{XC} and V_{XC} we would have the exact energy, as we would then know the difference in energies between the fictitious system and the actual system. Since DFT is formulated this way, it is an exact solution to the Schrödinger equation. This is an entry point to the problems that lay within DFT, as we have no way to describe this interaction without approximations.[29]

1.3.1 Approximate Functionals

Since there is no proper way, at least as of yet, to express 1.21 in an exact analytical sense, we require approximations to this expression. We then need to consider what is required for these approximations. E_{XC}, or the exchange correlation energy, recreates the difference between the classical and quantum mechanical electron-electron repulsion, and it also includes the difference in the kinetic energy between the non-interacting system and the real system. This can be handled through a few different ways, such as Local Density Approximation (LDA), a Generalized Gradient Approximation (GGA), meta-GGA, and Hybrid functionals. Each of these approximations takes a more complex approach than the last in trying to accommodate the short fallings of DFT.

The exchange correlation energy is a functional, E_{XC}, with a dependence on $\rho(r)$ expressed as an interaction between $\rho(r)$ and an energy density, ε_{XC}, that is also dependent on the electron density $\rho(r)$, 1.23.[18] LDA approaches solving for ε_{XC} by using ρ at an exact position, r, to compute ε_{XC}. In other words, the local value of ρ can be used to calculate ε_{XC}. LDA functionals are derived from a uniform electron gas approach, where the approximation to the density is uniform throughout. The correlation energy is the portion of E_{XC} that cannot be accounted for analytically, and therefore we require approximations to account for this missing term. Total energies of interacting uniform electron gases at various densities were calculated using quantum Monte Carlo simulations.[30] Using these full energies we can calculate the difference between the analytical exchange energy and the correlation energy. Using the results of Ceperley et al.[30] these results were fitted by Vosko, Wilk, and Nusair to estimate the local density and thusly get an approximation to the correlation energy.[31] This approximation is then called VWN, after the creators. VWN is used in chapter 4 of this thesis.

$$E_{XC}[\rho(r)] = \int \rho(r)\varepsilon_{XC}[\rho(r)]dr \quad (1.23)$$

Moving on to a less local theory the GGA theory attempts to address the shortcomings of using a uniform electron gas to approximate the changing density of each atom in a molecule; a uniform electron gas is not sufficient enough to calculate molecular density, as the density is not uniform for molecules. GGA takes the functional

expression for the density and does a "sort-of" Taylor expansion. This expansion accounts for the changing density in a molecule; most GGAs use this expansion with the LDA. The general form of the GGA can be seen in equation1.24.[18] These functionals are both dependent on the density and the gradient of the density, and this keeps track of how the density changes locally. This is used, for instance, in the Perdew, Burke, and Ernzerhof (PBE) functional, which is used in chapters 3 and 4 of this thesis.[32] The PBE functional is based on physical restrictions and not reliant on empirical parameters.

$$\varepsilon_{XC}{}^{GGA}[\rho(r)] = \varepsilon_{XC}{}^{LDA}[\rho(r)] + \Delta\varepsilon_{XC}{}^{GGA}\left[\frac{|\nabla\rho(r)|}{\rho^{\frac{4}{3}}(r)}\right] \quad (1.24)$$

The next logical step in improving functionals would be the second term of a "sort-of" Taylor expansion of the density, where we get the Laplacian of the density. These approaches are classified as the meta-GGA functionals. In modern meta-GGA functionals the standard practice is to use the kinetic-energy density, τ, which is not the Laplacian, but still includes second derivative character, and is instead based on the orbitals. Seen in 1.25, where $\psi(r)$ are the Kohn-Sham orbitals.[18,33] Within chapter 3, utilized is the M06-L meta-GGA functional.[33]

$$\tau(r) = \sum_{i}^{occupied} \frac{1}{2}|\nabla\psi_i(r)|^2 \quad (1.25)$$

The last type of functionals utilized in this thesis are the hybrid funcitonals, which exploit a percentage of Hartree-Fock (HF) exchange in order to correct the short-comings of the virtual Kohn-Sham exchange. This is done through the so-called adiabatic connection, where we want to find a smooth scaling between the fictitious density system

to the fully interacting system. This is seen as an integral expressed in equation 1.26, with $\lambda = 1$ being the full interacting system, and $\lambda = 0$ being the fictitious system.[18] Lambda is described as the interaction strength parameter.

$$E_{XC} = \int_0^1 \langle \psi(\lambda) | V_{XC} | \psi(\lambda) \rangle \, d\lambda \quad (1.26)$$

We want to express a certain percentage of the DFT exchange as a function of the interacting HF exchange, equation 1.27, in which a is an empirically determined parameter between 0 and 1. This is used in combination with GGAs to improve results. Within this thesis the B3LYP functional is utilized, this functional takes a combination of the B3 hybrid exchange as proposed by Becke, combined with the LYP correlation, equation 1.28, with a, b, and c equal to 0.20, 0.72, and 0.81, respectively.[34,35,36] The hybrid concept can be expanded upon further to make use of the second derivative based meta-GGA, called hybrid meta-GGAs, which also is used in chapter 3 in the form of M06-2X.[33]

$$E_{XC} = (1 - a)E_{XC}^{DFT} + aE_X^{HF} \quad (1.27)$$

$$E_{XC}^{B3LYP} = (1 - a)E_X^{LSDA} + aE_X^{HF} + b\Delta E_X^B + (1 - c)E_C^{LSDA} + E_C^{LYP} \quad (1.28)$$

Each of these functionals has their strengths and weaknesses. As each one becomes more complicated they require more computational time, and cost more computationally.[37] This, taking even more computation for the hybrid methods as one has to work with both the density and the HF exchange. All of these methods approach

finding the lowest energy state by making use of the SCF in order to get a self-consistent density. They also utilize the atomic orbitals in the form of basis sets, in order to get the lowest possible energy in the form of orbitals for the system.

1.3.2 Slater Type Orbitals and Gaussian Type Orbitals

Basis sets are used in order to express the atomic orbitals. Slater in 1930, devised a set of rules that valence orbitals have to follow in order to correctly represent the atomic orbitals.[38] This was done to find an analytical representation of the aforementioned one-electron hydrogenic orbitals, χ_i, that were taken from the wavefunction spectrum of hydrogen atoms. This expression of the orbitals takes the form of Slater Type Orbials, or STOs, and is mathematically formulated in 1.29.[18]

$$\chi(r,\theta,\phi;\zeta,n,l,m) = \frac{2\zeta^{n+1/2}}{[(2n)!]^{1/2}} r^{n-1} e^{-\zeta r} Y_l^m(\theta,\phi) \quad (1.29)$$

Where $Y_l^m(\theta,\phi)$ are the spherical harmonics, r,θ,ϕ are the polar coordinates, n is the principal quantum number, and l,m are the angular quantum numbers. STOs are useful for a number of reasons as they have the correct cusp at the nucleus, and the correct orbital decay at longer distances r. There is a downside associated with using STO, and that is they have no analytical solution to the Coulomb integral, and therefore advanced quadrature and density fitting must be used in order to allow the use of STOs. The Amsterdam Density Functional (ADF) program, which this thesis details the solvation implementation into, makes exclusive use of STOs. Mirko Franchini et al.'s implementation of a quadrature and fitting functions within the newest iteration of ADF

(2016) and ADF-BAND uses the Becke grid[39] in order to produce advanced, fast and smooth quadrature, as well as quick so-called zlm fitting procedure for density fitting, which uses spline functions.[40,39,41]

In order to get around the approximations required to employ STOs an old mathematical trick that squares the e^{-r} term to make it integrable was introduced by Boys in 1950, thus converting an STO into a Gaussian type function, or a GTO.[42] Which takes the form of 1.30, as presented by Cramer, in atom-centered Cartesian Coordinates.[18]

$$\chi(x,y,z;\alpha,i,j,k) = \left(\frac{2\alpha}{\pi}\right)^{3/4}\left[\frac{(2\alpha)^{i+j+k} i!\, j!\, k!}{(2i)!\,(2j)!\,(2k)!}\right]^{1/2} x^j y^i z^k e^{-\alpha\left(x^2+y^2+z^2\right)} \quad (1.30)$$

The Coulomb integral no longer requires numerical quadrature, when using GTOs. Though this also comes with a trade-off, as the GTOs do not abide by the rules originally devised by Slater, that is they don't have the proper cusp at the atomic nucleus, and the long-range r function drops off too quickly. Hehre, Stewart, and Pople in 1969 devised a way to properly incorporate several Gaussian type functions in tandem in order to reproduce the cusp and long-range failures. This formulation takes summation of several GTOs to properly fit them to the actual STO shape and consequently correct for the failures, while still maintaining its analytical nature.[43] Chapter 3 of this thesis goes into detail about the solvation effects as consequences of using STOs and GTOs in a comparison of the two types of approximations.

1.4 Solid-State Considerations

Modeling chemical systems can be further extended into the solid-state realm. Surface chemistry is a bourgeoning field, with great interest in controlling greenhouse gases, energy conversion, and harvesting solar energy. These types of systems require modeling of solid-state compounds, such as metal organic frameworks (MOF), surface catalysis, or solar cells. Moving to these larger systems, traditional molecular QM codes run into limitations since it is hard to model a semi-infinite system without running into extremely large basis set problems, and eventually it is not computationally cost effective to physically coordinate these systems. This can be more easily handled with the use of periodic boundary condition (PBC) codes, such as ADF-BAND.[44]

The aforementioned modeling problems may benefit from solvation (section 1.1 above), E.g., MOFs are more efficient at CO uptake within solution,[45] and molecular sieves show improvement on methane uptake through solvation as well.[46] Dye sensitized solar cells (DSSC) require to be solvated with a charge carrying electrolyte, and metal catalysis takes place exclusively within solution. The same problems mentioned in section 1.1 arise when we attempt to use explicit solvation within the confines of PBC, but the problems are only exacerbated as the surrounding unit cells are coordinated with the same amount of solvent molecules. This suggests the simple solution of using implicit continuum solvation models to model the solution on top of periodic systems.

1.4.1 Band theory

Next we require an understanding of how electrons will act within a large periodic system; this is the subject of the Band theory of solids. It describes the quantum states which electrons will occupy when within solids. Starting with molecular orbital (MO) theory, we know that if we take two atoms, for instance one with an electron in a 1s orbital and one with an interacting 1s orbital, as we bring these two orbitals closer together their electron distributions will begin to overlap. This interaction forms a chemical bond. The overlapping electron distribution then creates the molecular orbital, this also creates more space within the orbital for the electrons to move around and causes a lower energy. The creation of this molecular orbital causes the energies of the two 1s electrons to split, and we end up with a higher energy orbital as well as a lower energy one, only the overlap of the lower energy orbitals will then create the chemical bond.

Now, if we keep adding atoms to this system, we keep creating favorable and unfavorable configurations for the electrons to populate but at different energy levels from each other. When we bring an infinite amount of 1s orbitals together into a configuration, such as what we would expect when evoking PBCs, this creates two very distinct continua of energies. Between the lower and higher energy interactions we have a gap where there are no electronic states, excluding the case for conductors which is a full continuum, and no gap. This turns our system into a spectrum of two distinct ranges of energy called electronic bands, or bands. These bands are the MO configuration of an infinite amount of, in our example, 1s orbitals interacting, and the only allowed energy configurations are

either in the lower energy valence band or the higher energy conduction band. The energy associated with moving from the lower energy valence band to conduction band is then called the band gap energy, or the band gap.

Electrons within solid metals are treated as if they are within a finite potential well, they are trapped within the solids, we can then think of the metal as a box. The highest possible band region, or highest occupied energy quantum state at 0K, is then called the Fermi-energy level, or Fermi-level. The Fermi-level is defined as the energy required to add one electron to the system.

1.4.2 ADF-BAND Basis Sets

Moving to the periodic case we have to adapt our description of the basis sets to better describe our system, as single atomic basis sets that are repeated off into infinity are not useful for any practical purposes. This is done through Bloch's theorem, which writes the wavefunction in a periodic fashion.[47] Multiplying the wavefunction by a periodic function then produces a Bloch function. Using the explanation of the infinite series of potentials described in section 1.4.1, Bloch's theorem can describe how each potential well is described periodically with respect to the wavefunction. Bloch's theorem can be stated with two postulates:

1) A solid is periodic on the atomic scale, therefore a traveling wave solution affected by the symmetry of the lattice is required, 1.31.

2) The wavefunction at some point in the lattice $\psi(r)$, will have the same form, with an added phase factor, as another point in the lattice at position R, described by $\psi(r+R)$, 1.33.

$$\psi(r) = u_k(r)e^{ikr} \ (1.31)$$

Where,

$$u_k(r) = \ u_k(r+R) \ (1.32)$$

$$\psi(r+R) = C\psi(r) \ (1.33)$$

Using 1.33 we can then write 1.31 as:

$$\psi(r+R) = e^{ik(r+R)}u_k(r+R) \ (1.34)$$

Rearranging 1.31,

$$u_k(r) = \frac{\psi(r)}{e^{ikr}} \ (1.35)$$

1.34, through 1.32 then becomes

$$\psi(r+R) = e^{ik(r+R)}\frac{\psi(r)}{e^{ikr}} \ (1.36)$$

And we can write our periodic system as a function of the primitive cell, $\psi(r)$, 1.37.

$$\psi(r+R) = e^{ik(R)}\psi(r) \ (1.37)$$

Re-writing 1.37 to accommodate the orbitals established earlier, we then get equation 1.38.

$$\phi(k) = \sum_{R}^{periodic\ images} e^{ik(R)}\chi(r - R) \quad (1.38)$$

Here R is the sum over periodic images surrounding the primitive cell, and χ is the STO. The STOs utilized in BAND take a slightly different form. We use a numerical type orbital (NAO) to perfectly describe a spherical atom; these are augmented with STOs to appropriately map to real atomic functions, and utilized Zlm fitting mentioned previously, taking the form of 1.39.[44] Where χ_{nlm} is the Slater atomic orbital with the principal quantum numbers $n, l, m,$ and $Z_{lm}(\Omega)$ is the angular portion, calculated through splines.[41]

$$\chi_{nlm} = R_{nl}(r)Z_{lm}(\Omega) \quad (1.39)$$

1.5 Organization of this Thesis

This thesis is written in the sandwich style format. It comprises a published manuscript[48] in a peer-reviewed scientific journal (Chapter 3), with a second manuscript soon to be submitted (Chapter 4). The overall goal of this thesis is to explain in full the implementation of the SM12 solvation model into ADF and its periodic boundary condition sister code, ADF-BAND.

A quick overview of solvation, DFT, STOs, GTOs, Band theory, and periodic boundary conditions has been presented in Chapter 1. Chapter 2 details the theory behind the SM12 solvation model, which is central to this thesis. An implementation of solvation model 12 (SM12), into the Amsterdam Density Functional (ADF) Code, and a comparison with an ASC method, COSMO, is found within Chapter 3. This implementation is used to

explore the comparison of STOs and GTOs and the universality of SM12 and the underlying charge model CM5 , as well as some benchmarking on solvation within ADF. Chapter 4 is then a joint paper with Dr. Pier Philipsen, based out of SCM in The Netherlands, on the implementation of both SM12 and COSMO into the PBC code ADF-BAND.

Furthermore, Chapter 4 contains a proof of concept, explored are the applicability of solvation to PBCs and considerations within, and to explore novel surface chemistry. In Chapter 5, the overarching goal of modeling solvents is summarized, and how each chapter then plays an important role on understanding chemistry in condensed phases. Finally, future directions for the implementations explained and ideas for future surface chemistry in solutions are detailed.

1.6 References

(1) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3093.

(2) Orozco, M.; Luque, F. J. Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems. (Vol 100, Pg 4187, 2000). *Chem. Rev.* **2001**, *101*, 4187–4225.

(3) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.

(4) Brown, R. D.; Godfrey, P. D.; Storey, J. W. V.; Bassez, M. P. Microwave Spectrum and Conformation of Glycine. *J. Chem. Soc., Chem. Commun.* **1978**, *13* (13), 547–548.

(5) Suenram, R. D.; Lovas, F. J. Millimeter Wave Spectrum of Glycine. *J. Mol. Spectrosc.* **1978**, *72* (3), 372–382.

(6) Godfrey, P. D.; Brown, R. D. Shape of Glycine. *J. Am. Chem. Soc.* **1995**, *117* (7), 2019–2023.

(7) Iijima, K.; Tanaka, K.; Onuma, S. Main Conformer of Gaseous Glycine: Molecular Structure and Rotational Barrier from Electron Diffraction Data and Rotational Constants. *J. Mol. Struct.* **1991**, *246* (3-4), 257–266.

(8) Turbert-Brohman, I.; Sherman, W.; Repasky, M.; Beuming, T. Improved Docking of Polypeptides with Glide *J. Chem. Info. and Modeling*; **2013**; Vol. 53, pp 1689-1699.

(9) Sola, M.; Lledos, A.; Duran, M.; Bertran, J.; Abboud, J. L. M. Analysis of Solvent Effects on the Menshutkin Reaction. *J Am Chem Soc* **1991**, *113* (8), 2873–2879.

(10) Naejus, R.; Willmann, P. Battery Electrolyte Solutions. *Science (80-.).* **1998**, *43*, 275–284.

(11) Armand, M.; Endres, F.; MacFarlane, D. R.; Ohno, H.; Scrosati, B. Ionic-Liquid Materials for the Electrochemical Challenges of the Future. *Nat. Mater.* **2009**, *8* (8), 621–629.

(12) O'Regan, B.; Grätzel, M. A Low-Cost, High-Efficiency Solar Cell Based on Dye-Sensitized Colloidal TiO2 Films. *Nature* **1991**, *353* (6346), 737–740.

(13) Hagfeldt, A.; Boschloo, G.; Sun, L.; Kloo, L.; Pettersson, H. Dye-Sensitized Solar Cells. *Chem. Rev.* **2010**, *110* (11), 6595–6663.

(14) Jaque, P.; Marenich, A. V; Cramer, C. J.; Truhlar, D. G. Computational Electrochemistry : The Aqueous Ru 3 + | Ru 2 + Reduction Potential. *J. Phys. Chem. C* **2007**, *111* (15), 5783–5799.

(15) Aleksandr V. Marenich; Ryan M. Olson; Adam C. Chamberlin; Christopher J. Cramer, and; Truhlar, D. G. Polarization Effects in Aqueous and Nonaqueous Solutions. **2007**.

(16) Klamt, A.; Jonas, V. Treatment of the Outlying Charge in Continuum Solvation Models. *J. Chem. Phys.* **1996**, *105* (22), 9972.

(17) Schreckenbach, G.; Shamov, G. A. Theoretical Actinide Molecular Science. *Acc. Chem. Res.* **2010**, *43*, 19–29.

(18) Cramer, C. J. *Essentials of Computational Chemistry Theories and Models*; 2004; Vol. 2.

(19) Gu, W.; Schoenborn, B. P. Molecular Dynamics Simulation of Hydration in Myoglobin. *Proteins* **1995**, *22* (1), 20–26.

(20) Perkins, J.; Edwards, E.; Kleiv, R.; Weinberg, N. Molecular Dynamics Study of Reaction Kinetics in Viscous Media. *Mol. Phys.* **2011**, *109* (15), 1901–1909.

(21) Miertus, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of Ab Initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* **1981**, *55*, 117–129.

(22) Cancès, E.; Mennucci, B.; Tomasi, J. A New Integral Equation Formalism for the

Polarizable Continuum Model: Theoretical Background and Applications to Isotropic and Anisotropic Dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032.

(23) Chipman, D. M. Charge Penetration in Dielectric Models of Solvation. *J. Chem. Phys.* **1997**, *106*, 10194–10206.

(24) Klamt, A.; Schüürmann, G. 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* **1993**, *5*, 799–805.

(25) Schrödinger, E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys. Rev.* **1926**, *28* (6), 1049–1070.

(26) Slater, J. C. A Simplification of the Hartree-Fock Method. *Phys. Rev.* **1951**, *81* (3), 385–390.

(27) Hohenberg, P.; Kohn, W. Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136* (3B), B864–B871.

(28) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*. 1965.

(29) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; 2001; Vol. 3.

(30) Ceperley, D. M.; Alder, B. J. Ground State of the Electron Gas by a Stochastic Method. *Phys. Rev. Lett.* **1980**, *45* (7).

(31) Vosko, s. h.; Wilk A N, L.; Nusair, D. M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical analysis1. *J. Phys* **1980**, *58*.

(32) Perdew, J.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(33) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(34) Becke, A. D. A New Mixing of Hartree–Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98* (2), 1372.

(35) Becke, A. Density Functional Thermochemistry III The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(36) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(37) Perdew, J. P.; Schmidt, K. Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy. In *AIP Conference Proceedings*; AIP, 2001; Vol. 577, pp 1–20.

(38) Slater, J. C. Atomic Shielding Constants. *Phys. Rev.* **1930**, *36* (1), 57–64.

(39) Franchini, M.; Philipsen, P. H. T.; Visscher, L. The Becke Fuzzy Cells Integration Scheme in the Amsterdam Density Functional Program Suite. *J. Comput. Chem.* **2013**, *34* (21), 1819–1827.

(40) Becke, A. D. A Multicenter Numerical Integration Scheme for Polyatomic

Molecules. *J. Chem. Phys.* **1988**, *88* (1988), 2547.

(41) Franchini, M.; Philipsen, P. H. T.; Van Lenthe, E.; Visscher, L. Accurate Coulomb Potentials for Periodic and Molecular Systems through Density Fitting. *J. Chem. Theory Comput.* **2014**, *10* (5), 1994–2004.

(42) Boys, S. F. Electronic Wave Functions I. A General Method of Calculation for the Stationary States of Any Molecular System. *Proc. R. Soc. London* **1950**, *A200* (1063), 542–554.

(43) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *J. Chem. Phys.* **1969**, *51* (6), 2657.

(44) Te Velde, G.; Baerends, E. J. Precise Density-Functional Method for Periodic Structures. *Phys. Rev. B* **1991**, *44*, 7888–7903.

(45) Kornienko, N.; Zhao, Y.; Kley, C. S.; Zhu, C.; Kim, D.; Lin, S.; Chang, C. J.; Yaghi, O. M.; Yang, P. Metal–Organic Frameworks for Electrocatalytic Reduction of Carbon Dioxide. *J. Am. Chem. Soc.* **2015**, *137* (44), 14129–14135.

(46) Kim, J.; Maiti, A.; Lin, L.-C.; Stolaroff, J. K.; Smit, B.; Aines, R. D. New Materials for Methane Capture from Dilute and Medium-Concentration Sources. *Nat. Commun.* **2013**, *4*, 1694.

(47) Bloch, F. Über Die Quantenmechanik Der Elektronen in Kristallgittern. *Z. Phys.* **1929**, *52* (7-8), 555–600.

(48) Peeples, C. A.; Schreckenbach, G. Implementation of the SM12 Solvation Model into ADF and Comparison with COSMO. *J. Chem. Theory Comput.* **2016**,

acs.jctc.6b00410.

Chapter 2: SM12 Theory

SM12 is an implicit solvation model formulated by Marenich, Cramer, and Truhlar.[1] The SMx series of models are created with the purpose of trying to encapsulate both the electrostatics of solvation, along with approximating information lost by not explicitly incorporating the solvent structure. SM12 is the newest iteration of the Generalized Born Approximation solvation model in this series, which takes advantage of Hirshfeld Population Analysis[2] in order to be basis set independent.[3] SM12, before the work presented in Chapter 3 had only been used in conjunction with GTO basis sets. This chapter details the general theory behind SM12, and chapter 3 goes into details about the implementation of SM12 into the STO code, ADF.[4,5]

2.1 Implicit Solvation Model 12, Polarization

One can reduce the computational effort by approximating the Poisson equation in a more general way, see section 1.1.1 equation 1.1. This may be done through the Generalized Born Approximation (GBA), wherein the GBA essentially describes the charge associated with an atom as point charges placed at the centres of the nuclei of each atom. The GBA is a generalization of the Born Equation,[6] which describes a sphere with charge q, and has a radius α. The charge distribution, $p(s)$, over that surface can be described by Eq. 2.1. Where q is the charge, and α is the surface area.

$$p(s) = \frac{q}{4\pi\alpha^2} \quad (2.1)$$

Much like equation 1.3 (section 1.1.1), we can describe the potential, $V(r)$, outside the sphere as in 2.2. Where ϵ is the permittivity of the medium, $\|r\|$ is the distance vector from the centre to the edge of the sphere.

$$V(r) = -\frac{q}{\epsilon\|r\|} \quad (2.2)$$

Then to describe the work required for charging the sphere we integrate over the surface

$$w = \frac{1}{2}\int_S p(s)V(s)ds \quad (2.3)$$

$$= -\frac{q^2}{2\epsilon\alpha} \quad (2.4)$$

We then have to consider what happens when we insert the charged sphere into the condensed phase; this can be said to be the energy of polarization of both the solution and the solute. This can be described by the Born equation, 2.5. From this formulation it becomes apparent that, when the permittivity is 1, the solute is in gas-phase and the energy of polarization is 0 kcal/mol, and when the permittivity is different from 1 the charged sphere is doing work on the solution, or charging it.

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\frac{q^2}{\alpha} \quad (2.5)$$

The Born equation can be generalized to approximate more than just a single sphere, which is described by the GBA.[1] The GBA sees pairs of charged species interacting, and the associated Coulomb integrals between them, 2.6.

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right) \sum_{k,k'}^{atoms} q_k \gamma_{kk'} q_{k;} \quad (2.6)$$

Equation 2.6, takes advantage of Still et al.'s[7] approximation of the Coulomb integral, 2.7, in order to have appropriate boundary conditions. As the two atoms come closer together, equation 2.7, and subsequently 2.6, reduces slowly to the Born equation. If there is only one atom we have the exact Born equation. Likewise, it works for boundary conditions at far distances: As the two atoms move further apart equation 2.7 reduces to Coulombs law.

$$\gamma_{kk'} = \left(r_{kk'}^2 + \alpha_k \alpha_{k'} e^{\left(-r_{kk'}^2 / d\alpha_k \alpha_{k'}\right)} \right)^{-1/2} \quad (2.7)$$

In order to use the GBA appropriately we require a way to describe the charges in an atom accurately.[1] This can be done through partial charges. The concept of partial charges is not well-defined quantum mechanically but is a useful approach experimentally. Abnormally strong effects, such as hydrogen bonding, are described using the partial positive charge on hydrogen interacting with the partial negative charge on oxygen. It is difficult to describe accurately the partial charge using mathematical models, but it is something that comes intuitively to chemists.

Partial charges can be identified as belonging to 4 different general classes. Class I being empirical, or based on classical mechanics, much like what would be seen in force

fields, FFs. One formulation of this is the partial equalization of orbital electronegativity of Gasteiger and Marsilli.[8,9] Class II charges use partitioning of the wavefunction into atomic contributions. They are orbital based, such as for instance Mulliken population analysis.[10] A class III charge model makes use of a physical component computed from the wavefunction such as the density, for instance Löwdin Population Analysis,[11,12] or Hirshfeld Population Analysis (applied in this thesis).[2] Finally, class IV charge models are a mixture of class II or III with an empirical correction term in order to more appropriately map both the partial charges and the dipole of the system. An example is Charge Model 5 (CM5).[13]

For the purposes of the implicit solvation model, SM12, we require the use of empirical corrections along with a quantum mechanical model to portray a meaningful charge.[9] This is done through the use of a class IV charge model, namely Charge Model 5.[13] CM5, equation 2.8, has been shown to be very accurate in calculating dipole moments, without a heavy reliance on the choice of basis sets.[13] This is realized through the combination of Hirshfeld population analysis (HPA)[2] , q_k^{HPA}, and a single set of dipole-based parameters to accurately describe the quantum mechanical and empirical portions, respectively.[13]

$$q_k^{CM5} = q_k^{HPA} + \sum_{k\prime \neq k} T_{kk\prime} B_{kk\prime} \quad (2.8)$$

The empirical parameters take the form of $T_{kk'}$ and B_{kk}. $B_{kk'}$ is based on a bond order formulation that resembles a construct described by Pauling,[14,13] and can be seen in equation 2.9. While $T_{kk'}$ is purely parameter based, and seen in equation 2.10.

$$B_{kk'} = exp\left(-\alpha\left(r_{kk'} - R_{Z_k} - R_{Z_{k'}}\right)\right) \quad (2.9)$$

$$T_{kk'} = \begin{cases} D_{Z_k Z_{k'}} & Z_k \; and \; Z_{k'} = 1,6,7,8 \\ D_{Z_k} - D_{Z_{k'}} & All \; other \; cases \end{cases} \quad (2.10)$$

Where Z_k are the atomic numbers and k runs over all atoms, R_{Z_k} is the atomic covalent radius, $r_{kk'}$ is the interatomic distance, and α is parameterized to be 2.474 (Å^{-1}).[13]

$$D_{Z_k Z_{k'}} = -D_{Z_{k'} Z_k}$$

$$D_{Z_k Z_{k'}} = 0 \quad iff \quad Z_k = Z_{k'}$$

$$D_{Z_k} \equiv empirically \; determined \; parameter$$

These parameters are based on the parameters shown in *Table 1* in the work of Marenich et al.[13] They are parametrized for the entire periodic table, with special atom-atom interaction cases described by atomic numbers $Z_k and \; Z_{k'} = 1,6,7,8$ (Hydrogen, Carbon, Nitrogen, Oxygen).

In order to properly describe the atomic spheres that are required for the application of the GBA, we need to define the Born radius, α_k, in equation 2.11.

$$\alpha_k = \left(\frac{1}{R'} + \int_{\rho_{Z_k}}^{R'} \frac{A_k(r)}{4\pi r^4} dr \right)^{-1} \quad (2.11)$$

Here R' is the radius of a sphere at nuclear position k that is large enough to contain the entire molecule. ρ_{Z_k} is the Coulombic radius of atom k, and $A_k(r)$ is the solvent accessible surface area (SASA) of a sphere of radius r. $A_k(r)$ is calculated through the analytical surface area (ASA) algorithm of Liotard et al.[15] Here, we evaluate the integral numerically through Legendre-Gaussian Quadrature. Gaussian quadrature simplifies the integral by splitting it up into weighted sum of a function between a specified interval. Gaussian quadrature is expressed in equation 2.12. Legendre-Gaussian requires that the interval is between [-1, 1], therefore we require a change of variable, 2.13, in order to express our generic bounds.

$$\int_{-1}^{1} f(x)dx = \int_{-1}^{1} w(x)g(x) = \sum_{i=1}^{n} w_i g(x_i) \quad (2.12)$$

Where w_i is the weight associated with root x_i, and $g(x_i)$ is our function.

$$\int_a^b f(x)dx = \frac{b-a}{2} \int_{-1}^{1} f\left(\frac{b-a}{2} x + \frac{b+a}{2} \right) dx$$

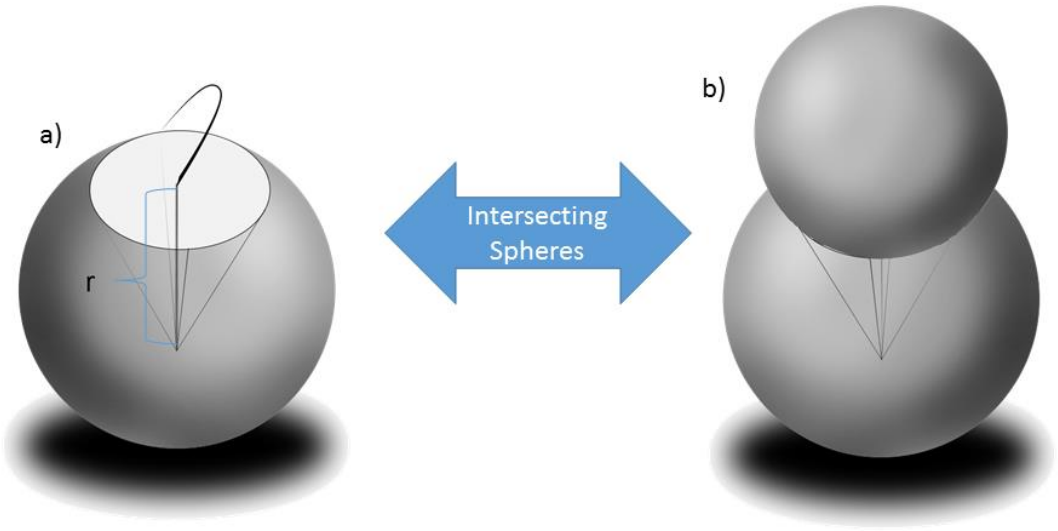
$$\int_a^b f(x)dx = \frac{b-a}{2} \sum_{i=1}^{n} w_i f\left(\frac{b-a}{2} x + \frac{b+a}{2} \right) \quad (2.13)$$

Then using equation 2.11 with the Gaussian-weighted quadrature from 2.12, and the change of variable supplied by 2.13, we can express our numerical integral as 2.14.

$$\int_{\rho_{Z_k}}^{R'} \frac{A_k(r)}{4\pi r^4}\,dr = \frac{R' - \rho_{Z_k}}{2}\sum_{i=1}^{n} w_i \frac{A_k\left(\frac{R' - \rho_{Z_k}}{2}r + \frac{\rho_{Z_k} + R'}{2}\right)}{4\pi\left(\frac{R' - \rho_{Z_k}}{2}r + \frac{\rho_{Z_k} + R'}{2}\right)^4} \quad (2.14)$$

This is counted for 16 iterations defined by the weights and roots given by Aramowitz, and Stegun.[16] The number of points was taken from the code originally supplied by Cramer et al.;[17] they used numbers based on the size of the system though here it is decided to hold it consistently at 16 for accuracy. The ASA function is based on the intersection of neighboring spheres, as defined by distance vectors between the nuclear coordinates of each sphere, *figure 2.1*. This algorithm is designed to indicate when a portion of a neighbouring sphere is located within the primitive sphere and cut off the associated surface area of the encapsulated spherical neighbour.

Figure 2.1: *Schematic illustration of the intersecting atomic spheres underlying the ASA. a) Cross section of where the two spheres meet; the radius from the centre to this intersection is defined by r. b) Full intersection.*



This algorithm is further designed to be analytical in order to reduce computational effort on the self-consistent field (SCF) cycle. The reduction is achieved through the analytical nature of the formulation along with the smoothing effect on numerical noise produced from using quadrature. The GBA is required for every SCF cycle and its contribution makes the SCF nonlinear, as it requires the density. The inclusion of the density, in the form of induced polarization, affects the Kohn-Sham matrix and is then the self-consistent reaction field (SCRF), which is described by equation 2.15.[18]

$$KS_{\mu,\nu}^{Mat} = KS_{\mu,\nu}^{Mat^\circ} + \frac{\partial G_P}{\partial \rho_{\mu,\nu}} \quad (2.15)$$

Where μ, ν are the basis set indices, $KS_{\mu,\nu}^{Mat^\circ}$ is an element of the gas-phase Kohn-Sham matrix, and $\rho_{\mu,\nu}$ is an element of the density matrix as defined by density functional theory (DFT), described in section 1.2.1.

2.1.1 SM12 Solvent Parameters

SM12[1] handles the energy associated with the solvent differently than other implicit solvation models. The ASC methods mostly only incorporate a term associated with the specific solvent through the inclusion of the permittivity, ε.[19] SM12 attempts to give a general description of empirically determined parameters that can be directly associated with the solvent one wishes to model. This is done to re-create some of the lost structural information that comes from the assumption that the liquid is a continuum. The solvent approximation is then handled by the so-called G_{CDS} term, equation 2.16. It incorporates (i) the cost of cavitation which comes from physically creating space for the insertion of a

molecule, (ii) a long-range interaction between atoms that describes the interaction between the solvent and the closely situated atoms within the solute; this resembles a sort of long-distance dispersion; as well as (iii) a description of experimentally measureable physical properties of each solvent. Thus, CDS stands for Cavitation, Dispersion, and Solvent. These parameters take the form of two general terms:

1. Atomic Surface Tension (AST), σ_k
2. Macroscopic Surface Tension (MST), $\sigma^{[M]}$

The AST and MST in combination with the SASA, A_k, provided by the ASA are seen as the sum in 2.16. The CDS energy only depends on the geometry of the molecule provided by the nuclear positions R, making this a post-SCF contribution to the solvation energy; only changing through calculations in which the geometry changes. The radii of individual atoms, R_{Z_k}, are based on Bondi's values,[20] and van der Waals radii from the CRC Handbook of Chemistry and Physics, 91st edition[21] for those not present in Bondi radii.[21] Added to each radius is a constant value of 0.4 Å to accommodate the space that would separate solute and solvent.

$$G_{CDS} = \sum_{k}^{atoms} \sigma_k A_k\left(R, \{R_{Z_k} + r_s\}\right) + \sigma^{[M]} \sum_{k}^{atoms} A_k\left(R, \{R_{Z_k} + r_s\}\right) \quad (2.16)$$

The MST, 2.17, parameterizes how the surface tension interaction of the solvent is affected by the entire solute. This leads to general terms describing the solvent and is only associated to the solute through the SASA. Therefore, it is independent of the atoms within the solute. The solvent descriptor used is γ, which is the macroscopic surface tension of

the solvent at the air/solvent interface, at room temperature. $\phi^2$is the carbon aromaticity of the solvent, which is the square of the fraction of aromatic carbon atoms that are not bonded to hydrogen atoms.[22] ψ^2 is the electronegative halogenicity of the solvent, defined as square fraction of halogen atoms that are not bonded to hydrogen atoms, excluding the halogens I, At, and element 117. Finally, β^2 is based on Abraham's basicity parameter of the solvent.[23,24] The remaining parameters $\tilde{\sigma}^{[\gamma]}$, $\tilde{\sigma}^{[\phi^2]}$, $\tilde{\sigma}^{[\phi^2]}$, and $\tilde{\sigma}^{[\beta^2]}$ are dependent on the choice of SMx model and CMx model used to study the solvation. For SM12 using CM5 the parameters are 0.17, -2.30, -7.02 and 6.69, respectively, and are used in equation 2.17.[1] The $\tilde{\sigma}^{[x]}$ parameters can be thought of as purely general parameters that need to be empirically obtained if one wishes to use a different flavor of charge model in combination with SM12.

$$\sigma^{[M]} = \tilde{\sigma}^{[\gamma]}(\gamma) + \tilde{\sigma}^{[\phi^2]}\phi^2 + \tilde{\sigma}^{[\psi^2]}\psi^2 + \tilde{\sigma}^{[\beta^2]}\beta^2 \quad (2.17)$$

AST then describes the attraction between each neighboring atom with the layer of solvent cavitating it, realized through equation 2.18. $\tilde{\sigma}_{Z_k}$is based on atomic number Z_k and in combination with equation 2.19 produces a formulation that ties atomic element *k* to the solvent, and is parameterized for specific main group elemental interaction. If there are two uniquely parameterized elements, described in *table 2.1* as HC, CC, etc., that are side by side, $\tilde{\sigma}_{Z_k}$ then becomes $\tilde{\sigma}_{Z_k Z_k}$, and describes the two-atom interaction with the solvent.

The elements parameterized are listed in *table 2.1*, and their respective parameters can be found in *Table 2* of Marenich et al.[1] The summation over all atoms, of equation 2.18, only incorporates the closest neighbours of each atom. This is ensured through a cut-

off function called "Cut Off Tan" (COT), $T_k(\{Z_{k\prime}, R_{kk\prime}\})$, which is described by 2.20. The COT function is further parameterized through $\Delta r_{Z_k Z_{k'}}$, and $r_{Z_k Z_{k\prime}}$ for atoms H, C, N, O, F, P, S, Cl, Br, I, and another generic X term for the rest of the periodic table. Furthermore, $R_{kk\prime}$ is the interatomic distance. The values for $\Delta r_{Z_k Z_{k'}}, r_{Z_k Z_{k\prime}}$, can be found in the Supporting Information of Marenich et al., *table S11*.[1] The solvent-based parameters utilized for 2.19 are the refractive index of the solvent, n, and again the Abraham hydrogen-bonding basicity, β, along with Abraham's hydrogen-bonding acidity, α.[23,24]

$$\sigma_k = \tilde{\sigma}_{Z_k} + \sum_{k}^{atoms} \tilde{\sigma}_{Z_k Z_{k\prime}} T_k(\{Z_{k\prime}, R_{kk\prime}\}) \quad (2.18)$$

$$\tilde{\sigma_\iota} = \tilde{\sigma}_i^{[n]} n + \tilde{\sigma}_i^{[\alpha]} \alpha + \tilde{\sigma}_i^{[\beta]} \beta \quad (2.19)$$

$$T(R_{kk\prime}, r_{HO}, \Delta r_{HO})$$

$$= \begin{cases} exp\left(\dfrac{\Delta r_{Z_k Z_{k\prime}}}{R_{kk\prime} - \Delta r_{Z_k Z_{k'}} - r_{Z_k Z_{k\prime}}}\right), & if\ R_{kk'} < \Delta r_{Z_k Z_{k\prime}} + r_{Z_k Z_{k\prime}} \\ 0, & else \end{cases} \quad (2.20)$$

Table 2.1: *A list of the elemental types parameterized for equations 2.18 & 2.19. *X is generically set up for the rest of the periodic table.*

$\tilde{\sigma}_{Z_k}$		$\tilde{\sigma}_{Z_k Z_{k\prime}}$
H		HC
C		CC
N		HN
O		CN
F		NC
Cl		NC 2
Br		HO
I		OC
Si		CO 2
P		ON
Si		OO
X*		OP
		OS

Scheme 2.1*: Breakdown of associated interactions with hydrogen and its nearest neighbors within the COT region.*

For $Z_k = H$

$$\sigma_k = \tilde{\sigma}_H + \tilde{\sigma}_{HC} \sum_{\substack{k\prime \\ Z_{k\prime}=C}}^{atoms} T(R_{kk\prime}, r_{HC}, \Delta r_{HC})$$

$$+ \tilde{\sigma}_{HN} \sum_{\substack{k\prime \\ Z_{k\prime}=N}}^{atoms} T(R_{kk\prime}, r_{HN}, \Delta r_{HN}) + \tilde{\sigma}_{HO} \sum_{\substack{k\prime \\ Z_{k\prime}=O}}^{atoms} T(R_{kk\prime}, r_{HO}, \Delta r_{HO})$$

Equation 2.18 uses the single atomic parameter, $\tilde{\sigma}_{Z_k}$, and also requires the use of an atomic-pair parameter $\tilde{\sigma}_{Z_k Z_{k\prime}}$. This leads to the summation of all components associated with element Z_k from *table 2.1*. Therefore, for the case of hydrogen, there is not only the $\tilde{\sigma}_{Z_k} = H, (or\ \tilde{\sigma}_H)$, but then there has to be addition of terms also associated with H, such as: HC ($\tilde{\sigma}_{HC}$), HO ($\tilde{\sigma}_{HO}$), and HN ($\tilde{\sigma}_{HN}$); this can be seen in *Scheme 2.1*.

This formulation is not limited to a linear relationship as can be seen with the case of $\tilde{\sigma}_C$, in *Scheme 2.2*. The list of all associations can be found in Marenich et al's supporting information, *Section S11*.[1] The parameters $\tilde{\sigma}_i^{[n]}, \tilde{\sigma}_i^{[\alpha]}, \tilde{\sigma}_i^{[\beta]}$ are further parameterized for the special case that the solvent is water, meaning these constraints are generalized for all solvents except water for which they are uniquely defined.

Scheme 2.2: *Breakdown of associated interactions with carbon and its nearest neighbors within the COT region.*

$$\sigma_k = \tilde{\sigma}_C + \tilde{\sigma}_{CC} \sum_{\substack{k' \\ Z_{k'}=C}}^{atoms} T(R_{kk'}, r_{CC}, \Delta r_{CC}) + \tilde{\sigma}_{CN} \left[\sum_{\substack{k' \\ Z_{k'}=N}}^{atoms} T(R_{kk'}, r_{CN}, \Delta r_{CN}) \right]^2$$
$$+ \tilde{\sigma}_{CO2} \left\{ \sum_{\substack{k' \\ Z_{k'}=O}}^{atoms} T(R_{kk'}, r_{CO}, \Delta r_{CO}) \left[\sum_{\substack{k'' \\ Z_{k'}=O \\ k''\neq k'}}^{atoms} T(R_{kk''}, r_{CO}, \Delta r_{CO}) \right] \right\}$$

2.1.2 Overall Solvation Energy

Combining both the energy associated with the SCRF polarization, post-SCF polarization, and the post-SCF surface tension energies there is now enough information to calculate the full solvation energy, which can be seen in equation 2.21.

$$\Delta G_S^{\otimes} = \Delta E_E + \Delta G_N + G_P + G_{CDS} + \Delta G_{conc.}^{\otimes} \quad (2.21)$$

The full energy associated with electrostatics is the combination of the first three terms of equation 2.21. They come together to produce the ΔG_{ENP}, which stands for electrostatic, nuclear, polarization, equation 2.22. ΔE_E is the relaxation energy when going from the gas-phase to a condensed-phase, and ΔG_N is associated to the change in energy due to solvation of a geometry optimization. These three terms together represent all the

electrostatics of solvation. Relating this back to the Schrödinger equation, section 1.2, we get the potential as described in equation 2.23.

$$\Delta G_{ENP} = \Delta E_E + \Delta G_N + G_P \quad (2.22)$$

$$\Delta G_{ENP} = \langle \psi^{(1)} | \hat{H} + V | \psi^{(1)} \rangle - \langle \psi^{(0)} | \hat{H} | \psi^{(0)} \rangle \quad (2.23)$$

Where (1) denotes the Slater determinant of the Kohn-Sham orbitals in solution, and (0) corresponds to gas-phase, and therefore, V is the potential due to the overall solvated system. If one does a single point calculation, the energy associated with ΔG_N is 0 kcal/mol. The last component of solvation is the energy associated with going from one concentration in gas-phase to a different one within solution, $\Delta G^{\otimes}_{conc.}$, and for the energies reported in this thesis is 0 kcal/mol, but is not limited to this.[25,26]

2.2 References

(1) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620.

(2) Hirshfeld, F. L. Bonded-Atom Freagments for Describing Molecular Charge Densities. *Theor. Chim. Acta* **1977**, *44*, 129–138.

(3) Peeples, C. A.; Schreckenbach, G. Implementation of the SM12 Solvation Model into ADF and Comparison with COSMO. *J. Chem. Theory Comput.* **2016**, acs.jctc.6b00410.

(4) Baerends, E. J.; Ziegler, T.; Autschbach, J.; Bashford, D.; Bérces, A.; Bickelhaupt,

F. M.; Bo, C.; Boerrigter, P. M.; Cavallo, L.; Chong, D. P.; Deng, L.; Dickson, R. M.; Ellis, D. E.; van Faassen, M.; Fan, L.; Fischer, T. H.; Fonseca Guerra, C.; Franchini, M.; Ghysels, A.; Giammona, A.; van Gisbergen, S. J. A.; Götz, A. W.; Groeneveld, J. A.; Gritsenko, O. V.; Grüning, M.; Gusarov, S.; Harris, F. E.; van den Hoek, P.; Jacob, C. R.; Jacobsen, H.; Jensen, L.; Kaminski, J. W.; van Kessel, G.; Kootstra, F.; Kovalenko, A.; Krykunov, M. V.; van Lenthe, E.; McCormack, D. A.; Michalak, A.; Mitoraj, M.; Morton, S. M.; Neugebauer, J.; Nicu, V. P.; Noodleman, L.; Osinga, V. P.; Patchkovskii, S.; Pavanello, M.; Philipsen, P. H. T.; Post, D.; Pye, C. C.; Ravenek, W.; Rodríguez, J. I.; Ros, P.; Schipper, P. R. T.; van Schoot, H.; Schreckenbach, G.; Seldenthuis, J. S.; Seth, M.; Snijders, J. G.; Solà, M.; Swart, M.; Swerhone, D.; te Velde, G.; Vernooijs, P.; Versluis, L.; Visscher, L.; Visser, O.; Wang, F.; Wesolowski, T. A.; van Wezenbeek, E. M.; Wiesenekker, G.; Wolff, S. K.; Woo, T. K.; A.L., Y. ADF. **2014**.

(5) Baerends, E. J.; Ros, P. Self-Consistent Molecular Hartree-Fock-Slater Calculations II. The Effect of Exchange Scaling in Some Small Molecules. *Chem. Phys.* **1973**, *2* (1), 52–59.

(6) Born, M. Volume and Hydration of Ions. *Z. Phys.* **1920**, *1*, 45–48.

(7) Still, C.; Tempczyk, A.; Hawley, R. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(8) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36* (22), 3219–3228.

(9) Cramer, C. J. *Essentials of Computational Chemistry Theories and Models*; 2004; Vol. 2.

(10) Mulliken, R. S. Electronic Population Analysis on LCAO[Single Bond]MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23* (10), 1833.

(11) Lowdin, P.-O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365–375.

(12) Löwdin, P. O. On the Nonorthogonality Problem. *Adv. Quantum Chem.* **1970**, *5* (C), 185–199.

(13) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.

(14) Pauling, L. Atomic Radii and Interatomic Distances in Metals. *J. Am. Chem. Soc.* **1947**, *69* (3), 542–553.

(15) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. Improved Methods for Semiempirical Solvation Models. *J. Comput. Chem.* **1995**, *16*, 422–440.

(16) M., A.; I., S. Handbook of Mathematical Functions. *Natl. Bur. Stand. Appl. Math. Ser.* **1954**.

(17) Cramer, C. Email Correspondance.

(18) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Density Functional Solvation Model Based on CM2 Atomic Charges. *J. Chem. Phys.* **1998**, *109*, 9117–9133.

(19) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3093.

(20) Bondi, A. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1965**, *68*, 441–451.

(21) Haynes, W. *CRC Handbook of Chemistry and Physics*, 91st ed.; Taylor & Francis, 2010.

(22) Cramer, C.; Truhlar, D. A Universal Approach to Solvation Modeling SM8. *Acc. Chem. Res.* **2008**, *41*, 760–768.

(23) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Duce, P. P.; Morris, J. J.; Taylor, P. J. Hydrogen-Bonding .7. A Scale of Solute Hydrogen-Bond Acidity Based on Log K-Values for Complexation in Tetrachloromethane. *J. Chem. Soc. Perkin Trans. 2* **1989**, No. 6, 699–711.

(24) Abraham, M. H. Scales of Solute Hydrogen-Bonding - Their Construction and Application to Physicochemical and Biochemical Processes. *Chem. Soc. Rev.* **1993**, *22* (2), 73–83.

(25) Martin, R. L.; Hay, P. J.; Pratt, L. R. Hydrolysis of Ferric Ion in Water and Conformational Equilibrium. *J. Phys. Chem. A* **1998**, *102* (2), 3565–3573.

(26) Shamov, G. A.; Schreckenbach, G. Density Functional Studies of Actinyl Aquo Complexes Studied Using Small-Core Effective Core Potentials and a Scalar Four-

Component Relativistic Method. *J. Phys. Chem. A* **2005**, *109* (48), 10961–10974.

Preface to Chapter 3

This chapter is based on a manuscript published in the "*Journal of Chemical Theory and Computation*" The full citation is as follows:

Craig A. Peeples and Georg Schreckenbach, "Implementation of the SM12 Solvation Model into ADF and Comparison with COSMO" *Journal of Chemical Theory and Computation*, **2016**, 12, pp 4033-4041.

General Born Solvation models, and previous charge models are heavily reliant on their choice of basis sets. With the newest iteration of SM12, Marenich et al formulate the model in such a way that it will forego the stringent basis set dependencies, through the use of basis set independent Hirshfeld population analysis. Before now, the SMx models have only been tested on Gaussian Type Orbitals, through the Gaussian program, and others, but have never been utilized within a Slater Type Orbital code, such as ADF. This paper details the implementation of SM12 into ADF, and a comparison of the native solvation model within ADF, COSMO. The paper then goes on to detail that this new iteration of charge model, CM5, and solvation model, SM12, both show little dependence on basis set, as well as outperform COSMO, while being comparable to COSMO-RS.

All calculations and implementations within the manuscript and chapter were performed by Craig Peeples. The manuscript was prepared together with Prof. Georg Schreckenbach.

Chapter 3: Implementation of the SM12 Solvation Model into ADF and Comparison with COSMO

Craig A. Peeples, Georg Schreckenbach*

Department of Chemistry, University of Manitoba

Winnipeg, MB, R3T 2N2, Canada

schrecke@cc.umanitoba.ca

3.1 Abstract

In this article, an implementation of the newest iteration of the Minnesota solvation model, SM12, into the Amsterdam Density Functional (ADF) computational package is presented. ADF makes exclusive use of Slater type orbitals (STO) which correctly represent the true atomic orbitals for atoms, whereas SM12 and the underlying Charge Model 5 (CM5) have previously only been tested on Gaussian type orbitals (GTO). This new implementation is used to prove the basis set independence of both CM5 and SM12. A detailed comparison of the SM12 and COSMO solvation models, as implemented in ADF, is also presented. We show that this new implementation of SM12 has a Mean Unsigned Error (MUE) of 0.68 kcal/mol for 272 molecules in water solvent, 4.10 kcal/mol MUE for 112 charged ions in water, and a MUE of 0.92 kcal/mol for 197 solvent calculations of various molecules. SM12 outperforms COSMO for all neutral molecules, performs as well as COSMO for cationic molecules, only falling short when anionic molecules are taken into consideration, likely due to CM5's use of Hirshfeld charges and their poor description of anionic molecules, though CM5 seems to improve upon this discrepancy.

3.2 Introduction

Modeling real-world chemistry often requires taking condensed phase effects, including solvation, into account. For quantum-chemical modeling, choices have to be made with respect to different levels of approximation. Earlier, we have proposed that these approximations can be described with three perpendicular axes: model chemistry (choice of correlation method and basis sets), relativistic effects, and finally the overall choice of reduced model for complex systems, including continuum solvation effects.[1] In order to properly describe a system with efficiency and accuracy one must find a way to optimize these axes. Within this paper, we focus on a newly implemented continuum solvation model, and the influence of basis sets and correlation methods on its accuracy.

Continuum solvation expresses the solvent molecules in a continuous distribution function.[2,3,4] In expressing the solvent in this way we reduce computational effort; this tradeoff comes at a cost of losing information about the solvent structure. The challenge is how to properly express the electrostatics of inserting a charged particle into a liquid medium; this is not a new problem but has been thoroughly studied through Poisson's equation. [2,3,4,5] Solving Poisson's equation will result in information about the electric potential for a given charge distribution. This type of calculation is currently handled by several continuum solvation models such as the Polarizable Continuum Model (PCM),[5] Integral Equation Formalism (IEF),[6] Surface and Volume Polarization for Electrostatic (SVPE and SS(V)PE)[7] the Conductor-like Screening Model (COSMO) of Klamt and

Schüürmann,[8] as well as others. [2] These methods are known as the Apparent Surface Charge (ASC) methods. [2] An alternative to the ASC methods can be sought through the use of an approximation of Poisson's equation, known as the Generalized Born Approximation (GBA). The GBA is a generalization of the Born equation, where the Born equation treats an atom as a constant charge over a perfect sphere of atomic radius, thus giving us the charge distribution over that atom. When treated this way the Born equation is exact for a point charge in a polarizable medium. In order to properly use this equation for real-world solutes it is required that the method be applicable to more than one atom and it is generalized through an approximation to incorporate multiple atoms. The GBA will act as the Born equation when working with a single atom, and when two atoms are at a far distance it simply reduces to Coulomb's law.

The recently published solvation model SM12 takes advantage of the charges arising from charge model 5 (CM5) as a fast and efficient way of solving the GBA equations.[9,10] CM5, in turn, makes use of Hirshfeld Population analysis, HPA.[11] Through use of the basis set independent HPA, there is a reduction in the stringent basis set dependence of most GBA methods, and consequently the resulting SMx models,[11,12] making SM12 a particularly attractive GBA solvation model, particularly in the context of quantum-chemical codes that do not use Gaussian type orbitals (GTO) as basis sets.

The Amsterdam Density Functional package (ADF)[13,14,15] is a code that uses Slater Type Orbital (STO) basis sets exclusively. STOs better approximate true atomic orbitals

of atoms.[14] Currently within ADF, the ASC model COSMO,[16] and COSMO-RS[17] are the methods for implicit solvation calculations, as well as the Quantum Mechanic/Molecular Mechanic (QM/MM) hybrid RISM[18] for a statistical mechanical approach to explicit/implicit solvation. Here we present a new alternative to modeling the effects of a solvent within ADF in the form of the SM12 method of Marenich, Cramer, and Truhlar.[9] An in-depth study of the basis set dependence of both CM5 and SM12 is presented with a comparison of calculations using varying ADF STO basis sets to the literature results of CM5 and SM12 using varying GTO basis sets and a modified Gaussian09 code.[10,9] In addition, a comparison of the performance of ADF's COSMO solvation model with the new SM12 implementation using neutral, cationic and anionic solutes as well as several calculations using solvents both models have in common are presented.

3.3 Theory

The SM12 model[9] is the newest GBA model in a series of solvation models generally named SMx that have been developed at the University of Minnesota. SM12 improves over the previous SMx[4,19,20,21] models by incorporating Hirshfeld charges into the new CM5, removing the reliance on the heavily basis set dependent Löwdin charges.[9,22,23] With the change from Löwdin to Hirshfeld, SM12 can be allowed to use STOs whereas the previous CMx[24,20] iterations were parameterized and optimized explicitly for GTOs only. Moreover, Löwdin population analysis becomes very unreliable as larger and more extended basis sets are utilized.[10,25] SM12 also improves upon the

previous iterations by incorporating a larger, more robust training set of molecules. This set includes heavier atoms and molecules up to and including I with a focus on the main group elements; this training set includes 2979 experimental solvation data.[9] The new parameterization also includes a generic term for elements not explicitly accounted for; this leads to an amelioration to the previous SMx cavitation terms.

SM12 breaks down the energy of solvation, $\Delta G_S^{\otimes}$, into several contributions.

$$\Delta G_S^{\otimes} = \Delta E_E + \Delta G_N + G_P + G_{CDS} + \Delta G_{conc.}^{\otimes} \quad (3.1)$$

The combination of the first three terms has been known in previous SMx models as the Electronic, Nuclear and Polarization term, or ΔG_{ENP}.

$$\Delta G_{ENP} = \langle \psi^{(1)} | \hat{H} + V | \psi^{(1)} \rangle - \langle \psi^{(0)} | \hat{H} | \psi^{(0)} \rangle \quad (3.2)^4$$

The potential added through solvation is V, the Slater determinant of the Kohn-Sham orbitals in solvent is $\psi^{(1)}$, and the gas phase is $\psi^{(0)}$. ΔG_N is the representation of the change in energy associated with displacement of the nuclear positions from gas to solvent. Therefore, in a single-point calculation this term vanishes. The calculations involved in this article are all done with optimized gas-phase geometries provided by the Minnesota Solvation Database[26], making ΔG_N zero within our study, though through a geometry optimization with SM12 this component will be non-zero. The polarization of the solute and solvent can be represented by the third term of equation (3.1), G_p. This is done through the GBA, which takes the form of equation 3.3, in atomic units. Equation 3.3 represents the polarization energy associated with the solvent re-orientating itself to properly

accommodate the induced dipole of the dielectric due to the charge distribution of the solute, creating a dipole favorable interaction.

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) \sum_{k,k'}^{atoms} q_k \gamma_{kk'} q_{k'} \quad (3.3)$$

Where ε is the permittivity of the medium, and γ is Still et al.'s[27] approximation of the Coulomb integral. Note the similarity between equation 3.3 and the well-known Born equation.[28] Finally the charges, q_k, are calculated using CM5.[10,27] CM5 is a class IV charge model. Class IV models incorporate a wavefunction- or density-based calculation (class III, HPA) corrected with a semi-empirical (class I) calculation.[29] The full form of CM5 is shown in equation 3.4.

$$q_k^{CM5} = q_k^{HPA} + \sum_{k' \neq k} T_{kk'} B_{kk'} \quad (3.4)$$

The HPA charges q_k^{HPA} are calculated using Hirshfeld's original conception,[11] and the results are corrected using a combination of an updated Pauli bond order, $B_{kk'}$, and a parameterized function, $T_{kk'}$.[10] ΔE_E is the electronic relaxation energy associated with the transition from gas phase to the condensed phase.[9] The combination of the ENP terms then expresses how the solute and solvent interact through their induced dipoles.

Finally, the G_{CDS} (cavitation, dispersion, solvent) term attempts to control for the energy that cannot be expressed through the electrostatic approximation. Such as the structure of the cavity of the solvent, the dispersion associated within the solute itself, and the structural information lost by assuming the solvent structure can be approximated by its permittivity. It is constructed from two contributions: one that encompasses how the solvent interacts with the full molecule, and the second how each solute atom affects the solvent.

$$G_{CDS} = \sum_{k}^{Atoms} \left(\sigma_k + \sigma^{[M]}\right) A_k \quad (3.5)$$

A_k represents the Solvent Accessible Surface Area (SASA) which is calculated through the Analytical Surface Area (ASA).[30] The ASA represents atoms as spheres with radii equal to each atom's respective van der Waals radii.[31] ASA uses the vectors between the nuclear centres of the atoms in order to find the surface area that is consumed by its closest neighbours; the ASA removes the area that is within the sphere of another atom. Equation 3.5 contains two independent terms $\sigma^{[M]}$ and σ_k.

The parameters of Equation 3.5 can also be customized by the end user. If the user has all the descriptors required for a solvent, in Table 3.1, they can create their own solvent. These descriptors are as follows: β is Abraham's hydrogen bond basicity parameter of the solvent; γ, the macroscopic surface tension of the solvent at the air/solvent interface at 298K. ϕ^2 is the square of the fraction of non-hydrogen atoms in the solvent molecule that

are aromatic carbons. Finally, ψ^2 is the square of the fraction of non-hydrogen atoms in the solvent that are halogens. The full form of $\sigma^{[M]}$ can be seen in equation 3.6.

$$\sigma^{[M]} = \sigma^{[\beta^2]}\beta^2 + \sigma^{[\gamma]}\gamma + \sigma^{[\phi^2]}\phi^2 + \sigma^{[\psi^2]}\psi^2 \quad (3.6)$$

$\sigma^{[M]}$ is the representation of how the structurally approximated solvent interacts with the bulk molecule. The $\sigma^{[x]}$ Parameters are dependent on two factors according to Marenich et al.[9] :

1) Solvent: β^2, γ, ϕ^2, and ψ^2

2) SMx:CMx model: $\sigma^{[\beta^2]}$, $\sigma^{[\gamma]}$, $\sigma^{[\phi^2]}$, and $\sigma^{[\psi^2]}$

The $\sigma^{[x]}$ terms are parameters that vary for the different generations of CMx and SMx. That is, if one uses SM12 with a different charge model (e.g. CM4) new parameters for $\sigma^{[x]}$ will be required.

The remaining parameter of G_{CDS}, σ_k, is represented in the same way as reported in the original work by Marenich et al.,[9] equation 3.7. The non-σ components are also customizable by the end user. Here, n is the refractive index of the solvent and α is Abraham`s hydrogen bond acidity parameter.

$$\sigma_k = \sigma_k^n n + \sigma_k^\alpha \alpha + \sigma_k^\beta \beta \quad (3.7)$$

$\sigma^{[x]}$ are parameters based on either aqueous or non-aqueous solvent. The last customizable parameter is the permittivity of the solvent, ε, which represents the polarity of the solvent.

Table 3.1: *Parameters used for describing the solvent (See Text).*[9] *Each parameter is customizable. *All 92 solvents are listed in Appendix 1.1 in the Supporting Information.*

Solvent Descriptors	
N	Index of refraction
A	Abraham's acidity parameter
B	Abraham's basicity parameter
γ	Macroscopic surface tension
φ	carbon aromaticity
ψ	electronegative halogencity
All solvents (92) parameterized by Marenich et al.[9] have been implemented into ADF*	

3.3.1 Implementation

Implementation of the empirical CDS term utilized the same parameters as were set by Marenich et al.. Each new iteration of the SMx family makes use of variations of how the CDS solvent-/parameter-based calculation is handled; therefore this implementation was done in a fashion that would allow for easy addition of the next generation of SMx CDS terms.

Re-programming and making use of the ASA algorithm of Liotard et al.[30] is the basis for all portions of the solvation model's energy calculation, as it is used to calculate the SASA. Also the ASA is called for each numeric integral step of the numerical Born calculation and therefore care must be taken. This procedure was adapted straight from the original implementation of Liotard et al.[30] The ASA was then updated from Fortran 60 to Fortran 95 in order to adapt it both to the advanced memory allocation and style guide of ADF.

The polarization term of SM12, G_P of Eq. (3.3), has also been fully implemented. The polarization term affects the energy from each SCF iteration, contributing to the SCRF and the post-SCF energy. The polarization term is calculated using the GBA, Eq. (3.3), making use of CM5's mapping of partial charges and the Coulomb approximation according to Still et al.[27]. CM5 makes use of the HPA. HPA charges were already incorporated in the ADF code and calculated for every job but this was done in a symmetry-specific fashion. Here, we require each atom to have its own Hirshfeld charge to properly map partial charges through CM5. Therefore, we needed to modify how HPA was calculated, and allow for this calculation to be done during the SCF in order to include an analysis for each atom. The same parameters put forth in the original paper[9] were utilized in the CM5 ADF implementation. This was done as charges from CM5 should be basis set independent and thus moving from the original GTOs to the STOs used by ADF will not cause significant differences in the charge mapping.

The ADF implementation of CM5 was tested using the same molecules used in Marenich et al.'s paper and their optimized coordinates.[10] This was done as the empirical component of CM5 is dependent on interatomic distances, and required testing the exact same geometries with the newly implemented ADF CM5. Testing the implementation in this way also allowed for direct comparison of the two implementations of CM5, as the only differences between them is how their HPA is calculated. This will result in a comparison of CM5 using GTOs and STOs, and thus give insight on how each partial charge mapping is dependent on their respective basis set.

The existing ADF implementation of HPA could only be applied in a post-SCF fashion, whereas we require an iterative calculation in order to properly incorporate the Self Consistent Reaction Field (SCRF). Density calculations can be easily performed for every cycle of the SCF with utilization of the newly implemented Becke grid.[32,33] HPA can be represented as a partitioning of the atomic density of a specific fragment, ρ^{at}, over the entire atomic density of the molecule, ρ^{pro}, better known as the Hirshfeld weights. In order to get the charge associated with an atom the Hirshfeld weights are applied to the molecular density, ρ^{mol}.

$$q_k^{HPA} = -\int \frac{\rho_k^{at}(r)\rho^{mol}(r)}{\rho^{pro}(r)} dv \quad (3.8)$$

The Hirshfeld calculation is now able to utilize the new density, ρ^{mol}, calculated from each iteration of the SCF cycle. This was done for two reasons:

1) We require, for each SCF cycle, a new CM5 calculation. This is done to properly represent the ever changing electronic density given by DFT, $\psi^{(1)}$, of equation 3.2.
2) For every SCF cycle we need to calculate, iteratively, the effect that the polarization has on the Fock Matrix, ($F_{\mu,\nu}^{Mat}$).

These are the two points in which basis set dependence, and thus the difference between GTOs and STOs, are relevant when using SM12. This is due to the basis sets being explicitly incorporated into the polarization of the molecule within this model. Polarization affects the Fock matrix iteratively through each SCF cycle; the application of the polarization of each atom on each SCF cycle depends on how the polarization changes with respect to the density matrix, $\frac{\partial G_P}{\partial \rho_{\mu,\nu}}$ in equation 3.9. This creates a correction to the Fock matrix, $F_{\mu,\nu}^{Mat^\circ}$ by $\frac{\partial G_P}{\partial \rho_{\mu,\nu}}$, which are both dependent on the basis set; this is better known as the SCRF.[24]

$$F_{\mu,\nu}^{Mat} = F_{\mu,\nu}^{Mat^\circ} + \frac{\partial G_P}{\partial \rho_{\mu,\nu}} \quad (3.9)$$

3.4 Computational Details

For calculating their CM5 results, Marenich et al. used the Minnesota basis set MG3S, which is equivalent to 6-311+G(3d2f,2df,2p),[10] along with the meta-GGA M06-L functional.[34] Within ADF we decided to test CM5 with a semi-large basis set in an attempt to both investigate the basis set independence, and the effect of switching from GTOs to STOs. The basis set chosen was then the TZ2P basis set of ADF. Using the fact that Hirshfeld population analysis isn't heavily reliant on the choice of basis set[12] we wanted to

ensure this was still the case when working with STO-based charges within CM5. We also want to investigate that STO-CM5 charge calculations are indeed basis set independent when comparing to their GTO-CM5 counterpart.[25] This was done with the molecules and geometries from the test set used by Marenich et al.[10] This test set consists of 721 molecules, or 5880 atomic point charges of changing molecular charges and multiplicities.

A solvation database[26] of 790 unique molecules with 3037 experimental solvation or transfer free energies was received from Minnesota in order to test the new SM12 implementation. These molecules range from simple methane to larger, chemically unique and polar molecules such as bensulfuron. This test set makes use of mostly organic molecules as the SMx series is focusing on exploring solvation of these types of compounds. All calculations were done within a locally modified version of the ADF[13,14,15] 2014 suite of programs, unless otherwise stated. Unless otherwise stated, the functional used was PBE,[35] and the basis sets explored are DZP, TZP, TZ2P. Next explored was how functionals affect the new solvation model. Investigation of pure Hartree-Fock, GGA PBE, hybrid B3LYP, [36,37] and metahybrid M06-2X[35,34] were performed using SM12. The 197 different solvents in the database were investigated using a TZP basis set with the hybrid B3LYP functional, as well as the metahybrid M06-2X. Finally, investigation of how SM12 effects charged molecules was performed. For this purpose, the ADF basis sets AUG-DZP, AUG-TZP, as well as AUG-TZ2P were used for anionic molecules, whereas, for cationic molecules, we used the AUG-TZP basis set. In order to compare the ADF STO results against a GTO version of SM12, we used the Q-Chem implementation.[38]

The majority of these calculations were repeated using ADF's original solvation model, COSMO,[16] and with a set of several solvents that COSMO and SM12 have in

common. This was done to get a general comparison of SM12 and COSMO within ADF as well as each model's strengths and weaknesses. All COSMO calculations use a solvent radius of 1.4 with cav0 = 1.321, cav1 = 0.0067639. For each calculation we utilize the Delley surface with triangles sized with ndiv = 4 and an overlap of 0.8, variational SCF always on, and we calculate the exact C-matrix (c-mat). All calculations, unless otherwise noted, were performed with a locally modified version of ADF2014.

3.5 Results and Discussion

Figure 3.1: *Comparison of CM5 in ADF (TZ2P STOs) to CM5 literature results obtained with Gaussian 09 (MG3S GTOs).*[10] *The data set comprises 721 molecular compounds (5880 charges), with varying charges and multiplicities. Results can be seen in Appendix 1.2*

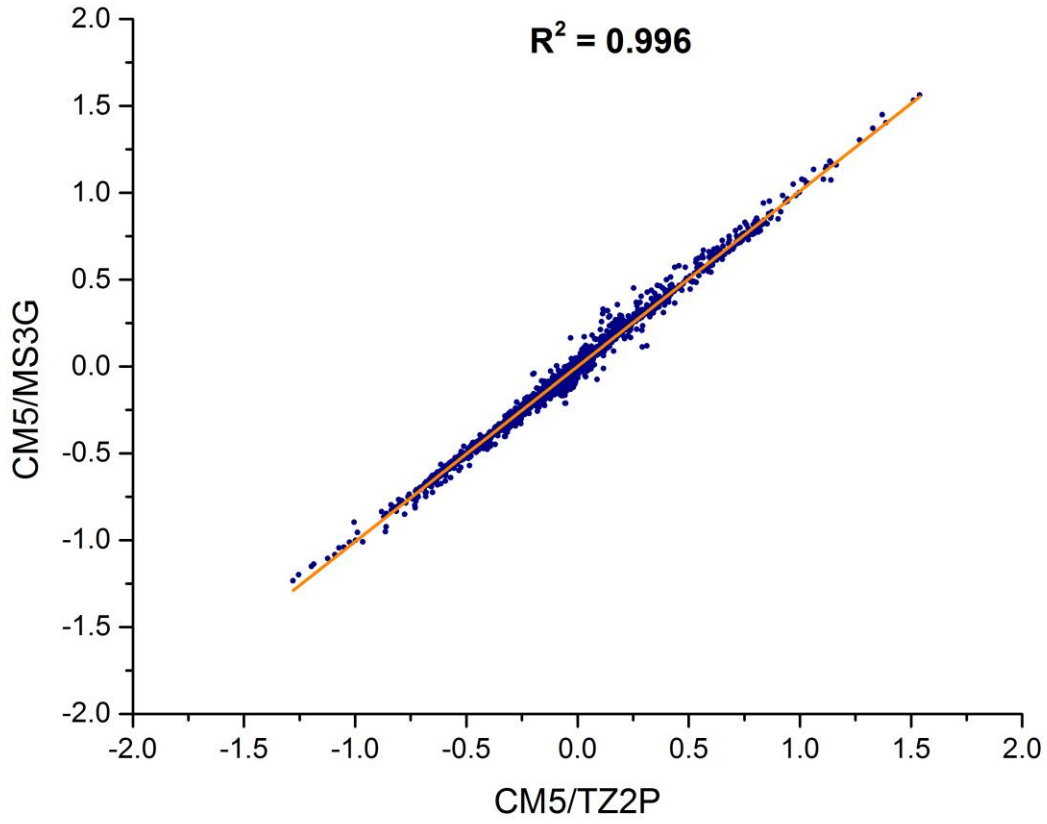
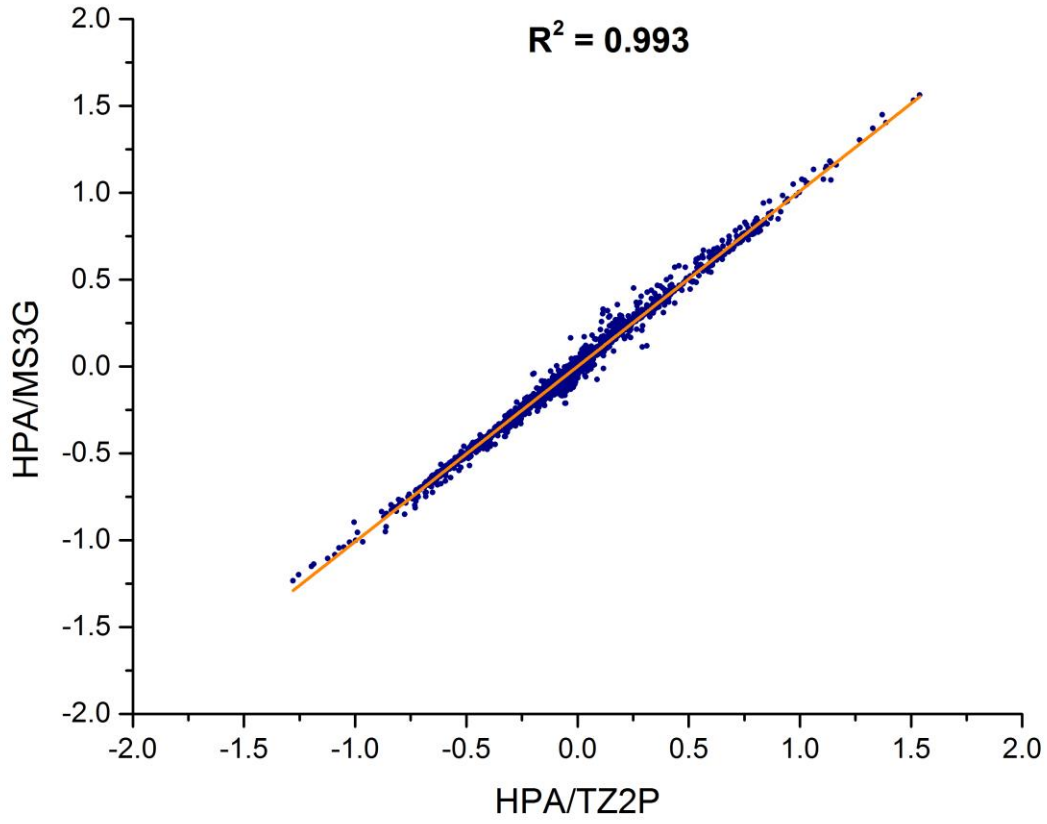


***Figure 3.2**: Comparison of HPA in ADF (TZ2P STOs) to HPA in Gaussian 09 (MG3S GTOs).*[10]

The data set comprises with 721 molecular compounds (5880 charges), with varying charges and multiplicities. Results can be seen in Appendix 1.2.



3.5.1 Charge Model 5

The dependence of the CM5 charges on the type of basis functions (STOs vs. GTOs) was explored by comparing the ADF STO results to the GTO results of Marenich et al., see Figure 3.1. We compared the results from ADF using STO-TZ2P and PBE to those of Marenich et al.,[10] which used MG3S. The dependence of the HPA on the type of basis functions was investigated with the same procedure, Figure 3.2. We conclude that the results from CM5 agree, with great precision, with the basis set independence assertion of Marenich et al.[10] The correlation coefficient put forth from comparison of CM5/6-31G(d) to CM5/MG3S[10] of $R^2 = 0.996$, agrees very well with our comparison of CM5/STO-TZ2P to CM5/MG3S with an $R^2 = 0.996$, see Figure 3.1. In both basis set cases a very strong correlation is seen. Comparison of HPA/MG3S with HPA/STO-TZ2P in Figure 3.2 shows an $R^2 = 0.993$. The full list of compounds can be found in Appendix 1.2, and the partial charges in Table S7.

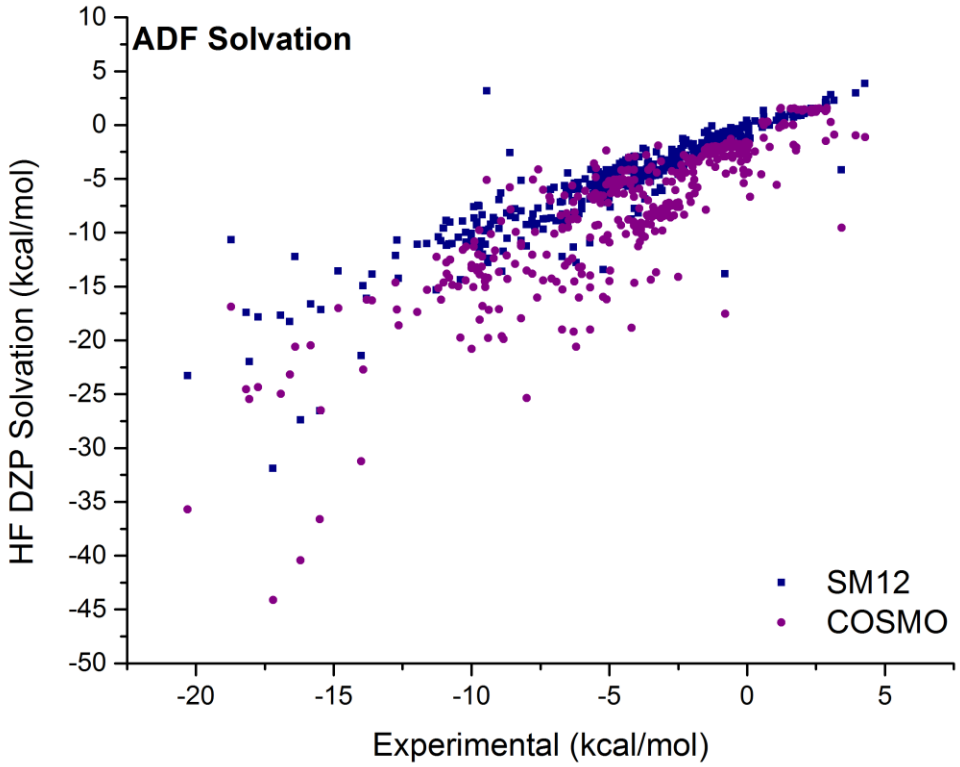
The original implementation of CM5 was done in a modified version of G09[10] with the 6-31G(d) and MG3S basis sets which both consist of GTOs. Moving into ADF we make use of STOs which are better representations of true atomic orbitals.[39] The results agree with little uncertainty to the results reported by Marenich et al. As we have found our graph is nearly identical to their comparison of two different basis sets with the same square of the correlation coefficient, $R^2 = 0.996$. This proves that indeed CM5 is not only basis set independent, but is code independent and stable for STOs as well. HPA has been studied thoroughly and has been proven to be basis set independent before.[12,40] With the parameterization from Marenich et al. of a very diverse set of molecules, we find that CM5

reduces the basis set dependence of HPA; making CM5 a very versatile tool for calculating and handling partial charges.

Table 3.2: *MUE of ADF based calculations and experiment of 381 molecules. Solvation energies are all calculated in water solvent. Single point calculations were done on optimized gas-phase geometries from the Minnesota Solvation database.*[26] *All energies are reported in kcal/mol, and the solvation free energy calculations of Hartree-Fock using DZP are in Appendix 1.4. *All Q-Chem calculations are from the Minnesota Solvation database.*

	ADF Solvation		
Basis Set	SM12	COSMO	SM12 Q-Chem
	PBE		
DZP	1.12	1.53	/
TZP	1.20	1.78	/
TZ2P	1.08	1.70	/
	Hartree-Fock		
STO DZP	1.20	3.71	/
GTO 6-31G*	/	/	1.21

Figure 3.3: *Comparison of SM12 (blue squares) and COSMO (violet circles) water solvation free energies with the GGA PBE using DZP.*



3.5.2 Solvation in Water

To fully explore the change and dependence of the solvation energy on the basis sets, Table 3.2 shows results for three different polarized Slater Type Orbital basis sets within ADF: DZP, TZP, and TZ2P, as well as PBE, and Hartree-Fock. Figure 3.3 provides a comparison of the DZP results using the PBE functional. Each of the three basis sets shows the ADF SM12 model outperforming the current COSMO model, Table 3.2. The results also demonstrate that, for both SM12 and COSMO, there is the very little basis set dependence – as the complexity increases with respect to basis sets the MUE remains relatively unchanged. STO-TZ2P achieves the lowest MUE for SM12 and STO-DZP for

COSMO. Solvation free energy calculations in water with Hartree-Fock using DZP for COSMO and SM12 are fully displayed in appendix 1.4.

As stated before, using the GBA requires the extensive use of partial charges, and thusly needs an appropriate mapping**.** Their use of Löwdin population analysis made previous iterations of SMx fairly poor at dealing with larger and diverse basis sets. Adequate results from CM5 using ADF`s STOs allows us to move onto calculation of the polarization term of the SM12 model. The Q-Chem reference data were done at the Hartree-Fock level and with the 6-31G* basis set. Therefore, utilized within ADF were HF calculations with the STO-DZP basis set as a relatively close substitute for the GTO basis set used in Q-Chem. As can be seen from Table 3.2, the STO implementation in ADF agrees fairly well with the GTO results from Q-Chem. These particular calculations were all done within a water solvent. The MUE seen from the ADF implementation shows a nearly identical MUE when compared to the Q-Chem results using HF, while COSMO performs poorly when compared to SM12 for these calculations. The results obtained from B3LYP and M06-2X in water from ADF, 0.68 and 0.69 kcal/mol (Table 3.3), show a slightly worse MUE than those of the modified Gaussian SM12[9] model, 0.59 and 0.63 kcal/mol, respectively. Both the ADF and Gaussian 09 implementations show a slight decrease in accuracy as the functional is changed from the hybrid B3LYP to the metahybrid M06-2X, but an improvement in accuracy over the Hartree-Fock method. 272 molecules[9] were tested for B3LYP and M06-2X, see Table 3.3 (figure 3.5), and the MUE was calculated in order to compare the SM12 implementations in ADF and Gaussian 09[9]. The 272 molecules from the original implementation[9] were also tested with COSMO in water

solvent with both functionals, The compounds and their respective solvation free energy results are collected in Appendix 1.3.

***Figure 3.4*:** *SM12 comparison to experiment of 376 compounds in a water solvent, HF DZP calculations*

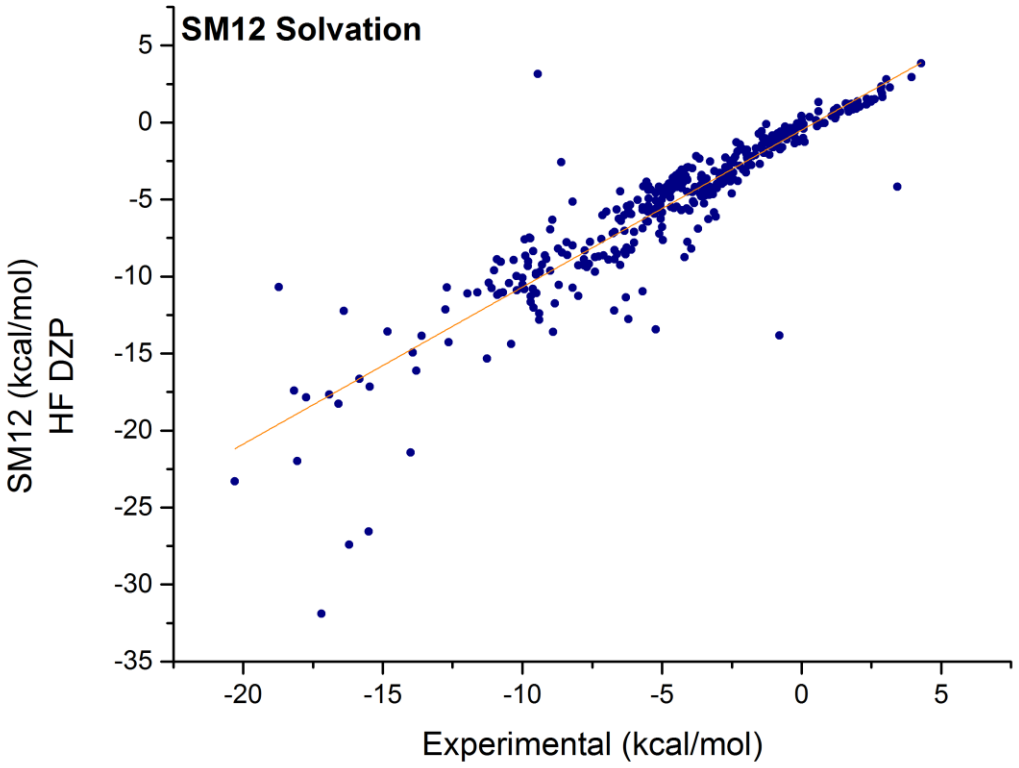
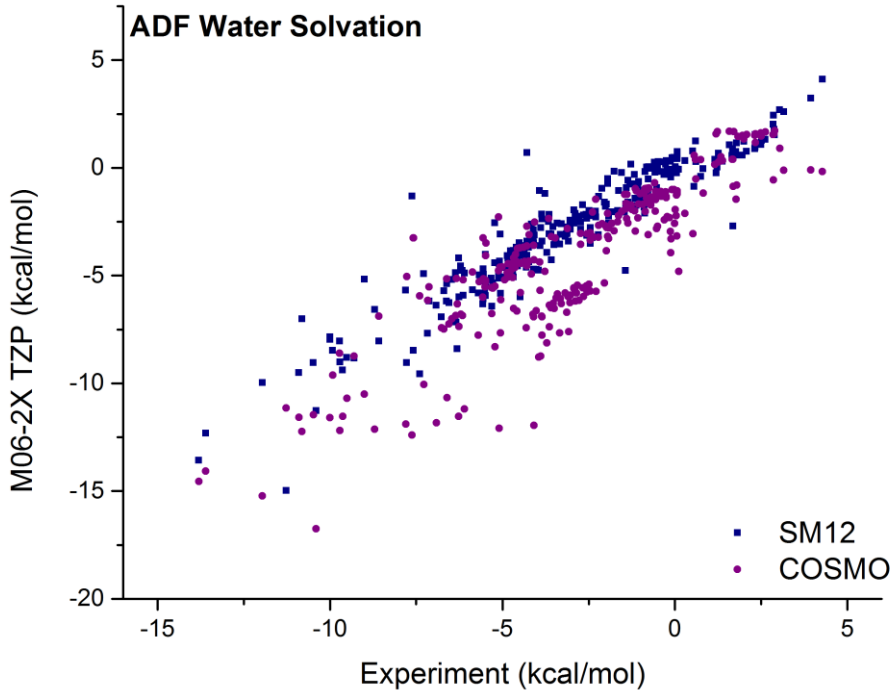


Table 3.3: *MUE of ADF based calculations and experiment of 272 molecules. Solvation energies are all calculated in water solvent. Calculations were done on optimized gas-phase geometries from the Minnesota Solvent database, and their implementation into Gaussian 09. All energies are reported in kcal/mol, and can be seen in Appendix 1.3.*

TZP *Functionals*	SM12 ADF	SM12 Marenich et al*	COSMO ADF
B3LYP	0.68	0.59	1.47
M06-2X	0.69	0.63	1.54

Figure 3.5: *SM12 comparison to experiment of 272 compounds in a water solvent, M06-2X TZP calculations.*



3.5.3 SM12 Solvents

In addition, we tested the 197 different solvents parameterized for SM12. For SM12 solvation within ADF we performed a total of 2141 calculations, with varying solvents and molecules (figure 3.6). The results of all solvent calculations are summarized in Table 3.4. There is a decrease in accuracy when we move from a GTO based code to the STO based ADF, although Table 3.5 shows a relative increase in accuracy when comparing SM12 to COSMO. Both COSMO and SM12 have varying solvents parameterized by default, and hence we tested three different solvents that they have in common; this was done to represent how each solvation model utilizes varying solvation environments, ranging from non-polar cyclohexane to a slightly higher permittivity

(benzylalcohol), to a more polar solvent (DMSO). The results are seen in Table 3.5. As can be seen when exploring less polar solvents both models perform essentially identically, but as the medium becomes vastly more polar we see that SM12 outperforms the COSMO. This agrees with the result from Table 3.3, with water's dielectric constant being around 78.4.

Table 3.4: *MUE of ADF based calculations and experiment of the 92 solvents available within SM12, and varying compounds. Calculations were done on optimized geometries from the Minnesota Solvent database, and their implementation into Gaussian 09. All energies are reported in kcal/mol. 2141 single point calculations were performed.*Marenich et al. had reported 2129 calculations*

Basis sets / *Functionals*	SM12 ADF TZP	SM12 Marenich et al.* MG3S	SM12 Marenich et al.* 6-31G(d)
B3LYP	0.93	0.67	0.60
M06-2X	1.02	0.65	0.60

Figure 3.6: *SM12 comparison to experiment of 2141 single point calculations of various compounds in the 97 available solvents, B3LYP TZP calculations.*

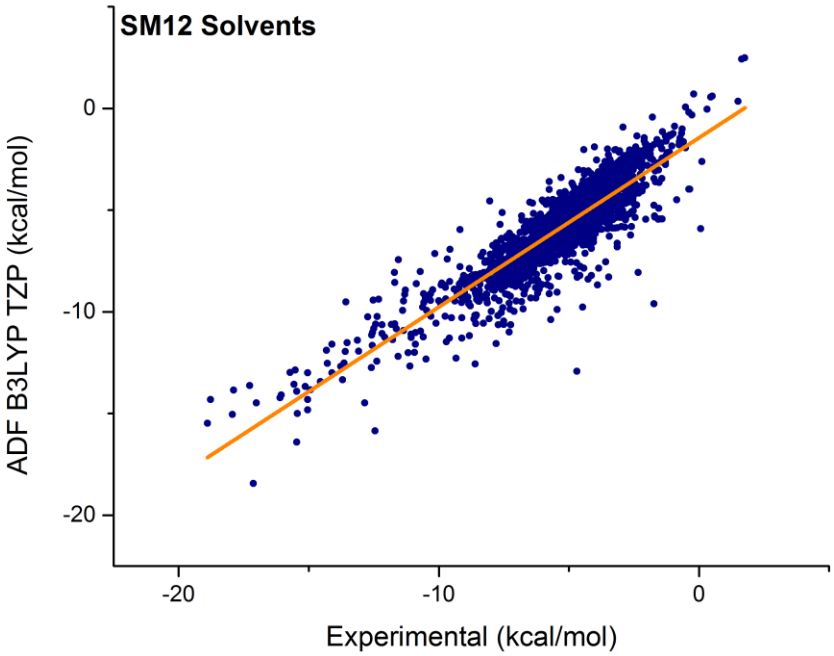


Table 3.5: *MUE of ADF based calculations and experiment of 109 varying compounds in three different solvents. Calculations were done on optimized geometries from the Minnesota Solvent database, and their implementation into Gaussian 09. All energies are reported in kcal/mol, and in Appendix 1.5.*

Solvents	***Benzyl Alcohol***	***Cyclohexane***	***DMSO***
ε	***12.5***	***2.0***	***46.8***
SM12			
B3LYP	2.1	1.9	2.0
M06-2X	2.2	1.8	2.0
COSMO			
B3LYP	1.8	2.9	3.5
M06-2X	1.9	2.8	3.7

3.5.4 Ionic Solvation

Solvation results discussed thus far are for neutral molecules. Also tested were the effects that COSMO and SM12 have both on cationic and anionic molecules. The results are summarized in Table 3.6. SM12 anionic calculations in ADF obtain MUE values close to those reported by Marenich et al.[9] and again show that there is very little dependence on basis sets when utilizing the SM12 model. COSMO, on the other hand, outperforms both GTO and STO SM12, achieving a lower MUE than the best performing calculation of Marenich et al.[9]. The better performance of COSMO, compared to SM12 when anionic compounds are taken into consideration, could possibly be attributed to the poor performance of HPA when calculating charged species.[25] Another reason why COSMO greatly outperforms SM12 could be attributed to the way COSMO handles the charge calculation of atoms. COSMO, as an ASC method, integrates the various charge calculations along the numerical integration grid points that define the atomic surfaces, and this could allow COSMO to account for the diffuse basis sets required by anionic calculations. Thus, COSMO samples the diffuse space more thoroughly and better accounts for the electronic charge distribution, while SM12, a GBA model, approximates the charges of each atom to be uniform and centred at the nucleus. The cationic results lead to an increase in the calculated MUE of about 1 kcal/mol for SM12 in ADF. When we compare COSMO to the work done by Marenich et al,[9] COSMO and the STO SM12 both have a dramatic increase in MUE of about 1 kcal/mol. The results are shown in Table 3.6, and solvation free energies and respective compounds are collected in Appendix 1.6. As can be seen the experimental list of ionic solvation free energies is extremely limited,

being less than a quarter of the amount of experimental data for neutral molecules for anionic and cationic systems individually, while being less than half the amount combined. This leads to a two-fold problem when modeling these more sensitive systems. First, there are simply not enough data points to properly parameterize any solvation model to be appropriately representative of the more diffuse space of anionic compounds, or even the cationic compounds. Second, ionic systems are extremely sensitive to solvation, as the solvation free energies are on average a magnitude of kcal/mol higher than the neutral compounds as can be seen by comparing the solvation free energies from Appendix 1.4 with Appendix 1.6.

Table 3.6: *MUE of ADF based calculations and experiment of 60 anions* and *MUE of ADF based calculations and experiment of 52 cations. Solvation energies are all calculated in water solvent. Calculations were done with optimizations from the Minnesota Solvent database, and their implementation into Gaussian 09. All energies are reported in kcal/mol; these results can also be seen in Appendix 1.6.*

Basis sets	SM12	SM12	SM12	COSMO	SM12	SM12
Functionals	ADF	ADF	ADF	ADF	Marenich et al.	Marenich et al.
	AUG-DZP	AUG-TZP	AUG-TZ2P	AUG-TZP	MG3S	6-31G(d)
Anionic						
B3LYP	4.1	4.0	4.0	2.8	3.7	3.1
M06-2X	3.8	3.7	3.8	2.4	3.5	2.9
Cationic						
B3LYP	4.6	4.6	4.6	4.4	3.5	3.0
M06-2X	4.6	4.5	4.6	4.5	3.5	2.9

3.6 Conclusions

Presented here is a new implementation of SM12 into the ADF suite of programs. In doing so, we extend SM12 and the underlying CM5 from the previous GTOs to STOs, and demonstrate basis set independence of both CM5 and SM12. Moreover, we present comparisons between the previous SM12 implementation of Marenich et al.[9], the current implementation of SM12, and the ASC COSMO already available within ADF. For neutral molecules and non-aqueous solvation energies, our results are within 0.1 kcal of the calculations put forth by Marenich et al.. This shows, furthermore, that the ADF STO-based SM12 results are comparable, and indeed on par, to the GTO SM12 implementation. For neutral aqueous and non-aqueous solvation calculations SM12 is the preferred choice in ADF and outperforms COSMO. For calculations using anionic compounds COSMO performs better than SM12, and cationic compounds result in equal performance for both SM12 and COSMO, possibly due to HPA being a relatively poor model of charged species. The implementation of SM12 into the periodic boundary condition part of the ADF suite, ADF-BAND, will be reported separately.

3.7 Acknowledgments

We would like to thank Stan van Gisbergen, Erik van Lenthe, Pier Philipsen, Fedor Goumans and all SCM for their continued support and help. We would also like to thank James Xidos for helpful discussion and insights. Finally, we thank Christopher Cramer and Donald Truhlar for providing us with the ASA code, the Minnesota Solvation Database, and their invaluable input and help. Funding from the Natural Sciences and

Engineering Research Council of Canada (NSERC, Discovery Grant) and from SCM Amsterdam (travel and in-kind support) are gratefully acknowledged.

3.8 Supporting Information (SI)

Further details on solvents that are implemented for SM12 in ADF (Appendix 1.1), the names and abbreviations of the compounds for CM5 calculations (Appendix 1.2), as well as the total partial charges for all compounds (Charge Model 5 Point Charges, see Supporting Information in text), the solvation free energy for calculations in water performed with M06-2X for COSMO and SM12 (Appendix 1.3), calculations performed with Hartree-Fock and DZP (Appendix 1.4), solvation free energies with DMSO, Benzyl Alcohol, Cyclohexane (Appendix 1.5), and ionic solvation free energies in water (Appendix 1.6).

3.9 References

(1) Schreckenbach, G.; Shamov, G. A. Theoretical Actinide Molecular Science. *Acc. Chem. Res.* **2010**, *43*, 19–29.

(2) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3093.

(3) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.

(4) Cramer, C.; Truhlar, D. A Universal Approach to Solvation Modeling SM8. *Acc. Chem. Res.* **2008**, *41*, 760–768.

(5) Miertus, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of Ab Initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* **1981**, *55*, 117–129.

(6) Cancès, E.; Mennucci, B.; Tomasi, J. A New Integral Equation Formalism for the Polarizable Continuum Model: Theoretical Background and Applications to Isotropic and Anisotropic Dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032.

(7) Chipman, D. M. Charge Penetration in Dielectric Models of Solvation. *J. Chem. Phys.* **1997**, *106*, 10194–10206.

(8) Klamt, A.; Schüürmann, G. 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* **1993**, *5*, 799–805.

(9) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620.

(10) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.

(11) Hirshfeld, F. L. Bonded-Atom Freagments for Describing Molecular Charge Densities. *Theor. Chim. Acta* **1977**, *44*, 129–138.

(12) Fonseca Guerra, C.; Handgraaf, J. W.; Baerends, E. J.; Bickelhaupt, F. M. Voronoi Deformation Density (VDD) Charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD Methods for Charge Analysis. *J. Comput. Chem.*

2004, *25*, 189–210.

(13) Baerends, E. J.; Ziegler, T.; Autschbach, J.; Bashford, D.; Bérces, A.; Bickelhaupt, F. M.; Bo, C.; Boerrigter, P. M.; Cavallo, L.; Chong, D. P.; Deng, L.; Dickson, R. M.; Ellis, D. E.; van Faassen, M.; Fan, L.; Fischer, T. H.; Fonseca Guerra, C.; Franchini, M.; Ghysels, A.; Giammona, A.; van Gisbergen, S. J. A.; Götz, A. W.; Groeneveld, J. A.; Gritsenko, O. V.; Grüning, M.; Gusarov, S.; Harris, F. E.; van den Hoek, P.; Jacob, C. R.; Jacobsen, H.; Jensen, L.; Kaminski, J. W.; van Kessel, G.; Kootstra, F.; Kovalenko, A.; Krykunov, M. V.; van Lenthe, E.; McCormack, D. A.; Michalak, A.; Mitoraj, M.; Morton, S. M.; Neugebauer, J.; Nicu, V. P.; Noodleman, L.; Osinga, V. P.; Patchkovskii, S.; Pavanello, M.; Philipsen, P. H. T.; Post, D.; Pye, C. C.; Ravenek, W.; Rodríguez, J. I.; Ros, P.; Schipper, P. R. T.; van Schoot, H.; Schreckenbach, G.; Seldenthuis, J. S.; Seth, M.; Snijders, J. G.; Solà, M.; Swart, M.; Swerhone, D.; te Velde, G.; Vernooijs, P.; Versluis, L.; Visscher, L.; Visser, O.; Wang, F.; Wesolowski, T. A.; van Wezenbeek, E. M.; Wiesenekker, G.; Wolff, S. K.; Woo, T. K.; A.L., Yakovlev. ADF. **2014**.

(14) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. a.; Snijders, J. G.; Ziegler, T.; Velde, G. T. E.; Guerra, C. F.; Gisbergen, S. J. V. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22*, 931–967.

(15) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. Towards an Order- N DFT Method. *Theor. Chem. Accounts Theory, Comput. Model. (Theoretica Chim. Acta)* **1998**, *99*, 391–403.

(16) Pye, C. C.; Ziegler, T. An Implementation of the Conductor-like Screening Model

of Solvation within the Amsterdam Density Functional Package. *Theor. Chem. Acc.* **1999**, *101*, 396–408.

(17) Pye, C. C.; Ziegler, T.; van Lenthe, E.; Louwen, J. N. An Implementation of the Conductor-like Screening Model of Solvation within the Amsterdam Density Functional Package — Part II. COSMO for Real Solvents 1. *Can. J. Chem.* **2009**, *87*, 790–797.

(18) Gusarov, S.; Ziegler, T.; Kovalenko, A. Self-Consistent Combination of the Three-Dimensional RISM Theory of Molecular Solvation with Analytical Gradients and the Amsterdam Density Functional Package. *J Phys Chem A* **2006**, *110*, 6083–6090.

(19) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Density Functional Solvation Model Based on CM2 Atomic Charges. *J. Chem. Phys.* **1998**, *109*, 9117–9133.

(20) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute-Water Clusters. *J. Chem. Theory Comput.* **2005**, 1133–1152.

(21) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. 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. *J. Phys. Chem. A* **2004**, *108*, 6532–6542.

(22) Lowdin, P.-O. On the Non-Orthogonality Problem Connected with the Use of

Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365–375.

(23) Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. More Reliable Partial Atomic Charges When Using Diffuse Basis Sets. *PhysChemComm* **2002**, *5*, 117–134.

(24) Zhu, T.; J., L.; G., H.; C., C.; D., T. Density Functional Solvation Model Based on CM2 Atomic Charges. *J. Chem Phys.* **1998**, *109*, 9117–9133.

(25) Davidson, E. R.; Chakravorty, S. A Test of the Hirshfeld Definition of Atomic Charges and Moments. *Theor. Chim. Acta* **1992**, *83*, 319–330.

(26) Marenich, A. V; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. Minnesota Solvation Database. *Minnesota Solvation Database - version 2012, Univ. Minnesota, Minneap.* **2012**, 1–28.

(27) Still, C.; Tempczyk, A.; Hawley, R. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(28) Born, M. Volume and Hydration of Ions. *Z. Phys.* **1920**, *1*, 45–48.

(29) Cramer, C. J. *Essentials of Computational Chemistry Theories and Models*; 2004; Vol. 2.

(30) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. Improved Methods for Semiempirical Solvation Models. *J. Comput. Chem.* **1995**, *16*, 422–440.

(31) Bondi, A. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1965**, *68*, 441–451.

(32) Becke, A. D. A Multicenter Numerical Integration Scheme for Polyatomic Molecules. *J. Chem. Phys.* **1988**, *88*, 2547.

(33) Franchini, M.; Philipsen, P. H. T.; Visscher, L. The Becke Fuzzy Cells Integration Scheme in the Amsterdam Density Functional Program Suite. *J. Comput. Chem.* **2013**, *34*, 1819–1827.

(34) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(35) Perdew, J.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(36) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(37) Becke, A. Density Functional Thermochemistry III The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(38) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V; Levchenko, S. V; O'Neill, D. P.; DiStasio, R. a; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A; Maslen,

P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Methods and Algorithms in a Modern Quantum Chemistry Program Package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

(39) Güell, M.; Lius, J. M.; Solà, M.; Swart, M. Importance of the Basis Set for the Spin-State Energetics of Iron Complexes. *J. Phys. Chem. A* **2008**, *112*, 6384–6391.

(40) Bultinck, P.; Ayers, P. W.; Fias, S.; Tiels, K.; Van Alsenoy, C. Uniqueness and Basis Set Dependence of Iterative Hirshfeld Charges. *Chem. Phys. Lett.* **2007**, *444*, 205–208.

(41) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235.

(42) Zhang, J.; Tuguldur, B.; van der Spoel, D. J. Correction Note. *J. Chem. Inf. Model* **2016**, 819–820.

3.10 Note Added In Proof

1. The original reference for COSMO-RS should have been cited, in addition to reference 17; See ref 41. 2. The comparison between COSMO and SM12, although based on the availability of solvation models within the ADF code proper, is not fully balanced: While SM12 parameterizes electrostatic and non-electrostatic terms of the solvation energy, COSMO contains only electrostatic terms, and it is its extension to COSMO-RS that also accounts for the non-electrostatic terms. Indeed, a recent benchmark study by Zhang et al. on a range of neutral organic solutes in organic solvents reports a root mean square deviation between calculated and experimental ΔG_{solv} of 2.3 $\pm$ 0.2 kj/mol for COSMO-RS as provided with the ADF suite of programs. This result is fully comparable to (slightly lower than) our SM12 MUE values in table 3.3 that were obtained for a wider benchmark set. Thus, the conclusions need to be amended such that, within the ADF suite of programs, the SM12 method incorporated in ADF and the freestanding COSMO-RS methods should yield comparable results.; see ref 42. The authors thank A. Klamt for drawing our attention to these points.

Preface to Chapter 4

This chapter is based on a manuscript to be submitted in the "*Journal of Chemical Theory and Computation*" The full title is:

Craig A. Peeples and Georg Schreckenbach, Pierre Herman Theodoor Philipsen "Periodic Solvation in ADF-BAND"

Solvation models have traditionally been applied only in molecular codes, and not much testing has been done with periodic systems. Recently, there has been an increase in interest in modeling solvation while evoking periodic boundary conditions (PBC). Within the Schreckenbach group, research is done on dye sensitized solar cells, which requires placement of an electrochemically active semi-conductor, which can be described as an infinitely periodic system, into an electrolyte (solvent). This is then the driving force behind the need for an accurate solvation method. There are a few concerns to consider when applying solvation to PBC, such as there is no experimental data on the quantitative effects of solvation energy for PBC compounds, therefore, great care must be taken when considering implementation, or the use of solvation in these systems. Solvation models create a SASA at a distance away from the surface making any network systems with large enough space in-between atoms become solvated, this becomes a cause for concern as it may create un-physical results. The following manuscript describes the theory, implementation, and overall goals of modeling periodic systems in liquid environments.

This manuscript is based on a collaborative effort between the Schreckenbach research group and the SCM (Scientific Computing & Modeling) company. This is the result of two separate works, the implementation of the COSMO solvation model into the

periodic boundary condition code, ADF-BAND, as done by Pierre Philipsen of SCM. The second half being the implementation of SM12 into the ADF-BAND code, as well. All results reported were conducted by Craig A. Peeples, and the manuscript was prepared together with Prof. Georg Schreckenbach and Dr. Pierre Philipsen.

Chapter 4: Periodic Solvation in ADF-BAND

Craig A. Peeples[a], Georg Schreckenbach[a]*, Pier Herman Theodoor Philipsen[b]

[a]Department of Chemistry, University of Manitoba

Winnipeg, MB, R3T 2N2, Canada

[b]Software for Chemistry & Modeling NV, Theoretical Chemistry, VU University Amsterdam, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands

schrecke@cc.umanitoba.ca, philipse@scm.com

4.1 Abstract

Periodic boundary condition solvation has been implemented into ADF-BAND and explored with two implicit solvation models: The Conductor like Screening Model (COSMO) [J. Chem. Soc. 1993, 5, 799-805], and Solvation Model 12 (SM12) [J. Chem. Theory. 2013, 9, 1, 609-620]. COSMO and SM12 are both approximate solutions to the Poisson equation of electrostatics, though they are approaching the approximations in vastly different ways. COSMO approximates the screening medium as a conductor, which allows analytical solutions to the Poisson equation. COSMO then calculates the surface charges of the molecule with small tesserae over the surface of each atom. SM12 makes use of the General Born Approximation, GBA, which approximates Poission's equation into simpler atomic functions, and makes use of Charge Model 5 to appropriately map partial atomic charges to each atom. BAND, much like its molecular counterpart ADF,

makes use of numerical Slater Type Orbitals (STOs) augmented with Numerical Atomic Orbitals (NAOs) as opposed to plane wave basis sets utilized in other periodic codes. BAND utilizes Bloch's theorem to incorporate these atomic basis sets into a periodic potential. Doing so allows BAND to calculate a fully periodic solid without having to worry about edge effects; these effects are however present in the Solvent Accessible Surface area of solvation and also have to be accounted for. Both solvation methods produce a potential of 1/r, therefore divergence at large distances is a concern and a screening function must be applied. Comparison between the molecular ADF code and a *zero*-dimensional cell is used to authenticate a single molecules transition to the BAND code. In addition, a carbon monoxide bonded on a single and double supercell copper bilayer is solvated and shown to have a lower energy upon solvation due to favourable dipole-dipole interactions.

4.2. Introduction

Solid-state chemistry and surface interactions are increasingly important parts of various areas of science, especially chemistry and materials science. For instance, Metal Organic Frameworks (MOF)s, and zeolites have been shown to be of great use in the uptake of greenhouse gases in the presence of solution.[1,2] Surface chemistry calculations are of great value, for instance towards understanding the feasibility of renewable energy sources.[3,4] Photovoltaics are based on absorbing photons from the sun to excite electrons to create a voltage across a solid-state compound such as dye sensitized,[5,6] silicon based,[7] or perovskite based solar cells.[8] Some solar cells require a liquid solvent in order to

properly stabilize them in various ways, as well as use solvents in their creation. Heterogeneous catalysis can also take place on the surface of metals and semiconductors that are in contact with a solution. There are ongoing research efforts to find cheaper metal combination for catalysis and more reliable ways to properly account for adsorption onto a metallic surface.[9]

In molecular calculations there has been many decades of research put into how to properly model the solvent-solute interaction, all while attempting to maintain control of calculation sizes and speeds.[10,11,12] Two prominent models have both been utilized and improved over the last decade, Klamt et al.'s Conductor like Screening model (COSMO)[13] and the newest in the series of SMx models, Solvation Model 12 (SM12),[14] put forth by Marenich et al. Both models have been shown repeatedly to reliably recreate experimental solvation energies,[15] while approaching the problem with different methodologies. We feel the inclusion of solvation into solid-state based codes will help further progress the research on renewable energy, among many other areas of surface science.

In the literature, there are already some examples of implementations of periodic boundary condition based solvation. This was originally proposed by Delley in the form of DMol[3] COSMO[16] for the calculation of infinite polymers, surfaces and cavities in 3D structures.[17] Others include the self-consistent continuum solvation (SCCS) method of Andreussi et al.[18], for use with plane waves. More recently the multigrid based continuum solvation model of Garcia-Ratés et al.[19] was implemented into the plane-wave pseudopotential based VASP code.[20,21,22,23] Compounds such as MOFs and Zeolites are of

particular use in the ADF-BAND code[24,25,26] as their cavities do not require the plane-waves associated with most periodic boundary condition (PBC) codes. Here we provide a detailed description of our adaptation from ADF to the Numerical Slater Orbital based PBC ADF-BAND of the aforementioned continuum solvation methods COSMO[27] and SM12.[28]

4.3. Theory

ADF-BAND[25] is a solid-state DFT code that is part of the Amsterdam Density Functional Suite of programs, ADF[24,29]. BAND makes use of periodic boundary conditions in order to model periodic 1D and 2D systems, as well as 3D solid compounds. BAND, much like ADF, makes use of numerical Slater Type Orbitals (STOs). However, for BAND the STOs are augmented with Numerical Atomic Orbitals (NAOs) as opposed to plane wave basis sets utilized in other periodic codes. The radial function of the NAO is numerically calculated and is exact for a spherical atom. BAND takes advantage of the correct shape of the atomic orbitals provided by STOs and improves the NAOs. NAOs with high quality basis sets will reproduce the same results as seen from STOs, but when considering smaller basis sets the NAOs are only useful for describing the core orbitals, therefore, we require the use of STOs in order to approximate the bonding region of atoms with smaller basis sets.[25] This provides a few advantages over plane wave codes that are realized, in particular, for systems with periodic cavities. In these cases, plane wave codes require all space to be expressed with basis functions while the atom-centered STO augmented NAOs allow these cavities to essentially be left empty.

Expanding from a molecular to a periodic code the system being modeled needs to be extended into further dimensions, which can be described as periodic images. This includes periodic extensions along the x-axis as a 1D polymer, the x- and y-axes as a 2D surface (slab), or in three spatial dimensions as a 3D solid structure. There needs to be careful considerations on how to handle the basis sets and structure. Simply extending the basis sets, in the same fashion as would be done for a molecular code will result in several problems. Molecular codes cannot properly model infinite systems as they require a physical limit to the molecules. This will lead to strange edge effects that are not physical within solids or surfaces. In addition, the potential will not behave correctly when approaching infinity. To properly combat this problem moving from ADF to ADF-BAND we make use of Bloch's theorem,[30] and a long-range screening function detailed later. Bloch's theorem states that if we have a solid that is periodic at the atomic scale then we get a traveling wave solution for our wave function that reflects the symmetry of the lattice. Bloch's theorem also states that as we move from point r in each periodic image we get the same value for the wave function (apart from a phase factor) at the same point in the next periodic image at point r+R, equation 4.1. There are a few problems that arise when we model potentials in a periodic system; the main concern within the context of this paper is edge effects. Block's theorem essentially ignores edge effects making the model treat the system as infinite.

$$\psi_k(r + R) = e^{i\vec{k}\cdot R}\psi_k(r) \quad (4.1)$$

When simulating real-world solids we can quickly think of situations that call for liquid being a part of our model, such as when binding a dye to the TiO_2 surface of dye sensitized solar cells, nanotubes, Metal Organic Framework (MOF), and even something as simple as dissolving solid NaCl into an aqueous solution. In molecular codes, there have been two principal ways for investigating this phenomenon. The first approach is explicit solvation where we physically incorporate the solvent molecules around the compound being studied; for surfaces this would require covering the surface with explicitly included solvent molecules. Explicit solvation quickly becomes a problem, as we increase the amount of atoms inserted into the model we increase number of basis functions, and in turn computational time and cost. This makes explicit solvation of limited use, especially within periodic boundary conditions. As the periodic images become more and more populated with increasing numbers of solvent molecules, they also need to be copied to neighbouring images, and quickly the system size becomes unreasonable and we run into a vastly larger problem than in a molecular code. The second way to handle the effects of solution is to use implicit solvation. [10, 11,12,13,14] Within ADF, there are several models available for solvation. Of these, COSMO[13,27] solvation and Solvation Model 12 (SM12)[14,28] are of the simplest form and suitable for a parallel implementation into the BAND program.

4.3.1 Continuum Solvation

Continuum solvation models are constructed of two components:

1) How the solvent interacts with the molecule, solvent effects;

2) How the molecule feels the effects of the solvent, polarization.

These two components are modeled in two different ways within ADF and now BAND through COSMO and SM12. The effect of electrostatics is comprised of how the solvent orientates itself to incorporate the polarization associated with an inserted charge distribution into the medium, and how the solute in turn is affected by the energy cost of the inserted charge. This interaction, in liquid, is solved with Poisson's equation.

$$-\vec{\nabla}\left[\epsilon(\vec{r})\,\nabla\,\mathrm{V}(\vec{r})\right] = 4\pi\rho_M(\vec{r}) \quad (4.2)$$

The Poisson equation expresses the potential, $V(\vec{r})$, within a cavity as a function of the molecular density, $\rho_M(\vec{r})$, as well as expressing the solvent outside the cavity as a function of its permittivity, $\epsilon(\vec{r})$. This then leads to a loss of information about the structure of the solvent, as we approximate the solvent as a whole with the solvents permittivity. The solvation models implemented solve this problem iteratively through a potential over the surface of the molecule defined by,

$$V_\sigma(\vec{r}) = \int_\Gamma \frac{\sigma(\vec{s})}{|\vec{r}-\vec{s}|} d^2s \quad (4.3)$$

Where we have the potential, $V_\sigma(\vec{r})$, due to the charge, $\sigma(\vec{s})$, over the cavity surface, Γ.

Both COSMO and SM12 take advantage of basic electrostatics. The Coulomb energy can described as two charges, q, separated at a distance, in atomic units:

$$E = \frac{q_1 q_2}{r_{12}} = \frac{1}{2}\sum q_i V_i \quad (4.4)$$

Then the potential due to a charge j at position i is

$$V_j(r_i) = \frac{q_j}{r_{ij}} \quad (4.5)$$

Which then allows the calculation of equation 4.3 to be related to the energy of the solvated system.

4.3.2 COSMO

COSMO makes use of changing the dielectric constant of the liquid medium to infinity in order to reproduce the effects of a conductor.[13,27] This greatly changes the effects at the limits of the Poisson equation. Namely, at the edges of the cavity surface the total potential from the polarization is cancelled out.[12] This then allows the surface charges to be calculated from the local potential on each surface. In solution, there needs to be an appropriate physical scaling for the polarization, and no solvent permittivity is in fact, infinity, therefore to recover the effects of the finite dielectric constant a scaling factor is required. This is done through an ideal unscreened charge density that is scaled by a proper function of the permittivity.

$$\sigma(s) = f(\epsilon)\sigma^*(s) \quad (4.6)$$

Where $\sigma(s)$ is the screened charge, $f(\epsilon)$, is the scaling function and $\sigma^*(s)$ is the unscreened charge. Each component is associated with the surface of the cavity, s. This

scaling function has to be empirically obtained. Klamt et al. suggest the following formula based on the dielectric energy screening for a given geometry: [13]

$$f(\epsilon) = \frac{\epsilon - 1}{\epsilon + k} \quad (4.7)$$

With a value of k=0.5. Within the ADF implementation Pye et al. suggested a value of k= 0.0 to accommodate Gauss' law.[27]

We need to define the surface of each atom, and show how COSMO handles the induced dipole interactions. COSMO segments the entire surface into small closed portions of area S_μ called tesserae. The closed surface is defined by the S_μ with constant surface charge density $\sigma(r_\mu)$. This can be used to express the atomic charge, q_μ, as a function of the screened charge over each tesserae, S_μ.

$$q_\mu = |S_\mu|\sigma_\mu \quad (4.8)$$

Using equation 4.8 in tandem with equation 4.4, we can express the electrostatic solvating energy as two portions.

$$E_S = E_{m\sigma} + E_{\sigma\sigma} \quad (4.9)$$

The first term ($E_{m\sigma}$) represents the interaction between the molecular density (nuclei and electron density) $\rho_m(r)$, and the surface charge $\sigma(r_S)$. This interaction can be further expressed as in equation 4.10. This requires an integral over the volume associated with the molecule, V, and a surface integral, over s, in order to incorporate each charge over its associated surface tesserae.

$$E_{m\sigma} = \int_V \int_S \frac{\rho_m(r)\sigma(r_S)}{|r - r_S|} dr dr_s \quad (4.10)$$

We want to be able to express the Coulomb potential from equation 4.10 at each point, r_μ. Therefore, if we use

$$C_\mu = \int \rho_m(r) v_\mu(r) dr \quad (4.11)$$

Along with

$$v_\mu(r) = \frac{1}{|r_\mu - r|} \quad (4.12)$$

Thus, the volume integral is equivalent to $V_C[\rho_m](r_s)$, and consequently we obtain equation 4.13 as a volume integral over these two parts.

$$V_C[\rho_m](r_S) = \int \rho_m(r) v_\sigma(r) dr \quad (4.13)$$

An advantage of this approach is that it does not require the fitted density used in ADF and ADF-BAND, and hence is not affected by fit errors.[24,25,31,32] A disadvantage is that the standard integration grid cannot formally integrate the cusps of the surface potential at the positions r_μ.

The second term of equation 4.9, ($E_{\sigma\sigma}$), represents the interaction of the surface charge distribution with itself. This requires a double surface integral to incorporate the surface of each charged and interacting tesserae, which is COSMO`s expression of equation 4.4.

$$E_{\sigma\sigma} = \frac{1}{2}\int_S \int_{S'} \frac{\sigma(r_S)\sigma(r_{S'})}{|r_S - r_{S'}|} dr_S dr_{S'} \quad (4.14)$$

In order to express the energy we approximate the integrals as finite sums. Therefore, in practice,

$$\int_S \sigma(r_S) f(r_S) d_{r_S} \approx \sum_\mu S_\mu \, \sigma(r_\mu) f(r_\mu) \quad (4.15)$$

Where μ is our associated tesserae, $f(r_\mu)$ is a general function of r_μ, S_μ is the surface area of tessera, and $\sigma(r_\mu)$ is the charge. Then from 4.15, and using equation 4.8, we get

$$\sum_\mu S_\mu \, \sigma(r_\mu) f(r_\mu) = \sum_\mu q_\mu f(r_\mu) \quad (4.16)$$

Furthermore, from equation 4.14, we get the simplified charge distribution energy associated with interacting surface tesserae

$$E_{\sigma\sigma} = \frac{1}{2}\sum_s \sum_{s'} q_s A_{s,s'} q_{s'} \quad (4.17)$$

$$A_{\mu\nu} = \frac{1}{r_{\mu\nu}} \; for \; \mu \neq \nu \quad (4.18)$$

$$A_{\mu\mu} = 1.07 \sqrt{\frac{4\pi}{S_\mu}} \quad (4.19) \; [27]$$

We can now write equation 4.10 in a simplified form that fits the energy equation, 4.9, and the molecule-surface interaction is calculated as

$$E_{m\sigma} = \vec{q} \cdot \vec{c} \quad (4.20)$$

With

$$C_{\mu} = V_c[\rho_m](r_{\mu}) \quad (4.21)$$

Finally, we can express equation 4.9 as a simplified formulation.

$$E_S = \vec{q} \cdot \vec{c} + \frac{1}{2}\vec{q}^T A\vec{q} \quad (4.22)$$

Solution to the COSMO equations. Next we have to take special care of the charges as the screening charge distribution is formulated in such a way that it minimizes the total energy. Therefore we search for optimal charges

$$\nabla_q E_S = 0 \quad (4.23)$$

When applied to equation 4.22, this results in

$$A\vec{q} = -\vec{C} \quad (4.24)$$

So that we can now calculate the energy at the optimum charges as

$$E_S[q = opt] = q \cdot C + \frac{1}{2} q \cdot Aq \quad (4.25)$$

$$= q \cdot C + \frac{1}{2} q \cdot (-c) \quad (4.26)$$

$$= \frac{1}{2} q \cdot c$$

$$E_S[q = opt] = -\frac{1}{2} q \cdot Aq \quad (4.27)$$

Generally, we want to constrain the surface charge to integrate to minus the charge of the molecule. In this way we can avoid the charge from being outside the cavity limits, and avoid the problem of outlying charge.[33] Therefore, we consider the constraint term

$$L^{constr}(\lambda, q) = \lambda \left(\sum q_i - Q_{mol} \right) \quad (4.28)$$

Making use of the Lagrangian, we can correct the outlaying charge

$$\nabla_q L = \lambda \vec{1} \quad (4.29)$$

Where $\vec{1} = [11 \ldots 1]^T$. This leads to the modified COSMO equation

$$A\vec{q}_{constr} = -\vec{c} - \lambda \vec{1} \quad (4.30)$$

The constraint is used to correct the charges through

$$\vec{q}_{constr} = \vec{q}_{un-con} + \lambda \vec{q}_{\lambda} \quad (4.31)$$

Plugging that into the variation equation, we get

$$A(\vec{q}_{un-con} + \lambda \vec{q}_{\lambda}) = -\vec{c} - \lambda \vec{1} \quad (4.32)$$

Which is satisfied if

$$A\vec{q}_{un-con} = -\vec{c} \quad (4.33)$$

Leading to,

$$A\vec{q}_\lambda = \vec{1} \quad (4.34)$$

and λ itself follows the from the constraint that the charges to sum to $-Q_{mol}$

$$Q_{uncon} + \lambda Q_\lambda = -Q_{mol} \quad (4.35)$$

$$\lambda = \frac{-Q_{mol} - Q_{uncon}}{Q_\lambda} \quad (4.36)$$

4.3.3 SM12

SM12 takes a very different approach to exploring the energy of solvation than COSMO.[14,28] SM12 solves the polarization problem, equation 4.3, with the Generalized Born Equation (GBA). This approximates the aforementioned Poisson equation. The GBAs expression of equation 4.4 can be seen in equation 4.38. Here we have the charge, q, the Coulomb integral between the two charges γ_{ij}, and the permittivity of the solvent, ϵ. In SM12, the electrostatics of polarization, G_P, consists of

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) \sum_{ij}^{atoms} q_i \gamma_{ij} q_j \quad (4.38)$$

Which involves two parts, the first being Charge Model 5 (CM5)[34]

$$q_i^{CM5} = q_i^{HPA} + \sum_{j \neq i} T_{ij} B_{ij} \quad (4.39)$$

CM5 is a class IV charge model. It utilizes Hirshfeld Charges[35] (q_i^{HPA}) to map the density, coupled with an empirical correction that helps improve the accuracy of describing partial

charges. The empirical portion is based on the so-called Pauling bond order B_{ij} and a purely empirical portion T_{ij}.[34]

The second electrostatics contribution is Still et al.'s approximation of the Coulomb integral γ_{ij}.[36]

$$\gamma_{ij} = \left(r_{ij}^2 + \alpha_i \alpha_j e^{\left(-r_{ij}^2 / d\alpha_i \alpha_j\right)} \right)^{-1/2} \quad (4.40)$$

This approximation allows us to take advantage of the electrostatics as a function of the atomic distance, and the Born area.[37]

In comparison with COSMO, there is a different expression that is used for the charges calculated within SM12. This is done through a combination of Still et al.'s[36] approximation to the Coulomb integral, as seen in equation 4.40, and the use of CM5, equation 4.39. The most important difference, for electrostatics, between the solvation models is the fact that SM12 uses point charges centred on the nucleus to describe the charge attributed to the entire atom (GBA). COSMO, on the other hand, separates the charges over the entire surface of each atom, utilizing the tesserae, to more thoroughly map the charge differences along the surface of each atom.

Within SM12 the energy calculation for each atom requires both, the Born surface area of the atom and the interatomic distance between neighbouring atoms.

$$\alpha_i = \left(\frac{1}{R'} + \int_{\rho_{z_i}}^{R'} \frac{A(r)}{4\pi r^2} dr \right)^{-1} \quad (4.41)$$

Here, R' is defined as a pro-sphere encompassing the other spheres of the molecule, ρ_{z_i} is defined as the Coulombic atomic radius of z, and $A(r)$ is calculated through Liotard, Hawkins and et al.'s analysis of spherical intersecting atoms, better known as the Analytical Surface Area (ASA).[38] Using the expression in Eq. 4.40, the GBA through equations 4.38, 4.39, and 4.41 can be used to calculate the energy associated to the polarization, represented as G_P in equation 4.38.[14]

SM12[14] handles the solvent in a very thorough way, using various experimental observables to empirically handle the solvent-molecule and solvent-atomic interactions. This is done through the nominally termed G_{CDS}, which stands for Cavitation, Dispersion and Solvent. This term is put in place to handle the subtle details of the solvent, in order to properly describe the effects of real-world solvation.

$$G_{CDS} = \sum_{i}^{Atoms} \left(\sigma_i + \sigma^{[M]}\right) A_i \quad (4.42)$$

Where we have the atomistic σ_i term, known as the atomic surface tension (AST), which describes how each atom interacts with its neighbours. AST has specific parameters for the majority of the main group elements while introducing a generic term for each element not fully characterized. AST consists of an atom-atom portion and an atom-solvent portion, equations 4.43 and 4.44, respectively.

$$\sigma_i = \widetilde{\sigma_i} + \sum_{j}^{atoms} \sigma_{Z_i Z_j} T_i\left(\{Z_i, R_{ij}\}\right) \quad (4.43)$$

$$\tilde{\sigma}_i = \tilde{\sigma}_i^n n + \tilde{\sigma}_i^\alpha \alpha + \tilde{\sigma}_i^\beta \beta \quad (4.44)$$

The ATS is also associated with a cut-off function, T_k. It ensures that the interaction of each neighbouring atom is acting only within a small sphere of influence. This is known as the Cut Off Tan function, which is fully explained in the SI of the paper by Marenich et al.[14] The parameters within equation 4.44 are all solvent specific and include n, the refractive index, as well as α and β that are the Abrahams acidity and basicity parameters, respectively.

The second term in equation 4.42, $\sigma^{[M]}$, approximates how the solvent interacts with the molecule as a whole. This term is known as the molecular surface tension (MST). MST is dependent on which flavour of the SMx and CMx models are chosen, as well as more solvent-based parameters. For the purpose of our implementation we used SM12 with CM5, but this is left open for straightforward adaptation to other, new SMx and CMx models in the future. Lastly, A_i is based on the solvent accessible surface area (SASA), which is calculated through the ASA.[38]

4.3.4 The Potential

The COSMO charges affect the density from the potential of the surface, expressed as

$$V^\sigma(r) = \sum_\mu \frac{q_\mu}{|r - r_\mu|} = \sum_\mu q_\mu v_\mu \quad (4.45)$$

While, SM12 affects the density via the change in the polarization term, G_P, with respect to the density on every SCF cycle.[39]

$$F_{\mu,\nu}^{Mat} = F_{\mu,\nu}^{Mat^\circ} + \frac{\partial G_P}{\partial \rho_{\mu,\nu}} \quad (4.46)$$

These are added to the potential in the SCF, independently, for each chosen solvation method.

4.3.5 Periodic Contribution to Electrostatics

We will now move on to the periodic case, and the consequences of the infinite surface. For COSMO to calculate the A_{ij} matrix, equation 4.18, we need to use a screening function, *h*, to cut off the long ranged diverging *1/r* term. This is done because the sum of *1/r* diverges as the sum goes towards infinity.[25]

$$\sum_{r=1}^{\infty} \frac{1}{r} = \infty \quad (4.47)$$

$$A_{ij} = \sum_{R} \frac{h\left(\left|r_i - r_j + R\right|\right)}{\left|r_i - r_j + R\right|} \quad (4.48)$$

$$A_{ii} = 1.07\sqrt{\frac{4\pi}{S_\mu}} + \sum_{R\neq 0} \frac{h(|R|)}{|R|} \quad (4.49)$$

When evaluating C_μ through Eq. (4.11) we get

$$C_\mu = \int \rho(r) \sum_R \frac{h(|r_i - r_j + R|)}{|r_i - r_j + R|} \quad (4.50)$$

Using the same substitutions as seen in (4.11), (4.12) and (4.13) the Coulomb potential becomes

$$V_C[\rho_m](r_S) = \int \rho(r) v_\mu(r) \quad (4.51)$$

With

$$v_\mu(r) \equiv \sum_R \frac{h(|r_i - r_j + R|)}{|r_i - r_j + R|} \quad (4.52)$$

Here, v_μ is the (screened) potential of a unit charge at a COSMO point r_μ. In practice, the screening function is limited for a very small $|r - r_\mu|$ and will not produce a significant difference for these distances.

Within SM12 we have to consider a similar situation. The approximation to the Coulomb potential exhibits the same divergent characteristic, *1/r*, as COSMO at long ranges. Therefore, we apply the same screening function to the formulation, equation 4.40, of Still et al..[36]

$$\gamma_{ij} \equiv \left[\left(h\left(r_i, r_j, R\right) \right)^2 + \alpha_i \alpha_j e^{\left(- \frac{h\left(r_i, r_j, R\right)^2}{d \alpha_i \alpha_j} \right)} \right]^{-1/2} \quad (4.53a)$$

$$h\left(r_i, r_j, R\right) = \sum_R \frac{h\left(\left| r_i - r_j + R \right|\right)}{\left| r_i - r_j + R \right|} (4.53b)$$

The screening function is then utilized to force the potential to behave correctly at extreme distances, as infinities are not physical. The screening essentially takes the distance matrix at the farthest separated atoms and forces them further apart, so as to not have any interaction anymore.

4.3.6 Overall Energy of Solvation

Taking together all the portions discussed above, we can fully calculate the total energy associated with solvation.

For COSMO, we have, from equation 4.9

$$E_S = E_{\sigma\sigma} + E_{m\sigma} \ (4.9)$$

From equation 4.14

$$E_{\sigma\sigma} = \frac{1}{2} \int_S \int_{S'} \frac{\sigma(r_S)\sigma(r_{S'})}{|r_S - r_{S'}|} dr_S dr_{S'} \ (4.14)$$

$$E_{\sigma\sigma} = \frac{1}{2} q \cdot Aq \text{ (4.27)}$$

From equation 4.10,

$$E_{m\sigma} = \int_S V[\rho_m](r_S)\sigma(r_S)dr_S \text{ (4.10)}$$

$$E_{m\sigma} = \vec{q} \cdot \vec{c} \text{ (4.20)}$$

In summary, we get,

$$E_S = \vec{q} \cdot \vec{c} + \frac{1}{2} q \cdot Aq \text{ (4.54)}$$

Likewise, for SM12, we have:

$$\Delta G_S^{\otimes} = \Delta E_E + \Delta G_N + G_P + G_{CDS} + \Delta G_{conc.}^{\otimes} \text{ (4.55)}$$

Where ΔG_N is the free energy contribution from the change in geometry as a consequence of being solvated; for a single-point calculation this term is 0 kcal/mol. Further, $\Delta G_{conc.}^{\otimes}$ is the change in concentration when moving from a molecule in a vacuum to the condensed phase, this is normally 0 kcal/mol, as well, or the term can be added as an explicit correction.[40.41] Finally, ΔE_E is the change in molecular energy when going from the gas phase to the solution phase as represented by,

$$\Delta G_{ENP} = \langle \psi^{(1)} | \hat{H} + V | \psi^{(1)} \rangle - \langle \psi^{(0)} | \hat{H} | \psi^{(0)} \rangle \text{ (4.56)}$$

Where (1) is the condensed phase, and (0) is the gas phase. Within COSMO and SM12, V is included with the addition to the potential within the SCF cycles.

4.4 Implementation

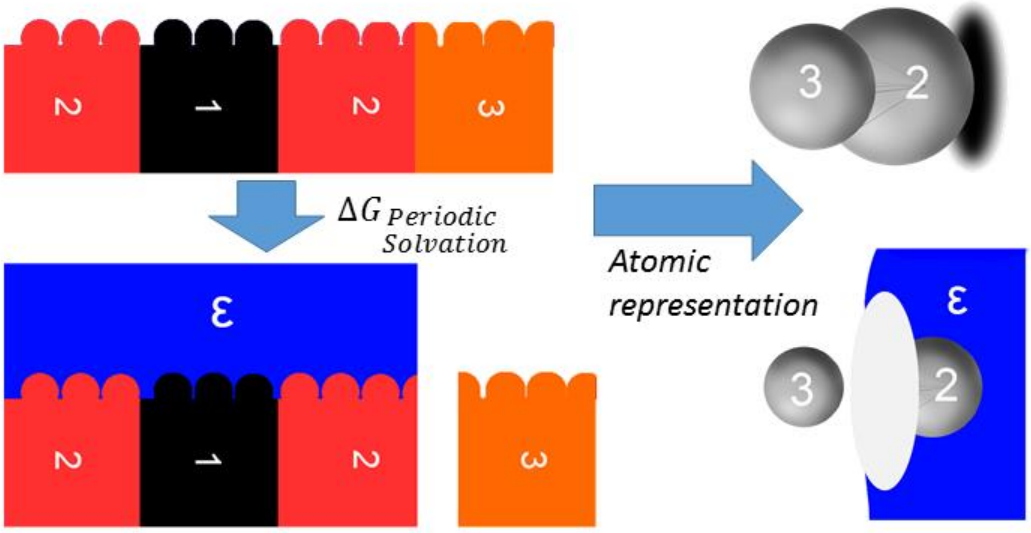
In moving from ADF to BAND there is a change in how the basis sets are handled. In BAND, the ADF Slater Type Orbitals (STO) are now augmented by using numerical type orbitals (NAOs). These tend to produce a difference in total bonding energy between ADF and BAND when using smaller basis sets. On the other hand, the TZP basis set is large, and accurate enough between the ADF and ADF-BAND codes that it should result in nearly identical energies. Therefore, the crucial test of the solvation model implementation is built on the idea that the energy of a Triple Zeta Polarized (TZP) basis set calculation in a *zero*-dimensional crystal in ADF-BAND should exactly replicate that of a TZP single-point calculation of the same compound in ADF, within numerical noise. Once we have the same energies for calculations on the 0-dimensional (0D) crystal, we need to carefully consider the consequences of moving from a 0D system to its 1D, 2D, and 3D counterparts.

4.4.1 Edge Effects on Solvation Accessible Surface Area (SASA)

Within BAND the first 3 periodic images, or replicas of the unit cell, are fully coordinated by default. This creates a problem for the two solvation models, as they both calculate their potentials in an atom centric fashion, COSMO using the numerical points of integration on the cavity surfaces and SM12 using the nuclear coordinates. Calculating the

Solvent Accessible Surface Area for both models requires calculation of the physical space taken up by each atom. When the calculation contains a periodic image that does not have fully coordinated atoms, the models assume there is none there, and will consider the edge as the full sphere of the outermost coordinated atom. The problem then is that this will result in an energy associated with a larger surface area than what is physical, or in worst-case scenarios we end up solvating a 2D slab as if it was a singular molecule, or for the 3D case a 3D cube (cluster) of a few unit cells.

Figure 4.1:. *Schematic representation of the split of the surface area between each image, in order to account for SASA edge effects. The solvent energy is only accounted for in the images 1 & 2, while not being applied to the third.*



Thus, we need to take into consideration which atoms are fully coordinated. This is done by fully coordinating the first 5 periodic dimensions of the unit cell for SM12, and 3 dimensions for COSMO solvation, and keeping track of them in our SASA calculation. For 1D this is a simple problem as BAND coordinates in a 4-5-1-2-3 fashion, 1 being the

primitive image. In a periodic solvation calculation, we should have image 1 being the only unique image in terms of SASA and energy, followed by image 2 being equal to image 5, and image 3 being equal to image 4. For clarity, we will call image 1 the α-image, images 2 and 5 the β-images, and image 3 and 4 the γ-images.

In order to get a proper representation of the SASA a calculation of the SASA of all 5 images is required. We need to cut off the surface area and energy associated with the outermost γ-images, as they are the last fully coordinated images and will have the aforementioned edge effects. Doing the calculation in this way leads to the γ-images contributing to the effects associated with the β-images, whereas the edge atoms associated with the γ-images will not be taken into consideration; yet, the β-images will still feel the γ-images as if they were present. This creates an approximation, which simply cannot be avoided without the inclusion of an infinite slab of fully coordinated atoms. A graphical representation of the separation between the β- γ images is represented in figure 4.1.

In practice doing the calculation in this way leads to a slightly lower energy for the α-image, while a slight increase in energy is seen for the β-images, and of course no energy is associated with the γ-images, by definition. The energies associated with the β-images should be exactly the same in a perfectly periodic calculation. The reason that there is some discrepancy between the α-image and the β-images is due to the α-image being close enough to feel the effects of all the β and γ images, while the β-images interaction with the γ-image on the opposite side of the slab is dampened through the screening function. Thus,

an average is taken over the SASA and energy associated with the α and β-images, and used as the SASA and energy.

4.5 Computational Details

All ADF calculations are done with a TZP basis set using the functional PBE.[42] PBE was chosen due to its cross implementation in both ADF and BAND, and will help minimize differences. TZP is chosen for the reasons stated earlier. In ADF and BAND, SM12 calculations were done with the keyword 'Solvation SM12' and 'SolvationSM12', respectively, with no additional keywords. In COSMO calculations with ADF and BAND a solvent radius of 1.4 was used with cav0 = 1.321, cav1 = 0.0067639. For each calculation, we utilized the Delley surface with triangles sized with ndiv = 4 and an overlap of 0.8, vibrational SCF always on, and we calculate the exact C-matrix (c-mat). All calculations were performed with a locally modified version of ADF2014, and as single point calculations.[28] The default solvent for both models is water. The validity of the SM12 and COSMO models in the *zero*-dimensional lattice were tested using the gas phase geometry optimizations present in the Minnesota solvation database.[43] The 2D systems were optimized in gas-phase with TZP at the LDA level within ADF-BAND. All calculations were done without symmetry, as it is not currently implemented for SM12 within ADF-BAND.

4.6 Results and Discussion

Both solvation models (COSMO and SM12) have been fully implemented and tested within BAND. To test the validity of the *zero*-dimensional crystal in comparison between ADF and BAND, we ran several calculations four times, one for each solvation model in ADF and then in BAND. The compounds chosen covered a broad spectrum chemically, and were chosen to exploit and test the significant components of the solvation models, polarization vs. cavitation. Each compound was run in ADF, followed by BAND using no lattice. The compound after each calculation provided the same electrostatic energy within numerical noise. A summary of a few representative results is seen in table 4.1; all other results from calculations can be seen in table A 2.1.

Table 4.1: *Periodic zero-dimensional solvation electrostatics: representative examples of the energy of solvation (kcal/mol). Calculations were done on optimized gas-phase geometries from the Minnesota Solvent database, and their implementation into Gaussian 09*[43,14] .

Model	*Ethane*	*Pyrrolidine*	*Bensulfuron*
	ADF		
COSMO	-0.36	-5.18	-36.46
SM12	-0.50	-2.59	-19.55
	ADF-BAND		
COSMO	-0.30	-4.96	-35.91
SM12	-0.50	-2.59	-19.40

A new problem arises for periodic systems, as there are no real ways of calculating the experimental energy required for rigorous solvation testing, such as done for SM12[14,28] for molecular systems, for periodic cells beyond the *zero*-dimension. The system chosen to demonstrate the performance of these implementations is the adsorption of molecular CO on a bilayer of a copper surface in a 1x1 super cell. In order to test this we performed separate calculations run in different environments, one in gas-phase, a second and third in condensed phase for both COSMO and SM12, and then we repeat each of these calculations for a super cell of copper four times the size of the original, i.e. 2x2. What we want to see from this is that the bonding energy for the 2x2 system is four times the bond energy of the 1x1 super cell. Essentially, we performed six computations:

1: Solvation Method = None, Super Cell = 1x1

2: Solvation Method = None, Super Cell = 2x2

3: Solvation Method = COSMO, Super Cell = 1x1

4: Solvation Method = COSMO, Super Cell = 2x2

5: Solvation Method = SM12, Super Cell = 1x1

6: Solvation Method = SM12, Super Cell = 2x2

From calculation 1 and 2 we learn if the ADF-BAND calculation is numerically accurate enough. From calculation 3 and 4 we learn whether COSMO works, and, similarly, from 5 and 6 we learn whether SM12 is correct. Combined with the fact that COSMO and SM12

both work for molecules, we can conclude that they are accurate for the higher dimensional crystal lattices as well. These results are seen in table 4.2.

Table 4.2: *Total bonding energy and energy of solvation. All energies are reported in kcal/mol. These calculations were based on gas phase optimizations, done within ADF-BAND at the LDA/TZP level of theory, of a bilayer of Cu with a CO layer adsorbed to the top.*

Total Bonding Energy of CO on Cu Bilayer

Cell Size	1x1	2x2	Ratio
Gas	-550.0	-2200	4.000
COSMO	-551.0	-2203	3.998
SM12	-553.0	-2210	3.996

Total Solvation Energy of CO on Cu Bilayer

COSMO	-3.090	-15.50	5.016
SM12	-2.300	-10.60	4.608

Table 4.2 demonstrates that we have a consistent bonding energy going from the 1x1 lattice to the 2x2 super cell. As the size of the supercell is increased by four times the bonding energy increases by the same increment. This shows that we have taken care of the *1/r* problem of equation 4.47, and that the boundary conditions are well behaved. Therefore, it can be concluded that our calculations and implementations are, at the very least, consistent within the context of this paper. Furthermore, an analysis of the total solvation energy associated with each solvation model is presented in table 4.2. The solvation energy ratios are higher than those of the total bonding energies. This is most likely a side effect of the SASA edge approximations. COSMO only approximates the unit

cell to 3 images for the calculation of the surface, while SM12 extends it to 5 surrounding images. From table 4.2 we can then see a direct correlation between the solvation energy ratios and the size of the system used to approximate it.

Next, to validate further, it was decided to investigate the solvation effect of the adsorption of the carbon monoxide to the copper bilayer, as shown schematically in figure 4.2. In theory, the addition of a solvent should result in an increase of the polarization of both the CO and Cu bilayer as well as increase with the polarization of the solvent involved. This should cause a favourable dipole-dipole interaction, and for the solvated systems we should see the energy of the system become lower. The results can be seen in table 4.3.

Figure 4.2. *Reaction cycle of solvating the adsorption of CO onto a Cu bilayer.*

Table 4.3: *Energy of adsorption and energy of solvation for adsorption. All energies are reported in kcal/mol. These calculations were based on gas-phase optimizations, done within ADF-BAND at the LDA/TZP level of theory, of a bilayer of CO on Cu.*

Energy of Solvation for Adsorption

Compound	Cu	CO	Cu + CO	Adsorption
COSMO	-0.07	-0.19	-3.09	-2.83
SM12	-0.02	-0.53	-2.30	-1.75
	Total Bonding Energy for Adsorption			
Gas	-170	-351	-550	-28.8
COSMO	-170	-352	-551	-29.9
SM12	-170	-352	-553	-30.6

The results show that we have an energy decrease (increased stabilization) of 0.4 kcal/mol, and 1.8 kcal/mol for COSMO and SM12, respectively. As discussed above, this stabilization is expected to happen when we solvate a system. We can further look at the breakdown of the energies associated with each model in table 4.3. What is of note here is that CO with SM12 produces a much lower energy than that of COSMO by 0.34 kcal/mol. When looking at solvation of the metal Copper, the lower energy for Cu can be attributed to COSMOs tesserae description[13] of the Cu surface, better describing it, while SM12 has only limited parameterization for metal atoms.[14] The opposite can be seen for the CO, as

SM12 is specially parameterized for main group elements, while COSMO only describes them in the same general way. Another interesting note is that the solvation energies for SM12 are actually higher than those of COSMO in table 4.3, but there is a lower total energy associated with the adsorption study. This essentially means that SM12 has induced a dipole within the reactants, products, and solvent stronger than that of COSMO.

4.7 Conclusion

Presented here is an adaptation of two of the solvation models available within ADF to its periodic boundary condition sister code ADF-BAND, the models being COSMO and SM12. An in depth discussion of how both models are treated in the PBC code and an explanation of the required formulations to properly account for periodic solvation is reported. Such considerations include edge effects that arise from using Bloch's theorem and the divergent (infinite) potential, as *1/r* is divergent in an infinite system. Furthermore, we discuss how the transition from a molecular to a PBC has been achieved successfully through showing that the transition from a molecule in ADF to a *zero*-dimensional crystal (ADF-BAND) results in nearly identical energies for these solvation methods, with a TZP basis set. As there are no experimental data for solvation in the solid state, we postulate an applicable way of ensuring that our implementation is consistent when moving to higher dimensional crystals. We do so by showing that the energy of a large system of CO on a Cu bilayer is consistent as the size is increased, and that the energies associated with this adsorption are decreased upon solvating the systems due to the induced dipole effect.

4.8. Acknowledgments

CAP and GS would like to thank Stan van Gisbergen, Erik van Lenthe, Pier Philipsen, Fedor Goumans and all SCM for their continued support and help. We would also like to thank James Xidos for helpful discussion and insights. Finally, we thank Christopher Cramer and Donald Truhlar for providing us with the ASA code, the Minnesota Solvation Database, and their invaluable input and help. Funding from the Natural Sciences and Engineering Research Council of Canada (NSERC, Discovery Grant, GS) and from SCM Amsterdam (travel and in-kind support, CAP, GS) are gratefully acknowledged.

4.9. Supporting Information

Further details on solvation free energies for comparison of ADF and ADF-BAND for *zero*-dimensional cells (Table A2.1 in Appendix 2).

4.10 References

(1) Kornienko, N.; Zhao, Y.; Kley, C. S.; Zhu, C.; Kim, D.; Lin, S.; Chang, C. J.; Yaghi, O. M.; Yang, P. Metal–Organic Frameworks for Electrocatalytic Reduction of Carbon Dioxide. *J. Am. Chem. Soc.* **2015**, *137* (44), 14129–14135.

(2) Kim, J.; Maiti, A.; Lin, L.-C.; Stolaroff, J. K.; Smit, B.; Aines, R. D. New Materials for Methane Capture from Dilute and Medium-Concentration Sources. *Nat. Commun.* **2013**, *4*, 1694.

(3) Asaduzzaman, A. M.; Schreckenbach, G. Computational Studies on the Interactions among Redox Couples, Additives and TiO2: Implications for Dye-Sensitized Solar Cells. *Phys. Chem. Chem. Phys.* **2010**, *12* (43), 14609–14618.

(4) Asaduzzaman, A. M.; Schreckenbach, G. Computational Studies of the Interactions of I- and I3- with TiO2 Clusters: Implications for Dye-Sensitized Solar Cells. *Theor. Chem. Acc.* **2011**, *129* (2), 199–208.

(5) O'Regan, B.; Grätzel, M. A Low-Cost, High-Efficiency Solar Cell Based on Dye-Sensitized Colloidal TiO2 Films. *Nature* **1991**, *353* (6346), 737–740.

(6) Hagfeldt, A.; Boschloo, G.; Sun, L.; Kloo, L.; Pettersson, H. Dye-Sensitized Solar Cells. *Chem. Rev.* **2010**, *110* (11), 6595–6663.

(7) Saga, T. Advances in Crystalline Silicon Solar Cell Technology for Industrial Mass Production. *NPG Asia Mater.* **2010**, *2* (3), 96–102.

(8) Jeon, N. J.; Noh, J. H.; Kim, Y. C.; Yang, W. S.; Ryu, S.; Seok, S. Il. Solvent Engineering for High-Performance Inorganic-Organic Hybrid Perovskite Solar Cells. *Nat. Mater.* **2014**, *13* (9), 897–903.

(9) Phan, N. T. S.; Van Der Sluys, M.; Jones, C. W. On the Nature of the Active Species in Palladium Catalyzed Mizoroki-Heck and Suzuki-Miyaura Couplings - Homogeneous or Heterogeneous Catalysis, a Critical Review. *Adv. Synth. Catal.*

2006, *348* (6), 609–679.

(10) Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.

(11) Orozco, M.; Luque, F. J. Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems. (Vol 100, Pg 4187, 2000). *Chem. Rev.* **2001**, *101*, 4187–4225.

(12) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3093.

(13) Klamt, A.; Schüürmann, G. 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* **1993**, *5*, 799–805.

(14) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620.

(15) Guerard, J. J.; Arey, J. S. Critical Evaluation of Implicit Solvent Models for Predicting Aqueous Oxidation Potentials of Neutral Organic Compounds. *J. Chem. Theory Comput.* **2013**, *9*, 5046–5058.

(16) Andzelm, J.; Kölmel, C.; Klamt, A. Incorporation of Solvent Effects into Density Functional Calculations of Molecular Energies and Geometries. *J. Chem. Phys.* **1995**, *103* (21), 9312.

(17) Delley, B. The Conductor-like Screening Model for Polymers and Surfaces. *Mol. Simul.* **2006**, *32* (2), 117–123.

(18) Andreussi, O.; Marzari, N. Electrostatics of Solvated Systems in Periodic Boundary Conditions. *Phys. Rev. B* **2014**, *90* (24), 245101.

(19) Garcia-Ratés, M.; Lopez, N. Multigrid-Based Methodology for Implicit Solvation Models in Periodic DFT. *J. Chem. Theory Comput.* **2016**, acs.jctc.5b00949.

(20) Kresse, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B* **1993**, *47* (1), 558–561.

(21) Kresse, G.; Hafner, J. Ab Initio Molecular-Dynamics Simulation of the Liquid-Metalamorphous- Semiconductor Transition in Germanium. *Phys. Rev. B* **1994**, *49* (20), 14251–14269.

(22) Kresse, G.; Furthm??ller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mater. Sci.* **1996**, *6* (1), 15–50.

(23) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54* (16), 11169–11186.

(24) Baerends, E. J.; Ziegler, T.; Autschbach, J.; Bashford, D.; Bérces, A.; Bickelhaupt, F. M.; Bo, C.; Boerrigter, P. M.; Cavallo, L.; Chong, D. P.; Deng, L.; Dickson, R. M.; Ellis, D. E.; van Faassen, M.; Fan, L.; Fischer, T. H.; Fonseca Guerra, C.; Franchini, M.; Ghysels, A.; Giammona, A.; van Gisbergen, S. J. A.; Götz, A. W.; Groeneveld, J. A.; Gritsenko, O. V.; Grüning, M.; Gusarov, S.; Harris, F. E.; van den Hoek, P.; Jacob, C. R.; Jacobsen, H.; Jensen, L.; Kaminski, J. W.; van Kessel,

G.; Kootstra, F.; Kovalenko, A.; Krykunov, M. V.; van Lenthe, E.; McCormack, D. A.; Michalak, A.; Mitoraj, M.; Morton, S. M.; Neugebauer, J.; Nicu, V. P.; Noodleman, L.; Osinga, V. P.; Patchkovskii, S.; Pavanello, M.; Philipsen, P. H. T.; Post, D.; Pye, C. C.; Ravenek, W.; Rodríguez, J. I.; Ros, P.; Schipper, P. R. T.; van Schoot, H.; Schreckenbach, G.; Seldenthuis, J. S.; Seth, M.; Snijders, J. G.; Solà, M.; Swart, M.; Swerhone, D.; te Velde, G.; Vernooijs, P.; Versluis, L.; Visscher, L.; Visser, O.; Wang, F.; Wesolowski, T. A.; van Wezenbeek, E. M.; Wiesenekker, G.; Wolff, S. K.; Woo, T. K.; A.L., Y. ADF. **2014**.

(25) Te Velde, G.; Baerends, E. J. Precise Density-Functional Method for Periodic Structures. *Phys. Rev. B* **1991**, *44*, 7888–7903.

(26) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. a.; Snijders, J. G.; Ziegler, T.; Velde, G. T. E.; Guerra, C. F.; Gisbergen, S. J. V. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22*, 931–967.

(27) Pye, C. C.; Ziegler, T. An Implementation of the Conductor-like Screening Model of Solvation within the Amsterdam Density Functional Package. *Theor. Chem. Acc.* **1999**, *101*, 396–408.

(28) Peeples, C. A.; Schreckenbach, G. Implementation of the SM12 Solvation Model into ADF and Comparison with COSMO. *J. Chem. Theory Comput.* **2016**, acs.jctc.6b00410.

(29) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22* (9), 931–967.

(30) Bloch, F. Über Die Quantenmechanik Der Elektronen in Kristallgittern. *Z. Phys.* **1929**, *52* (7-8), 555–600.

(31) Baerends, E. J.; Ros, P. Self-Consistent Molecular Hartree-Fock-Slater Calculations II. The Effect of Exchange Scaling in Some Small Molecules. *Chem. Phys.* **1973**, *2* (1), 52–59.

(32) Baerends, E. J.; Ros, P. Self-Consistent Molecular Hartree-Fock-Slater Calculations. *Chem. Phys.* **1975**, *8*, 412–418.

(33) Klamt, A.; Jonas, V. Treatment of the Outlying Charge in Continuum Solvation Models. *J. Chem. Phys.* **1996**, *105* (22), 9972.

(34) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.

(35) Hirshfeld, F. L. Bonded-Atom Freagments for Describing Molecular Charge Densities. *Theor. Chim. Acta* **1977**, *44*, 129–138.

(36) Still, C.; Tempczyk, A.; Hawley, R. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

(37) Born, M. Volume and Hydration of Ions. *Z. Phys.* **1920**, *1*, 45–48.

(38) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. Improved Methods for Semiempirical Solvation Models. *J. Comput. Chem.* **1995**, *16*, 422–440.

(39) Zhu, T.; J., L.; G., H.; C., C.; D., T. Density Functional Solvation Model Based on CM2 Atomic Charges. *J. Chem Phys.* **1998**, *109*, 9117–9133.

(40) Martin, R. L.; Hay, P. J.; Pratt, L. R. Hydrolysis of Ferric Ion in Water and Conformational Equilibrium. *J. Phys. Chem. A* **1998**, *102* (2), 3565–3573.

(41) Shamov, G. A.; Schreckenbach, G. Density Functional Studies of Actinyl Aquo Complexes Studied Using Small-Core Effective Core Potentials and a Scalar Four-Component Relativistic Method. *J. Phys. Chem. A* **2005**, *109* (48), 10961–10974.

(42) Perdew, J.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(43) Marenich, A. V; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. Minnesota Solvation Database. *Minnesota Solvation Database - version 2012, Univ. Minnesota, Minneap.* **2012**, 1–28.

Chapter 5: Summary and Future developments and studies of SM12 in both ADF and ADF-BAND

5.1 Summary

Solvation has a profound effect on many physical aspects of chemistry, from stabilizing compounds, to creating new reaction pathways. ADF, previously only had use of one solvation model within the code proper, that is directly dependant on solving Poission's equation. With the current implementation of SM12, the ADF user is now allowed access to a separate, simple, accurate model of solvation. This can also be of use for quick calculations that are not reliant on the choice of large, computationally expensive, basis sets that are usually required in order to get good accuracy. This is all achieved while maintaining a more-accurate description of orbitals through STOs, and still performing as well as the original GTO implementation. Solvation is extended to the NAOs used within ADF-BAND with confidence that these orbitals, too, will be independent of basis set choice. Thusly allowing both an alternative and fairly accurate periodic solvation model to the periodic COSMO.

The implementation of SM12 in to ADF proves the theory is applicable across different codes, and still allows for accurate modeling of solvation energies. This implementation also allows for the end user a more in depth study using liquids within ADF. As one can now approach modeling the solvation with two separate models, to get

a more robust look into how the system of interest is affected. With both models having their strengths and weaknesses one can calculate solvation energies using each model to better understand the system of interest, this also allows for a compare and contrast between methods.

5.2 Future Work

Gradients- There needs to be several considerations when moving forward into research with both implementations. First and foremost is the requirement of analytical gradients for both ADF and ADF-BAND, as transition states can be stabilized with the use of solvation.[13] Also of importance is ground-state geometries, as some compounds require optimizations in solution in order to reach the correct geometries, such as glycine.[14,15,16,17] The groundwork is laid out for this implementation, but the derivative of the pro-molecule (entire molecule), as required by the derivative of HPA, section 3.3.1, proves to be a difficult process. As each geometry step is taken we need to recalculate how the derivative of the pro-molecule changes at each grid-point in space, this introduces significant complications as these points will appear and disappear for each geometry configuration.

Ions, and Radicals- The next consideration proposed for future studies is a further look into COSMO's performance when using negatively charged ions, as the mean unsigned error is very low and is suspect. This could be for many different reasons, perhaps COSMO performs particularly well for this test set, or has fortuitous cancelation

of errors associated with anions. Further inspection into how charged compounds are calculated has an ongoing problem when modeling them with solvation. There are not enough data points in order to accurately describe the solvation of ions, there has not been enough experimental research on ions in solution to properly test the validity of the models, for these charged systems. This is a major problem within the implementation of solvation model; as charged ions have larger than normal polarization. Introduction of this large polarization into a liquid system will cause the polarization of the entire solvent to increase dramatically, and thus the overall energy of solvation. This leads to the foremost difficulty in modeling ionic systems in solution; these systems are extremely sensitive to solution. Therefore, this ionic environment is when solvation is most required, as the energy of solvating a charged ion is an order of magnitude larger than neutral systems.

The solvation models cannot be parameterized nor can a statistically relevant study be fully completed with the limited data. This could also be of interest for an experimental group. This could be done in several ways; one is with the rate constant between two phases. Equation 5.1 describes x as the concentration of either gas or liquid, the rate constant may be defined by K and described by equation 5.2, which in turn may be used with 5.3 to calculate the Gibbs free energy, ΔG_{Gibbs} , of the system, where R is the gas constant, and T is the temperature of the system. Though, doing so, with ionic systems could prove to be challenging, as ions are usually unstable in gas phase.

$$x_{gas} \rightleftharpoons x_{liquid} \quad (5.1)$$

$$K = \frac{[x_{liquid}]}{[x_{gas}]} \quad (5.2)$$

$$\Delta G_{Gibbs} = -RTln(K) \quad (5.3)$$

Periodic Solvation- The application of the periodic solvation model to interesting systems is also to be done in the future. As most systems are stabilized by the introduction of solvation, the band-gap should be affected by the introduction of a liquid, as it should lower the lowest un-occupied molecular orbital. It is of interest to also investigate the effects of solvation on the molecular cavities mentioned earlier, as the model will solvate the cavities found within, if the space is large enough. Dr. S Patchkovskii, through discussion at the 99th Chemistry exhibition and conference, had brought up an interesting idea of solvating both a cationic atom and anionic molecule within a molecular sieve, as the increased polarization could have an interesting effect in these periodic solvated compounds. This has an interesting consequence, solid-state modeling with ions requires a non-physical charge-balancing counter-ion. Without this the charge of the system will be modeled as infinity. Modeling a real-world system with this type of neutral cationic-anionic interaction could give insight onto the consequences of the aforementioned non-physical counter-ion.

5.3 References

(1) Hohenberg, P.; Kohn, W. Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136* (3B), B864–B871.

(2) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*. 1965.

(3) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620.

(4) Bloch, F. Über Die Quantenmechanik Der Elektronen in Kristallgittern. *Z. Phys.* **1929**, *52* (7-8), 555–600.

(5) Baerends, E. J.; Ziegler, T.; Autschbach, J.; Bashford, D.; Bérces, A.; Bickelhaupt, F. M.; Bo, C.; Boerrigter, P. M.; Cavallo, L.; Chong, D. P.; Deng, L.; Dickson, R. M.; Ellis, D. E.; van Faassen, M.; Fan, L.; Fischer, T. H.; Fonseca Guerra, C.; Franchini, M.; Ghysels, A.; Giammona, A.; van Gisbergen, S. J. A.; Götz, A. W.; Groeneveld, J. A.; Gritsenko, O. V.; Grüning, M.; Gusarov, S.; Harris, F. E.; van den Hoek, P.; Jacob, C. R.; Jacobsen, H.; Jensen, L.; Kaminski, J. W.; van Kessel, G.; Kootstra, F.; Kovalenko, A.; Krykunov, M. V.; van Lenthe, E.; McCormack, D. A.; Michalak, A.; Mitoraj, M.; Morton, S. M.; Neugebauer, J.; Nicu, V. P.; Noodleman, L.; Osinga, V. P.; Patchkovskii, S.; Pavanello, M.; Philipsen, P. H. T.; Post, D.; Pye, C. C.; Ravenek, W.; Rodríguez, J. I.; Ros, P.; Schipper, P. R. T.; van Schoot, H.; Schreckenbach, G.; Seldenthuis, J. S.; Seth, M.; Snijders, J. G.; Solà, M.; Swart, M.; Swerhone, D.; te Velde, G.; Vernooijs, P.; Versluis, L.; Visscher, L.; Visser, O.; Wang, F.; Wesolowski, T. A.; van Wezenbeek, E. M.; Wiesenekker, G.; Wolff, S. K.; Woo, T. K.; A.L., Y. ADF. **2014**.

(6) Baerends, E. J.; Ros, P. Self-Consistent Molecular Hartree-Fock-Slater Calculations II. The Effect of Exchange Scaling in Some Small Molecules. *Chem.*

Phys. **1973**, *2* (1), 52–59.

(7) Baerends, E. J.; Ros, P. Self-Consistent Molecular Hartree-Fock-Slater Calculations. *Chem. Phys.* **1975**, *8*, 412–418.

(8) Klamt, A.; Schüürmann, G. 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* **1993**, *5*, 799–805.

(9) Pye, C. C.; Ziegler, T. An Implementation of the Conductor-like Screening Model of Solvation within the Amsterdam Density Functional Package. *Theor. Chem. Acc.* **1999**, *101*, 396–408.

(10) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.

(11) Te Velde, G.; Baerends, E. J. Precise Density-Functional Method for Periodic Structures. *Phys. Rev. B* **1991**, *44*, 7888–7903.

(12) Garcia-Ratés, M.; Lopez, N. Multigrid-Based Methodology for Implicit Solvation Models in Periodic DFT. *J. Chem. Theory Comput.* **2016**, acs.jctc.5b00949.

(13) Sola, M.; Lledos, A.; Duran, M.; Bertran, J.; Abboud, J. L. M. Analysis of Solvent Effects on the Menshutkin Reaction. *J Am Chem Soc* **1991**, *113* (8), 2873–2879.

(14) Brown, R. D.; Godfrey, P. D.; Storey, J. W. V.; Bassez, M. P. Microwave Spectrum and Conformation of Glycine. *J. Chem. Soc., Chem. Commun.* **1978**, *13*

(13), 547–548.

(15) Suenram, R. D.; Lovas, F. J. Millimeter Wave Spectrum of Glycine. *J. Mol. Spectrosc.* **1978**, *72* (3), 372–382.

(16) Godfrey, P. D.; Brown, R. D. Shape of Glycine. *J. Am. Chem. Soc.* **1995**, *117* (7), 2019–2023.

(17) Iijima, K.; Tanaka, K.; Onuma, S. Main Conformer of Gaseous Glycine: Molecular Structure and Rotational Barrier from Electron Diffraction Data and Rotational Constants. *J. Mol. Struct.* **1991**, *246* (3-4), 257–266.

Appendix 1

Contents:

A 1.1 : Solvents available in ADF for SM12 Page A 2

A 1.2: Compounds for CM5 Page A 3

A 1.3: Solvation energies in water with M06-2X. Page A 19

A 1.4: Solvation energies in water with Hartree Fock Page A 29

A 1.5: Varying solvents with calculated energies with B3LYP Page A 41

A 1.6: Ionic solvation energies for COSMO and SM12 with M06-2X Page A 45

A 1.1 Solvents as reported.

Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620.

Solvents Available for SM12 in ADF		
Acetic Acid	Dibutyl ether	Methylene Chloride
Acetonitrile	*o*-dicholrobenzene	N-methylformamide
Acetophenone	1,2-dichloroethane	4-methyl-2-pentanone
Aniline	Diethyl ether	2-methylpyrididne
Anisole	Diisopropyl ether	Nitrobenzene
Benzene	N,N-dimethylacetamide	Nitroethane
Benzonitrile	N,N-dimethylformamide	Nitromethane
Benzyl Alcohol	2,6-dimethylpyridine	*o*-nitrotoluene
Bromobenzene	Dimethyl sulfoxide	Nonane
Bromoethane	Dodecane	Nonanol
Bromoform	Ethanol	Octane
Bromooctane	Ethoxybenzene	Octanol
n-butanol	Ethyl acetate	Pentadecane
sec-butanol	Ethylbenzene	Pentane
Butanone	Fluorobenzne	Pentanol
Butyl acetate	1-fluoro-n-octane	Perfluorobenzene
n-butylbenzene	Heptane	Phenyl ether
sec-butylbenzene	Heptanol	Propanol
t-butylbenzene	Hexadecane	Pyridine
Carbon Disulfide	Hexadecyl iodide	Tetrachloroethene
Carbon Tetrachloride	Hexane	Tetrahydrofuran
Chlorobenzne	Hexanol	Tetrahydrothiophene dioxide
Chloroform	Iodobenzene	Tetralin
Chlorohexane	Isobutanol	Toluene
m-cresol	Isooctane	Tributylphosphate
Cyclohexane	Isopropanol	Trimethylamine
Cyclohexanone	Isopropylbenzene	1,2,4-trimethylbenzene
Decalin (*cis/trans* mixture)	p-isopropyltolunene	Undecane
Decane	Mesitylene	Water
Decanol	Methanol	Xylene (*m/o/p* mixture)
1,2-dibromoethane	Methoxyethanol	

A 1.2 Charge Model 5 Compound List.

Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541. SI

Label	Compound Name
mol001	fulvene
mol002	propane
mol003	isobutane
mol004	cyclopropene
mol005	propene
mol006	propyne
mol007	1,3-pentadiyne, methyldiacetylene
mol008	ammonia
mol009	water
mol010	Z-diazene
mol011	hydroxylamine
mol012	nitric oxide
mol013	nitrogen dioxide
mol014	nitrosyl hydride
mol015	methylamine
mol016	dimethylamine
mol017	trimethylamine
mol018	ethylamine
mol019	propylamine
mol020	cyclopropylamine
mol021	isopropylamine
mol022	aziridine
mol023	piperidine (NH equatorial)
mol024	piperidine (NH axial)
mol025	1,2,5,6-tetrahydropyridine (NH equatorial)
mol026	1,2,5,6-tetrahydropyridine (NH axial)
mol027	aniline
mol028	pyrrole
mol029	pyridine
mol030	2-methylpyridine
mol031	4-methylpyridine
mol032	indole
mol033	quinoline

mol034	isoquinoline
mol035	imidazole
mol036	pyridazine
mol037	pyrimidine
mol038	2-methylpyrimidine
mol039	hydrogen cyanide
mol040	ethanonitrile, acetonitrile
mol041	dicyanomethane
mol042	propanonitrile, propionitrile
mol043	butanonitrile (anti)
mol044	butanonitrile (gauche)
mol045	cyclopropane carbonitrile
mol046	isobuteronitrile
mol047	cyclobutane carbonitrile
mol048	t-butyl cyanide
mol049	pentanenitrile
mol050	benzonitrile
mol051	propyleneimine (cis)
mol052	propyleneimine (trans)
mol053	E-acetaldimine
mol054	Z-acetaldimine
mol055	Z-N-ethylidene methanamine
mol056	N-methylformaldimine
mol057	diazomethane
mol058	methyl azide
mol059	2-butenenitrile (trans)
mol060	Z-2-butenenitrile
mol061	cyanoacetylene
mol062	acrylonitrile
mol063	methacrylonitrile
mol064	2-cyanopyridine
mol065	3-cyanopyridine
mol066	4-cyanopyridine
mol067	cyanimide
mol068	cyano allene
mol069	cyclopentadiene-1-carbonitrile
mol070	methylaminonitrile
mol071	methanol
mol072	ethanol
mol073	1,2-ethanediol
mol074	propargyl alcohol
mol075	2-propanol
mol076	1-propanol (trans)
mol077	1-propanol (gauche)
mol078	1,2-propanediol (CH_3 gauche)
mol079	1,2-propanediol (CH_3 anti)

mol080	1-butanol
mol081	Cyclopropanol
mol082	Cyclobutanol
mol083	Phenol
mol084	methyl propyl ether
mol085	dimethyl ether, methoxymethane
mol086	diethyl ether, ethoxyethane
mol087	vinyl methyl ether
mol088	Tetrahydrofuran
mol089	anisole, methoxybenzene
mol090	Tetrahydropyran
mol091	1,3-dioxane
mol092	3,4-dihydro-2,4-pyran
mol093	Oxetane
mol094	3-methyleneoxetane
mol095	Formaldehyde
mol096	ethanal, acetaldehyde
mol097	Propanal
mol098	Butanal
mol099	E-2-butenal
mol100	propanone, acetone
mol101	2-butanone
mol102	Cyclopentanone
mol103	methyl phenyl ketone, acetophenone
mol104	cyclobutane-1,2-dione
mol105	Cyclobutanone
mol106	Cyclopropanone
mol107	cyclopentadienone
mol108	4-cyclopentene-1,3-dione
mol109	3-cyclopentenone
mol110	Cyclohexanone
mol111	formic acid (s-cis)
mol112	formic acid (s-trans)
mol113	ethanoic acid, acetic acid
mol114	formylformic acid (trans)
mol115	propanoic acid (cis)
mol116	acrylic acid ({C-C}-s-cis)
mol117	acrylic acid ({C-C}-s-trans)
mol118	2-methoxyethanoic acid (gauche)
mol119	acetoacetic acid
mol120	methyl ethanoate, methyl acetate
mol121	methyl methanoate, methyl formate
mol122	ethyl methanoate (cis)
mol123	ethyl methanoate (trans)
mol124	ethyl ethanoate, ethyl acetate
mol125	pentyl formate

mol126	diethyl carbonate
mol127	maleic anhydride
mol128	beta-propiolactone
mol129	gamma-butyrolactone
mol130	2-methoxyethanol (gauche)
mol131	benzyl alcohol
mol132	Diketene
mol133	3-oxetanone
mol134	2(5H)-furanone
mol135	Ketene
mol136	Methylketene
mol137	Formamide
mol138	Ethanamide
mol139	N-methylformamide
mol140	N,N-dimethylformamide
mol141	E-N-methylacetamide
mol142	Z-N-methylacetamide
mol143	N,N-dimethylacetamide
mol144	Benzamide
mol145	Phenylurea
mol146	Nitromethane
mol147	Nitroethane
mol148	1-nitropropane
mol149	2-nitropropane
mol150	Nitrobenzene
mol151	methyl nitrate
mol152	acetyl cyanide
mol153	dimethylnitrosamine
mol154	Formaldoxime
mol155	fulminic acid
mol156	3-imino-2,3-dihydroisoxazole
mol157	3-iminofuran
mol158	Isoxazole
mol159	Morpholine
mol160	Uracil
mol161	HeH^+
mol162	$HeLi^+$
mol163	HeF^+
mol164	$HeCl^+$
mol165	$HeOH^+$
mol166	$HeSH^+$
mol167	$HeNH_2^+$
mol168	$HePH_2^+$
mol169	$HeCH_3^+$
mol170	$HeSiH_3^+$
mol171	LiF

mol172	$LiO\ (^2\Pi)$
mol173	$LiS\ (^2\Pi)$
mol174	$LiCl$
mol175	$LiBr$
mol176	LiI
mol177	$LiOH$
mol178	$LiSH$
mol179	$LiNH_2$
mol180	$LiCH_3$
mol181	$LiCH_3O$
mol182	$LiOCH_3$
mol183	LiC_2H
mol184	vinyl lithium
mol185	LiC_3H_5
mol186	LiC_2H_2F
mol187	LiC_2H_2Cl
mol188	LiC_2H_6N
mol189	BeF^+
mol190	$BeCl^+$
mol191	BeO
mol192	BeS
mol193	$Be{=}NH$
mol194	$Be{=}PH$
mol195	$Be{=}NF$
mol196	$Be{=}NCl$
mol197	$BeClF$
mol198	$BeBrF$
mol199	$BeBrCl$
mol200	$Be{=}CH_2$
mol201	$Be{=}SiH_2$
mol202	$BeFCH_3$
mol203	$BeClCH_3$
mol204	$BeBrCH_3$
mol205	BF_2H
mol206	BH_2Cl
mol207	BH_2OH
mol208	BF_2OH
mol209	$BH(OH)_2$
mol210	$BF(OH)_2$
mol211	BH_2NH_2
mol212	BF_2NH_2
mol213	BF_2CH_3
mol214	BF_2CCH
mol215	BF_2CHCH_2
mol216	$BF_2CH_2CH_3$
mol217	$BH(NH_2)_2$

mol218	BH_3CO
mol219	BH_3NH_3
mol220	BF_3NH_3
mol221	HF
mol222	fluorocyclohexane (axial)
mol223	fluorocyclohexane (equatorial)
mol224	1,2-difluoroethane (gauche)
mol225	1-fluoropropane (gauche)
mol226	1-fluoropropane (trans)
mol227	fluoromethane
mol228	1,1-difluoroethane
mol229	Fluoroethane
mol230	2-fluoropropane
mol231	difluoromethane
mol232	trifluoromethane
mol233	1,1,1-trifluoroethane
mol234	pentafluoroethane
mol235	1,1,1,2,2,3,3-heptafluoropropane
mol236	2-fluoro-2-methylpropane
mol237	1,2-difluoroethene
mol238	3-fluoropropene (eclipsed)
mol239	3-fluoropropene (gauche)
mol240	1,1-difluoro-1-propene
mol241	fluoroethene
mol242	2-fluoropropene
mol243	cis-1-fluoro-1-propene
mol244	1,1-difluoroethene
mol245	difluoroallene
mol246	trifluoroethene
mol247	3,3,3-trifluoropropene
mol248	fluoroallene
mol249	fluorobenzene
mol250	trifluoromethyl-benzene
mol251	ortho-fluorotoluene
mol252	meta-fluorotoluene
mol253	para-fluorotoluene
mol254	1,3-difluorobenzene
mol255	1,2,3,4-tetrafluorobenzene
mol256	1,2-difluorobenzene
mol257	fluoroacetylene
mol258	acetyl fluoride
mol259	3,3,3-trifluoropropyne
mol260	tetrafluoropropyne
mol261	carbonyl fluoride
mol262	bis(trifluoromethyl) ether
mol263	cyanogen fluoride

mol264	trifluroacetonitrile
mol265	1-fluoro-4-nitrobenzene
mol266	NeH^+
mol267	$NeLi^+$
mol268	NeF^+
mol269	$NeCl^+$
mol270	$NeOH^+$
mol271	$NeSH^+$
mol272	$NeNH_2^+$
mol273	$NePH_2^+$
mol274	$NeCH_3^+$
mol275	$NeSiH_3^+$
mol276	NaF
mol277	NaO ($^2\Pi$)
mol278	NaS ($^2\Pi$)
mol279	NaCl
mol280	NaBr
mol281	NaI
mol282	NaOH
mol283	NaSH
mol284	$NaNH_2$
mol285	$NaCH_3$
mol286	$NaCH_3O$
mol287	$NaOCH_3$
mol288	NaC_2H
mol289	vinyl sodium
mol290	NaC_3H_5
mol291	NaC_2H_2F
mol292	NaC_2H_2Cl
mol293	NaC_2H_6N
mol294	MgF^+
mol295	$MgCl^+$
mol296	MgO
mol297	MgS
mol298	Mg=NH
mol299	Mg=PH
mol300	Mg=NF
mol301	Mg=NCl
mol302	MgClF
mol303	MgBrF
mol304	MgBrCl
mol305	$Mg=CH_2$
mol306	$Mg=SiH_2$
mol307	$MgFCH_3$
mol308	$MgClCH_3$
mol309	$MgBrCH_3$

mol310	AlN
mol311	AlF
mol312	AlP
mol313	AlCl
mol314	AlBr
mol315	AlH_2OH
mol316	AlH_2SH
mol317	AlH_2NH_2
mol318	AlH=O
mol319	AlH=S
mol320	AlH=NH
mol321	AlH=PH
mol322	SiH
mol323	SiS
mol324	SiH_3
mol325	CH_2SiH_3
mol326	methylsilane
mol327	dimethylsilane
mol328	Ethylsilane
mol329	trimethylsilane
mol330	vinylsilane
mol331	phenylsilane
mol332	silicon monoxide
mol333	Disiloxane
mol334	methylsilyl ether
mol335	$((CH_3)_3Si)_2O$
mol336	$CH_3OSi(CH_3)_3$
mol337	CH_3SiH_2OH
mol338	$H_2Si(OH)_2$
mol339	H_2SiO
mol340	H_3SiOH
mol341	Fluorosilane
mol342	difluorosilane
mol343	trifluorosilane
mol344	fluoromethylsilane
mol345	difluoromethylsilane
mol346	trifluoromethylsilane
mol347	SiH_3SiH_2F
mol348	1,1,1-trifluorodisilane
mol349	chlorosilane
mol350	dichlorosilane
mol351	trichlorosilane
mol352	CH_3SiH_2Cl
mol353	trichloromethylsilane
mol354	CH_2ClSiH_3
mol355	$CH_3CH_2SiHCl_2$

mol356	$Si(CH_3)_3Cl$
mol357	dichloromethylsilane
mol358	Cl_3FSi
mol359	iodomethylsilane
mol360	phosphine
mol361	methylphosphine
mol362	dimethyl methylphosphonate
mol363	OPH_2OH
mol364	trimethylphosphine
mol365	dimethylphosphine
mol366	PH
mol367	CH_2PH_2
mol368	$CH_3CH(PH_2)CH_3$
mol369	CH_3PH
mol370	$(CH_3)_3PH_2$
mol371	$CH_3CH(PH)CH_3$
mol372	phosphorus nitride
mol373	phosphorus oxychloride
mol374	thiophosphoryl fluoride
mol375	hypophosphorous acid
mol376	methylphosphonic acid
mol377	OPH_3
mol378	OPH
mol379	PH2OH
mol380	$(CH_3)_3PO$
mol381	PO
mol382	phosphorus trichloride
mol383	phosphorus trifluoride
mol384	phosphoryl fluoride
mol385	FCP
mol386	$CH_3P(O)(OCH_3)(SCH_3)$
mol387	$CH_3P(O)(SCH_3)_2$
mol388	$OP(OCH_3)(SCH_3)_2$
mol389	$(CH_3)_3PS$
mol390	PS
mol391	$SP(CH_3)(OCH_3)_2$
mol392	SH
mol393	CH2SH
mol394	$H_2C=CHSH$
mol395	methanethiol
mol396	ethanethiol (gauche)
mol397	ethanethiol (trans)
mol398	1-propane thiol
mol399	2-methyl-2-propane thiol
mol400	HSSH
mol401	CH3SSH

mol402	isopropyl thiol
mol403	$CH_3(CS)CH_3$
mol404	CH_3S
mol405	dithiane
mol406	$(CH_3)_3CS$
mol407	ethylmethyl sulfide
mol408	ethylmethyl sulfide
mol409	dimethyl sulfide
mol410	diethyl sulfide
mol411	hydrogen sulfide
mol412	thietane
mol413	thiacyclohexane
mol414	dimethyl disulfide
mol415	thiophene
mol416	2-methylthiophene
mol417	2,5-dihydrothiophene
mol418	3-methylthiophene
mol419	carbon monosulfide
mol420	methylisothiocyanate
mol421	carbonoxy sulfide
mol422	thioformaldehyde
mol423	propanethial S-oxide
mol424	thioacetaldehyde
mol425	methylthiocyanate
mol426	thiazole
mol427	methane sulfonic acid
mol428	methyl methane sulfenate
mol429	methyl methane sulfinate
mol430	methane sulfonamide
mol431	dicyanogen sulfide
mol432	dimethyl sulfone
mol433	dimethyl sulfoxide
mol434	HCl
mol435	chloromethane, methyl chloride
mol436	dichloromethane
mol437	trichloromethane, chloroform
mol438	chloroethane
mol439	1,1-dichloroethane
mol440	1,1,1-trichloroethane
mol441	pentachloroethane
mol442	cyclopropyl chloride
mol443	1-chloropropane (gauche)
mol444	1-chloropropane (trans)
mol445	1,3-dichloropropane
mol446	2-chloropropane, isopropyl chloride
mol447	1-chloro-2-methylpropane

mol448	2-chlorobutane
mol449	1-chlorobutane
mol450	2-chloro-2-methylpropane
mol451	1-chloropentane
mol452	chlorocyclohexane (axial)
mol453	chlorocyclohexane (equatorial)
mol454	chloroethene, vinyl chloride
mol455	1,1-dichloroethene
mol456	Z-1,2-dichloroethene
mol457	1-chloropropene (cis)
mol458	1-chloropropene (trans)
mol459	2-chloropropene
mol460	3-chloropropene, allyl chloride
mol461	chlorobenzene
mol462	1,2-dichlorobenzene
mol463	1,3-dichlorobenzene
mol464	ortho-chlorotoluene
mol465	para-chlorotoluene
mol466	chloroacetylene
mol467	3-chloropropyne
mol468	chlorofluoromethane
mol469	chlorodifluoromethane
mol470	fluorotrichloromethane
mol471	chloropentaflouroethane
mol472	1-chloro-1-fluoroethane
mol473	1,1-dichloro-2-fluoropropene
mol474	chlorotrifluoromethane
mol475	dichlorofluoromethane
mol476	1-chloro-1,1-difluoroethane
mol477	acetyl chloride
mol478	carbonyl chloride
mol479	para-chlorophenol
mol480	cyanogen chloride
mol481	1-chloro-2-nitrobenzene
mol482	1-chloro-3-nitrobenzene
mol483	1-chloro-4-nitrobenzene
mol484	ArH^+
mol485	$ArLi^+$
mol486	ArF^+
mol487	$ArCl^+$
mol488	$ArOH^+$
mol489	$ArSH^+$
mol490	$ArNH_2^+$
mol491	$ArPH_2^+$
mol492	$ArCH_3^+$
mol493	$ArSiH_3^+$

mol494	propyne-argon complex
mol495	KH
mol496	KLi
mol497	KF
mol498	KNa
mol499	KCl
mol500	KBr
mol501	KI
mol502	KO
mol503	KS
mol504	KOH
mol505	KSH
mol506	KCN
mol507	KNH_2
mol508	KPH_2
mol509	KCH_3
mol510	$KSiH_3$
mol511	CaF^+
mol512	$CaCl^+$
mol513	CaO
mol514	CaS
mol515	Ca=NH
mol516	Ca=PH
mol517	Ca=NF
mol518	Ca=NCl
mol519	$Ca=CH_2$
mol520	$Ca=SiH_2$
mol521	$CaFCH_3$
mol522	$CaClCH_3$
mol523	ZnH^+
mol524	ZnN^+
mol525	ZnP^+
mol526	ZnO
mol527	ZnS
mol528	ZnF^+
mol529	$ZnCl^+$
mol530	$ZnCN^+$
mol531	$ZnOH^+$
mol532	$ZnSH^+$
mol533	Zn=NH
mol534	Zn=PH
mol535	$Zn=CH_2$
mol536	$Zn=SiH_2$
mol537	$Zn\text{-}NH_2^+$
mol538	$Zn\text{-}PH_2^+$
mol539	$Zn\text{-}CH_3^+$

mol540	$Zn\text{-}SiH_3^+$
mol541	GeH
mol542	GeC
mol543	GeSi
mol544	GeN
mol545	GeP
mol546	GeO
mol547	GeS
mol548	GeSe
mol549	GeH_2
mol550	GeF_2
mol551	$GeCl_2$
mol552	GeH_3F
mol553	GeH_3Cl
mol554	GeH_3N_3
mol555	AsH
mol556	AsC
mol557	AsSi
mol558	AsN
mol559	AsP
mol560	AsO
mol561	AsS
mol562	AsF
mol563	AsCl
mol564	OAsH
mol565	FCAs
mol566	AsH_3
mol567	AsF_3
mol568	$AsCl_3$
mol569	$OAsH_3$
mol570	SeH
mol571	SeC
mol572	SeSi
mol573	SeN
mol574	SeP
mol575	SeO
mol576	SeS
mol577	SeF
mol578	SeCl
mol579	H_2Se
mol580	COSe
mol581	CSSe
mol582	SeO_2
mol583	SeS_2
mol584	CH_2Se
mol585	F_4Se

mol586	Cl_4Se
mol587	HBr
mol588	bromomethane
mol589	dibromomethane
mol590	tribromomethane, bromoform
mol591	bromoethane
mol592	1-bromopropane
mol593	2-bromopropane
mol594	1-bromobutane
mol595	2-bromobutane
mol596	1-bromopentane
mol597	1-bromoheptane
mol598	bromoethene
mol599	bromobenzene
mol600	bromoacetylene
mol601	3-bromopropyne
mol602	bromotrifluoromethane
mol603	dibromodifluoromethane
mol604	HI
mol605	FI
mol606	ClI
mol607	BrI
mol608	CIN
mol609	CH_3I
mol610	C_2HI
mol611	C_2H_3I
mol612	C_2H_5I
mol613	n-C_3H_7I
mol614	C_6H_5I
mol615	1,1,1,3,3,3-hexafluoropropan-2-ol (-1)
mol616	1,2-ethanediol (-1)
mol617	1-aminonaphthalene (+1)
mol618	1-propanethiol (-1)
mol619	1-propanol (-1)
mol620	2,2,2-trifluoroethanol (-1)
mol621	2-butanol (-1)
mol622	2-chlorophenol (-1)
mol623	2-methoxyethanol (-1)
mol624	2-methylaniline (+1)
mol625	2-methylphenol (-1)
mol626	2-nitrophenol (-1)
mol627	3-aminoaniline (+1)
mol628	3-chloroaniline (+1)
mol629	3-hydroxyphenol (-1)
mol630	3-methylaniline (+1)
mol631	3-methylphenol (-1)

mol632	3-nitrophenol (-1)
mol633	3-pentanone (-1)
mol634	4-chloroaniline (+1)
mol635	4-chlorophenol (-1)
mol636	4-hydroxyphenol (-1)
mol637	4-methoxyaniline (+1)
mol638	4-methylaniline (+1)
mol639	4-methyl-N,N-dimethylaniline (+1)
mol640	4-methylphenol (-1)
mol641	4-nitroaniline (+1)
mol642	4-nitroaniline (-1)
mol643	4-nitrophenol (-1)
mol644	acetaldehyde (-1)
mol645	acetamide (+1)
mol646	acetamide (-1)
mol647	acetic acid (-1)
mol648	acetone (+1)
mol649	acetone (-1)
mol650	acetonitrile (+1)
mol651	acetonitrile (-1)
mol652	acetophenone (+1)
mol653	acetylene (-1)
mol654	acrylic acid (-1)
mol655	allyl alcohol (-1)
mol656	allylamine (+1)
mol657	aniline (+1)
mol658	aniline (-1)
mol659	azacycloheptane (+1)
mol660	azetidine (+1)
mol661	aziridine (+1)
mol662	benzamide (+1)
mol663	benzoic acid (-1)
mol664	benzyl alcohol (-1)
mol665	chloroacetic acid (-1)
mol666	cyanamide (-1)
mol667	cyclohexanamine (+1)
mol668	diallylamine (+1)
mol669	dichloroacetic acid (-1)
mol670	diethyl ether (+1)
mol671	diethylamine (+1)
mol672	dimethyl ether (+1)
mol673	dimethyl sulfide (+1)
mol674	dimethyl sulfoxide (+1)
mol675	dimethyl sulfoxide (-1)
mol676	dimethylamine (+1)
mol677	di-n-propylamine (+1)

mol678	diphenylamine (-1)
mol679	ethanethiol (-1)
mol680	ethanol (+1)
mol681	ethanol alcohol (-1)
mol682	ethyl hydroperoxide (-1)
mol683	formic acid (-1)
mol684	hexanoic acid (-1)
mol685	hydrazine (+1)
mol686	hydrogen cyanide (-1)
mol687	hydrogen peroxide (-1)
mol688	hydrogen sulfide (-1)
mol689	isopropanol (-1)
mol690	isopropylamine (+1)
mol691	methanethiol (-1)
mol692	methanol (+1)
mol693	methanol (-1)
mol694	methyl hydroperoxide (-1)
mol695	methylamine (+1)
mol696	morpholine (+1)
mol697	N,N-diethylaniline (+1)
mol698	N,N-dimethylaniline (+1)
mol699	N-ethylaniline (+1)
mol700	nitromethane (-1)
mol701	N-methylaniline (+1)
mol702	n-propylamine (+1)
mol703	phenol alcohol (-1)
mol704	piperazine (+1)
mol705	piperidine (+1)
mol706	propanoic acid (-1)
mol707	pyridine (+1)
mol708	pyrrole (+1)
mol709	pyrrolidine (+1)
mol710	pyruvic Acid (-1)
mol711	quinoline (+1)
mol712	t-butanol (-1)
mol713	t-butylamine (+1)
mol714	thiophenol (-1)
mol715	trichloromethane (-1)
mol716	triethylamine (+1)
mol717	trifluoroacetic acid (-1)
mol718	trimethylamine (+1)
mol719	tri-n-propylamine (+1)
mol720	water (+1)
mol721	water (-1)

A 1.3 Solvation free energies calculated in ADF with TZ2P with M06-2X, for COSMO and SM12 in water. Reported in kcal/mol. Based on solvation free energies reported by Marenich et al.

Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620. SI

Marenich, A. V; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. Minnesota Solvation Database. *Minnesota Solvation Database - version 2012, Univ. Minnesota, Minneap.* **2012**, 1–28

Nicholls et al. *J. Med. Chem.* **2008**, 51, 769

Guthrie, *J. Phys. Chem. B* **2009**, 113, 4501

Gaballe et al., *J. Comput. Aided Mol. Des.* **2010**, 24, 259

Zhu et. al., *J. Chem. Phys.* **1998**, 109, 9117

Label	Solute	SM12	COSMO
0407tet	1,1,1,2-tetrachloroethane	-0.86	-2.94
0212hex	1,1,1,3,3,3-hexafluoropropan-2-ol	-1.19	-4.81
0165tri	1,1,1-trichloroethane	-0.31	-1.34
0211tri	1,1,1-trifluoropropan-2-ol	-2.61	-4.93
0206tri	1,1,2-trichloro-1,2,2-trifluoroethane	0.74	-1.46
0166tri	1,1,2-trichloroethane	-0.72	-3.18
0154dif	1,1-difluoroethane	-0.16	-3.3
0401amia	1,1-dimethyl-3-phenylurea	-9.38	-11.53
n006	1,1-dimethylhydrazine	-5.99	-5.79
0175odi	1,2-dichlorobenzene	-1.6	-1.7
0171Zdi	Z-1,2-dichloroethene	-0.31	-2
0172Edi	E-1,2-dichloroethene	0.04	-0.94
0066dim	1,2-dimethoxyethane	-4.19	-4.98
0046eth	1,2-ethanediol	-8.83	-8.74

0176pdi	1,4-dichlorobenzene	-1.39	-1.2
n016	1,2-ethanediamine	-8.03	-8.61
0062dio	1,4-dioxane	-5.83	-6.12
0203bro	1-bromo-1,2,2,2-tetrafluoroethane	0.77	-3.06
0201bro	1-bromo-1-chloro-2,2,2-trifluoroethane	0.29	-2.96
0202bro	1-bromo-2-chloroethane	-0.5	-2.11
0184bro	1-bromobutane	0.26	-1.1
0418bri	1-bromo-isobutane	0.47	-1.01
0185bro	1-bromopentane	0.46	-1.01
0182bro	1-bromopropane	0.1	-1.09
0049but	1-butanol	-4.54	-4.45
0025buta	1-butene	0.3	0.33
0032but	1-butyne	-0.07	-2.31
0205chl	1-chloro-2,2,2-trifluoroethane	0.41	-3.16
0209chl	1-chloro-2,2,2-trifluoroethyldifluoromethylether	-0.07	-4.8
0410clp	1-chloropentane	0.5	-1.07
0167chla	1-chloropropane	0.16	-1.12
0058hep	1-heptanol	-3.69	-4.33
0054hex	1-hexanol	-3.9	-4.35
0029hex	1-hexene	0.88	0.4
0034hex	1-hexyne	0.33	-2.12
0403thi	1-methylthymine	-11.27	-16.75
0133nit	1-nitrobutane	-3.41	-7.6
0131nit	1-nitropropane	-3.54	-7.67
0236oct	1-octanol	-3.49	-4.27
0052pen	1-pentanol	-4.07	-4.44
0027pen	1-pentene	0.57	0.39

0033pen	1-pentyne	0.04	-1.94
0138pro	1-propanethiol	-0.64	-1.69
0047pro	1-propanol	-4.69	-4.45
0207tri	2,2,2-trifluoroethanol	-4.62	-7.43
0214tri	2,2,2-trifluorethylvinylether	-0.43	-3.94
0416dcl	2,2',3'-trichlorobiphenyl	-2.66	-3.85
0014tri	2,2,4-trimethylpentane	2.02	1.54
0414dcl	2,2'-dichlorobiphenyl	-2.86	-3.55
0433pho	2,2-dichloroethenyldimethylphosphate	-5.38	-10.67
0011dim	2,2-dimethylpropane	1.08	1.47
0415dcl	2,3-dichlorobiphenyl	-2.78	-3.02
0013dim	2,4-dimethylpentane	1.77	1.59
0123dim	2,4-dimethylpyridine	-4.07	-4.87
0124dim	2,5-dimethylpyridine	-3.94	-4.7
0426dcl	2,6-dichlorobenzonitrile	-2.55	-8.3
0427dcl	2,6-dichlorothiobenzamide	-7.01	-12.23
0125dim	2,6-dimethylpyridine	-3.69	-4.34
0183bro	2-bromopropane	0.28	-1.31
0076but	2-butanone	-3.05	-6.38
0409clb	2-chlorobutane	0.44	-1.2
0411chp	2-chloropentane	0.75	-1.08
0168chl	2-chloropropane	0.32	-1.38
0230eth	2-ethylpyrazine	-5.82	-5.52
0082hep	2-heptanone	-2.44	-6.25
0080hex	2-hexanone	-2.6	-6.3
0146met	2-methoxyethanol	-6.92	-7.42
0147met	2-methoxyethanamine	-6.26	-7.25

0135met	2-methyl-1-nitrobenzene	-3.08	-6.49
n009	2-methylaniline	-4.69	-5.18
0012met	2-methylpentane	1.26	1.61
0010met	2-methylpropane	0.87	1.51
0024met	2-methylpropene	0.38	0.16
0117met	2-methylpyrazine	-6.32	-6.01
0119met	2-methylpyridine	-4.27	-4.61
0132nit	2-nitropropane	-3.07	-6.7
0239oct	2-octanone	-2.24	-6.19
0028Epe	E-2-pentene	0.32	0.51
0078pen	2-pentanone	-2.73	-6.29
0081dim	3,3-dimethylbutanone	-2.12	-5.56
0471dim	3,4-dimethylpyridine	-4.41	-5.49
0425dbr	3,5-dibromo-4-hydroxybenzonitrile	-5.16	-10.51
0571dim	3,5-dimethylpyridine	-4.12	-5.08
n015	3,aminoaniline	-8.47	-9.62
0417brp	3-bromopropene	-0.17	-2.03
0170chl	3-chloropropene	-0.2	-2.01
n010	3-methylaniline	-4.97	-5.28
0120met	3-methylpyridine	-4.49	-5
0079pen	3-pentanone	-2.57	-5.86
0428ami	4-amino-3,5,6-trichloropyridine-2-carboxylicacid	-9.97	-15.23
0574eth	4-ethylpyridine	-4.49	-5.04
0083hep	4-heptanone	-1.96	-5.57
0121met	4-methylpyridine	-4.63	-5.19
n011	4-methylaniline	-4.95	-5.15
0085non	5-nonanone	-1.72	-5.57

0402adn	9-methyladenine	-12.32	-14.08
0070eth	acetaldehyde	-3.57	-6.52
0233ethb	acetamide	-9	-12.19
0086eth	aceticacid	-5.71	-7.49
0075pro	acetone	-3.4	-6.92
0126eth	acetonitrile	-2.78	-8.74
0084met	acetophenone	-4.46	-6.65
0145pro	allylalcohol	-4.33	-4.62
0216amm	amonia	-4.07	-4.55
0118ani	aniline	-5.26	-5.23
0068ani	anisole	-3.5	-3.35
0042ant	anthracene	-4.36	-3.11
0105aze	azetidine	-6.08	-3.26
0074ben	benzaldehyde	-4.72	-6.63
n008	benzamide	-9.5	-11.58
0035ben	benzene	-2.1	-0.96
0129ben	benzonitrile	-2.88	-6.92
0213bis	bis(2-chloroethyl)sulfide	-1.05	-4.38
0186bro	bromobenzene	-2.05	-1.24
0180bro	bromoethane	-0.03	-1.36
0177bro	bromomethane	0.02	-1.42
0419brt	bromotoluene	-2.23	-3.21
0423brt	bromotrichloromethane	-1.08	-1.7
0197bro	bromotrifluoromethane	1.15	-0.8
0072but	butanal	-3.18	-5.81
0088but	butanoicacid	-4.87	-6.86
0128butb	butanonitrile	-2.17	-7.38

0148but	butenyne	-0.05	-2.22
0099but	butylacetate	-2.35	-5.76
0110but	butylamine	-3.41	-3.62
0174chl	chlorobenzene	-1.74	-1.07
0199chl	chlorodifluoromethane	-0.04	-2.72
0163chl	chloroethane	0	-1.4
0169chl	chloroethene	0.25	-0.7
0198chl	chlorofluoromethane	-0.52	-3.11
0162tri	chloroform	-0.79	-2.32
0160chl	chloromethane	0.07	-1.47
0424clp	chloropentaflouroethane	2.43	-0.56
0412clt	chlorotoluene	-2.25	-3.28
0020cis	cis-1,2-dimethylcyclohexane	1.05	1.69
0016cyc	cyclopropane	-0.41	0.38
0017cyc	cyclopentane	-0.23	1.57
0018cyc	cyclohexane	0.02	1.68
0026cyc	cyclopentene	-0.87	0.58
0051cyc	cyclopentanol	-4.02	-4.07
0077cyc	cyclopentanone	-3.83	-6.52
0178dib	dibromomethane	-0.96	-2.6
0161dic	dichloromethane	-0.58	-2.56
0438pho	diethyl2,4-dichlorophenylthiophosphate	-2.15	-6.89
0447pho	diethyl4-nitrophenylthiophosphonate	-4.18	-11.53
0223die	diethyldisulfide	-1.06	-2.22
0060dim	dimethylether	-2.57	-2.79
0140dim	dimethylsulfide	-0.22	-2.02
0111die	diethylamine	-3.12	-2.52

0421dfl	difluorodichloromethane	-2.7	-0.86
0444pho	dimethyl2,4,5-trichlorophenylthiophosphate	-3.07	-7.67
0445pho	dimethyl4-bromo-2,5-dichlorophenylthiophosphate	-5.81	-7.77
0440pho	dimethyl5-(4-chloro)bicyclo[3.2.0]heptylphosphate	-4.91	-10.06
0441pho	dimethyl4-nitrophenylthiophosphate	-1.31	-12.4
0141dim	dimethyldisulfide	-2.05	-2.58
0063die	diethylether	-0.16	-2.93
0142die	diethylsulfide	-4.76	-1.91
0104dim	dimethylamine	0.7	-2.72
0143dip	dipropylsulfide	0.16	-1.73
0115dip	dipropylamine	-2.5	-2.36
0002eth	ethane	0.58	1.46
0137ethb	ethanethiol	-0.76	-1.78
0045eth	ethanol	-5.06	-4.6
0021eth	ethene	0.67	0.28
0449pho	ethyl4-cyanophenylphenylthiophosphonate	-4.43	-12.09
0095eth	ethylacetate	-2.74	-5.97
0092ethb	ethylformate	-2.2	-5.98
n019	ethylperoxide	-5.12	-5.56
0246eth	ethylphenylether	-3.11	-3.23
0103eth	ethylamine	-3.77	-3.72
0037eth	ethylbenzene	-1.53	-1.03
0030eth	ethyne	0.41	-2.58
0157flu	fluorobenzene	-1.13	-1.63
0153flu	fluoromethane	-0.14	-2.26
0422ftc	fluorotrichloromethane	-0.05	-1.18
0408hex	hexachloroethane	-0.93	-2.35

0405hex	hexafluoroethane	3.23	-0.1
0090hex	hexanoicacid	-4.56	-6.79
0229hyd	hydrazine	-5.99	-7.37
0400hyd	H2	1.53	1.17
n017	hydrogenperoxide	-8.04	-6.88
0219hyd	hydrogensulfide	-1.5	-1.73
0048pro	isopropanol	-3.86	-4.48
0242dii	isopropylether	-0.93	-2.82
0056mcr	m-cresol	-5.94	-5.29
0001met	methane	1.22	1.34
0136met	methanethiol	-0.74	-1.94
0437pho	methyl3-methyl-4-thiomethoxyphenylthiophosphate	-6.38	-11.83
0093met	methylacetate	-3.09	-6.08
0240met	methylbenzoate	-3.62	-5.69
0096met	methylbutanoate	-2.26	-5.45
0091met	methylformate	-2.51	-6.15
0100met	methylhexanoate	-1.99	-5.41
0065met	methylisopropylether	-1.78	-2.7
0238met	methyloctanoate	-1.59	-5.35
0098met	methylpentanoate	-2.17	-5.48
n018	methylperoxide	-5.35	-5.58
0094met	methylpropanoate	-2.57	-5.6
0064met	methylpropylether	-1.98	-2.5
0228met	methylamine	-4.56	-3.75
0019met	methylcyclohexane	0.62	1.68
n005	methylhydrazine	-6.43	-6.77
0150mhy	m-hydroxybenzaldehyde	-8.8	-10.7

0149mor	morpholine	-7.68	-6.16
0039mxy	m-xylene	-1.54	-1.03
n014	N,N-dimethylaniline	-4.27	-3.2
0114NNd	N,N'-dimethylpiperazine	-8.47	-3.25
0041nap	naphthalene	-3.24	-2.05
0004nbu	n-butane	0.76	1.55
n013	N-ethylaniline	-4.39	-4.01
0006nhe	n-hexane	1.15	1.62
0007nhe	n-heptane	1.32	1.67
0134nit	nitrobenzene	-3.6	-6.79
0130nit	nitroethane	-3.93	-8.13
0506nit	nitromethane	-4.75	-8.79
234ENmb	E-N-methylacetamide	-7.85	-11.59
235ZNmb	Z-N-methylacetamide	-7.96	-11.59
0122Nme	N-methylaniline	-5.07	-4.18
0227Nme	N-methylmorpholine	-7.15	-5.14
0112Nme	N-methylpiperazine	-9.05	-5.05
0008noc	n-octane	1.51	1.73
0003pro	n-propane	0.58	1.5
0005npe	n-pentane	0.95	1.57
0413clt	o-chlorotoluene	-1.42	-0.97
0055ocr	o-cresol	-5.66	-4.83
0406oct	octafluoropropane	4.1	-0.17
0237oct	octanal	-2.29	-5.73
0442pho	O-ethylO'-4-bromo-2-chlorophenylS-propylphosphorothioate	-4.71	-11.95
0038oxy	o-xylene	-1.66	-1.18
0420pbr	p-bromotoluene	-1.79	-1.23

0215pbr	p-bromophenol	-6.18	-5.53
0057pcr	p-cresol	-5.92	-5.19
0187dib	p-dibromobenzene	-2.01	-1.46
0073pen	pentanal	-2.85	-5.91
0089pen	pentanoicacid	-4.76	-6.86
0101pen	pentylacetate	-2.16	-5.72
0113pen	pentylamine	-3.22	-3.59
0053phe	phenol	-6.19	-5.15
0218pho	phosphine	1.24	0.38
0151phy	p-hydroxybenzaldehyde	-9.05	-11.46
0109pip	piperazine	-9.56	-5.95
0225pipa	piperidine	-4.82	-2.28
0071proa	propanal	-3.28	-6.04
0087pro	propanoicacid	-5.19	-7
0022pro	propene	0.35	0.27
0127pro	propionitrile	-2.37	-7.76
0097pro	propylacetate	-2.5	-5.81
0106pro	propylamine	-3.61	-3.68
0031pro	propyne	-0.31	-2.42
0040pxy	p-xylene	-1.5	-0.95
0116pyr	pyridine	-4.82	-5.09
0108pyr	pyrrolidine	-5.4	-3.49
0023str	s-trans-1,3-butadiene	0.28	-0.52
0050met	t-butanol	-3.48	-4.41
0067but	t-butylmethylether	-1.33	-2.67
0204tet	tetrachloroethene	-0.23	-1.27
0200tet	tetrafluoromethane	2.6	-0.11

0061tet	tetrahydrofuran	-3.52	-3.25
0244tet	tetrahydropyran	-3.05	-2.82
0939tet	tetramethylsilane	2.68	0.91
0144thi	thioanisole	-2.52	-3.05
0245thi	thiophene	-1.69	-1.65
0139thi	thiophenol	-3.14	-3.06
0036tol	toluene	-1.82	-1.05
0173tri	trichloroethene	-0.17	-1.32
0179tri	bromoform	-1.69	-2.78
0221tri	triethylphosphate	-5.68	-11.89
0220tri	trimethylphosphate	-6.57	-12.13
0222tri	tripropylphosphate	-4.88	-11.19
0217wat	Water	-8.39	-6.32
n007	urea	-13.57	-14.55
c000	waterdimer	-14.97	-11.15
0044met	methanol	-5.12	-4.77

A 1.4 Solvation free energies calculated in ADF with DZP using Hartree Fock, for COSMO and SM12 in water. Reported in kcal/mol. Based on solvation energies reported by Marenich et al.

Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620. SI

Marenich, A. V; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. Minnesota Solvation Database. *Minnesota Solvation Database - version 2012, Univ. Minnesota, Minneap.* **2012**, 1–28

Label	Solute (short name)	SM12	COSMO
0001met	methane	1.21	1.33
0002eth	ethane	0.55	1.41
0003pro	n-propane	0.54	1.42
0004nbu	n-butane	0.72	1.46
0005npe	n-pentane	0.92	1.47
0006nhe	n-hexane	1.12	1.51
0007nhe	n-heptane	1.3	1.54
0008noc	n-octane	1.5	1.59
0010met	2-methylpropane	0.82	1.37
0011dim	2,2-dimethylpropane	1.02	1.23
0012met	2-methylpentane	1.23	1.45
0013dim	2,4-dimethylpentane	1.74	1.39
0014tri	2,2,4-trimethylpentane	2	1.27
0016cyc	cyclopropane	-0.46	0.32
0017cyc	cyclopentane	-0.3	1.48
0018cyc	cyclohexane	-0.03	1.57
0019met	methylcyclohexane	0.56	1.54
0020cis	cis-1,2-dimethylcyclohexane	1.01	1.5
0021eth	ethene	0.6	-0.03
0022pro	propene	0.28	-0.09
0023str	s-trans-1,3-butadiene	0.14	-1.19
0024met	2-methylpropene	0.27	-0.24
0025buta	1-butene	0.22	0.02
0026cyc	cyclopentene	-0.97	0.31
0027pen	1-pentene	0.49	0.05
0028Epe	E-2-pentene	0.26	0.17

0029hex	1-hexene	0.82	-0.01
0030eth	ethyne	0.24	-3.17
0031pro	propyne	-0.52	-2.71
0032but	1-butyne	-0.25	-2.65
0033pen	1-pentyne	-0.12	-2.28
0034hex	1-hexyne	0.18	-2.46
0035ben	benzene	-2.41	-1.94
0036tol	toluene	-2.19	-2.14
0037eth	ethylbenzene	-1.84	-2.15
0038oxy	o-xylene	-2.06	-2.25
0039mxy	m-xylene	-1.94	-2.11
0040pxy	p-xylene	-1.9	-1.99
0041nap	naphthalene	-3.93	-3.72
0042ant	anthracene	-5.3	-5.45
0044met	methanol	-5.73	-5.54
0045eth	ethanol	-5.6	-5.2
0046eth	1,2-ethanediol	-9.81	-10.13
0047pro	1-propanol	-5.19	-5.07
0048pro	isopropanol	-4.31	-5.21
0049but	1-butanol	-5.04	-5.08
0050met	t-butanol	-3.94	-5.02
0051cyc	cyclopentanol	-4.45	-4.84
0052pen	1-pentanol	-4.53	-5.26
0053phe	phenol	-6.79	-6.53
0054hex	1-hexanol	-4.36	-5.18
0055ocr	o-cresol	-6.24	-6.11
0056mcr	m-cresol	-6.57	-6.7
0057pcr	p-cresol	-6.52	-6.65
0058hep	1-heptanol	-4.14	-5.18
0060dim	dimethylether	-3.35	-3.42
0061tet	tetrahydrofuran	-4.21	-3.88
0062dio	1,4-dioxane	-6.87	-7.22
0063die	diethylether	-2.62	-3.41
0064met	methylpropylether	-2.62	-3

0065met	methylisopropylether	-2.4	-3.1
0066dim	1,2-dimethoxyethane	-5.2	-5.75
0067but	t-butylmethylether	-1.9	-3.02
0068ani	anisole	-4.31	-4.67
0070eth	acetaldehyde	-5.16	-9.26
0071proa	propanal	-4.72	-8.53
0072but	butanal	-4.59	-8.41
0073pen	pentanal	-4.26	-8.47
0074ben	benzaldehyde	-6.23	-9.95
0075pro	acetone	-4.84	-9.09
0076but	2-butanone	-4.33	-8.38
0077cyc	cyclopentanone	-5.13	-8.82
0078pen	2-pentanone	-3.95	-8.32
0079pen	3-pentanone	-3.71	-7.61
0080hex	2-hexanone	-3.84	-8.37
0081dim	3,3-dimethylbutanone	-3.27	-7.5
0082hep	2-heptanone	-3.65	-8.29
0083hep	4-heptanone	-2.99	-7.5
0084met	acetophenone	-5.79	-9.32
0085non	5-nonanone	-2.78	-7.5
0086eth	aceticacid	-6.8	-9.5
0087pro	propanoicacid	-6.1	-8.85
0088but	butanoicacid	-5.73	-8.75
0089pen	pentanoicacid	-5.63	-8.76
0090hex	hexanoicacid	-5.42	-8.69
0091met	methylformate	-3.9	-8.61
0092ethb	ethylformate	-3.48	-8.39
0093met	methylacetate	-4.27	-7.95
0094met	methylpropanoate	-3.59	-7.32
0095eth	ethylacetate	-3.82	-7.78
0096met	methylbutanoate	-3.21	-7.22
0097pro	propylacetate	-3.54	-7.65
0098met	methylpentanoate	-3.12	-7.27
0099but	butylacetate	-3.36	-7.63

0100met	methylhexanoate	-2.93	-7.21
0101pen	pentylacetate	-3.17	-7.62
0103eth	ethylamine	-4.01	-4.27
0104dim	dimethylamine	-5.08	-3.01
0105aze	azetidine	-6.43	-3.59
0106pro	propylamine	-3.83	-4.24
0107tri	trimethylamine	-5.23	-1.9
0108pyr	pyrrolidine	-5.71	-4
0109pip	piperazine	-10.01	-6.01
0110but	butylamine	-3.62	-4.18
0111die	diethylamine	-3.33	-2.91
0112Nme	N-methylpiperazine	-9.56	-5.08
0113pen	pentylamine	-3.42	-4.18
0114NNd	N,N'-dimethylpiperazine	-9.06	-4.12
0115dip	dipropylamine	-2.66	-2.78
0116pyr	pyridine	-5.68	-6.39
0117met	2-methylpyrazine	-7.24	-6.93
0118ani	aniline	-5.75	-6.54
0119met	2-methylpyridine	-5.09	-5.6
0120met	3-methylpyridine	-5.36	-6.09
0121met	4-methylpyridine	-5.55	-6.35
0122Nme	N-methylaniline	-5.6	-5.33
0123dim	2,4-dimethylpyridine	-4.96	-5.88
0124dim	2,5-dimethylpyridine	-4.76	-5.62
0125dim	2,6-dimethylpyridine	-4.52	-5.14
0126eth	acetonitrile	-4.63	-10.85
0127pro	propionitrile	-3.97	-9.67
0128butb	butanonitrile	-3.66	-9.19
0129ben	benzonitrile	-4.34	-9.34
0130nit	nitroethane	-5.54	-10.4
0131nit	1-nitropropane	-4.98	-9.87
0132nit	2-nitropropane	-4.45	-8.63
0133nit	1-nitrobutane	-4.84	-9.82
0134nit	nitrobenzene	-4.81	-8.99

0135met	2-methyl-1-nitrobenzene	-4.12	-8.5
0136met	methanethiol	-0.76	-2.19
0137ethb	ethanethiol	-0.78	-2.11
0138pro	1-propanethiol	-0.66	-2.06
0139thi	thiophenol	-3.47	-4.46
0140dim	dimethylsulfide	-0.24	-2.32
0141dim	dimethyldisulfide	-1.35	-3.3
0142die	diethylsulfide	-0.2	-2.09
0143dip	dipropylsulfide	0.14	-2.02
0144thi	thioanisole	-2.88	-4.3
0145pro	allylalcohol	-4.8	-5.49
0146met	2-methoxyethanol	-7.84	-8.4
0147met	2-methoxyethanamine	-7.04	-8.2
0148but	butenyne	-0.41	-2.87
0149mor	morpholine	-8.44	-6.68
0150mhy	m-hydroxybenzaldehyde	-10.45	-14.13
0151phy	p-hydroxybenzaldehyde	-10.91	-14.99
0153flu	fluoromethane	-1	-3.42
0154dif	1,1-difluoroethane	-1.02	-4.59
0157flu	fluorobenzene	-1.6	-2.92
0160chl	chloromethane	-0.1	-2.1
0161dic	dichloromethane	-0.78	-3.48
0162tri	chloroform	-0.88	-3.17
0163chl	chloroethane	-0.14	-2.04
0165tri	1,1,1-trichloroethane	-0.37	-1.95
0166tri	1,1,2-trichloroethane	-0.89	-4.38
0167chla	1-chloropropane	0.03	-1.78
0168chl	2-chloropropane	0.18	-2.02
0169chl	chloroethene	0.16	-1.29
0170chl	3-chloropropene	-0.35	-2.75
0171Zdi	Z-1,2-dichloroethene	-0.45	-2.74
0172Edi	E-1,2-dichloroethene	0	-1.72
0173tri	trichloroethene	-0.22	-2.06
0174chl	chlorobenzene	-2.03	-2.15

0175odi	1,2-dichlorobenzene	-1.87	-2.77
0176pdi	1,4-dichlorobenzene	-1.57	-2.26
0177bro	bromomethane	-0.04	-2
0178dib	dibromomethane	-1.01	-3.52
0179tri	bromoform	-1.7	-3.86
0180bro	bromoethane	-0.08	-1.97
0182bro	1-bromopropane	0.05	-1.72
0183bro	2-bromopropane	0.23	-1.92
0184bro	1-bromobutane	0.22	-1.72
0185bro	1-bromopentane	0.43	-1.65
0186bro	bromobenzene	-2.31	-2.41
0187dib	p-dibromobenzene	-2.17	-2.78
0197bro	bromotrifluoromethane	0.94	-2.09
0198chl	chlorofluoromethane	-1.19	-4.52
0199chl	chlorodifluoromethane	-0.66	-4.2
0200tet	tetrafluoromethane	0	-0.89
0201bro	1-bromo-1-chloro-2,2,2-trifluoroethane	0	-4.18
0202bro	1-bromo-2-chloroethane	-0.62	-3.17
0203bro	1-bromo-1,2,2,2-tetrafluoroethane	0.3	-4.58
0204tet	tetrachloroethene	-0.23	-1.9
0205chl	1-chloro-2,2,2-trifluoroethane	-0.15	-4.44
0206tri	1,1,2-trichloro-1,2,2-trifluoroethane	0.42	-2.37
0207tri	2,2,2-trifluoroethanol	-5.72	-8.8
0209chl	1-chloro-2,2,2-trifluoroethyldifluoromethylether	-0.94	-6.68
0211tri	1,1,1-trifluoropropan-2-ol	-3.22	-6.13
0212hex	1,1,1,3,3,3-hexafluoropropan-2-ol	-1.85	-6.47
0213bis	bis(2-chloroethyl)sulfide	-1.24	-5.91
0214tri	2,2,2-trifluorethylvinylether	-1.26	-5.41
0215pbr	p-bromophenol	-6.66	-7.05
0216amm	amonia	-4.35	-5.13
0217wat	water	-9.01	-7.18
0218pho	phosphine	1.32	-0.03
0219hyd	hydrogensulfide	-1.51	-1.89
0220tri	trimethylphosphate	-9.05	-14.32

0221tri	triethylphosphate	-7.73	-13.82
0222tri	tripropylphosphate	-6.74	-13.24
0223die	diethyldisulfide	-1.09	-2.9
0225pipa	piperidine	-5.1	-2.38
0227Nme	N-methylmorpholine	-7.94	-5.64
0228met	methylamine	-4.84	-4.19
0229hyd	hydrazine	-6.49	-8.12
0230eth	2-ethylpyrazine	-6.64	-6.36
0233ethb	acetamide	-10.83	-13.89
0234ENmb	E-N-methylacetamide	-9.57	-13.29
0235ZNmb	Z-N-methylacetamide	-9.76	-12.99
0236oct	1-octanol	-3.94	-5.12
0237oct	octanal	-3.67	-8.34
0238met	methyloctanoate	-2.54	-7.18
0239oct	2-octanone	-3.46	-8.26
0240met	methylbenzoate	-4.72	-7.74
0242dii	isopropylether	-1.4	-3.07
0244tet	tetrahydropyran	-3.67	-3.37
0245thi	thiophene	-1.82	-2.73
0246eth	ethylphenylether	-3.79	-4.41
0400hyd	H2	1.53	1.15
0401amia	1,1-dimethyl-3-phenylurea	-11.16	-13.16
0402adn	9-methyladenine	-14.11	-16.3
0403thi	1-methylthymine	-13.34	-19.76
0405hex	hexafluoroethane	3.07	-0.97
0406oct	octafluoropropane	3.96	-1.14
0407tet	1,1,1,2-tetrachloroethane	-0.98	-3.9
0408hex	hexachloroethane	-0.93	-2.72
0409clb	2-chlorobutane	0.31	-1.8
0410clp	1-chloropentane	0.39	-1.72
0411chp	2-chloropentane	0.64	-1.72
0412clt	chlorotoluene	-2.69	-5.07
0413clt	o-chlorotoluene	-1.69	-1.97
0414dcl	2,2'-dichlorobiphenyl	-3.32	-5.26

0415dcl	2,3-dichlorobiphenyl	-3.22	-4.73
0416dcl	2,2',3'-trichlorobiphenyl	-3.11	-5.61
0417brp	3-bromopropene	-0.26	-2.94
0418bri	1-bromo-isobutane	0.43	-1.6
0419brt	bromotoluene	-2.55	-4.94
0420pbr	p-bromotoluene	-2.06	-2.39
0421dfl	difluorodichloromethane	0.51	-1.84
0422ftc	fluorotrichloromethane	-0.18	-2.05
0423brt	bromotrichloromethane	-1.09	-2.47
0424clp	chloropentaflouroethane	2.25	-1.48
0425dbr	3,5-dibromo-4-hydroxybenzonitrile	-6.47	-13.67
0426dcl	2,6-dichlorobenzonitrile	-4	-10.66
0427dcl	2,6-dichlorothiobenzamide	-8.03	-14.19
0428ami	4-amino-3,5,6-trichloropyridine-2-carboxylicacid	-11	-17.36
0433pho	2,2-dichloroethenyldimethylphosphate	-7.28	-13.06
0437pho	methyl3-methyl-4-thiomethoxyphenylthiophosphate	-7.66	-14.58
0438pho	diethyl2,4-dichlorophenylthiophosphate	-3.12	-9.41
0440pho	dimethyl5-(4-chloro)bicyclo[3.2.0]heptylphosphate	-7.6	-12.07
0441pho	dimethyl4-nitrophenylthiophosphate	-6.84	-16.03
0442pho	O-ethylO'-4-bromo-2-chlorophenylS-propylphosphorothioate	-6.35	-14.68
0444pho	dimethyl2,4,5-trichlorophenylthiophosphate	-3.88	-10.24
0445pho	dimethyl4-bromo-2,5-dichlorophenylthiophosphate	-4.25	-10.47
0447pho	diethyl4-nitrophenylthiophosphonate	-5.82	-14.54
0449pho	ethyl4-cyanophenylphenylthiophosphonate	-6.3	-16.21
0471dim	3,4-dimethylpyridine	-5.35	-6.59
0506nit	nitromethane	-6.79	-11.26
0571dim	3,5-dimethylpyridine	-5.03	-6.16
0574eth	4-ethylpyridine	-5.37	-6.17
0939tet	tetramethylsilane	2.71	0.26
c000	waterdimer	-16.02	-12.25
n005	methylhydrazine	-6.95	-7.27
n006	1,1-dimethylhydrazine	-6.61	-6.39
n007	urea	-15.36	-16.23
n008	benzamide	-11.07	-13.82

n009	2-methylaniline	-5.19	-6.45
n010	3-methylaniline	-5.51	-6.68
n011	4-methylaniline	-5.44	-6.54
n013	N-ethylaniline	-4.84	-5.06
n014	N,N-dimethylaniline	-4.92	-4.17
n015	3,aminoaniline	-9.19	-11.3
n016	1,2-ethanediamine	-8.41	-9.76
n017	hydrogenperoxide	-8.7	-7.85
n018	methylperoxide	-6.13	-6.68
n019	ethylperoxide	-5.8	-6.56
n191	uracil	-16.97	-23.19
n200	5-fluorouracil	-16.04	-24.96
n201	5-trifluoromethyluracil	-15.2	-26.51
n202	5-chlorouracil	-16.01	-24.36
n203	5-bromouracil	-15.86	-24.55
test0001	1,3-diacetyloxypropan-2-ylacetate(glyceroltriacetate)	-10.45	-19.88
test0004	m-bis(trifluoromethyl)benzene	0.84	-5.56
test0005	N,N-dimethyl-p-methoxybenzamide	-9.8	-14.64
test0006	N,N,4-trimethylbenzamide	-7.55	-12.04
test0007	bis(2-chloroethyl)ether	-2.57	-5.49
test0008	1,1-diacetoxyethane	-6.7	-13.51
test0009	1,1-diethoxyethane	-2.87	-4.43
test0011	diethylpropanedioate	-7.18	-13.18
test0012	dimethoxymethane	-4.15	-5.23
test0013	ethyleneglycoldiacetate	-7.6	-12.38
test0014	1,2-diethoxyethane	-4.46	-5.29
test0016	phenylformate	-5.08	-10.18
test0017	imidazole	-9.7	-13.24
test1001	nitroglycol	-5.43	-13.98
test1002	1,2-dinitroxypropane	-5.39	-14.49
test1003	butylnitrate	-2.47	-6.61
test1004	2-butylnitrate	-2.05	-5.81
test1005	isobutylnitrate	-2.08	-6.34
test1006	ethyleneglycolmononitrate	-7.74	-11.23

test1007	2-chloro-N-(2,6-diethylphenyl)-N-(methoxymethyl)acetamide (**alachlor**)	-4.96	-10.98
test1008	2-methyl-2-(methylthio)propanalO-(N-methylcarbamoyl)oxime (**aldicarb**)	-8.02	-13.55
test1009	2-(ethylamino)-4-isopropylamino-6-methyl-thio-s-triazine (**ametryn**)	-8.83	-9.94
test1010	O,O-dimethylS-[(4-oxo-1,2,3-benzotriazin-3(4H)-yl)methyl]dithiophosphate (**azinphosmethyl**)	-8.38	-20.79
test1011	N-butyl-N-ethyl-2,6-dinitro-4-(trifluoromethyl)aniline (**benefin**)	-1.95	-14.36
test1012	a-[(4,6-dimethoxypyrimidin-2-ylcarbamoyl)sulfamoyl]-o-toluicacid (**bensulfuron**)	-30.68	-44.11
test1013	5-Bromo-3-sec-butyl-6-methyl-uracil (**bromacil**)	-9.72	-18.08
test1014	3a,4,7,7a-tetrahydro-2-[(trichloromethyl)thio]-1H-isoindole-1,3(2H)-dione (**captan**)	-9.2	-17.11
test1015	1-naphthylmethylcarbamate (**carbaryl**)	-11.08	-15.06
test1016	2,3-dihydro-2,2-dimethyl-7-benzofuranylmethylcarbamate (**carbofuran**)	-11.53	-16.81
test1017	S-4-chlorophenylthiomethylO,O-diethylphosphorodithioate (**carbophenothion**)	-4.17	-9.51
test1018	octachloro-4,7-methanohydroindane (**chlordane**)	-1.09	-8.09
test1019	phosphoricacid[(E)-2-chloro-1-(2,4-dichlorophenyl)vinyl]diethylester (**chlorfenvinphos**)	-7.53	-14.26
test1020	ethyl2-(4-chloro-6-methoxypyrimidin-2-ylcarbamoylsulfamoyl)benzoate (**chlorimuronethyl**)	-18.86	-31.24
test1021	trichloro(nitro)methane (**chloropicrin**)	-3.2	-7.88
test1022	O,O-diethylO-3,5,6-trichloro-2-pyridylphosphorothioate (**chlorpyrifos**)	-5	-10.91
test1023	S-(RS)-2-chloro-1-phthalimidoethylO,O-diethylphosphorodithioate (**dialifor**)	-9.32	-19
test1024	diethoxy-[(2-isopropyl-6-methyl-4-pyrimidinyl)oxy]-thioxophosphorane (**diazinon**)	-8.03	-12.72
test1025	3,6-dichloro-2-methoxybenzoicacid (**dicamba**)	-8.52	-10.81
test1027	N1,N1-diethyl-2,6-dinitro-4-trifluoromethyl-m-phenylenediamine (**dinitramine**)	-3.42	-15.08
test1028	(RS)-2-sec-butyl-4,6-dinitrophenol (**dinoseb**)	-10.32	-20.59
test1029	6,7,8,9,10,10-hexachloro-1,5,5a,6,9,9a-hexahydro-6,9-methano-2,4,3-benzodioxathiepine-3-oxide (**endosulfanalpha**)	-6.76	-18.83

	(1R,4S,4aS,5S,6S,7R,8R,8aR)-1,2,3,4,10,10-hexachloro-1,4,4a,5,6,7,8,8a-octahydro-6,7-epoxy-1,4:5,8-dimethanonaphthalene		
test1030	(**endrin**)	-5	-9.34
	O,O,O',O'-tetraethylS,S'-methylenebis(phosphorodithioate)		
test1031	(**ethion**)	-3.44	-16.03
	1,4,5,6,7,8,8-heptachloro-3a,4,7,7a-tetrahydro-4,7-methanoindene		
test1033	(**heptachlor**)	-1.42	-6.04
	3,5,5-trimethyl-2-cyclohexen-1-one		
test1034	(**isophorone**)	-4.86	-10.63
	1,2,3,4,5,6-hexachlorocyclohexane		
test1035	(**lindane**)	-1.47	-9.27
	2-(dimethoxyphosphinothioylthio)butanedioicaciddiethylester		
test1036	(**malathion**)	-9.31	-17.95
	N-methylcarbamicacid[1-(methylthio)ethylideneamino]ester		
test1037	(**methomyl**)	-10.03	-14.86
	methyl2-(4-methoxy-6-methyl-1,3,5-triazin-2-ylcarbamoylsulfamoyl)benzoate		
test1039	(**metsulfuronmethyl**)	-23.53	-36.63
	4-mesyl-2,6-dinitro-N,N-dipropylaniline		
test1040	(**nitralin**)	-6.87	-25.36
test1041	nitroxyacetone	-6.14	-13.85
	O,O-diethyl-O-4-nitro-phenylthiophosphate		
test1043	(**parathion**)	-5.99	-15.28
	S-propylbutyl(ethyl)thiocarbamate		
test1044	(**pebulate**)	-2.91	-6.99
	O,O-diethylS-ethylthiomethylphosphorodithioate		
test1045	(**phorate**)	-2.1	-10.14
	N-cyclopropylmethyl-2,6-dinitro-N-propyl-4-trifluoromethylaniline		
test1046	(**profluralin**)	-2.23	-14.11
	N2,N4-diisopropyl-6-methylthio-1,3,5-triazine-2,4-diamine		
test1047	(**prometryn**)	-7.92	-9.95
	N-(3,4-dichlorophenyl)propanamide		
test1048	(**propanil**)	-8.22	-12.05
	5-amino-4-chloro-2-phenyl-3-(2H)-pyridazinone		
test1049	(**pyrazon**)	-11.66	-20.61
	6-chloro-N,N'-diethyl-1,3,5-triazine-2,4-diamine		
test1050	(**simazine**)	-10.5	-11.31
	methyl2-(4,6-dimethylpyrimidin-2-ylcarbamoylsulfamoyl)benzoate		
test1051	(**sulfometuron-methyl**)	-20.43	-35.69
	3-t-butyl-5-chloro-6-methyluracil		
test1052	(**terbacil**)	-9.12	-16.22
test1053	N2-tert-butyl-N4-ethyl-6-methylthio-1,3,5-triazine-2,4-diamine(terbutryn)	-8.32	-9.69
	3-(4-methoxy-6-methyl-1,3,5-triazin-2-ylcarbamoylsulfamoyl)thiophene-2-carboxylicacid		
test1054	(**thifensulfuron**)	-24.27	-40.42

test1055	dimethyl(RS)-2,2,2-trichloro-1-hydroxyethylphosphonate (**trichlorfon**)	-10.2	-17.13
test1056	a,a,a-trifluoro-2,6-dinitro-N,N-dipropyl-p-toluidine (**trifluralin**)	-1.74	-13.69
test1057	S-propyldipropyl(thiocarbamate) (**vernolate**)	-3.07	-6.69
test1058	4-amino-4'-nitroazobenzene	-10.25	-15.17
test1059	1-amino-4-anilinoanthraquinone	-10.63	-14.43
test1060	1,4,5,8-tetraminoanthraquinone	-14.75	-19.64
test1061	1-amino-anthraquinone	-9.46	-13.53
test1063	(2-dimethylamino-5,6-dimethyl-pyrimidin-4-yl)N,N-dimethylcarbamate(pirimor,pirimicarb)	-12.31	-14.22
test2001	acetylsalicylicacid	-10.51	-15.07
test2003	butylparaben	-8.27	-12.11
test2004	caffeine	-13.4	-18.61
test2006	6-chlorouracil	-14.82	-20.46
test2007	cyanuricacid	-20.78	-25.45
test2010	diflunisal	-12.99	-19.79
test2011	ethylparaben	-8.72	-12.37
test2013	flurbiprofen(racemic)	-8.4	-12.92
test2015	hexachlorobenzene	-0.89	-4.16
test2017	ibuprofen(racemic)	-6.17	-10.1
test2019	ketoprofen(racemic)	-10.91	-16.9
test2020	methylparaben	-9.23	-12.51
test2021	naproxen	-10.8	-14.55
test2022	4-nitroaniline	-9.02	-14.45
test2023	octafluorocyclobutane	3.47	-5.12
test2024	pentachloronitrobenzene	-3.27	-9.53
test2025	phthalimide	-12.83	-15.98
test2026	propylparaben	-8.43	-12.22
test2027	sulfolane	-6.64	-17.18
test2029	trimethylorthotrifluoroacetate	-2.63	-5.79
test3001	paracetamol	-14.12	-17.54
test3002	N-(3-hydroxyphenyl)acetamide	-13.93	-17
test3003	fenbufen	-15.33	-22.73
test3004	N-(2-hydroxyphenyl)acetamide	-12.35	-14.65

test3005	phenacetin	-11.22	-15.33
test3007	2-methoxybenzoicacid	-9.36	-12.76
test3014	4-methoxybenzoicacid	-9.3	-11.59
test3015	3-methoxybenzoicacid	-9.32	-11.67
test3019	tolfenamicacid	-7.08	-8.93
test3020	diclofenacacid	-11.74	-19.01
test3021	flufenamicacid	-10.49	-19.21

A 1.5. Varying solvents. Solvation free energies calculated in ADF with TZP using B3LYP, for COSMO and SM12. Reported in kcal/mol. Based on solvation energies reported by Marenich et al.

Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620. SI

Marenich, A. V; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. Minnesota Solvation Database. *Minnesota Solvation Database - version 2012, Univ. Minnesota, Minneap.* **2012**, 1–28

Solvent	Label	Solute	SM12	COSMO
benzylalcohol	0008noc	n-octane	-2.28	-5.8
benzylalcohol	0036tol	toluene	-10.17	-4.94
benzylalcohol	0045eth	ethanol	-3.81	0.69
benzylalcohol	0062dio	1,4-dioxane	-5.41	-1.21
benzylalcohol	0069met	formaldehyde	-6.12	-4.35
benzylalcohol	0076but	2-butanone	-6.11	-5.1
benzylalcohol	0086eth	aceticacid	-5.15	-5.74
benzylalcohol	0110but	butylamine	-7.48	-6.79
benzylalcohol	0506nit	nitromethane	-4.76	-3.74
benzylalcohol	0533ben	benzylalcohol	-4.75	-7.6
cyclohexane	0003pro	n-propane	-8.36	-4.79
cyclohexane	0004nbu	n-butane	-4.62	-3.58
cyclohexane	0005npe	n-pentane	-4.35	-2.09
cyclohexane	0008noc	n-octane	-6.83	-1.94
cyclohexane	0018cyc	cyclohexane	-2.03	0.19

cyclohexane	0035ben	benzene	-2.6	0.23
cyclohexane	0036tol	toluene	-3.17	0.27
cyclohexane	0037eth	ethylbenzene	-4.89	0.39
cyclohexane	0038oxy	o-xylene	-4.39	0.31
cyclohexane	0039mxy	m-xylene	-4.7	-0.55
cyclohexane	0041nap	naphthalene	-5	-0.51
cyclohexane	0044met	methanol	-5.44	-0.48
cyclohexane	0045eth	ethanol	-5.29	-0.53
cyclohexane	0047pro	1-propanol	-5.29	-0.47
cyclohexane	0048pro	isopropanol	-7.13	-0.77
cyclohexane	0049but	1-butanol	-2.69	-2.16
cyclohexane	0050met	t-butanol	-3.8	-1.99
cyclohexane	0052pen	1-pentanol	-4.34	-1.94
cyclohexane	0053phe	phenol	-3.8	-1.9
cyclohexane	0054hex	1-hexanol	-4.92	-1.91
cyclohexane	0055ocr	o-cresol	-3.96	-1.77
cyclohexane	0056mcr	m-cresol	-5.36	-1.9
cyclohexane	0057pcr	p-cresol	-6.29	-2.08
cyclohexane	0058hep	1-heptanol	-5.94	-1.86
cyclohexane	0062dio	1,4-dioxane	-6.47	-1.85
cyclohexane	0063die	diethylether	-6.59	-2.05
cyclohexane	0068ani	anisole	-6.58	-2.01
cyclohexane	0074ben	benzaldehyde	-6.5	-1.82
cyclohexane	0075pro	acetone	-5.13	-2.35
cyclohexane	0076but	2-butanone	-3.51	-1.13
cyclohexane	0078pen	2-pentanone	-5.19	-1.31
cyclohexane	0079pen	3-pentanone	-6.15	-2.41
cyclohexane	0080hex	2-hexanone	-3.92	-2.69
cyclohexane	0081dim	3,3-dimethylbutanone	-4.41	-2.38
cyclohexane	0082hep	2-heptanone	-4.88	-2.32
cyclohexane	0084met	acetophenone	-4.87	-2.07
cyclohexane	0086eth	aceticacid	-5.49	-2.29
cyclohexane	0087pro	propanoicacid	-4.42	-2.05
cyclohexane	0093met	methylacetate	-6.06	-2.25
cyclohexane	0094met	methylpropanoate	-6.8	-2.43
cyclohexane	0095eth	ethylacetate	-4.76	-3.18
cyclohexane	0097pro	propylacetate	-5.18	-2.9
cyclohexane	0098met	methylpentanoate	-3.71	-2.53

cyclohexane	0099but	butylacetate	-4.14	-2.26
cyclohexane	0100met	methylhexanoate	-4.59	-2.38
cyclohexane	0101pen	pentylacetate	-5.21	-2.31
cyclohexane	0103eth	ethylamine	-5.23	-2.17
cyclohexane	0107tri	trimethylamine	-5.78	-2.27
cyclohexane	0111die	diethylamine	-5.8	-2.12
cyclohexane	0116pyr	pyridine	-6.36	-2.23
cyclohexane	0118ani	aniline	-2.13	-1.76
cyclohexane	0119met	2-methylpyridine	-2.99	-0.82
cyclohexane	0120met	3-methylpyridine	-3.07	-1.15
cyclohexane	0121met	4-methylpyridine	-4.5	-1.93
cyclohexane	0122Nme	N-methylaniline	-5.27	-2.1
cyclohexane	0125dim	2,6-dimethylpyridine	-4.82	-1.73
cyclohexane	0126eth	acetonitrile	-4.8	-1.93
cyclohexane	0129ben	benzonitrile	-4.84	-2.01
cyclohexane	0131nit	1-nitropropane	-5.79	-1.61
cyclohexane	0134nit	nitrobenzene	-5.16	-1.47
cyclohexane	0135met	2-methyl-1-nitrobenzene	-2.23	-3.42
cyclohexane	0138pro	1-propanethiol	-4.64	-2.72
cyclohexane	0144thi	thioanisole	-4.36	-2.98
cyclohexane	0150mhy	m-hydroxybenzaldehyde	-5.72	-2.72
cyclohexane	0151phy	p-hydroxybenzaldehyde	-5.83	-2.53
cyclohexane	0157flu	fluorobenzene	-3.41	-0.95
cyclohexane	0165tri	1,1,1-trichloroethane	-5.88	-1.2
cyclohexane	0173tri	trichloroethene	-7.72	-3.9
cyclohexane	0174chl	chlorobenzene	-7.94	-4.12
cyclohexane	0176pdi	1,4-dichlorobenzene	-4.09	-0.89
cyclohexane	0186bro	bromobenzene	-4.38	-0.96
cyclohexane	0207tri	2,2,2-trifluoroethanol	-3.61	-0.93
cyclohexane	0215pbr	p-bromophenol	-5.24	-0.66
cyclohexane	0217wat	water	-5.74	-0.76
cyclohexane	0220tri	trimethylphosphate	-5.83	-0.72
cyclohexane	0221tri	triethylphosphate	-3.65	-3.06
cyclohexane	0222tri	tripropylphosphate	-7.43	-2.28
cyclohexane	0240met	methylbenzoate	-3.97	-2.96
cyclohexane	0244tet	tetrahydropyran	-6.22	-4.45
cyclohexane	0246eth	ethylphenylether	-8.86	-4.14
cyclohexane	0421dfl	difluorodichloromethane	-10.63	-3.87

cyclohexane	0422ftc	fluorotrichloromethane	-6.44	-2.19
cyclohexane	0425dbr	3,5-dibromo-4-hydroxybenzonitrile	-4.88	-1.17
cyclohexane	0506nit	nitromethane	-6.01	-1.09
cyclohexane	0515dim	N,N-dimethylformamide	-2.47	-0.79
cyclohexane	0579pyy	pyrrole	-3.73	-0.94
cyclohexane	0582qui	quinoline	-7.45	-4.13
cyclohexane	n008	benzamide	-3.57	-3.45
cyclohexane	n009	2-methylaniline	-6.37	-0.77
cyclohexane	n010	3-methylaniline	-5.54	-1.95
cyclohexane	n011	4-methylaniline	-5.56	-2.05
cyclohexane	test4001	iodobenzene	-5.52	-2.01
dimethylsulfoxide	0008noc	n-octane	-7.01	-12.9
dimethylsulfoxide	0036tol	toluene	-3.09	0.68
dimethylsulfoxide	0045eth	ethanol	-4.49	-1.36
dimethylsulfoxide	0062dio	1,4-dioxane	-7.08	-4.85
dimethylsulfoxide	0076but	2-butanone	-6.63	-5.73
dimethylsulfoxide	0503dim	dimethylsulfoxide	-5.32	-6.54
dimethylsulfoxide	0506nit	nitromethane	-5.13	-8.48

A 1.6. Ionic solvation free energies calculated in ADF with AUG TZP using M06-2X, for COSMO and SM12. Reported in kcal/mol. Based on solvation energies reported by Marenich et al.

Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620. SI

Marenich, A. V; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. Minnesota Solvation Database. *Minnesota Solvation Database - version 2012, Univ. Minnesota, Minneap.* **2012**, 1–28

Cationic

Label	Compound	SM12	COSMO
i029	1-aminonaphthalene	-62.18	-66.92
i019	2-methylaniline	-63.32	-67.15
i023	3-aminoaniline	-65.35	-67.76
i125	3-chloroaniline	-66.54	-72.87
i093	4-methoxyaniline	-63.88	-68.51
i021	4-methylaniline	-63.68	-67.25
i126	4-chloroaniline	-66.23	-73.38
i027	4-methyl-N,N-dimethylaniline	-49.95	-53.41
i094	4-nitroaniline	-73.75	-87
i098	acetamide	-72.32	-68.1
c054	acetone	-61.29	-57.98
i040	acetonitrile	-70.28	-68.27
i056	acetophenone	-54.51	-55.35
i008	allylamine	-69.91	-69.35
c047	ammonia	-76.95	-70
i018	aniline	-66.28	-69.07
i034	azacycloheptane	-57.1	-58.69
i031	azetidine	-66.45	-64.18
i030	aziridine	-72.51	-67.24
i099	benzamide	-61.43	-61.87
i007	cyclohexanamine	-63.34	-64.41
i012	diallylamine	-54.78	-59.36
i053	diethylether	-59.29	-58.35
i010	diethylamine	-58.68	-59.58
i052	dimethylether	-69.36	-65.82
i106	dimethylsulfide	-62.21	-63.78
i112	dimethylsulfoxide	-61.22	-63.39

i009	dimethylamine	-68	-65.56
i011	di-n-propylamine	-54.6	-58.07
c051	ethanol	-66.27	-61.3
i048	hydrazine	-83.93	-79.44
i005	isopropylamine	-67.88	-66.45
c050	methanol	-70.57	-64.33
i003	methylamine	-78.21	-72.67
i095	morpholine	-66.6	-67.54
i026	N,N-dimethylaniline	-52.07	-54.9
i028	N,N-diethylaniline	-47.07	-50.09
i025	N-ethylaniline	-54.61	-58.23
i024	N-methylaniline	-58.3	-61.35
i004	n-propylamine	-69.92	-68.65
i039	piperazine	-64.12	-62.83
i033	piperidine	-59.77	-60
i036	pyridine	-60.29	-59.41
i035	pyrrole	-63.61	-61
i032	pyrrolidine	-62.95	-61.88
i037	quinoline	-52.45	-54.22
i006	tert-butylamine	-63.87	-63.61
i013	trimethylamine	-60.11	-59.44
i014	triethylamine	-50.01	-51.29
i015	tri-n-propylamine	-45.12	-49.56
c088	water	-81.52	-71.35
i020	3-methylaniline	-63.81	-67.36

Anionic

Label	Compound	SM12	COSMO
c122	1,1,1,3,3,3-hexafluoropropan-2-ol	-57.32	-61.73
c078	1,2-ethanediol	-77.64	-78.98
i109	1-propanethiol	-64.82	-75.21
c067	1-propanol	-72.75	-76.22
c121	2,2,2-trifluoroethanol	-65.66	-67.66
c069	2-butanol	-69.4	-72.26
i123	2-chlorophenol	-62.27	-65.65
c073	2-methoxyethanol	-73.71	-77.6
i075	2-methylphenol	-64.17	-67.9
i100	2-nitrophenol	-59.6	-64.8
i080	3-hydroxyphenol	-69.01	-71.19
i076	3-methylphenol	-65.42	-70.5
i101	3-nitrophenol	-58.02	-61.33

i124	4-chlorophenol	-61.58	-63.52
i081	4-hydroxyphenol	-70	-74.7
i077	4-methylphenol	-65.55	-70.93
i104	4-nitroaniline	-56.82	-57.16
i102	4-nitrophenol	-56.84	-56.67
i084	acetaldehyde	-77.74	-73.7
i105	acetamide	-78.44	-74.8
i059	aceticacid	-76.35	-77.3
i041	acetonitrile	-72.57	-68.78
c001	acetylene	-74.39	-73.34
i062	acrylicacid	-72.14	-75.88
i044	aniline	-65.75	-67.59
i064	benzoicacid	-66.2	-74.08
c072	benzylalcohol	-67.08	-72.13
i119	chloroaceticacid	-68.92	-71.01
i043	cyanamide	-74.77	-73.46
i120	dichloroaceticacid	-65.35	-66.41
i113	dimethylsulfoxide	-66.27	-71.67
i045	diphenylamine	-55.2	-58.66
i108	ethanethiol	-67.31	-75.27
c066	ethanol	-74.99	-76.47
c083	ethylhydroperoxide	-73.67	-78.41
i058	formicacid	-79.7	-75.58
i061	hexanoicacid	-68.5	-76.3
c116	hydrobromicacid	-61.77	-65.44
c115	hydrochloricacid	-72.58	-70.47
c114	hydrofluoricacid	-86.46	-83.19
c046b	hydrogencyanide	-74.35	-67.88
c090	hydrogenperoxide	-80.14	-74.83
c111	hydrogensulfide	-71.05	-68.48
c091	hydroperoxylradical	-79.05	-69.98
c068	isopropanol	-71.77	-74.79
i107	methanethiol	-71.89	-75.81
c065	methanol	-79.45	-77.63
c082	methylhydroperoxide	-76.08	-78.03
i103	nitromethane	-74.49	-76.37
i074	phenol	-66.86	-69.74
i060	propanoicacid	-72.77	-76.39
i063	pyruvicacid	-71.12	-72.07
c070	t-butanol	-68.66	-73.41
i110	thiophenol	-60.69	-67.17
i118	trifluoroaceticacid	-64.06	-63.29

c089	water	-89.61	-83.82
i086	3-pentanone	-66.45	-70.47
i085	acetone	-73.54	-74.13
c071	allylalcohol	-71.91	-75.6
i117	chloroform	-55.82	-53.35

Appendix 2

Periodic Solvation in ADF-BAND

Craig A. Peeples[a], Georg Schreckenbach[a]*, Pier Herman Theodoor Philipsen[b]

[a]Department of Chemistry, University of Manitoba

Winnipeg, MB, R3T 2N2, Canada

[b]Software for Chemistry & Modeling NV, Theoretical Chemistry, VU University Amsterdam, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands

schrecke@cc.umanitoba.ca, philipse@scm.com

Contents:

Table A 2.1: Free energies of solvation comparison for molecular ADF and periodic *zero*-dimensional ADF-BAND.

Table A 2.1. Solvation free energies calculated in ADF-BAND with TZP with PBE, for COSMO and SM12 in water. Reported in kcal/mol. Based on solvation free energies reported by Marenich et al.

Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620. SI

Marenich, A. V; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. Minnesota Solvation Database. *Minnesota Solvation Database - version 2012, Univ. Minnesota, Minneap.* **2012**, 1–28

Nicholls et al. *J. Med. Chem.* **2008**, 51, 769

Guthrie, *J. Phys. Chem. B* **2009**, 113, 4501

Gaballe et al., *J. Comput. Aided Mol. Des.* **2010**, 24, 259

Zhu et. al., *J. Chem. Phys.* **1998**, 109, 9117

Label	Solute	**SM12 (ADF)**	**SM12 (BAND)**	COSMO (ADF)	COSMO (BAND)
0001met	**Methane**	**-0.27**	**-0.28**	-0.32	-0.29
0002eth	**Ethane**	**-0.50**	**-0.50**	-0.36	-0.30
0045eth	**Ethanol**	**-3.64**	**-3.64**	-5.78	-5.77
0075pro	**Acetone**	**-3.51**	**-3.50**	-8.05	-7.99
0108pyr	**Pyrrolidine**	**--2.59**	**-2.59**	-5.14	-4.96
0125dim	**2,6-dimethylpyridne**	**-3.93**	**-3.89**	-5.71	-5.80
0128butb	**Butanonitrile**	**-5.18**	**-5.18**	-9.37	-9.47

0129ben	**Benzonitrile**	**-5.22**	**-5.23**	-9.15	-8.57
0217wat	**Water**	**-5.10**	**-5.10**	-7.37	-7.31
0233ethb	**Acetamide**	**-6.85**	**-6.85**	-12.98	-12.91
n014	**N,N-dimethylaniline**	**-4.48**	**-4.45**	-4.54	-4.82
n017	**Peroxide**	**-5.55**	**-5.54**	-7.87	-7.64
test1012	**Bensulfuron**	**-19.55**	**-19.40**	-35.24	-35.91