University of Strathclyde

Department of Chemical and Process Engineering

Improving Predictions of Solvation Free Energies from Non-Polarisable Models by Applying Polarisation Corrections

Doctoral Thesis

Author:
Maria Cecilia Barrera

Supervisor:
Dr. Miguel Jorge

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

August 5, 2021

Abstract

Classical non-polarisable models, normally based on a combination of Lennard-Jones (LJ) sites and point charges, are extensively used to model thermodynamic properties of fluids. An important shortcoming of this class of models is that they do not explicitly account for polarisation effects - i.e. a description of how the electron density responds to changes in the molecular environment. Instead, polarisation is implicitly included into the parameters of the model, usually by fitting to pure liquid properties (e.g. density). A problem arises when trying to describe thermodynamic properties that involve a change of phase (e.g. enthalpy of vaporisation), solutions/mixtures (e.g. solvation free energies), or properties that directly depend on the electronic response of the medium (e.g. dielectric constant). Fully polarisable models present a natural route for addressing these limitations, but at the price of a much higher computational cost. The main goal of this thesis is to obtain a non-polarisable force field for alcohols, amines and ketones able to predict both pure liquid properties and solvation free energies with a high degree of accuracy through the use of *post-facto* polarisation corrections. These corrections are applied to the properties computed using the non-polarisable force field in order to account for the effects of polarisation, and thus, increase the model's accuracy while maintaining its computational efficiency.

This work is part of a larger project which end goal is to predict the solubility of drug molecules (e.g. paracetamol). These molecules usually contain hydroxyl, amino and carbonyl groups, and thus, this thesis focuses on molecules with these functional groups. Aromatic rings are another functional group present in most drug molecules, however, they are not studied here due to time limitations. Alcohols and amines are interesting

from a fundamental point of view as they are the simplest molecules that combine a hydrophobic moiety with a hydrogen-bonding functional group. Also, alcohols and ketones are widely used as solvents and amines are used in CO_2 adsorption/desorption processes designed to decrease CO_2 emissions.

The model developed in this thesis is called PolCA, standing for "Polarisation-Consistent Approach", and it is an extension of the modified TraPPE force field for hydrocarbons proposed by Jorge [1] that eliminates systematic deviations from experimental solvation free energies. The new amino, hydroxyl and carbonyl parameters were fitted to several pure-component experimental properties including the density and enthalpy of vaporisation, and in some cases also self-solvation free energies. The optimization was carried out using meta-models that predict how the simulated properties change with the input parameters, allowing for a better exploration of the force field parameters' space.

The PolCA force field for alcohols can accurately predict methanol to decanol's densities, diffusion constants (except for methanol), enthalpies of vaporisation, free energies of self-solvation, dielectric constants and solvation free energies in hexadecane. PolCA also does a very good job at predicting the densities, enthalpies of vaporisation and free energies of self-solvation of linear and branched primary amines, and its predicted solvation free energies of linear primary amines in hexadecane are in very good agreement with experimental data. However, it greatly overpredicts the dielectric constant of methylamine and significantly overpredicts other linear and branched amines' dielectric constants. Furthermore, PolCA can accurately predict the densities, enthalpies of vaporisation, diffusion constants and self-solvation free energies of propanone to 2-decanone (except for butanone and 2-pentanone's densities which are underpredicted), however, it greatly overpredicts ketones' dielectric constants and solvation free energies in hexadecane (more negative values).

Lastly, PolCA was used to calculate solvation free energies of amines and ketones in octanol and multifunctional compounds' densities and dielectric constants to test its transferability. Our force field was not able to simultaneously predict solvation free energies in hexadecane and octanol, and thus a re-parameterisation will be carried out

in future work using polarisation corrections obtained with more accurate methods. Nonetheless, the results obtained in this work show that the approach proposed here is very promising since they significantly improve agreement with experiment for the dielectric constant and solvation free energies of alcohols and amines in hexadecane.

Acknowledgements

I would like to especially thank my supervisor Miguel who is not only a very smart and knowledgeable academic but also a great person who believed in me even when I did not. Also, I would like to dedicate this work to my partner David who constantly makes me laugh, encourages me to pursue my dreams and helps me fight my migraines by finding the best tinted glasses and migraine and gastritis-friendly foods.

I am extremely grateful for my family (René, Ines, Pablo, Pao, Vero and Tere) who have supported me during this journey, and I would like to particularly thank René and Vero for financially helping me during the start of my PhD. In addition, I would like to thank David's family for making me feel as one of their own, and I am particularly indebted to Anne and Alex for letting me stay at their home during the pandemic.

Furthermore, I would like to thank Andrew for helping me during the start of my PhD and Russell, Georgia, João and Neret for providing me with a place to stay when I needed it.

I would also like to thank my reviewers Javier and Carlos for the interesting discussion and constructive feedback during what was a very pleasant viva experience.

Last but not least, I would like to thank the Funds for Women Graduates for partially funding my PhD since I would not have been able to complete this work without their financial support.

Contents

List of Figures

List of Tables

List of Symbols and Acronyms

Symbol	Meaning
$< ... >$	Average over all possible quantum states of a system (ensemble).
A	Helmholtz free energy.
a	Attractive term (EoS).
b	Mixture parameter.
c	Molar concentration.
$C6$	Attractive LJ parameter.
$C12$	Repulsive LJ parameter.
$C_{Berendsen}$	Berendsen correction.
C_n	Fourier coefficients.
C_{dist}	Distortion term of the *post-facto* polarisation correction.
C_{elec}	Electronic term of the *post-facto* polarisation correction.
C_{pol}	*Post-facto* polarisation correction.
C_{kl}	Prefactor that guarantees that the Mie potential's minimum is equal to $-\epsilon_{kl}$.
D	Variance.
D_A	Self-diffusion constant.
E	Energy of the system.
$\mathrm{e}^{-\beta H}$	Boltzmann factor.
f	Fugacity.
$\mathbf{F}_i$	Net force acting on particle i.
$F(X)$	Objective function.
$f(x)$	Fermi function.
f_k	Value predicted by the meta-model.
G_{CDS}	Cavitation-dispersion-solvent-reorganisation contribution to the solute-solvent interaction energy.
H	Same symbol is used for the Hamiltonian and the enthalpy.
h	Planck's constant.
h^E	Excess enthalpy.
k	Same symbol is used for the dipole moment scaling factor, the kinetic energy and the inverse Debye screening length of the medium outside the cut-off.
k_r	Force constant for the bonds.
k_θ	Force constant for the angles.
k_B	Boltzmann constant.
$K_{kl,ab}$	Bonding volume used to describe $r^c_{kl,ab}$.

L	Simulation box length.
M	Total dipole moment of the system.
M_w	Molecular weight.
N	Number of particles in the system.
n_D	Refraction index of the medium.
N_f	Number of degrees of freedom.
N_{mol}	Number of molecules used in the optimisation.
N_{prop}	Target properties used in the optimisation.
NPT	Ensemble with constant pressure, temperature and number of particles.
NVE	Ensemble with constant volume, energy and number of particles.
NVT	Ensemble with constant volume, temperature and number of particles.
P	Pressure.
p	Momentum of a particle.
P_c	Critical pressure.
P_{ow}	Octanol/water partition coefficient.
$P_S(\sigma)$	Charge density distribution (σ-profile).
Q	Partition function.
q	Partial charge.
Q_0	Bare point charges.
Q_2	Predictivity coefficient.
R	Ideal gas constant.
r	Coordinates of a particle.
r_0	Equilibrium bond-length.
r_{ij}	Separation for the pair of atoms i and j.
r_c	Cut-off distance for intermolecular interactions.
r_{kl}	Distance between the centres of segments k and l.
$r_{kl,ab}$	Distance between the sites of type a and b located in segments k and l, respectively.
$r^c_{kl,ab}$	Cut-off range for the interaction between sites of type a and b located in segments k and l, respectively.
S	Entropy of the system.
s	Relative entropy.
S_0	Intrinsic solubility.
S_i	First-order Sobol's indices.
S_T	Total Sobol's indices.
T	Temperature.
t	Same symbol is used for time and step length.
T_c	Critical temperature.
T_m	Melting point.
U	Potential energy of a system.
u	Potential energy of a system.
u^{RF}	Reaction-field correction.
V	Volume.
V_m	Molar volume of the crystal.
X	Coordinates of the coded values used in the optimisation.
x	Mole fraction.
x_c	Initial guess used in the optimisation routine.

$\overline{y}$	Mean value of $y_{sim}(X_i)$ on the whole sample.
y_{exp_k}	Experimental value for property k.
y_{sim}	Simulated value obtained using MD.
α	Electronic polarisability.
β	$1/(k_B\ T)$.
ΔG	Gibbs free energy.
$\Delta G_{dissolution}$	Dissolution free energy.
$\Delta G_{hydration}$	Hydration free energy.
ΔG_{sol}	Solvation free energy.
$\Delta G_{solvation}$	Solvation free energy.
$\Delta G_{sublimation}$	Sublimation free energy.
$\Delta G_{transfer}$	Transfer free energy of the solute from an organic solvent to water.
$\Delta \mathrm{H_{vap}}$	Enthalpy of vaporisation.
$\Delta H_{fus}(T_m)$	Solute's enthalpy of fusion at its normal melting point.
ϵ	Static dielectric constant.
ϵ_0	Vacuum permittivity.
ϵ_1	Relative permittivity of the medium.
ϵ_2	Relative permittivity of the medium outside the cut-off.
ϵ_{el}	High-frequency permittivity.
ϵ_{kl}	Unlike dispersion energy between groups k and l.
$\epsilon^{HB}_{kl,ab}$	Unlike hydrogen bond energy between sites a and b located in groups k and l, respectively.
$\mathrm{\varepsilon_{ij}}$	LJ well-depth for the pair of atoms i and j.
ϵ_{sim}	Simulated dielectric constant.
ϵ_{sol}	Static dielectric constant of the solvent.
λ	Coupling parameter used to calculate solvation free energies.
λ^a_{kl}	Attractive exponent of the interactions between groups k and l.
λ^r_{kl}	Repulsive exponent of the interactions between groups k and l.
μ	Same symbol is used for the dipole moment and the chemical potential.
μVT	Ensemble with constant volume, temperature and chemical potential.
∇	Gradient.
$\Omega(E,V,N)$	Number of eigenstates with energy E of a system of N particles in a volume V.
ϕ	Dihedral angle.
$\mathrm{\varrho}$	Density.
$\mathrm{\sigma_{ij}}$	LJ size for the pair of atoms i and j.
σ_{kl}	Unlike segment diameter between groups k and l.
θ	Measured bending angle.
θ_0	Equilibrium bending angle.
γ	Activity coefficient.
γ^C_i	Combinatorial part of the activity coefficient.
γ^R_i	Residual part of the activity coefficient.
γ^∞_i	Activity coefficient at infinite dilution.

Acronym	Meaning
ATB	Automated Topology Builder.
ANNs	Artificial neural networks.
API	Active pharmaceutical ingredient.
COSMO	Conductor-like screening model.
COSMO-RS	Conductor-like screening model for real solvents.
DNNs	Deep neural networks.
GC	Group Contribution.
CG-MD	Coarse-grained molecular dynamics.
GC-EoS	Group contribution equation of state.
GROMOS	The GROningen MOlecular Simulation computer program package.
GSE	General solubility equation.
FQ	Fluctuating charges.
LJ	Lennard-Jones.
MC	Monte Carlo.
MD	Molecular dynamics.
MLL	Multiple model learning.
MLR	Multiple linear regression.
OPLS	Optimized Potentials for Liquid Simulations.
PCM	Polarisable continuum model.
PME	Particle-Mesh Ewald.
PolCA	Polarisation-Consistent Approach.
PSRK	Predictive Soave-Redlich-Kwong.
QM	Quantum mechanics.
QSPR	Quantitative Structure Property Relations.
RMSD	Root-mean-square-deviation.
SAFT	Statistical associating fluid theory.
SLE	Solid-liquid equilibria.
SMD	Solvation model based on density.
SRK	Soave-Redlich-Kwong.
TraPPE	Transferable Potentials for Phase Equilibria.
UA	United-atom.
UNIFAC	Universal quasi-chemical functional group activity coefficient.
VLE	Vapour-liquid equilibria.

Chapter 1

Introduction

Solubility is a fundamental property in the manufacture and formulation of drugs [2]. For example, drug absorption from the gastrointestinal tract depends strongly on the solubility of orally administered drugs. Nowadays, most drug candidates have a low solubility in water, and as a consequence, their absorption has to be improved through a suitable formulation approach. About 40 % of drugs in the market and nearly 90 % of drugs in development are poorly water-soluble [3]. Cosolvents can be used to enhance the solubility of non-polar solutes by several orders of magnitude. These compounds commonly contain hydrogen bonding groups, which interact strongly with water, and non-hydrogen bonding groups. Dimethylacetamide, ethanol and propylene glycol are widely used in the pharmaceutical industry as cosolvents [4]. Consequently, knowing the solubility of the drug in water and other solvents during its formulation process is extremely important [5, 6]. In addition, the octanol/water partition coefficient is a very important property in the pharmaceutical industry, and for example, it can be used to describe a drug's ability to diffuse through lipids [7].

Experimental determination of drug molecules' solubility is very laborious and time-consuming, especially when the solubility is very low [8, 9]. Brown et al. [10] have presented a workflow for the manufacture of active pharmaceutical ingredients (API) using seeded cooling continuous crystallisation, where the challenges of solvent selection are discussed. At the early stages of the process design, possible solvents are selected

from solvent libraries based on different properties like API's solubility, boiling point, viscosity, environmental impact, cost and reactivity with the solute. Afterwards, these solvents' solubility curves and stabilities are studied using solutions of known concentration at different temperatures to find the most suitable candidates, which will go through to the validation stage. The properties considered during the solvent selection include the yield, operating temperature range, likelihood of fouling, formation of agglomerates, particle shape, purity and crystal morphology, among others. Due to all these requirements, it is not always possible to find suitable single-component solvents, and hence, solvent mixtures are used instead. These steps are time-consuming, and thus, there is usually a compromise between the resources spent and the accuracy obtained. For example, relevant impurities' solubilities are generally not included during the selection process, even though crystallisation is used as a purification technique. [10] As a result, much time and effort could be saved by accurately predicting the solubility using computational methods [11].

Computational chemistry is already being used in the design and optimisation of new drugs to help save time and money. For example, Ngo et al. [12] studied 33 drugs using molecular dynamics (MD) simulations to detect potential inhibitors of the coronavirus SARS-CoV-2's main protease (Mpro). Thirteen of these compounds, most of which are already used to treat various diseases in humans, were predicted to have a strong binding affinity to SARS-CoV-2 Mpro. One of these molecules was delamanid, an antituberculosis drug, and thus, it has a higher probability of successfully treating COVID-19 since both tuberculosis and COVID-19 are lung diseases.

1.1 Thermodynamics of Solubility

Solubility is the amount of solute that can be dissolved in a specific solvent. For solid solutes, the molecule needs to dissociate from its solid form before dissolving in the solvent. The change in the Gibbs free energy (ΔG) associated with the sublimation process is always positive, and its magnitude depends on the intermolecular bonds' strength and complex interaction patterns present in the solid. Also, the solvent needs to create a cavity able to accommodate the solute, and the energy needed to accomplish

this step increases with the cavity's size. Finally, the solute can be inserted into the cavity and interact with the surrounding solvent molecules (negative ΔG). [13]

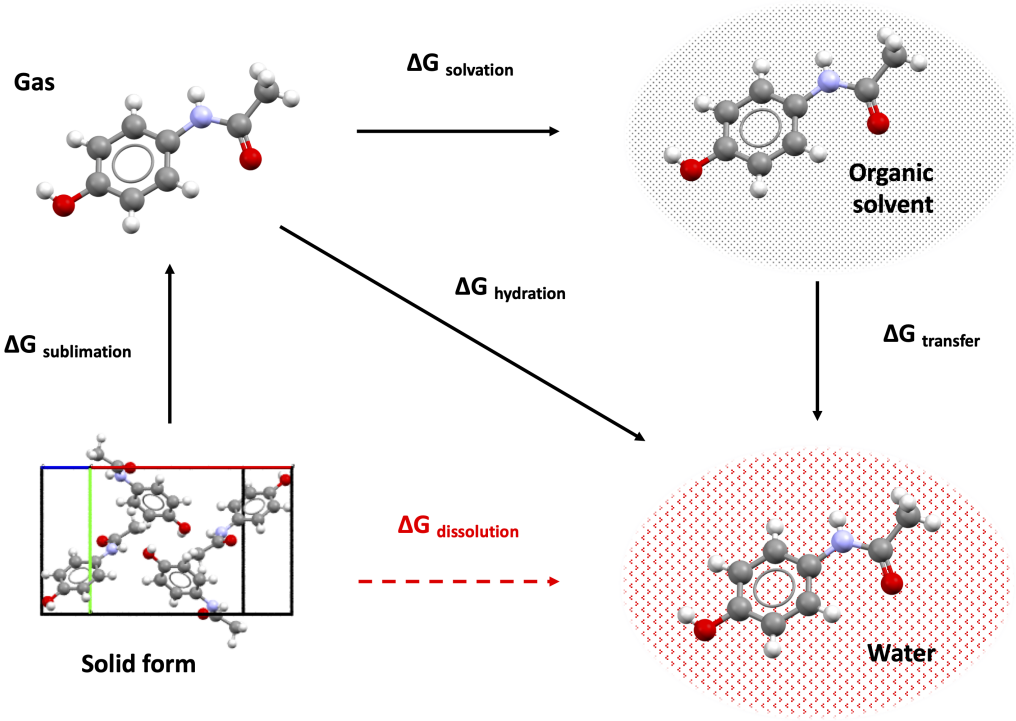

Figure 1.1: Thermodynamic cycle used to calculate the solubility of solid solute. This figure was adapted from reference [13].

To better understand the underlying mechanisms of solubility and increase the computational methods' predictability, researchers can model solubility using the thermodynamic cycle presented in Figure 1.1. The dissolution free energy ($\Delta G_{dissolution}$) of a solid solute in a solvent is the sum of the free energy changes due to the sublimation ($\Delta G_{sublimation}$) and solvation processes ($\Delta G_{solvation}$). [13] The intrinsic solubility (S_0), solubility of a fully non-ionic molecule, is related to the dissolution free energy by:

$$\Delta G_{dissolution} = -RT \ln (S_0 \, V_m) \tag{1.1}$$

where V_m is the molar volume of the crystal [14].

The free energy of solvation in water is called free energy of hydration ($\Delta G_{hydration}$), and the transfer free energy of the solute from an organic solvent to water ($\Delta G_{transfer}$)

is the difference between $\Delta G_{hydration}$ and $\Delta G_{solvation}$. The transfer free energy from octanol to water can easily be obtained from the widely studied octanol/water partition coefficient (P_{ow}) through Equation 1.2:

$$\Delta G_{transfer} = 2.303\, RT \log P_{ow} \tag{1.2}$$

Consequently, a molecule's hydration free energy can be calculated from its solvation free energy in octanol and its octanol/water partition coefficient. [14]

Another approach commonly used to calculate the solubility of a pure solid solute in a liquid solvent is through Equation 1.3:

$$\ln x_1\, \gamma_1 = -\frac{\Delta H_{fus}\,(T_m)}{RT}\left(1 - \frac{T}{T_m}\right) \tag{1.3}$$

where $\Delta H_{fus}(T_m)$ is the solute's enthalpy of fusion at its normal melting point (T_m), and x_1 and γ_1 are the solute's mole fraction in the liquid phase and activity coefficient, respectively. This equation is an approximation since it assumes that the solute's molar heat capacity is not a function of the temperature and is the same in the liquid and solid states. [15, 16]

Relative Solubilities

If the solute's fugacity in the solid state (f_A^S) is known, the solubility of a solute in a particular solvent can be obtained using the following equation:

$$\ln c_A^\alpha = -\beta\, \mu_A^{\alpha,res}(T, P, x_A) - \ln\,(RT) + \ln f_A^S(T, P) \tag{1.4}$$

where c_A^α, x_A^α and $\mu_A^{\alpha,res}$ are the molar concentration, mole fraction and residual chemical potential, respectively, of solute A in solvent α. The residual chemical potential is multiplied by β to make it dimensionless, where β is the inverse of the product between the Boltzmann constant (k_B) and the temperature (T). In addition, P is the pressure and R is the gas constant. [17] The residual chemical potential is the difference between

the chemical potential of the real fluid and the chemical potential of an ideal gas at the same temperature and pressure [18].

However, fugacities of solids can be very difficult to calculate, therefore, it is common to calculate relative solubilities of a solute in two different solvents instead of absolute values [17]. Since $f_A^S(T,P)$ is a solute dependent property that is not affected in any way by the solvent, the term $\ln f_A^S(T,P)$ in Equation 1.4 disappears when we take the difference between $\ln c_A^\alpha$ and $\ln c_A^\zeta$, where ζ is a different solvent at the same T and P. The same occurs to the term $\ln(RT)$, and thus, we get:

$$\ln\left(\frac{c_A^\alpha}{c_A^\zeta}\right) = \beta\,\mu_A^{\zeta,res}(T,P,x_A) - \beta\,\mu_A^{\alpha,res}(T,P,x_1) \tag{1.5}$$

At infinite dilution, the residual chemical potential is the same as the Gibbs free energy of solvation of a single solute molecule, and Equation 1.5 turns into Equation 1.6.

$$\ln\left(\frac{c_A^\alpha}{c_A^\zeta}\right) = \beta\,\Delta G_{A,solv}^{\zeta,\infty}(T,P) - \beta\,\Delta G_{A,solv}^{\alpha,\infty}(T,P) \tag{1.6}$$

Consequently, relative solubilities of a solute in different solvents can be predicted using solvation free energies at infinite dilution if the solubilities are low enough that solute-solute interactions can be neglected. [17]

1.2 Computational methods for solubility prediction

Some of the most common computational methods used in solubility predictions are briefly described here, and a table that summarises the advantages and disadvantages of each method has been included at the end of this section. For a more exhaustive review, the reader can see reference [14].

1.2.1 Quantitative Structure Property Relations (QSPR)

QSPR methods are commonly used in the pharmaceutical industry to predict drug molecules' aqueous solubility, and they are based on the idea that molecules with similar structure will have similar properties. These methods use training data sets, molecular

descriptors and statistical methods to create a mathematical model able to predict the property of interest, such as the aqueous solubility. These molecular descriptors are chosen to implicitly incorporate physical, topological, chemical or energetic properties of the molecules in the system [14]. For example, the melting point (T_m) of a molecule is directly related to the solid-state molecule's structure and stability, and thus, it is normally used to describe the dissociation of the solute from the solid form. In addition, the solute's molecular weight can be used to represent the energy penalty associated with the cavity formation in the solvent, and the logarithm of the octanol/water partition coefficient ($\log P_{ow}$) is commonly used to describe solute-water interactions. For example, the modified general solubility equation (GSE) [19], Equation 1.7, uses T_m and $\log P_{ow}$ to calculate the solubility (S_0) of a non-ionised solute in water. [13]

$$\log S_0 = 0.5 - 0.01\,(T_m - 25) - \log P_{ow} \tag{1.7}$$

Multiple linear regression (MLR) is a very important and straightforward statistical method widely used to model the relationship between multiple input parameters and their dependent variables. However, it can easily overfit the data when small training sets are used [14]. This model only uses additive linear terms, but nonlinear terms can be included to increase its performance (e.g. squaring or taking the logarithm of the original data) [20].

More elaborated algorithms like artificial neural networks (ANNs) are sometimes needed to predict complex properties. ANNs have layers formed by simple operators (neurons) that map the previous layer's input parameters onto a nonlinear function and feed the output to the next layer. The layer or layers that are not directly in contact with the model's input or output are called hidden, and the complexity of the model increases with the number of hidden layers. Deep neural networks (DNNs) have multiple hidden layers and can extract features from unprocessed raw data. [20]

Reference [21] contains a review of studies from 2009 to 2019 that used QSPR methods to predict water solubilities, octanol/water partition coefficients and vapour pressures of pesticides. For example, Kherouf et al. [22] used MLR and ANN to predict wa-

ter solubilities of phenol derivatives using features from their optimised 3D structure. Additionally, Chinta and Rengaswany [23] have recently proposed a multiple model learning (MML) method that uses structural parameters to predict drug solubilities. MML methods use different linear models for different regions of the domain.

These methods are easy to use, computationally fast and can produce good results for molecules similar to those included in the training set. However, they require a large amount of input data, their parameters may not have simple physical interpretations, and results tend to be quite poor when the studied molecule differs too much in molecular structure from those used in the training set [8]. Thus, it is not always easy to understand why a particular prediction is made or what has to be done to change the solubility in a specific direction [17]. Also, the model of an ANN is not explicitly available and thus, they act almost as a "black box" [23].

1.2.2 Group contribution methods

Group Contribution (GC) methods use empirical parameters to describe interactions between molecules' structural units, like -CH_3, -$COCH_2$, etc. [24]. The universal quasi-chemical functional group activity coefficient (UNIFAC) [25] is one of the most important GC methods used in the design of chemical processes, especially separation processes [26]. This method assumes that the natural logarithm of the activity coefficient ($\ln \gamma_i$) is the sum of two contributions: a combinatorial part ($\ln \gamma_i^C$) due to the molecules' size and shape and a residual part ($\ln \gamma_i^R$) that depends on the functional groups' interactions:

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \tag{1.8}$$

The binary group interaction parameters used to calculate $\ln \gamma_i^R$ were fitted to experimental vapour-liquid equilibrium (VLE) data, while the parameters used to obtain $\ln \gamma_i^C$ were taken from the values proposed by Bondi [24, 27]. This method tends to have a poor performance for mixtures outside the 290 K - 400 K temperature range and sometimes struggles to predict activity coefficients when the solute's concentration is

very low. Therefore, several modifications have been proposed to solve these problems, like the modified UNIFAC (Dortmund) [28]. This model uses temperature-dependent parameters to capture the temperature-activity coefficient relationship, and activity coefficients at infinite dilution (γ_i^∞) were included in its parameterisation. [26] Furthermore, UNIFAC has been extended to model ionic liquids in contact with liquid [29, 30] and gas phases [31].

In a study done by Grensemann and Gmehling [32] in 2005, modified UNIFAC (Dortmund) provided the most reliable results when compared to the original UNIFAC and a modified COSMO-RS (see section 1.2.3) for the VLE curve, γ_i^∞ and excess enthalpy (h^E). The calculation of the 1362 VLE data set in this study took only a few minutes for UNIFAC and several hours for COSMO-RS. More recent applications of UNIFAC include VLE predictions of systems with maximum and minimum azeotrope boiling temperatures [33] and the estimation of edible fats' solid-liquid equilibria (SLE) [34]. In 2021, UNIFAC and UNIFAC (Dortmund)'s binary interaction parameters for nicotinamide, an organic acid that belongs to the vitamin B3 family, were parameterised by Silveira et al. [15]. These parameters were fitted to nicotinamide's solubility in two mixtures: methanol + ethanol and methanol + 2-propanol at four different temperatures.

GC methods are extremely fast and can produce good results in many cases, however, they require a high amount of experimental data to calculate the contributions of each group [17]. Besides, they cannot differentiate between a molecule's different isomers. [32]

1.2.3 Based on QM continuum solvation models

These methods use quantum mechanics (QM) calculations to obtain the solute's charge distribution in a continuous medium that represents the solvent. Continuum solvation models place the solute in a cavity inside the solvent medium, and the cavity's size and shape is calculated using different techniques (see reference [35] for some examples) designed to match the solute's shape and the solvent-accessible surface area (SASA). Once the solvent cavity is defined, the solute-solvent interaction energy can be calcu-

lated as the sum of two contributions: the cavitation-dispersion-solvent-reorganisation (G_{CDS}) and the electrostatic interactions between the solute and solvent charges. The first term can be calculated using an empirical equation and arises from the cavity formation, the solute-solvent dispersion forces and the solvent's reorganisation due to the perturbation of the solvent-solvent interactions. On the other hand, the electrostatic potential is obtained from the solute's charge distribution which simultaneously depends on and alters the polarised cavity wall. Polarisable continuum models (PCM) like IEF-PCM [36], called this way due to the cavity's polarisation, calculate the electrostatic contribution in an iterative way since the solute polarises the cavity wall which consequently polarises the solute and so on. This calculation can be simplified if a conducting medium ($\epsilon = \infty$) is used instead of a dielectric medium (i.e. nonconducting) and the solvent's dielectric constant is applied as a correction factor, as implemented in the conductor-like screening model (COSMO). [37] For an extensive review about continuum solvation models the reader is referred to reference [38].

Continuum solvation methods assume that solute-solvent electrostatic interactions are not dependent on the solvent's structure and that the dielectric response of the medium is uniform and linear outside the cavity. However, this assumption is not valid when there are strong specific interactions, like hydrogen bonds, between the solute and the solvent molecules in the first solvation shell. [39]

The conductor-like screening model for real solvents (COSMO-RS), developed by Klamt et al. [40], is the most popular method in this category and is described below. The solvation model based on density (SMD) [39] is another popular method that uses QM calculations in an implicit solvent model, and it was parameterised using 2821 solvation free energies of neutral and ionic solutes in water and other solvents. This method belongs to the SMx solvent models but instead of using discrete partial charges it uses the electron density function.

COSMO-RS is somewhat similar to GC activity-coefficient methods (i.e. it uses Equation 1.8), but the molecule surface is divided into segments with equal area instead of subdividing the molecule into functional groups [41]. These segments' charge densities are obtained using QM calculations, and then, they are used to calculate the charge

density distribution $P_S(\sigma)$. The charge density distribution, also called σ-profile, is the probability of finding a segment with a specific screening charge density σ, and it is used to calculate the residual part of the activity coefficient. Meanwhile, the activity coefficient's combinatorial component is obtained using the molecular volume and area from the calculated surfaces. [32] Also, COSMO-RS calculates both solvent and solute's cavities and surface charges, and it uses eight general parameters and two empirical parameters for each element to implicitly account for G_{CDS}. [37]

In the early 2000s, COSMO-RS was used to predict aqueous solubilities of 107 pesticides with good results [42]. In the last two years, COSMO-RS has been used to successfully predict the solubility of CO_2 in seven deep eutectic solvents [43] and to study benorilate's solubility in twelve different solvents [44], to name two examples. With respect to the latter case, COSMO-RS significantly underpredicts benorilate's solubility in polar protic solvents, except for isopropanol, and greatly overpredicts its solubility in dipolar aprotic systems [44].

These methods' main advantages are that they treat electrostatic components to a high degree of accuracy and that only a few basic parameters need to be parameterised from experimental data, which are in principle transferable to other systems. However, as explained above, these simulations are significantly slower than GC and, in some cases, also less accurate. COSMO-RS' poor performance in some systems, like alcohol-ether mixtures, could be due to the way these methods model dispersive interactions (i.e. these interactions are treated as a function of the molecular segments' size only and not their type) [32]. Also, implicit models cannot explicitly consider entropy changes in the solvent due to solvation.

1.2.4 Group contribution equations of state

Group contribution equations of state (GC-EoSs) apply the GC concept within an equation of state (EoS). Equations of state are mathematical functions that describe the relationship between the temperature, pressure and volume of the system, and thus, they can calculate single-phase and critical properties [18]. Two important group contribution equations of state are described below as an example.

Predictive Soave-Redlich-Kwong

The predictive Soave-Redlich-Kwong (PSRK) [45] group contribution equation of state is a combination of the Soave-Redlich-Kwong (SRK) [46] equation of state and the UNIFAC model. The PSRK model uses Equation 1.9 to correlate the pressure of a system, with its volume (V) and temperature:

$$P = \frac{R\,T}{V - b} - \frac{a(T)}{V\,(V - b)} \tag{1.9}$$

where the mixture parameter b is obtained using the mole fractions, critical temperatures (T_c) and critical pressures (P_c) of the components, as it is done in the SRK equation of state. On the other hand, the attractive term $a(T)$ from the SRK equation of state has been modified to incorporate the components' activity coefficients, which are obtained using the UNIFAC model with temperature-dependent interaction parameters. Also, PSRK uses parameters fitted to experimental vapour-pressure data, when available, instead of the components' acentric factors, in the calculation of the $a(T)$ term to improve the description of polar molecules. [45]

This method is widely used to predict vapour-liquid equilibria at a large range of conditions, including supercritical systems, and is able to predict gas solubilities in solvents with strong electrolytes [47].

SAFT-γ-Mie

SAFT-γ-Mie [48] incorporates the GC concept within the statistical associating fluid theory (SAFT) [49], and it models molecules as chains of different types of fused spherical segments. These segments are selected based on the chemical functional groups within the molecule (see Figure 1.2), and they interact with each other through a Mie potential of variable range.

The Mie potential between segments k and l is described by Equation 1.10:

$$\Phi_{kl}^{Mie}(r_{kl}) = C_{kl}\,\epsilon_{kl}\left[\left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{kl}^{r}} - \left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{kl}^{a}}\right] \tag{1.10}$$

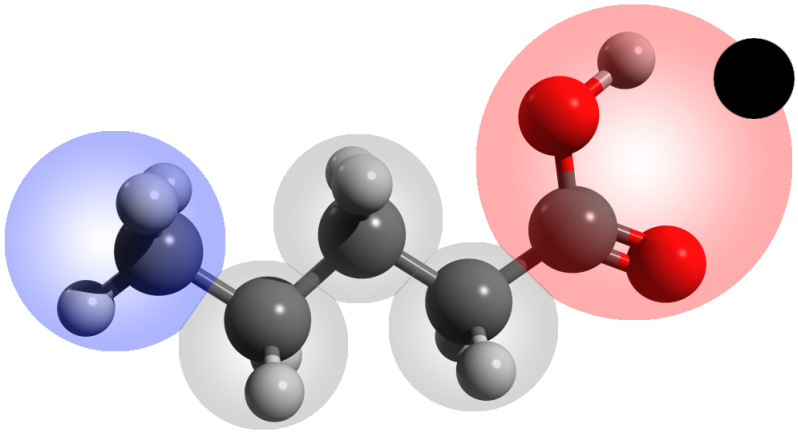

Figure 1.2: Pentanoic acid's representation using the SAFT-γ-Mie approach. The COOH group (shaded red) has an association site shown as a black sphere. This figure was recreated from reference [16].

$$C_{kl} = \frac{\lambda_{kl}^r}{\lambda_{kl}^r - \lambda_{kl}^a} \left(\frac{\lambda_{kl}^r}{\lambda_{kl}^a}\right)^{\lambda_{kl}^a/(\lambda_{kl}^r - \lambda_{kl}^a)} \tag{1.11}$$

where r_{kl} is the distance between the segments' centres, σ_{kl} and ϵ_{kl} are the unlike segment diameter and the dispersion energy between groups k and l, respectively, and C_{kl} is a prefactor that guarantees that Equation 1.10's minimum is equal to $-\epsilon_{kl}$. Additionally, λ_{kl}^a and λ_{kl}^r, are the attractive and repulsive exponents of the unlike segment-segment interactions, respectively. When the repulsive and attractive exponents are 12 and 6, respectively, the potential is called Lennard-Jones. As an example, Figure 1.3 shows how λ_{kl}^r and λ_{kl}^a affect the Mie potential. Increasing λ_{kl}^r while keeping λ_{kl}^a equal to 6, decreases the width of the well and shifts the long-range tail upwards. The attractive exponent is commonly fixed to 6 [50–52] since this is the value suggested by London to describe dispersion forces of molecules without permanent multipoles [53]. However, carbon dioxide has multipolar interactions and thus, its optimum exponents are $\lambda_{kl}^a = 6.66$ and $\lambda_{kl}^r = 23$ [54]. For this reason, a 23-7 Mie potential has been included in Figure 1.3 to show how increasing λ_{kl}^a by one unit while maintaining λ_{kl}^r constant, very slightly shifts the long-range tail upwards.

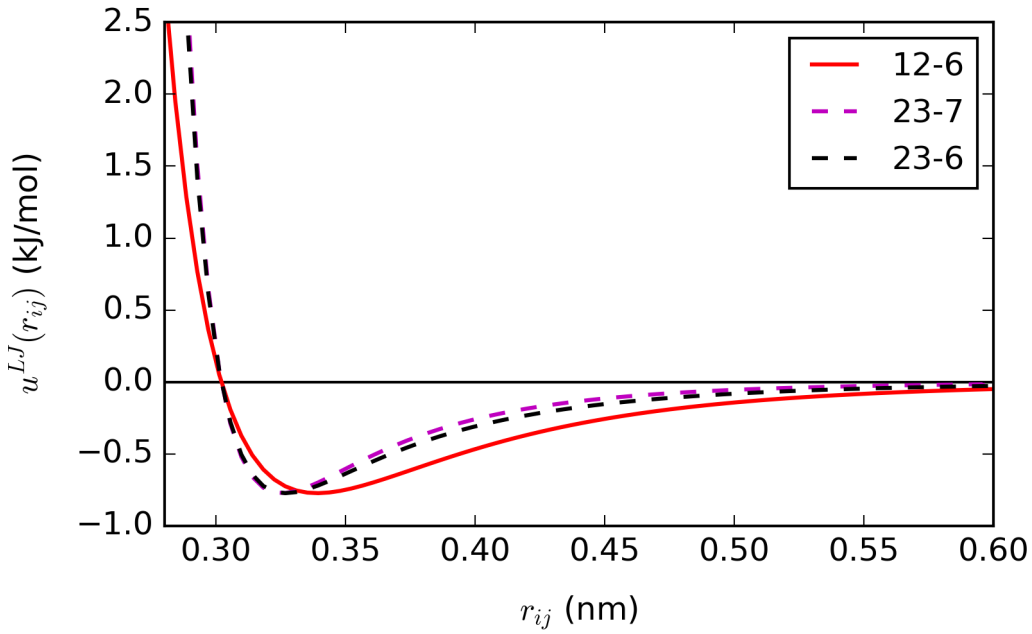

Figure 1.3: 12-6, 23-7 and 23-6 Mie potentials of two interacting particles with $\sigma_{kl} = 0.302$ nm and $\epsilon_{kl} = 0.773$ kJ/mol.

The unlike parameters are normally estimated using the combining rules given by Equations 1.12 to 1.14 and then optimised to improve agreement with experimental data if necessary. Equation 1.14 can be used with λ^a_{kl} or λ^r_{kl}. [16]

$$\sigma_{kl} = \frac{\sigma_{kk} + \sigma_{ll}}{2} \tag{1.12}$$

$$\epsilon_{kl} = \frac{\sqrt{\sigma^3_{kk} + \sigma^3_{ll}}}{\sigma^3_{kl}} \sqrt{\epsilon_{kk}\, \epsilon_{ll}} \tag{1.13}$$

$$\lambda_{kl} = 3 + \sqrt{(\lambda_{kk} - 3)\,(\lambda_{ll} - 3)} \tag{1.14}$$

Furthermore, SAFT-γ-Mie uses association sites to model strong polar interactions like hydrogen bonds. Each segment can have more than one site type, and each association site interacts with other segments' sites through a square-well potential. Equation 1.15 calculates the interaction between two association sites placed in segments k and l:

$$\Phi_{kl}^{HB}(r_{kl,ab}) = \begin{cases} -\epsilon_{kl,ab}^{HB} & \text{if } r_{kl,ab} \leq r_{kl,ab}^{c} 1 \\ 0 & \text{otherwise} \end{cases} \tag{1.15}$$

where $r_{kl,ab}$ and $r_{kl,ab}^{c}$ are the distance between the sites and the interaction's cutoff range, respectively. The site in segment k is type a while the one in segment l is type b, and $a \neq b$. In addition, each site is placed at a certain distance from the segment's centre, which is usually equal to 0.4 times the segment's diameter. [16, 52]. The unlike parameter $\epsilon_{kl,ab}^{HB}$ can be estimated using Equation 1.16 if there is no experimental data for its optimisation. Also, the bonding volume ($K_{kl,ab}$) shown in Equation 1.17 is normally used to describe $r_{kl,ab}^{c}$. [16]

$$\epsilon_{kl,ab}^{HB} = \sqrt{\epsilon_{kk,aa}^{HB}\, \epsilon_{ll,bb}^{HB}} \tag{1.16}$$

$$K_{kl,ab} = \left(\frac{\sqrt[3]{K_{kk,aa}}\, \sqrt[3]{K_{ll,bb}}}{2} \right)^{3} \tag{1.17}$$

The parameters that describe interactions between groups of the same type (i.e. σ_{kk}, ϵ_{kk}, etc.) are obtained from experimental properties of molecules with that functional group. Vapour pressures, saturated liquid densities and compressed liquids densities are among the most common properties used during this method's parameterisation. However, fluid-phase behaviour and excess properties of mixtures are also included when extra data is required. [16]

SAFT-γ-Mie has been recently used to determine ibuprofen's optimal solvent and anti-solvent mixtures at several temperatures [55] and to successfully predict solubilities of several pharmaceutical compounds in water, butanol and acetone, as well as their partition coefficients [52].

GC-EoSs are computationally fast and have the same advantages as GC activity-coefficient approaches without some of their limitations, like the vapour phase ide-

ality assumption and the inability to predict bulk properties (e.g. density and heat capacity) [50]. On the other hand, these methods cannot be used to study structural, interfacial, and transport properties. Hence, SAFT-γ-Mie parameters have been used in the development of coarse-grained molecular dynamics (CG-MD) models (explained below). The parameters' transferability between these two approaches arises from the way particle's interactions are calculated. SAFT-γ-Mie and MD simulations can both describe intermolecular interactions in the system using the Mie potential. However, MD simulations calculate Coulomb interactions when modelling polar molecules while SAFT-γ-Mie uses a square-well association sites approach for highly polar interactions. [56]. For example, Rahman et al. proposed a CG-MD model for linear alkanes obtained using parameters from SAFT-γ-Mie and TraPPE-UA to describe non-bonded and bonded interactions, respectively [51]. Tasche et al. developed SAFT-γ-Mie parameters for mixtures of polybutadiene and squalane and used these parameters to create a coarse-grained molecular dynamics force field [57]. Other examples can be found in references [58, 59].

1.2.5 Molecular simulations

Molecular simulations can model molecules at an atomic level based on the potential energy between the particles. Unlike QM calculations, which are based on the Schrödinger equation to describe the electrons' behaviour in a molecule, molecular simulations represent the molecules' atoms as interacting particles that follow the laws of classical mechanics. The set of equations and parameters used to calculate the potential energy of the system is called force field and is normally parameterised against experimental and/or QM data (see section 2.6). There are different types of force fields depending on the way they represent the molecules (all-atom or coarse-grained) and how they model electrostatic interactions (non-polarisable or polarisable). All-atom force fields treat each atom in the system as an interacting particle while coarse-grained models group some atoms together to reduce the number of interacting particles and decrease the simulation time [60]. The differences between polarisable and non-polarisable force fields are described in section 1.4 since this classification is highly relevant to this work, and thus, it deserves its own section. Furthermore, molecular dynamics simulations are

explained in detail in sections 2.4 and 2.5.

There has been some work done on predicting the solubility of drugs in water using molecular simulations. For example, Westergren and his group studied the solubility of a drug molecule into water [11] and a drug melt at 673.15 K [8]. The simulations were performed at this temperature because most drugs at room temperature form glasses rather than liquids, and the rigidity of those systems would have complicated the simulations. Later, they studied the insertion of a drug molecule into a pure amorphous drug phase at 278.15 K by cooling the pure liquid drug from 673.15 to 278.15 K [5]. They studied between 46 and 48 drug molecules, including acetylsalicylic acid, caffeine, ibuprofen, lidocaine and morphine. Lastly, they compared different force fields and concluded that good predictions of amorphous drug solubility are obtained when OPLS-AA is combined with partial charges from the COMPASS force field [61]. Other examples are the work by Schniederes et al. [62], who studied n-alkylamides' solubilites using a polarisable force field, and the study of crizotinib's solubility by Fan et al. [63]

The crystal structure and fugacity of solids are very difficult to simulate [17], and hence, obtaining sublimation free energies is very challenging. Consequently, some researches have focused on the prediction of relative solubilities, partition coefficients and excess solubilities. For example, Garrido et al. predicted free energies of solvation of n-alkanes (methane to octane) in 1-octanol and water using three different force fields (TraPPE, GROMOS and OPLS-AA) and the MSPC/E water model [64]. Bannan et al. [65] studied octanol/water and cyclohexane/water partition coefficients using GAFF and the dielectric corrected GAFF (GAFF-DC), and they found that both force fields have the same overall accuracy (RMSD = 1.2 log units) even though GAFF-DC performs better than GAFF at predicting hydration free energies. They also noticed that alcohols' partition coefficients between cyclohexane and water were systematically overpredicted with GAFF and systematically underpredicted with GAFF-DC, something that the authors concluded could be due to the limitations of fixed charged force fields. Zhang et al. also tested several force fields' performance at predicting octanol/water, chloroform/water and cyclohexane/water partition coefficients and noticed an overall overestimation of

these parameters [66]. Mobley and co-workers calculated acetanilide, acetaminophen, phenacetin, benzocaine and caffeine's excess solubilities in water/ethanol mixtures using GAFF, and they found good qualitative agreement between predicted and simulated data [67]. Liu et al. compared MD simulations' performance at predicting relative solubilities of poorly soluble solutes against UNIFAC, UNIFAC (Dortmund) and SMD, and found that molecular simulations were the most accurate method overall [17].

These simulations are very time consuming, and thus, an implicit solvent model can be used instead of explicitly simulating the solvent. In this case, the solute molecule is represented at an atomistic level while the solvent is treated as an isotropic continuous medium, using methods like COSMO. [13] However, as explained in section 1.2.3, implicit methods cannot explicitly consider entropy changes in the solvent due to solvation, and they are less accurate than simulations with explicit solvents [68].

On the other hand, molecular simulations have a great potential for transferability [69] and provide a higher level of detail about the solvation process. Their parameters have a physical meaning, which is not always the case for statistical methods, and a force field that has been extensively validated against solubility measurements can in principle predict the solubility of other molecules without using experimental data as input parameters. [13] Furthermore, molecular simulations take into account the functional groups' arrangement, unlike GC methods and EoSs, and thus, they can be used to study isomers and structural properties. Also, even though they treat electrostatic components to a lower degree of accuracy than COSMO-RS, they are better at modelling dispersion forces and can explicitly account for solute-solvent interactions and the cavity formation. For these reasons, molecular simulations are the method of choice in this thesis, and the project will focus only on predicting free energies of solvation to avoid the difficulties associated with solid-phase simulations. One of the biggest problems of solid-phase simulations is finding a force field able to accurately model the different crystal structures of a solute.

To summarise the techniques described in this section, each method's advantages and disadvantages are presented in Table 1.1.

Table 1.1: Advantages and disadvantages of different methods used to predict solubility.

QSPR Methods	• Easy to use and computationally fast. • Good results for molecules similar to those included in the training set.
	• Require a large amount of training data. • Parameters may not have simple physical interpretations. • Results tend to be quite poor when the studied molecule differs too much in molecular structure from those used in the training set.
Group Contribution Methods	• Extremely fast and can produce good results in many cases.
	• Require a high amount of experimental data to calculate the contributions of each group. • Cannot differentiate between a molecule's different isomers. • Inability to predict bulk properties.
Solvation Models Based on QM	• Treat electrostatic components to a high degree of accuracy. • Only a few basic parameters need to be parameterised from experimental data.
	• Significantly slower than GC methods. • Cannot explicitly consider entropy changes in the solvent due to solvation.
Group Contribution Equations of State	• Computationally fast. • Same advantages as GC activity-coefficient approaches without some of their limitations.
	• Cannot be used to study structural, interfacial, and transport properties.
Molecular Simulations	• Great potential for transferability. • Higher level of detail about the solvation process. • Parameters have a physical meaning. • Can be used to study isomers and structural properties.
	• Very time consuming.

1.3 Solvation free energy predictions using molecular simulations

Most solvation free energy studies use water as a solvent due to its biological importance and the large amount of experimental data available in the literature [70]. For example, Shivakumar et al. simulated hydration free energies of 239 small molecules using OPLS, GAFF and CHARMM with three different water models [71]. Mobley et al. calculated this property for 504 neutral small organic molecules using GAFF and the TIP3P water model, and they found an overall root-mean-square-deviation (RMSD) of 1.24 kcal/mol between simulated and experimental data [68]. In addition, Wu and Kieffer used a hybrid cluster/continuum model (e.i. implicit solvent with some explicit molecules) to obtain four ionic molecular groups's hydration free energies [72], and Dasari and Mallik

simulated a cardiovascular drug's solvation free energies in water and seven ionic liquids using GAFF with SPC/E [73].

Studies that consider free energies of solvation in hydrophobic solvents are less common; however, they are starting to gain popularity since many biomolecular processes involve both hydrophilic and hydrophobic interactions [74] (e.g. membrane permeabilities can be predicted based on water/oil transfer free energy [75]). Garrido et al. showed that TraPPE can successfully predict propane, benzene, ethanol and acetone's solvation free energies in six solvents of different polarity (including alkanes, 1-octanol and water) with average deviations that ranged from 0.2 to 1.2 kJ/ mol [76]. Horta et al. parameterised a GROMOS-compatible parameter set using small organic molecules' liquid densities, enthalpies of vaporisation and solvation free energies in water and cyclohexane [77–80]. However, they were not able to simultaneously predict hydration free energies and solvation free energies in cyclohexane of alcohols and carboxylic acids [77]. Jorge et al. simulated solvation free energies of linear, branched and cyclic alkanes in apolar solvents (mainly hexadecane) using three different force fields, and they found small but systematic deviations [81]. Based on these results, Jorge proposed a new set of alkyl parameters that accurately predict solvation free energies in hydrophobic solvents [1]. Also, Stroet et al. studied the performance of the Automated Topology Builder (ATB) 3.0 at predicting 685 molecules' hydration free energies (RMSD = 5.5 kJ/mol) and 218 compounds' solvation free energies in hexane (RMSD = 5.7 kJ/mol) and discussed about the uncertainties and inconsistencies associated with experimental solvation free energy databases normally used in force fields' validation and calibration [82].

Mobley et al. have found that classical molecular dynamics simulations have not yet surpassed the accuracy of much less computationally expensive approaches like quantitative structure-property relationships or group contribution methods in blind tests of solvation free energy predictions [83]. This is likely due to some fundamental shortcoming of current force fields. The main hypothesis of this thesis is that this failure is mainly due to the inability of classical force fields to accurately account for polarisation effects.

1.4 Polarisable and non-polarisable models

Nowadays, most force fields used for biomolecular simulations (AMBER, CHARMM, GROMOS and OPLS) are non-polarisable models [84]. These force fields use fixed partial charges to describe electrostatic interactions, and thus, they cannot explicitly account for polarisation effects (i.e. change in the charge distribution of a molecule due to its environment). Despite this approximation, non-polarisable force fields can accurately predict many properties and have been used to study drug-like compounds, proteins, DNA, and even entire viruses [85]. However, these models struggle to predict properties that directly involve electronic polarisation, like solvation free energies and dielectric constants, especially in low-polar and non-polar solvents. [86]. For example, OPLS-AA and CHARMM36 with twelve different water models overestimate water transfer free energies from water to hexadecane (i.e. simulated values are more positive than experimental transfer free energies, making the process less favourable) [75], and simulated dielectric constants of common solvents seem to be systematically underpredicted [87]. Hence, one would expect significant improvements in classical force fields' performance if polarisation effects are correctly accounted for.

Polarisation effects can be included explicitly using different types of polarisable models: fluctuating charges, Drude oscillators, inducible dipoles and methods including multipole electrostatics. Fluctuating charges (FQ) models are similar to classical non-polarisable models but with atomic partial charges that are allowed to vary their magnitude throughout the simulation. Meanwhile, Drude oscillators use a pair of point fixed-charges with one of these charges represented as a massless particle (Drude particle) attached to the nucleus by a string and the other charge located on the atomic nucleus. The Drude particle can freely move anywhere around the nucleus in response to an external field, and as a consequence, generate a new dipole moment. Drude oscillators force fields are chemically intuitive and better at representing polarisation effects of planar molecules, like benzene, than the first approach. However, they are also more computationally expensive due to the addition of many extra charges. These two approaches use Coulomb potentials to describe electrostatic interactions, unlike the other two polarisable models described below. Inducible dipoles models, like AMBER

ff02 [88], use fixed partial charges and inducible point dipoles located on the atomic nuclei or bonds between atoms. Consequently, they need to add additional terms to the potential energy function to account for charge-dipole and dipole-dipole interactions. Lastly, some polarisable force fields, like AMOEBA [89, 90], represent atoms using anisotropic multipole moments able to capture any non-spherical components of the atomic charge density. [85]

Polarisable models are harder to parameterise than non-polarisable models since they contain more individual parameters per atom (e.g. two charges per atom for Drude oscillators, dipoles for inducible dipoles and multipoles for AMOEBA) and more complicated potential functions (e.g. charge-dipole and dipole-dipole interactions) [85]. For these reasons and the iterative minimisation approach traditionally used to calculate induced multipole moments, polarisable force fields have a high computational cost which can be from 3 to 10 times the cost of non-polarisable force fields [60]. Furthermore, well known non-polarisable models have been extensively tested and refined throughout the years, and thus, they can sometimes provide similar or more accurate results than the more complex polarisable models. [85] For example, Mohamed et al. simulated free energies of solvation of 21 small molecules in four different solvents (toluene, chloroform, acetonitrile and dimethylsulfoxide) using AMOEBA and GAFF and found that GAFF had a better performance overall with a mean absolute error of 0.66 kcal/mol compared to 1.22 kcal/mol for AMOEBA. [70]

In the following section, we will assess to what extent the lack of polarisation effects in classical simulations can be taken into account by inexpensive correction schemes.

1.5 Accounting for polarisation effects

As explained above, the dipole moment (or any other multipole moment) of a molecule in a non-polarisable force field is not able to respond to its environment, and thus, the modelled molecule always has the same fixed dipole regardless of the polarity of the medium. In reality, a molecule's dipole moment can significantly vary with respect to its surroundings, and thus, problems tend to arise when molecules are simulated in environments that significantly differ from those used during the model's parametrisa-

tion [91]. For example, the dipole moment of a water molecule in vacuum is 1.85 D, while in the liquid bulk state, the dipole is close to 3 D [92]. In 1987, Berendsen et al. proposed a correction that accounts for the re-polarisation, or distortion, of a molecule when it moves from the liquid to the gas phase, and they used this correction to develop a new water model called SPC/E [93]. Since then, most "modern" water models [94–97] and a few methanol models [98, 99] have been parametrised using this correction. However, the issue has been mostly neglected in the development of generic force fields like OPLS-AA [100], GROMOS[101], TraPPE-UA [102] and GAFF [103].

In 2014, Leontyev and Stuchebrukhov mathematically proved that, within certain approximations, polarisable molecular systems can be described by equivalent non-polarisable fixed-charge models if the molecules have a well-defined average charge distribution and the electronic polarisability of the system is macroscopically homogeneous [104]. Since in the classical approximation the bare point charges Q_0 can be considered to be in a continuum polarisable medium with a relative high-frequency permittivity equal to ϵ_{el}, meant to describe the purely electronic contribution to the polarisation, the Coulomb interaction between them can be calculated using Equation 1.18, where ϵ_0 is the vacuum permittivity.

$$u^{qq}(r_{ij}) = \frac{1}{\epsilon_{el}} \frac{Q_{0i}\, Q_{0j}}{4\pi\epsilon_0 r_{ij}} \tag{1.18}$$

Most force fields do not include the relative permittivity of the medium when calculating electrostatic interactions. Thus, screening effects (i.e. reduction of electrostatic interactions in a dielectric medium to those in vacuum) can be taken into account implicitly by using scaled partial charges $q_i^{eff} = Q_{0i}/\sqrt{\epsilon_{el}}$. The original charges Q_0 are not the same as the atomic partial charges of the molecule in vacuum since they correspond to the charges in the real liquid, and therefore, they are challenging to identify *a priori*, except for ions or ionised groups in a protein. [84] For example, the charge +1 of an Na^+ cation in solution should be scaled by $1/\sqrt{\epsilon_{el}}$. For most organic materials, ϵ_{el} is close to 2; consequently, Leontyev and Stuchebrukhov compared interactions between ions in a continuum with dielectric $\epsilon_{el} = 2$ to those obtained using CHARMM with charges scaled

by ≈ 0.7. CHARMM with scaled charges can accurately reproduce *ab initio* interaction calculations while the original force field significantly overestimates the interaction energy [86]. On the other hand, empirical partial charges of common non-polarisable force fields can also be thought of as scaled charges since they can describe reasonably well interatomic interactions without explicitly introducing the electronic screening [84]. For example, the TIP3P water model [105] has an effective dipole moment of 2.35 D, and thus, the actual dipole moment should be taken as $\mu = \mu^{eff}\sqrt{\epsilon_{el}} = 3.14$ D [86]. This value is much closer to estimates obtained from experimental data [106] and from liquid-phase QM calculations [107]. However, these models are not completely consistent with the charge scaling concept described here because they simulate ions and ionised groups with unscaled charges, leading to an overestimation of the interaction's strength. Nonetheless, many of these force fields can still do a good job at predicting hydration free energies due to a fortuitous compensation of errors since they also ignore the pure electronic contribution to the electrostatic free energy, which represents more than half of the solvation free energy for ions [86]. For example, high-level QM calculations of liquid water's distortion and purely electronic contributions to the polarisation energy show that these terms, which have opposite sign, nearly cancel each other out [108]. This pure electronic contribution is not included in non-polarisable models because it arises in response to the external field induced by the electronic clouds of the surrounding solvent molecules, and it is present even when no rearrangement of the molecules occurs. Meanwhile, the other component of the polarisation, called nuclear or inertial, is due to the interaction between the medium molecules themselves and is assumed to be implicitly included in the force field through properly parameterised atomic partial charges and other non-electrostatic parameters. [84]. The limitations of common non-polarisable force fields become clear when simulating free energies of solvation in non-polar solvents since, in these cases, polarisation effects are almost exclusively determined by the pure electronic component. [86]

When calculating properties that directly involve electronic polarisation, like phase transitions and dielectric properties, the polarisation's pure electronic component needs to be added explicitly [84]. Leontyev and Stuchebrukhov argued that Berendsen dis-

tortion correction is, in most cases, underestimated since it uses the effective dipole moment of the model to represent the liquid-phase, while the (often much larger) actual dipole in the liquid-phase should be used instead. Besides, this correction should be supplemented by another term that represents the interaction of the polarised molecule with the purely electronic continuum surrounding it [92]. The nature of this new term can be better understood from Figure 1.4, which explains the vaporisation process of water in the electronic dielectric continuum ϵ_{el}. First, a polarised water molecule in bulk is separated from other water molecules, keeping its dipole constant and remaining in the electronic continuum (step-1). Then, this molecule is moved from the electronic continuum to vacuum (step-2), and finally, the molecule's dipole is relaxed to its gas-phase value (step-3). The energy change associated with step-1 is obtained from MD simulations, while the other two steps need to be added explicitly. Step-3 can be calculated using the Berendsen correction but with the actual liquid dipole, while a new correction term is used to obtain the energy change from step-2. [92] These corrections are explained in more detail in section 2.7.2.

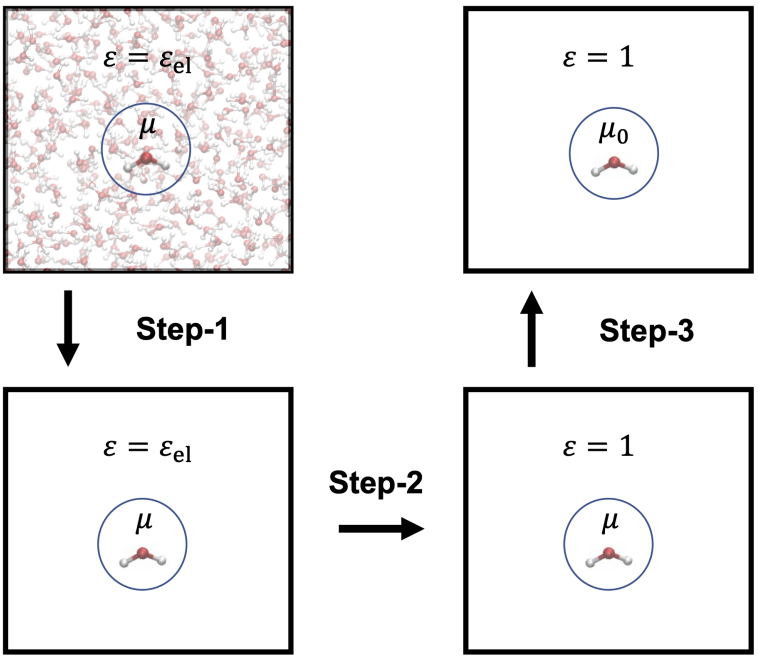

Figure 1.4: Vaporisation process of a water molecule in the electronic dielectric continuum ϵ_{el}. This figure was adapted from reference [92].

As mentioned above, simulated static dielectric constants (ϵ) should be corrected to account for the polarisation's pure electronic component. Multiplying the simulated value by ϵ_{el} gives a good estimate of the static dielectric constant in different systems [84]. However, the dipole moments of many popular non-polarisable models of water calculated using this approach are nearly twice as large as the experimental value. Therefore, in 2018, Farahvash et al. proposed a new approach that considers dynamic corrections when calculating the dielectric constant. However, theoretically, the static dielectric constant should not depend on dynamics. [109] A year later, Jorge and Lue [87] proposed a simple and straightforward way to correct the simulated dielectric constant. This correction uses a simple dipole moment scaling factor, previously presented in Vega's work [110], and includes an additional term that accounts for the polarisation change of each instantaneous configuration of the system due to the external field ($\epsilon_{el} - 1$). This extra additive term has a small relative effect on the dielectric constant of polar liquids; however, its impact is significant on dielectric constants of non-polar and weakly polar liquids. For example, this term represents less than 1 % of water's total dielectric constant and more than 50 % of cyclohexane's dielectric constant. When this approach was used to correct dielectric constants simulated using non-polarisable force fields, agreement with experimental data was significantly improved for a wide variety of molecules, and previously observed systematic deviations were eliminated. [87] This thesis uses this correction to obtain dielectric constants, and its equation will be described in section 2.7.4.

Some authors have taken other approaches to account for polarisation effects in non-polarisable force fields. Kramer et. al [75] optimised the LJ parameters σ_{ij} and ε_{ij} between TIP3P water and CHARMM36 [111] to reproduce the experimental transfer free energy of a water molecule from bulk water to hexadecane. The modified parameters improved the target property and the interfacial tension of water/hexadecane, however, they also highly overestimated the diffusion constant of water in hexadecane and worsened the models' performance at predicting transfer free energies of alkanes from liquid water to hexadecane. The GROMOS force field [101] does something similar since it uses specific van der Waals repulsion parameters for pairs of atoms with

ionic or hydrogen-bond interactions.

1.6 Objectives and outline of this thesis

The work by Leontyev and Stuchebrukhov [84, 86, 92, 104] and the promising results obtained by Jorge and Lue [87] are the foundation of this project. We hypothesise that marked improvements in the performance of classical non-polarisable models can be obtained if accurate polarisation corrections are employed consistently in force field parameterisation and validation stages.

Here, a new force field for alcohols, amines and ketones is developed, considering polarisation corrections during the parametrisation and validation stages. This force field is called PolCA, standing for "Polarisation-Consistent Approach", and it is an extension of the model for alkanes, alkenes and alkynes proposed by Jorge [1]. The latter was proposed after observing that GROMOS, OPLS-UA and TraPPE-UA presented small but systematic deviations when predicting hydrocarbon solvation free energies in hydrophobic systems [81]. Out of the three force fields tested, TraPPE-UA was the one with the best performance, and thus, its parameters were chosen as the starting point for our parameterisation. The new models for alcohols and ketones borrow all bonded parameters from the TraPPE-UA force field without modifications, and only non-bonded parameters were optimised. Unfortunately, TraPPE does not have UA parameters for amines, and thus, bonded parameters for our amine force field were taken from TraPPE-EH, OPLS-AA and AUA4, as described in detail in Chapter 4.

The methods used in this work and the theory behind them are explained in Chapter 2. This chapter serves as an introduction to statistical mechanics, the foundation of molecular modelling, and contains the information needed to run and analyse molecular dynamics simulations, including free energy calculations. It also describes the set of parameters and equations commonly present in a force field and the algorithm used here for their optimisation. Meanwhile, the three following chapters correspond to the development and validation of the PolCA force field for alcohols (Chapter 3), amines (Chapter 4) and ketones (Chapter 5). Finally, PolCA's transferability is analysed in Chapter 6. In this chapter, free energies of solvation of alkanes, ketones and amines in

octanol obtained using PolCA and other well-known force fields are compared against experimental data. Also, densities of several multifunctional compounds predicted using PolCA are presented there as an additional transferability test.

The thesis ends with general conclusions and suggestions for future work (Chapter 7).

There is one publication associated with this work:

Maria Cecilia Barrera and Miguel Jorge (**2020**), *"A Polarisation-Consistent Model for Alcohols to Predict Solvation Free Energies"*, Journal of Chemical Information and Modeling 60 (3), 1352-1367.

And three publications in preparation:

Miguel Jorge, José RB Gomes and Maria Cecilia Barrera, *"The dipole moment of alcohols in the pure liquid phase and in solution".*

Maria Cecilia Barrera, Jordan Cree, José RB Gomes and Miguel Jorge, *"Polarisation-consistent force field for ketones".*

Maria Cecilia Barrera, José RB Gomes and Miguel Jorge, *"Polarisation-consistent force field for alkylamines and alkanolamines".*

Chapter 2

Methodology

2.1 Introduction

As explained in Chapter 1, the main goal of this thesis is to develop an improved approach for optimising molecular models to better predict solvation free energies using molecular dynamics. This chapter contains an overview of molecular simulation methods and provides the necessary theoretical background to understand the subsequent chapters. Although the description is intended to be general, particular emphasis is placed on aspects that are directly relevant to the topic of this thesis.

The majority of the material in sections 2.2 to 2.4 was based on the book "Understanding Molecular Simulations, From Algorithms to Applications" by Frenkel and Smit [112], and section 2.5 was based on the book by Allen and Tildesley called "Computer Simulation of Liquids" [60]. The reader is referred to these publications for additional details.

The first section of this chapter explains the fundamentals of statistical mechanics, on which all molecular simulations are based. Afterwards, two different simulation techniques are explained: Monte Carlo and molecular dynamics, emphasising on the latter since it is more relevant to this work. The following two sections explain the set of equations and parameters, known as force field, used by these techniques to calculate the particles' interactions. Finally, the last three sections give background information

as well as specific details about the simulations run here and the method used for the new force field's optimisation. Section 2.7 describes simulations run in the gas and liquid phases from which densities, diffusion constants, enthalpies of vaporisation and dielectric constants were obtained; while section 2.8 contains information about the free energy calculations.

2.2 Statistical Mechanics

Statistical mechanics connects the information generated using computer simulation of a microscopic state or microstate (atomic and molecular positions, velocities, etc.) into macroscopic terms (pressure, internal energy, etc.) [113]. To better understand the fundamental laws of statistical mechanics, it is essential to look at some simple quantum mechanical concepts. A quantum mechanical system can be found in different states, which are eigenvectors of the Hamiltonian of the system, and the probability of finding a system with fixed volume, energy and number of particles in any of its $\Omega(E)$ eigenstates is assumed to be the same. This statement is the main postulate of statistical mechanics. The average over all possible quantum states of a system is called an ensemble and is denoted by $< ... >$ to differentiate it from a time average denoted by an overbar. If a system is ergodic, the average of any given function of the coordinates and momenta of a many-particle system computed by averaging over all initial phase coordinates is equivalent to the value obtained by averaging over the time-evolved phase space coordinates. In other words, a time average is the same as an ensemble average. [112]

The number of eigenstates with energy E of a system of N particles in a volume V is denoted by $\Omega(E, V, N)$ and this variable can be related to the thermal entropy S through Equation 2.1, where k_B is the Boltzmann constant (1.38066 10^{-23} J/K). [112]

$$S(N, V, E) = k_B \ln \Omega(N, V, E) \tag{2.1}$$

The temperature (T) can easily be related to the entropy and therefore, to the number of eigenstates using Equations 2.2 and 2.3.

$$\left(\frac{\partial S}{\partial E}\right)_{V,N} = \frac{1}{T} \tag{2.2}$$

$$\left(\frac{\partial \ln \Omega(E,V,N)}{\partial E}\right)_{N,V} = \frac{1}{k_B T} = \beta \tag{2.3}$$

where we have defined β as the scaled reciprocal temperature. If we now have a system A that is in thermal equilibrium with a large heat bath (B), at temperature T, and the total energy of the system is fixed ($E = E_A + E_B$), we can use Equation 2.4 to obtain the probability of finding system A in one specific quantum state i with energy E_i [112].

$$P_i = \frac{\Omega_B(E - E_i)}{\sum_j \Omega_B(E - E_j)} = \frac{\exp(-E_i/(k_B T))}{\sum_j \exp(-E_j/(k_B T))} \tag{2.4}$$

We can see that the second part of the right-hand side of Equation 2.4 is the Boltzmann distribution for a system at temperature T. Additionally, the average energy $< E >$ of the system is:

$$< E > = \sum_i E_i P_i = -\frac{\partial \ln Q}{\partial \beta} \tag{2.5}$$

where Q is the partition function and in this case is equal to $\sum_i \exp(-E_i/(k_B T))$. The derivation of Equation 2.5 can be found in Frenkel and Smit's book [112]. Furthermore, the partition function Q is related to the Helmholtz free energy A as shown in Equation 2.6.

$$< A > = -k_B \, T \ln Q \tag{2.6}$$

The partition function Q is the sum over all possible states and energy levels, and its expression varies from ensemble to ensemble depending on which variables are kept constant in the system of interest. This function acts as a normalisation factor and is independent of the actual state of the system. An ensemble from a system where

the number of particles, the volume and the energy are fixed is called microcanonical, while an ensemble where the number of particles, the volume and the temperature are kept constant is called canonical. Other important ensembles are the NPT (constant pressure, temperature and number of particles) and the Grand Canonical (fixed volume, temperature and chemical potential). All the simulations in this work, except those in the gas phase, were carried out in the NPT ensemble. In this ensemble, the density of a liquid can easily be obtained by dividing the constant number of molecules by the ensemble average of the volume. The partition functions for the most important ensembles are shown in Table 2.1 [60].

Table 2.1: Partition function of the most important ensembles in the quantum realm.

Ensemble	**Fixed variables**	**Partition function**	**Simulation examples**
Microcanonical	NVE	Ω	Heat exchange between two subsystems
Canonical	NVT	$Q = \Omega \exp(-\beta E)$	Heat capacity at constant volume
Isothermal-Isobaric	NPT	$\Delta = Q \exp(-\beta PV)$	Density
Grand Canonical	μVT	$\Xi = Q \exp(\beta \mu N)$	Adsorption

2.2.1 Classical statistical mechanics

The thermal average of some observable X could, in theory, be computed using Equation 2.7 since we now know the probability that a system at temperature T will be found in an energy eigenstate with energy E_i.

$$< X >= \frac{\sum_i \exp(-E_i/(k_B T)) < i|X|i >}{\sum_j \exp(-E_j/(k_B T))} \tag{2.7}$$

where $< i|X|i >$ is the expected value of X in quantum state i. Unfortunately, the number of quantum states that contribute to the average is astronomically large and thus, this approach is not practical. Luckily, we can express the equations described so far in the classical limit for most systems, where interacting atoms are the basic units instead of describing all particles as wavepackets. In order to do this, the Hamiltonian operator (H) should be used to represent the total energy of the system, and the sum over states should be replaced by an integration over all coordinates (r) and momenta (p). Consequently, the classical partition function for 3D systems is defined by Equation 2.8:

$$Q_{classical} = \frac{1}{h^{3N} N!} \int \exp\left[-\beta H(\mathbf{p}^N, \mathbf{r}^N)\right] \mathrm{d}\mathbf{p}^N \mathrm{d}\mathbf{r}^N \tag{2.8}$$

where h is the Planck's constant and the factor $1/N!$ takes into account the indistinguishability of particles that correspond to the same quantum state but differ in their labelling. Also, r^N and p^N stand for the coordinates and momenta of all N particles, respectively.

Finally, the thermal average of the observable X can be obtained using Equation 2.9:

$$< X >= \frac{\int X(\mathbf{p}^N, \mathbf{r}^N) \exp\left[-\beta H(\mathbf{p}^N, \mathbf{r}^N)\right] \mathrm{d}\mathbf{p}^N \mathrm{d}\mathbf{r}^N}{\int \exp\left[-\beta H(\mathbf{p}^N, \mathbf{r}^N)\right] \mathrm{d}\mathbf{p}^N \mathrm{d}\mathbf{r}^N} \tag{2.9}$$

where, $\mathrm{e}^{-\beta H}$ is the Boltzmann factor, and the observable X has been expressed as a function of coordinates and momenta. Additionally, the Hamiltonian is the sum of the kinetic (k) and potential (U) energy of the system. The kinetic energy is a quadratic function of the momenta ($\sum_i \frac{p_i^2}{2\, m_i}$) and thus, the integration over momenta can be done analytically. On the other hand, the multidimensional integral over particle coordinates, defining the potential energy of the system, cannot be computed analytically, except in a few cases, and numerical techniques need to be used. Two important techniques can be used to solve Equation 2.9. These techniques are Monte Carlo (MC) and molecular dynamics (MD), which are explained in the following two sections.

2.3 Monte Carlo simulations

The simplest Monte Carlo technique is called random sampling or "brute force" Monte Carlo. With this technique, one can evaluate the integral of a function $f(x)$ over the interval [a,b] using Equation 2.10:

$$\int_a^b f(x)\, \mathrm{d}x = (b-a)\ < f(x) > \tag{2.10}$$

where $< f(x) >$ is the unweighted average of $f(x)$ over the interval [a,b] and is ob-

tained by evaluating $f(x)$ at a large number of x_i values randomly distributed over this interval. However, using brute force Monte Carlo to solve Equation 2.9 is extremely inefficient since most of the computing time may be spent on points that contribute negligibly to the average. Consequently, it is better to sample many points in regions where the Boltzmann factor is significant and few elsewhere. This is the idea behind importance sampling– the x_i values are no longer randomly distributed but instead they are chosen according to some probability density function. Unfortunately, we cannot generate points with an absolute probability density proportional to the Boltzmann factor without analytically computing the system's partition function. Thus, the simple importance sampling scheme cannot be used to solve Equation 2.9. Metropolis et al. [114] created a method that can solve Equation 2.9 by generating points in configuration space with a *relative* probability proportional to the Boltzmann factor. In other words, a random walk is created through the region of phase space where the numerator of Equation 2.9 is non-negligible. The sequence of trials satisfies the two conditions of a Markov chain: each trial's outcome belongs to a finite set of outcomes, and the outcome of each trial depends only on the outcome of the preceding trial. [60].

To carry out a Monte Carlo simulation, we must first prepare a system in a configuration $\mathbf{r}^N$ with a non-zero Boltzmann factor and then add a small random displacement to generate a new trial configuration. This trial move is always accepted if it is energetically favourable, and if it is not, it is accepted with a probability equal to $e^{-\beta(U_{new}-U_{old})}$. Some unfavourable moves need to be accepted to allow for fluctuations at equilibrium and avoid a frozen system. In equilibrium, the average number of accepted moves from a state o to any other state n is exactly cancelled by the number of reverse moves.

Monte Carlo simulations present a few advantages over molecular dynamics simulations. They do not have to calculate the forces between all the particles in a system like it is necessary for MD simulations. This step is very time-consuming, as will be explained in section 2.4. Furthermore, the MC technique allows the user to efficiently cross over energy barriers, making it easier to equilibrate complicated systems with high energy barriers. Additionally, MC simulations in the NVT and NPT ensembles can be run without using thermostats and barostats, which are necessary for MD

simulations.

On the other hand, MD simulations give information about the evolution of the system over time and thus, transport properties (e.g. diffusion and viscosity) can easily be obtained using MD. In this work, the diffusion constant is used to optimise and validate the new PolCA model and therefore, MD is the method of choice. Another reason to choose MD is convenience since many freely available software packages contain all the algorithms necessary to run the desired MD simulations. It is important to mention that both MD and MC simulations normally have the same setup, which consists of representing the molecules as collections of interaction sites centred in the atoms, and they both use force fields for the potential energy terms [115]. Classical force fields are explained in section 2.6.

2.4 Molecular dynamics simulations

Molecular dynamics simulations are used to calculate the equilibrium and transport properties of a classical many-body system. This technique consists of selecting a model system of N particles, computing the forces on all particles, and solving Newton's equations of motion (Equation 2.11) for this system until its properties no longer change with time (equilibrium has been reached) [112].

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i \,, \;\; i = 1....N \tag{2.11}$$

The forces in Equation 2.11 are the negative derivatives of a potential function $U(r_1, r_2, .., r_N)$, which will be described in detail in section 2.5:

$$\mathbf{F}_i = -\frac{\partial U}{\partial \mathbf{r}_i} \tag{2.12}$$

The equations are solved simultaneously in small discrete time steps, and the coordinates are written to an output file at regular intervals. These coordinates as a function of time are the trajectory of the system [116]. Different algorithms can be used to integrate Newton's equations of motion such as Verlet [117] and leap-frog [118]. The choice

between them depends on a balance between computational speed and accuracy. The Verlet algorithm uses the positions and velocities of all particles at time t to predict the positions at the next time step based on the forces acting on the particles, as shown in Equations 2.13 and 2.14 [60]:

$$r(t + \Delta t) = r(t) + \Delta t\, v(t) + \frac{\Delta t^2}{2\, m}\, \mathbf{F}(t) \tag{2.13}$$

$$v(t + \Delta t) = v(t) + \frac{\Delta t}{2\, m}\, [\mathbf{F}(t) + \mathbf{F}(t + \Delta t)] \tag{2.14}$$

The leap-frog algorithm uses positions r at time t and velocities v at time $t - 0.5\, \Delta t$ (Equations 2.15 and 2.16):

$$v(t + 0.5\, \Delta t) = v(t - 0.5\, \Delta t) + \frac{\Delta t}{m}\, \mathbf{F}(t) \tag{2.15}$$

$$r(t + \Delta t) = r(t) + \Delta t\, v(t + 0.5\, \Delta t) \tag{2.16}$$

This algorithm is algebraically equivalent to Verlet's. However, it minimises the loss of precision because, unlike Verlet's, it does not need to take the difference between the coordinates at time $t + \Delta t$ and $t - \Delta t$ to obtain the velocity at time t. Instead, the velocity at time t is obtained as the average between the velocities at times $t + 0.5\, \Delta t$ and $t - 0.5\, \Delta t$ [60]. Another alternative is to use the velocity Verlet algorithm. This algorithm is also equivalent to the original Verlet's form but it calculates the velocities at time step $t + 0.5\, \Delta t$ as a mid-step.

Constraints can be added to connect the coordinates of bonded atoms during the simulation. Bond lengths can be kept fixed using different methods. The SHAKE algorithm, proposed by Ryckaert et al. [119], uses Lagrange multipliers to get the forces along the bonds needed to keep the bond lengths constant. These multipliers are determined after each timestep based on the constraint conditions, and the atoms' positions are corrected accordingly. This method was designed to be used with the Verlet algorithm,

but there is another version called RATTLE [120], which works with velocity Verlet. In this work, the LINCS [121] algorithm is used because it is faster and more stable than SHAKE. LINCS resets bonds to their correct lengths using two steps. First, the new bonds' projections to the old bonds are set to zero, and the bond length after this projection is called l_i. Then, the projection of the bond on the old direction (p_i) is corrected using Equation 2.17:

$$p_i = \sqrt{2\,d_i^2 - l_i^2} \tag{2.17}$$

where d_i is the old bond's length. This equation accounts for the lengthening of the bond due to rotation.

The initial positions of each particle have to be specified before starting a simulation. These particles should not be placed at positions that result in an appreciable overlap of the atomic or molecular core to avoid large forces that can eventually crash the simulation [112]. Therefore, it is good practice to obtain the coordinates of a compound from a molecule editor software like Avogadro [122] and to add additional molecules to the system using the proper tools. These tools assure there will not be an overlap between the molecules, and the tool used in this work is the command "gmx insert-molecules" from the Gromacs software [123].

Once the system has been equilibrated, many macroscopic properties can be extracted from the output file by averaging over a trajectory that has reached equilibrium. Unfortunately, some thermodynamic properties (e.g. entropy and free energy) depend on the total partition function and hence, cannot be directly measured in a simulation. In these situations, special techniques need to be used, as explained in section 2.8 [112].

The number of particles in an MD or MC simulation is small, normally ranging between 10 and 10,000, and a large fraction of them lie at the surface of the simulation box. For example, 488 molecules out of 1000 molecules are at the surface in a simple cubic crystal [60]. Consequently, the boundary conditions' choice affects the system's properties since the molecules in contact with vacuum or a wall experience very different forces from those in bulk. Periodic boundary conditions (PBC) [124] in the three dimensions

are generally implemented to overcome surface effects when simulating bulk liquids. This method mimics the presence of an infinite bulk by placing the atoms of the system into a unit cell that is surrounded by translated copies of itself. Consequently, when a molecule leaves the central box, one of its images enters the cell through the opposite direction, conserving the number density of the system (Figure 2.1).

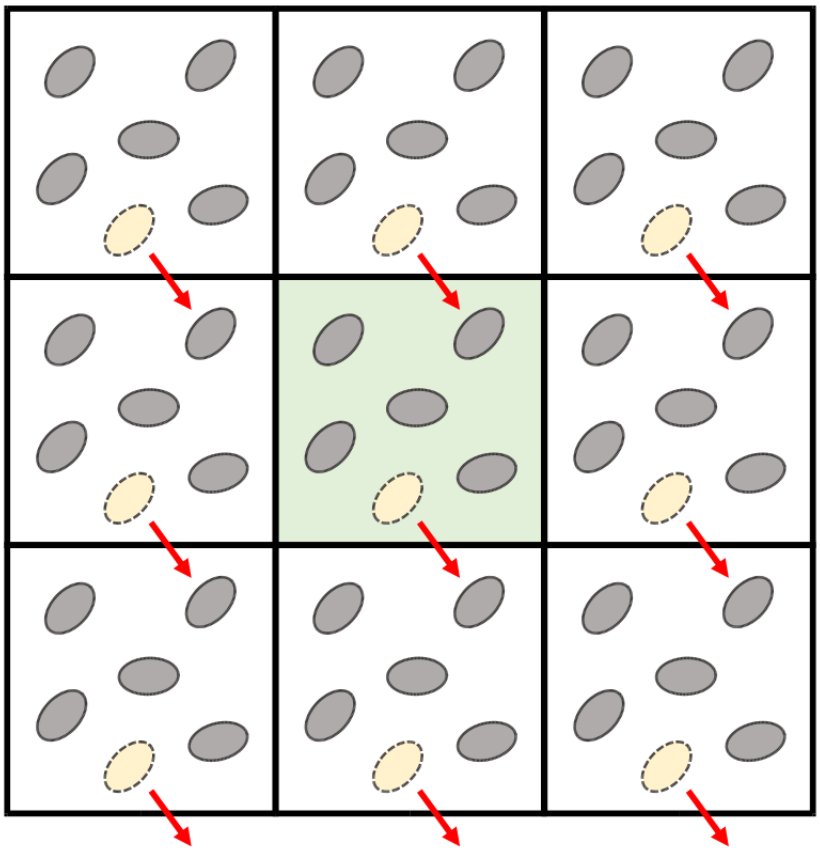

Figure 2.1: Periodic boundary conditions usually used to simulate liquids. Here, while the yellow molecule leaves the central box (coloured in green), its periodic image enters in it from the top.

The most time-consuming part of all MD simulations is computing the forces since interactions with all surrounding particles need to be considered when calculating the force on a particle i. In practice, short-range potential functions are commonly employed due to the large number of interactions. Short-range means that all intermolecular interactions beyond a specific cut-off distance r_c are truncated. When the cut-off radius is less than half the periodic box's shortest size, only the interactions of a given particle i with the nearest periodic image of any other particles j in the simulated box

are calculated (this is called the minimum image convention).

2.4.1 Thermostats

The default ensemble for MD simulations is NVE, however, most of the quantities we are interested in need to be measured at a constant temperature. To simulate an NVT ensemble, we need to add a thermostat that simulates constant temperature (the system is in thermal contact with a large heat bath). The instantaneous temperature of a classical many-body system is obtained from the kinetic energy using Equation 2.18, where N_f is the number of degrees of freedom.

$$T(t) = \sum_{i}^{N} \frac{m_i \, v_i^2(t)}{k_B \, N_f} \tag{2.18}$$

The most commonly used thermostats are Berendsen [125], Andersen [126] and Nosé-Hoover [127, 128]. The Berendsen thermostat mimics weak coupling with first-order kinetics to an external heat bath with temperature T_0. This algorithm scales the velocities per time step by adding friction terms proportional to $(T_0/T - 1)$ in the equations of motion. Consequently, a deviation of the temperature of the system from the desired temperature T_0 decays exponentially with a time constant τ according to Equation 2.19 [125].

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{T_0 - T}{\tau} \tag{2.19}$$

Unfortunately, the Berendsen thermostat does not simulate a proper canonical ensemble because fluctuations of the kinetic energy are suppressed, and thus, a modified algorithm, called velocity rescaling [129], is customarily used instead for a more accurate simulation. This thermostat uses an additional stochastic term that ensures a correct kinetic energy distribution [116].

The Andersen thermostat simulates collisions with an imaginary heat bath by periodically changing the velocities of randomly selected particles to values taken from the Maxwell-Boltzmann distribution. This method can slow down the kinetics of the sys-

tem and should not be used when calculating transport properties. Another approach that also mimics the coupling of the system to a thermal bath is the Langevin thermostat. This method modifies the equations of motion by adding a random force and a frictional force proportional to the particle velocities [60].

The last thermostat explained here is Nosé-Hoover, which belongs to the extended-Lagrangian methods. This approach uses artificial coordinates (s) for the thermal bath with their associated momentum (p_s) and the desired temperature to calculate the Hamiltonian of an "extended system". This setup allows the energy to flow between the system and reservoir, where the reservoir has an associated "thermal inertia" Q_s. Thus, the coordinates of the particles are obtained using Equations 2.20 and 2.21. [60, 116]

$$\frac{\mathrm{d}^2 r_i}{\mathrm{d}t^2} = \frac{\mathbf{F}_i}{m_i} - \frac{p_s}{Q_s}\frac{\mathrm{d}r_i}{\mathrm{d}t} \tag{2.20}$$

$$\frac{\mathrm{d}p_s}{\mathrm{d}t} = (T - T_0) \tag{2.21}$$

2.4.2 Barostats

Similarly, as with the temperature, the system can be coupled to a "pressure bath" to simulate an NPT ensemble. Again, the simplest way to do this is by using the Berendsen pressure coupling scheme [125]. This algorithm rescales the coordinates and box vectors every n_{PC} steps with a matrix $\boldsymbol{\mu}$ and has the same effect of a first-order kinetic relaxation of the pressure towards a reference pressure P_0 [112]. Each element of the matrix $\boldsymbol{\mu}$ is calculated using Equation 2.22:

$$\mu_{ij} = \delta_{ij} - \frac{n_{PC}\Delta t}{3\tau_p}\,\beta_{ij} P_{0ij} - P_{ij}(t) \tag{2.22}$$

where, β is the isothermal compressibility of the system. A rough estimate of β is enough to get a good average of the pressure and therefore, the value for water can be used to simulate other liquids. For water, this value is 4.6 x 10^{-5} bar $^{-1}$ at 300 K and 1

atm, and most other liquids have similar values.

The Berendsen pressure control algorithm does not generate the exact NPT ensemble, and thus, other barostats need to be used to simulate the true NPT ensemble. One example is the Parrinello-Rahman barostat [130] used in this work, which is similar to the Nosé-Hoover thermostat explained in the previous subsection.

2.5 Interaction Potentials

As explained above, knowing the potential energy of a system is essential to run molecular simulations. MC simulations use this quantity to accept or reject a move, while MD simulations use the potential to calculate the forces that govern the entire time-evolution of the system. The potential energy U of an N particles system can be calculated using Equation 2.23:

$$U = \sum_i V_1(r_i) + \sum_i \sum_{j>i} V_2(r_i, r_j) + \sum_i \sum_{j>i} \sum_{k>j} V_3(r_i, r_j, r_k) + ... \tag{2.23}$$

where, r are the coordinates of the particles and $\sum_i \sum_{j>i}$ represents a summation over all pairs i and j, counting each ij pair only once. Additionally, the term $V_1(r_i)$ represents the effect of an external field, and the other terms correspond to particle interactions. The second term is called pair potential and is the most important. On the other hand, the third term in Equation 2.23 is already very time-consuming to compute, and thus, it is not explicitly accounted for in most molecular simulations despite being significant at liquid densities. Instead, this last term, as well as any higher-order terms, is partially included using an "effective" pair potential:

$$U \approx \sum_i V_1(r_i) + \sum_i \sum_{j>i} V_2^{eff}(r_{ij}). \tag{2.24}$$

The electrons of an atom are continually moving, and their specific positions create an instantaneous dipole moment that only lasts a tiny fraction of a second. These instantaneous dipoles cancel each other out over time, and thus, the average dipole

moment is zero for non-polar molecules. However, each instantaneous dipole can create temporary dipoles in the atoms near it, producing attractive forces between the atoms or molecules. These forces are called dispersion or London forces, and they are responsible for the condensation of non-polar gases [131]. As the distance between the atoms decreases, the potential energy also decreases until a minimum is reached where the balance between attractive and repulsive forces is the most favourable. If the distance between atoms is further decreased, the potential energy will rapidly increase due to electron-electron repulsion forces. The most common way of describing these two forces is using the Lennard-Jones (LJ) 12-6 potential (Equation 2.25):

$$u^{LJ}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.25}$$

where r_{ij}, ε_{ij} and σ_{ij} are the separation, LJ well-depth and LJ size, respectively for the pair of atoms i and j. This potential has an attractive tail of the form $(-1/r^6)$ which corresponds to interactions between instantaneous dipoles, and a steeply rising repulsive wall when the distance between the particles is lower than σ. This repulsive part is due to the overlap of electron clouds. An example of this potential is presented in Figure 2.2 for a particle with $\sigma = 0.302$ nm and $\varepsilon = 0.773$ kJ/mol.

This potential can also be written using the coefficients $C12$ and $C6$ instead of σ and ε:

$$u^{LJ}(r_{ij}) = \frac{C12_{ij}}{{r_{ij}}^{12}} - \frac{C6_{ij}}{{r_{ij}}^{6}} \tag{2.26}$$

where, $C12$ and $C6$ relate to σ and ε through Equations 2.27 and 2.28:

$$C12 = 4\ \epsilon\sigma^{12} \tag{2.27}$$

$$C6 = 4\ \epsilon\sigma^{6} \tag{2.28}$$

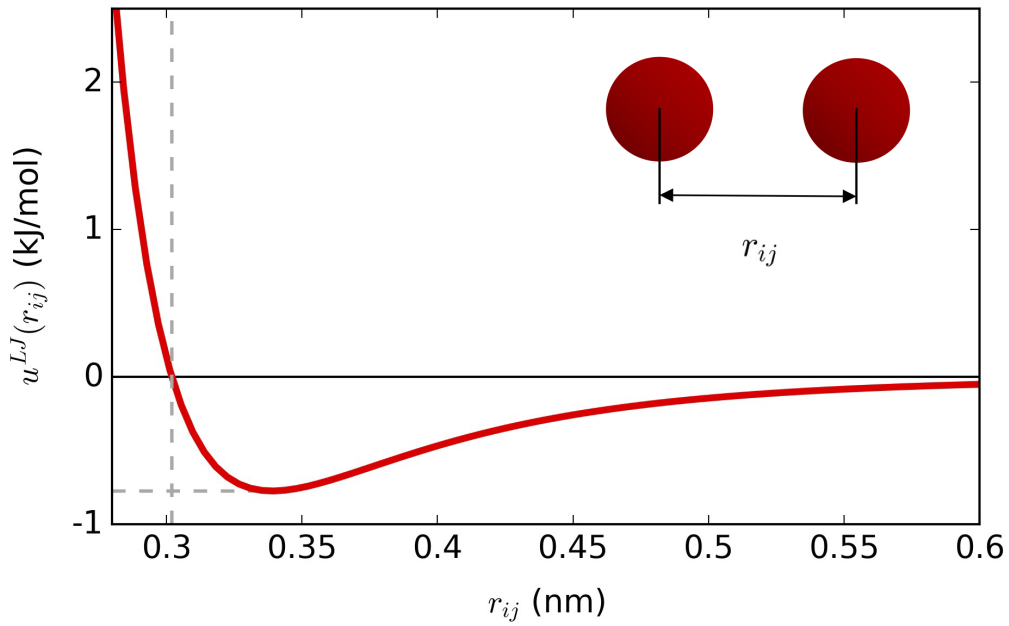

Figure 2.2: LJ potential of a particle with $\sigma = 0.302$ nm (vertical grey line) and $\epsilon = 0.773$ kJ/mol (horizontal grey line).

Additionally, the Lennard-Jones potential is a particular case of the Mie potential with $n = 12$ and $m = 6$, as can be seen from Equation 2.29.

$$u^{Mie}(r_{ij}) = \left(\frac{n}{n-m}\right)\left(\frac{n}{m}\right)^{\frac{m}{n-m}} \epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{n} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{m}\right] \tag{2.29}$$

The Mie potential only describes the interaction between uncharged particles (van der Waals forces), and thus, another equation needs to be added to the potential when simulating charged particles. Equation 2.30 is used to calculate the long-range Coulomb interactions:

$$u^{qq}(r_{ij}) = \frac{q_i\, q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \tag{2.30}$$

where q_i and q_j are the partial charges of particles i and j, and ϵ_0 and ϵ_1 are the vacuum permittivity and the relative permittivity of the medium, respectively.

As explained in section 2.4, intermolecular interactions usually are truncated when the

distance between particles is larger than a specific cut-off distance. In a 3D system, the contribution of the truncated tail to the potential converges if the potential function decays faster than r^{-3}. Dispersion forces follow this condition and thus, tail corrections can be easily estimated to compensate for the missing part of the potential. On the other hand, electrostatic interactions are long-range forces (they do not decay faster than r^{-3}) and therefore, the tail correction diverges. For this reason, electrostatic interactions cannot be calculated using a straightforward spherical truncation, as can be done with the dispersion forces, and other techniques need to be used such as Particle-Mesh Ewald (PME) [132]. This method accounts for the interaction of a particle with all its periodic images by assigning the charges in the simulation cell to a fine, regular mesh and using a fast Fourier transform to calculate the long-range part of the forces. [60] Another way to account for electrostatic interactions beyond the cut-off is using a reaction-field correction (u^{RF}) which represents the interaction of atom i with the dielectric medium outside the cut-off distance r_c [101]:

$$u^{RF}(r_{ij}) = \sum_{pairs\ i,j} \frac{q_i\, q_j}{4\pi\epsilon_0\epsilon_1} \left[\frac{-0.5\ C_{rf}\ r_{ij}^2}{r_c^3} + \frac{-(1 - 0.5\ C_{rf})}{r_c} \right] \tag{2.31}$$

where,

$$C_{rf} = \frac{(2\epsilon_1 - 2\epsilon_2)\ (1 + k\ r_c) - \epsilon_2\ (k\ r_c)^2}{(\epsilon_1 + 2\epsilon_2)\ (1 + k\ r_c) + \epsilon_2\ (k\ r_c)^2} \tag{2.32}$$

and ϵ_2 and k are the relative permittivity and inverse Debye screening length of the medium outside the cut-off sphere defined by r_c, respectively.

When simulating molecular systems, molecules are normally represented using interaction sites, which are typically centred on the molecule's nuclei but not always. For example, nitrogen can be thought of as two LJ atoms separated by a fixed bond length. Furthermore, the molecular charge distribution can be described using fictitious partial charges placed around the molecule. These sites are usually placed at the atom positions, the centre of a bond or within a lone pair. Also, dipoles and quadrupoles can be added to these sites if a better representation of the electrostatic potential is

required [133].

A significant problem of simulating charged particles is that induction interactions (induced dipoles) are not pairwise additive, and consequently, including them in a simulation is challenging. An induced dipole appears when a polar molecule or an ion distorts the electron distribution of an atom or a non-polar molecule in its proximity. The magnitude of this induced dipole depends on the strength of the external dipole or ion's charge and the atom or non-polar molecule's polarisability. These polarisation effects are very expensive to calculate because they cannot be broken down to a sum over pair interactions, and for this reason, they used to be often ignored. Nowadays, different approaches can be used to include polarisation in a model explicitly, like the induced point multipole model [134], the fluctuating charge model [135] and the Drude oscillator model [136] (see section 1.4). However, these polarisable models are very computationally expensive, and thus, non-polarisable models are commonly used instead. Non-polarisable models implicitly include induction effects in the fixed charge values, and therefore, they are faster but less accurate, especially for properties that involve a change of phase. Adding *post-facto* polarisation corrections to non-polarisable models is an excellent compromise between accuracy and speed. The *post-facto* polarisation corrections proposed by Leontyev and Stuchebrukhov [92] are used in this work to generate a new model, and they are explained in more detail in section 2.7.2.

When simulating molecular systems, additional terms need to be added to the potential to account for bonds, angles and dihedrals. These terms will be described in section 2.6. Bonds can be treated as classical harmonic springs or as fixed lengths. In most cases, the second approach is prefered because bond vibrations are difficult to handle and do not affect many liquid properties.

Furthermore, coarse-grained potential models can be used to reduce computational time. These models have interaction sites that represent more than one atom, reducing the number of explicit pairs needed for the calculations. One example is the MARTINI model [137] which was developed to model bilayers and proteins. In this model, the hydrogen atoms bonded to a heavier atom, like C, N and O, are not simulated explicitly, and instead, they are combined with the heavier atom forming a pseudo-atom with its

own interaction site. These pseudo-atoms are then combined to create larger beads that interact with each other through Lennard-Jones and Coulomb potentials.

United-atom (UA) models are a simple example of coarse-grained potentials, and these models treat all hydrogen atoms united to a carbon atom as a single interaction site. This means that CHx groups are treated as pseudo-atoms located at the sites of the carbon atoms. UA models have fewer interaction sites than all-atom models, and consequently, they are less computationally expensive and more straightforward to parameterise. Additionally, UA can use a longer time step than all-atom models during the integration process of alkanes since hydrogens require a smaller time step for an accurate integration due to their small mass. However, larger time steps cannot be used when the molecule contains non-aliphatic carbons. Polar and aromatic hydrogens need to be represented explicitly to simulate hydrogen bonding and π-stacking accurately [138]. On the other hand, an all-atom model is more realistic and more appropriate for simulations of solids and high-density liquids [102]. The force field developed here is an UA model because simplicity and efficiency were two essential features desired in the new PolCA force field.

2.6 Force Fields

Both Monte Carlo and molecular dynamics calculations use a force field to determine the evolution of the system. A force field is the set of equations and parameters that generate the total energy (MC) and forces (MD) [115]. Different sets of parameters can be used with the same set of equations; however, it is dangerous to make changes in a subset of parameters since all these parameters are usually correlated. Every change should be documented, compared against experimental data and published in a peer-reviewed journal before its use [116]. Force fields are usually created using an iterative process which searches for the set of parameters that minimise the difference between simulated and experimental results. Typically, bonded parameters are obtained using quantum mechanical (QM) calculations, and only non-bonded parameters are further optimised. QM simulations are often used to obtain an initial guess for the partial charges.

Some widely used general-purpose force fields are OPLS [139], CHARMM22 [140], AMBERff99 [141], GAFF [103], GROMOS [142] and TraPPE [102], to name a few. General-purpose force fields can be used to simulate a large range of compounds, while specific force fields have been parameterised for only one or a small subset of compounds. A model's transferability and precision depend, among other factors, on the amount of interaction-site types and the molecules used during its development. For example, using a common set of charges for the hydroxyl group of alcohols ignores how the alkyl chain affects its electronic structure, and thus, more precise values can be obtained if different charges are used for each alcohol. However, this approach reduces the transferability of the model and increases its complexity. [143] Most force fields, including those mentioned above, are constantly updated and extended.

Some researchers classify force fields in three classes [60]. Class I force fields calculate the potential function using Equation 2.33, where the sums $\sum_{bonds}$, $\sum_{angles}$, $\sum_{torsion}$ and $\sum_{non-bonded}$ are over all bonds, angles, dihedrals and non-bonded pairs separated by more than three bonds, respectively. Some force fields also include LJ and Coulomb 1-4 interactions, but they use different parameters to describe these interactions.

$$U^I = \sum_{bonds} u_{bonds} + \sum_{angles} u_{angles} + \sum_{torsion} u_{torsion} + \sum_{non-bonded} u_{non-bonded} \tag{2.33}$$

where,

$$u_{bonds} = \frac{k_r}{2}(r_{ij} - r_0)^2 \tag{2.34}$$

$$u_{angles} = \frac{k_\theta}{2}(\theta - \theta_0)^2 \tag{2.35}$$

$$u_{torsion} = c_0 + c_1[1 + \cos(\phi)] + c_2[1 - \cos(2\phi)] + c_3[1 + \cos(3\phi)] \tag{2.36}$$

$$u_{non-bonded} = \frac{q_i\, q_j}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.37}$$

here, r_0, θ, θ_0, k_r and k_θ are the equilibrium bond-length, measured bending angle, equilibrium bending angle, and force constants, respectively. Also, ϕ is the dihedral angle and c_0, c_1, c_2 and c_3 are the Fourier coefficients.

The torsion term can include improper dihedrals to enforce planarity around sp^2 centres. Improper dihedrals are formed by four atoms that are not all connected by covalent bonds. Additionally, the term $u_{torsion}$ could be expressed using the Ryckaert-Bellemans convention:

$$u^{RB}_{torsion} = \sum_{n=0}^{5} C_n(cos(\phi - 180^\circ))^n \tag{2.38}$$

The parameters of Equation 2.36 can be converted to Ryckaert-Bellemans parameters using the following equations:

$$C_0 = 2\ c_2 + c_1 + c_3 + c_0 \tag{2.39}$$

$$C_1 = -c_1 + 3\ c_3 \tag{2.40}$$

$$C_2 = -2\ c_2 + 8\ c_4 \tag{2.41}$$

$$C_3 = -4\ c_3 \tag{2.42}$$

$$C_4 = -8\ c_4 \tag{2.43}$$

Furthermore, Equation 2.37 is actually the sum of Equation 2.25 and 2.30 with $\epsilon_1 = 1$, and the LJ cross-interaction parameters are normally obtained using the Lorentz-Berthelot combining rules (Equations 2.44 and 2.45). There are other combining rules like for example using also a geometric average to obtain sigma [144].

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \tag{2.44}$$

$$\epsilon_{ij} = (\epsilon_{ii}\epsilon_{jj})^{1/2} \tag{2.45}$$

The other two classes of force fields are not used in this work, and therefore, they are only briefly mentioned here. Class II force fields, such as COMPASS [145], contain extra cubic or anharmonic terms in the stretching potentials to represent the dependence between bonds, angles and dihedrals in a molecule. Lastly, a class III force field includes the particles' polarisability in the calculations, and thus, it can represent the electrostatic interactions between particles more accurately.

The force field developed in this work was based on TraPPE-UA, and for this reason, TraPPE is explained in more detail below. This model was the starting point for most of the work presented here because it performs better than OPLS-UA and GROMOS at predicting hydrophobic solvation [81], which is one of the properties we are interested in.

2.6.1 TraPPE-UA Force Field for Alcohols and Ketones

The TraPPE-UA (Transferable Potentials for Phase Equilibria - United Atom) force field treats the CHx groups as pseudo-atoms located at the carbon atoms sites while it models all other atoms explicitly [69]. Non-bonded interactions are described by pairwise-additive LJ 12-6 potentials (Equation 2.25) and Coulombic interactions of partial charges (Equation 2.30 with $\epsilon_1 = 1$) [69]. The LJ parameters for the interaction between two different atoms are calculated using the Lorentz-Berthelot combining rules (Equations 2.44 and 2.45). Additionally, intramolecular interactions are described using fixed bond lengths, a harmonic potential for bond angle bending (Equation 2.35) and torsional potentials that restrict the dihedral rotations around bonds (Equation 2.36).

The bonded and non-bonded parameters for TraPPE-UA for alcohols [69] and ketones [146] are listed in Table 2.2. The Lennard Jones parameters were determined by fitting the single component vapour-liquid phase equilibria of methanol, ethanol, 2-propanol, 2-methylpropan-2-ol and acetone. The charges were taken from the OPLS-

UA force field [100, 147]. In the case of ketones, the carbonyl oxygen's parameters were taken from the TraPPE force field for carbon dioxide, and only ε and σ of the carbonyl carbon were optimised. [146]

Table 2.2: Bonded and non-bonded parameters for the TraPPE-UA force field [69].

stretch	r_0 [nm]			
CH_x-CH_y	0.154			
CH_x-OH	0.143			
O-H	0.0945			
C=O	0.1229			
bend (eq 2.35)	θ_0 [deg]	k_0/k_b [K]		
CH_x-(CH_2)-CH_y	114	62,500		
CH_x-(CH)-CH_y	112	62,500		
CH_x-(C)-CH_y	109.47	62,500		
CH_x-(CH_y)-O	109.47	50,400		
CH_x-(O)-H	108.5	55,400		
CH_x-C(=O)-CH_y	117.2	62,500		
CH_x-C=O	121.4	62,500		
torsion (eq 2.36)	c_0/k_b [K]	c_1/k_b[K]	c_2/k_b [K]	c_3/k_b [K]
CH_x-(CH_2)-(CH_2)-CH_y	0	335.03	-68.19	791.32
CH_x-(CH_2)-(CH_2)-OH	0	176.62	-53.34	769.93
CH_x-(CH_2)-(O)-H	0	209.82	-29.17	187.83
CH_x-(CH)-(O)-H	215.96	197.33	31.46	-173.92
CH_x-(C)-(O)-H	0	0	0	163.56
CH_x-CH_y-C=O	2035.58	-736.90	57.84	-293.23
non-bonded (eq 2.37)	ϵ/k_b [K]	σ [nm]	q	
CH_x-(**O**)-H	93	0.302	-0.700	
O-(**H**)	0	0	+0.435	
(**CH_3**)-OH	98	0.375	+0.265	
(**CH_3**)-CH_x	98	0.375	0	
CH_x-(**CH_2**)-OH	46	0.395	+0.265	
$(CH_x)_2$-(**CH_2**)	46	0.395	0	
$(CH_x)_2$-(**CH**)-OH	10	0.433	+0.265	
$(CH_x)_3$-(**CH**)	10	0.468	0	
$(CH_x)_3$-(**C**)-OH	0.5	0.580	+0.265	
$(CH_x)_4$-(**C**)	0.5	0.640	0	
C(=O)	40.0	0.382	+0.424	
O (sp^2)	79.0	0.305	-0.424	

The non-bonded parameters correspond to the atoms in bold.

Jorge [1] proposed new LJ parameters for alkanes, alkenes and alkynes that correct systematic deviations observed when using TraPPE-UA. These new parameters (Table 2.3) give better predictions of the solvation free energies of hydrocarbons in hydrophobic systems while simultaneously providing a better description of pure liquid properties.

Table 2.3: LJ Parameters proposed by Jorge for hydrocarbons [1].

Molecule type	Site	σ [nm]	ε/k_b [K]
Alkanes (sp^3)	CH_4	0.371	144.33
	CH_3	0.379	100.19
	CH_2 (linear and branched)	0.399	47.15
	CH_2 (cyclic)	0.392	54.12
	CH	0.473	10.22
	C	0.646	0.51
Alkenes (sp^2)	identical to TraPPE-UA		
Alkynes (sp)	CH	0.3315	75.53
	C	0.390	45.70

2.7 Bulk Properties and Polarisation corrections

2.7.1 Simulation details

MD simulations were run with Gromacs 5.1.2 [123]. First, N molecules were inserted in a cubic box of approximately 3 nm on each side, and the number of molecules for each simulation can be found in the corresponding results chapters. Each value of N was selected to maintain an approximately constant box size (circa 3 nm) to minimise the use of computational resources. It has been previously shown that the results (except for the self-diffusion coefficient; see below) are independent of system size provided long-range corrections are employed [81]. After this, a steepest descent minimisation run was performed, followed by a 100 ps NVT simulation at 298.15 K using the V-rescale thermostat [129] to accurately simulate the kinetic energy distribution. Subsequently, a 100 ps NPT simulation was carried out using the Berendsen barostat [125] with a target pressure of 1 bar to avoid having very large box oscillations when switching to the more accurate Parrinello-Rahman barostat [130]. Lastly, a 25 ns NPT simulation was run where the temperature was kept at 298.15 K using the V-rescale thermostat, with a time constant of 0.1 ps, and the pressure was controlled using the Parrinello-Rahman barostat, with a time constant of 2 ps and isothermal compressibility of 4.5 x 10^{-5} bar^{-1}. A leap-frog algorithm [118] was used to integrate Newton's equations of motion with a time step of 2 fs. All bonds were constrained using the LINCS algorithm [121]. The Verlet scheme was chosen for neighbour searching, with a cut-off radius of 1 nm for both van der Waals and electrostatic interactions, and long-range dispersion corrections for

energy and pressure were applied. Long-range electrostatic interactions were calculated using PME [132] with a Fourier spacing of 0.16 nm. Additionally, periodic boundary conditions were always applied unless stated otherwise.

In total, ten independent simulations were run for each compound and error bars were calculated to give a 95 % confidence interval using a Students' t equal to 2.262 [148]. In some cases, the error bars were too small to be plotted and therefore, they are not visible in the graphs.

2.7.2 Enthalpy of vaporisation

The enthalpies of vaporisation were obtained using Equation 2.46:

$$\Delta H_{vap} = E_{\text{gas}} - E_{\text{liq}} + RT + C_{pol} \tag{2.46}$$

where E_{gas} is the molar total energy in the vapour phase, E_{liq} is the molar total energy in the liquid phase, R is the ideal gas constant and T is the temperature. To estimate the total energy in the vapour phase (i.e. the intramolecular energy of an isolated molecule), a 50 ns simulation of a single molecule in vacuum with no boundary conditions was run for each molecule, and the first 10 ns were discarded. For these simulations, the cut-off scheme selected for neighbour searching was "group", and the cut-off radii were all set to 0.

The correction term (C_{pol}) was obtained using Equation 2.47 [92] and was proposed by Leontyev and Stuchebrukhov (see section 1.5):

$$C_{pol} = -\frac{(\mu_l - \mu_g)^2}{2\alpha} + \frac{6(\epsilon_{el} - 1)^2 \mu_l^2}{\pi(2\epsilon_{el} + 1)(\epsilon_{el} + 2)\alpha} \tag{2.47}$$

where, μ_l and μ_g are the liquid and gas phase dipole moments, respectively. Additionally, α and ϵ_{el} are the electronic polarisability of the molecule in the gas phase and the experimental high-frequency dielectric permittivity of the solvent, respectively. The first term on the right-hand side of Equation 2.47 is the distortion energy, which is favourable when moving from the liquid to the gas phase (as in a vaporisation pro-

cess). The second term is the interaction energy of the molecule with the surrounding electronic continuum, described here by a simple Kirkwood-Onsager model [149–151] for a dipole in a spherical cavity. The cavity's radius is treated self-consistently, and thus, it is eliminated from the polarisation energy equations [92]. As mentioned in section 1.5, liquid water's distortion and electronic contributions to the polarisation energy estimated from high-level quantum mechanical calculations nearly cancel each other out [108], however, it is not yet clear to what extent this effect is observed for other molecules of lower polarity. Finally, it is important to clarify that the correction C_{pol} is independent of the parameters of the model, unlike the correction proposed by Berendsen et al., which depends on the effective dipole of the model (μ_{model}) as can be seen from Equation 2.48 [93].

$$C_{Berendsen} = -\frac{(\mu_{model} - \mu_g)^2}{2\alpha} \tag{2.48}$$

The theory of Leontyev and Stuchebrukhov [92] is based on decoupling the contributions to the dielectric response of the solvent coming from the "fast" electronic degrees of freedom and those arising from the "slow" nuclear degrees of freedom, following the Born-Oppenheimer approximation. As such, the liquid state can be thought of as nuclear charges q_i moving in a polarisable electronic continuum with a high-frequency dielectric permittivity (ϵ_{el}) equal to n_D^2 (where n_D is the refraction index of the medium), while the relative permittivity of the gas phase is close to 1 ($\epsilon_{el} \approx 1$ for a perfect vacuum). This purely electronic continuum interacts with the polarised molecule and also screens its charges, resulting in effective charges that are much lower than those of the real liquid, i.e. $q_i^{eff} = q_i/\epsilon_{el}^{0.5}$. Hence, in order to consistently account for polarisation effects, the value of μ_l to be used in Equation 2.47 should be the dipole of the real liquid, which is usually much higher than the dipole moment of non-polarisable fixed-charge models due to the charge screening effect [92, 108]. The original Berendsen correction uses the model's dipole moment as a proxy for the real liquid dipole, hence leading to strongly underestimated distortion corrections [92]. Unfortunately, there is no experimental estimate of the liquid phase dipole moment for most molecules, and Equation 2.49, which was also proposed by Leontyev and Stuchebrukhov, was used to

approximate the liquid dipole. This expression provides a good estimate of the dipole moment of liquid water [108].

$$\mu_l = \mu_g \left(1 - \frac{12(\epsilon_{el} - 1)(\epsilon_{sol} - 1)}{\pi(\epsilon_{el} + 2)(2\epsilon_{sol} + 1)}\right)^{-1} \tag{2.49}$$

In the above equation, ϵ_{sol} is the static dielectric constant of the solvent and ϵ_{el} can be estimated from the square of the experimental refraction index at the Sodium D-line frequency.

2.7.3 Diffusion constant

The diffusion constant was calculated using the Einstein relation (Equation 2.50). The left-hand side of Equation 2.50 was obtained by linear regression of the mean square displacement (MSD), where the times were weighted according to the number of reference points. In every case, the fitting was done between t= 100 ps and t = 500 ps, where t is the time from the reference positions and not simulation time. The estimated error of each simulation was the difference of the diffusion coefficients obtained from fits over the two halves of the fit interval [116].

$$\lim_{t\to\infty} \langle \| \mathrm{r}_i(t) - \mathrm{r}_i(0) \|^2 \rangle_{i\in A} = 6\ D_A\ t \tag{2.50}$$

D_A is the self-diffusion constant of particles of type A and the left-hand side of Equation 2.50 is the limit of the mean square displacement when time tends to infinity.

It has been shown that the apparent self-diffusion coefficients depend significantly on the system size, and thus, it is important to correct for the deviations observed when comparing to an infinite system [152]. The diffusion coefficient depends linearly on the inverse box length ($1/L$), and consequently, the diffusion constant for an infinite system can be calculated by extrapolating a straight-line fit [152] (see Figure 2.3). For this reason, simulations were run for four different box lengths (3nm, 4nm, 5nm, 6nm), and a correction term was calculated for each molecule. This correction was obtained by subtracting the simulated diffusion constant obtained using a cubic box of 3nm each

side from the intercept of the fitted line (see Figure 2.3 for an example).

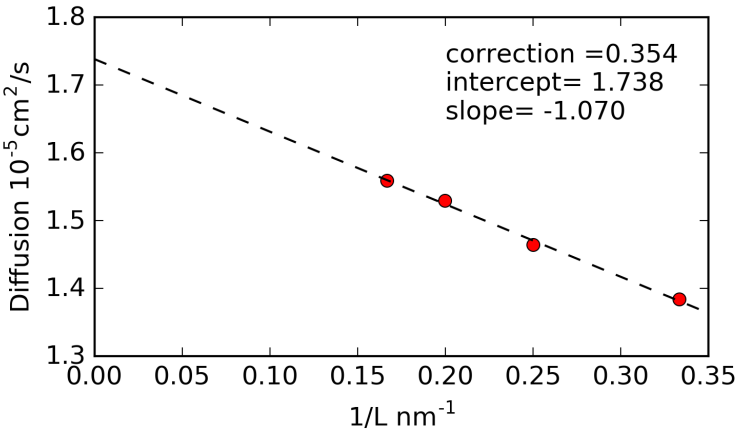

Figure 2.3: Diffusion coefficient of methanol as a function of the inverse box length, obtained using the new PolCA force field.

2.7.4 Dielectric constant

Gromacs calculates the static dielectric constant using Equation 2.51 [153]:

$$\epsilon_{sim} = 1 + \frac{\langle M^2 \rangle - \langle M \rangle^2}{3\epsilon_0 k_B T \langle V \rangle} \tag{2.51}$$

where V is the volume of the simulation box and M is the total dipole moment of the system.

Jorge and Lue showed that agreement with experimental data dramatically improves when a simple dipole moment scaling factor is used to take into account the purely electronic response of the liquid, which is not accounted for in non-polarisable models, as well as the charge screening caused by the presence of this electronic continuum [87]. Therefore, Equation 2.52 was used to correct the value obtained from the simulation.

$$\epsilon_{scaled} = \epsilon_{el} + (\epsilon_{sim} - 1)k^2 \tag{2.52}$$

where ϵ_{sim} is the dielectric constant obtained from Equation 2.51 and k is the ratio between the dipole moment of the real liquid (estimated from Equation 2.49) and the

dipole moment of the non-polarisable model. A $k = 1$ was used when the model's dipole was higher than the estimated liquid dipole.

2.8 Free energy calculations

In classical molecular simulations, three categories of free energy differences can be distinguished: conformational (between two distinct conformational states of the same system), alchemical (between two states differing on their Hamiltonian), and thermodynamic (between two thermodynamic state points) [154]. Alchemical free energy differences are the relevant ones for this work. The free energy, entropy, and related quantities are not simply averages of functions of the phase space coordinates of the system, and consequently, they cannot be measured directly in a simulation. These properties are called thermal quantities, and they cannot be measured directly in experiments either. Experiments always determine a derivative of the free energy, like the pressure, and then integrate the derivative along a path that links the state under consideration to a state of known free energy [112]. The same idea is applied in simulations; alchemical free energy differences between two molecular species can be calculated by slowly changing the Hamiltonian (H) of a the system from one that describes a system A to one that describes a system B. To accomplish this, H is made a function of a coupling parameter λ in such a way that λ=0 describes system A and λ=1 describes system B. In the NPT ensemble, the derivative of the Hamiltonian with respect to λ equals the derivative of the Gibbs free energy with respect to λ. [116]

The free energy of solvation is the free energy difference between the solute in the solution phase and in the gas phase [155], and since it is a state function, it can be calculated using the thermodynamic cycle shown in Figure 2.4, which includes non-physical transformations.

First, the solute in the gas phase is transformed into a dummy molecule. After this, the dummy molecule is transferred into the solution, and finally, all non-bonded interactions are turned on again. A dummy molecule is a molecule that does not interact with its environment, and as a result, the free energy difference associated with its insertion into the solution is equal to zero ($\Delta G_2 = 0$) [64]. Then, the free energy of solvation can

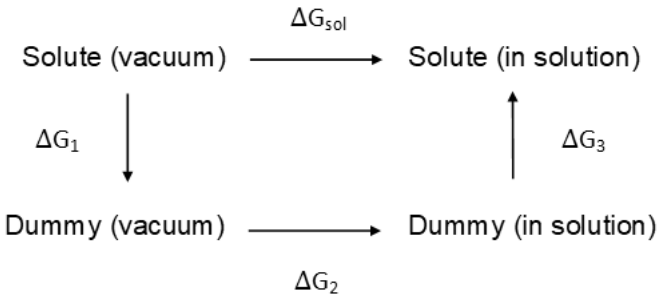

Figure 2.4: Thermodynamic cycle to calculate ΔG_{sol} using non-physical transformations.

be obtained using the following equation:

$$\Delta G_{sol} = \Delta G_1 + \Delta G_3 \tag{2.53}$$

ΔG_1 and ΔG_3 can be calculated using the coupling parameter λ. When calculating ΔG_1, λ=0 describes a fully interacting molecule in the gas phase, and λ=1 represents a dummy molecule. On the contrary, the last transformation (ΔG_3) goes from a non-interacting molecule (λ=0) to a fully interacting solute in the solution (λ=1). An schematic explanation on how the coupling parameter λ is used to calculate ΔG_3 can be seen in Figure 2.5.

It is also possible to calculate the solvation free energy in one step instead of using the cycle shown in Figure 2.4 if the solute's intra-molecular interactions are kept on when solute-solvent interactions are turned off. In this case, the decoupled state of the molecule corresponds to the proper vacuum state without periodicity effects and not to a dummy molecule [116]. This approach assumes that the contribution of the solute's intra-molecular degrees of freedom to the solvation free energy is the same in the gas and liquid phases. It has been shown that this assumption is accurate for the solvation of alkanes up to hexadecane [81]. However; for relatively large molecules, where intra-molecular non-bonded interactions might lead to kinetically trapped vacuum conformations, it might be better to calculate ΔG_{sol} using Equation 2.53 [116]. Therefore, to confirm the validity of this assumption, the free energy of self-solvation of decanol obtained using a one-step simulation was compared to that obtained using a full thermodynamic cycle (i.e., a standard two-step simulation), and no significant

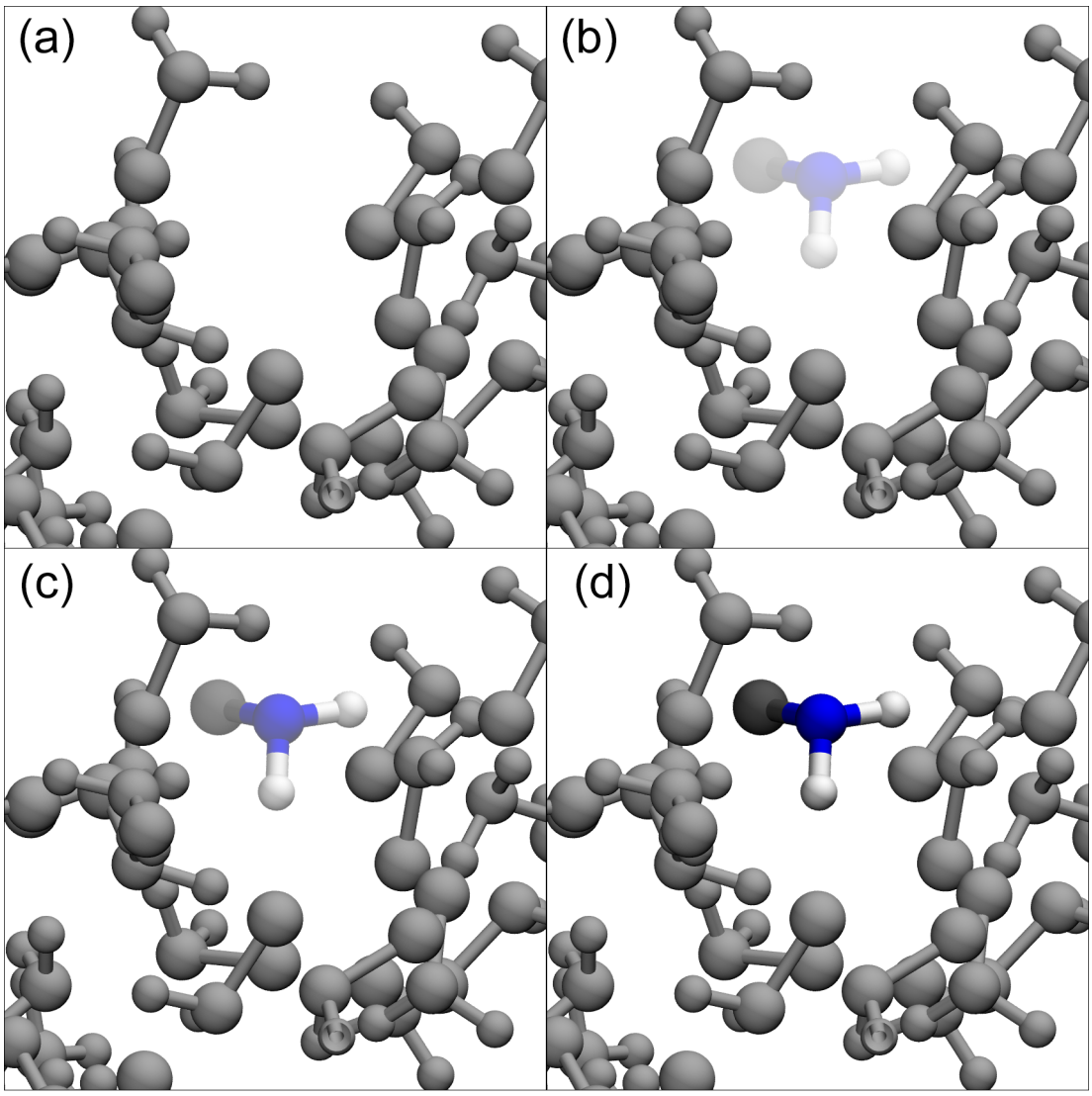

Figure 2.5: Schematic of the coupling parameter λ used to calculate solvation free energies. (a) λ = 0, the solute does not interact with the solvent. (b) λ = 0.3, interactions between the solute and solvent are partially turned on (c) λ = 0.7, interactions between the solute and solvent are partially turned on but to a higher degree than for λ = 0.3 (d) λ = 1, interactions between the solute and solvent are fully turned on.

differences were observed (see Table 2.4).

It is better to decouple electrostatic and non-electrostatic transformations separately to avoid atoms with residual charges from getting too close, and thus, create very large forces in the system and instabilities when integrating the equations of motion [155, 156]. Besides, electrostatic interactions are usually very well behaved once decoupled from other transformations [155]. For these reasons, all the calculations were done decoupling electrostatic and Lennard Jones interactions separately and then adding

Table 2.4: Comparison between the LJ component of the free energy of self-solvation of decanol calculated using the full two-step procedure and results obtained from 2 independent 5 ns simulations using the option couple-intramol = no (intra-molecular interactions are kept on when the inter-molecular interactions are turned off). ΔG_1 is the free energy difference between a non-interacting decanol molecule in vacuum and a decanol molecule also in vacuum (simulation time = 50 ns), while ΔG_3 is the free energy difference between decanol in solution and a non-interacting decanol molecule in solution (5 ns simulation).

Two-step calculation			One-step calculation
ΔG_1 (kJ/mol)	ΔG_3 (kJ/mol)	ΔG_{total} (kJ/mol)	ΔG (kJ/mol)
2.989 +/- 0.5978	-27.5018 +/- 0.1355	-24.5128 +/- 0.6129	-24.4621 +/- 0.0504 -24.1792 +/- 0.1480

them together to get the total free energy of solvation.

Furthermore, huge potential fluctuactions can occur when interaction sites are created or removed during a free energy calculation (λ close to zero or one). To avoid the endpoint singularity effect, the distance r in the nonbonded potential equation can be replaced by an effective distance $r_{eff}(\lambda)$ that generates smooth potential curves [157]. The soft-core potentials (V_{sc}) used in this work are presented in Equation 2.54, where V^A and V^B are the van der Waals or electrostatics potentials in state A (λ= 0) and state B (λ= 1) respectively, α_{sc} is the soft-core paramater and p_{sc} is the soft-core λ power [116]:

$$V_{sc}(r) = (1-\lambda)\, V^A(r_A) + \lambda\, V^B(r_B) \tag{2.54}$$

$$r_A = (\alpha_{sc}\, \sigma_A^6\, \lambda^{p_{sc}} + r^6)^{1/6} \tag{2.55}$$

$$r_B = (\alpha_{sc}\, \sigma_B^6\, (1-\lambda)^{p_{sc}} + r^6)^{1/6} \tag{2.56}$$

if σ is zero, an input parameter called *sc-sigma* is included in these equations.

2.8.1 Thermodynamic integration and the Bennett's acceptance ratio method

Two methods commonly used to estimate free energy differences are thermodynamic integration (TI) and the Bennett's acceptance ratio (BAR) method. The TI method integrates the derivative of the Hamiltonian with respect to λ from λ= 0 to λ=1 [64]:

$$\Delta G = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda \tag{2.57}$$

where the angular brackets indicate an ensemble average at a particular value of λ. The function $\langle \frac{\partial H}{\partial \lambda} \rangle_\lambda$ is unknown, but the derivatives of the free energy with respect to each λ can be evaluated using molecular modelling. Different methods can be used to evaluate the integral in Equation 2.57 numerically; like for example trapezoidal rule, Simpson's rule or fitting the entire data to a function that can be integrated analytically. However, it has been demonstrated that polynomial and cubic spline interpolation techniques tend to be more accurate than the trapezoidal rule [158, 159].

On the other hand, BAR calculates the Hamiltonian difference between two states λ_i and λ_j from configurations in the trajectory of λ_i and from configurations in the trajectory of λ_j, and uses these two values to predict the free energy difference between the two stages [160, 161]. Bennett showed that the free energy difference $\Delta G(\lambda_i \rightarrow \lambda_j)$ is given by [161]:

$$\Delta G(\lambda_i \rightarrow \lambda_j) = k_B T \left(\ln \frac{\langle f(U(\lambda_i) - U(\lambda_j) + C) \rangle_{\lambda_j}}{\langle f(U(\lambda_j) - U(\lambda_i) - C) \rangle_{\lambda_i}} \right) + C \tag{2.58}$$

where $f(x)$ denotes the Fermi Function:

$$f(x) = \frac{1}{1 + \exp\left(\frac{x}{k_b T}\right)} \tag{2.59}$$

When the term inside the natural logarithm equals 1, the first term of Equation 2.58 becomes zero and $\Delta G(\lambda_i \rightarrow \lambda_j) = C$, therefore to find $\Delta G(\lambda_i \rightarrow \lambda_j)$ the value of C,

for which Equation 2.60 is true, needs to be found:

$$\langle f(U(\lambda_i) - U(\lambda_j) + C)\rangle_{\lambda_j} = \langle f(U(\lambda_j) - U(\lambda_i) - C)\rangle_{\lambda_i} \tag{2.60}$$

BAR treats the two states λ_i and λ_j completely equivalently. The free energy difference between two physical states 0 and 1 with n intermediate states λ_l is obtained using Equation 2.61:

$$\Delta G = \sum_{l=1}^{n-1} \Delta G(\lambda_l \rightarrow \lambda_{l+1}) \tag{2.61}$$

The error is estimated using block averaging. The data is split into several blocks, and the value of $\Delta G(\lambda_i \rightarrow \lambda_j)$ is calculated for each block k. The average of $\Delta G(\lambda_i \rightarrow \lambda_j)$ and its error are calculated using Equations 2.62 and 2.64, respectively:

$$\Delta G_{av}(\lambda_i \rightarrow \lambda_j) = \frac{1}{n} \sum_{k=1}^{n} \Delta G_k(\lambda_i \rightarrow \lambda_j) \tag{2.62}$$

$$\sigma_m = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (\Delta G_k(\lambda_i \rightarrow \lambda_j)^{\,2} - \Delta G_{av}(\lambda_i \rightarrow \lambda_j)^{\,2})} \tag{2.63}$$

$$\sigma_{av} = \frac{\sigma_m}{\sqrt{n-1}} \tag{2.64}$$

The final error of the total free energy difference is the average error of each block's total free energy difference and not the error propagation of the sum of each $\Delta G(\lambda_i \rightarrow \lambda_j)$. The total error cannot be computed as the error propagation of each interval since each interval's variance is correlated to the variance of the neighbouring intervals. For example, the free energy differences between states 1 and 2 and between 2 and 3 both contain statistical information from state 2 [162].

It has been shown, by Brucker and Boresh, that TI may require significantly longer simulation times than BAR at each λ-state when the convergence of $\langle \frac{\partial H}{\partial \lambda} \rangle_\lambda$ is slow [163].

Besides, the BAR method is incorporated in the Gromacs 5.1.2 software and therefore, the free energy difference can be predicted using only one command. This command also calculates the relative entropy of both states in each other's ensemble, which is a measure of phase space overlap [116] and therefore, provides an estimate of the precision of the calculation.

2.8.2 Phase space overlap and simulation time

Phase space is the 6N-dimensional space formed from the 3N configurations and 3N momentum components of all N atoms (see section 2.2.1). When momentum contributions are included analytically, each point in the phase space represents a unique configuration. The configurations that must be sampled well by a free energy algorithm to produce accurate results constitute the important phase space of a system [164].

The BAR method is at its best when there is an overlap between the important phase spaces for λ_i (system A) and λ_j (system B) since one of the requirements of this method is that the two studied systems need to be defined by potentials acting on the same configuration space [160]. If there is not sufficient overlap between these two systems, Equation 2.60 does not converge or it produces inaccurate results [161]. The overlap between them can be measured using ideas from information theory, specifically relative entropy. Relative entropy (s) is a measure of how one probability distribution diverges from a second and it is defined by Equation 2.65 [164]. In this equation, $p_A(\gamma)$ and $p_B(\gamma)$ are the probability densities for phase-space points ($\gamma \in \Gamma$) sampled in the A and B systems. When two distributions are identical, sA and sB are equal to zero. Larger values of sA and sB means more difference or more distance between the two distributions (increasing degree of non-overlap) [164].

$$\begin{aligned} s_A &= \int_\Gamma \mathrm{d}\gamma p_A(\gamma) \ln\left[\frac{p_A(\gamma)}{p_B(\gamma)}\right] \\ s_B &= \int_\Gamma \mathrm{d}\gamma p_B(\gamma) \ln\left[\frac{p_B(\gamma)}{p_A(\gamma)}\right] \end{aligned} \tag{2.65}$$

When the two ensembles are widely separated, it is better to add intermediate states that connect the two of them [160]. Thus, analysing the phase space overlap of a

particular system helps determine the number of lambdas needed and their values. This can be better understood using alcohols as an example. Table 2.5 shows methanol's self-solvation free energy calculated using 21 and 11 equidistant lambdas for the LJ and electrostatic components, respectively, and it is clear that the number of intermediate lambdas can be reduced while still getting a good overlap (consecutive rows of the same colour can be grouped together).

Table 2.5: Lambdas' relative entropies (s) for the LJ (left) and Electrostatic (right) components of the free energy of self-solvation of methanol, obtained using the TraPPE-UA force field and PME for the van der Waals interactions.

λ_A	λ_B	sA	sB
0	0.05	0.0209	0.021
0.05	0.1	0.0322	0.0321
0.1	0.15	0.0601	0.0591
0.15	0.2	0.0831	0.0787
0.2	0.25	0.1323	0.1199
0.25	0.3	0.1387	0.1165
0.3	0.35	0.15	0.122
0.35	0.4	0.1155	0.0904
0.4	0.45	0.1001	0.0786
0.45	0.5	0.0769	0.06
0.5	0.55	0.0685	0.0553
0.55	0.6	0.0568	0.0468
0.6	0.65	0.0484	0.0394
0.65	0.7	0.0436	0.036
0.7	0.75	0.0389	0.0328
0.75	0.8	0.0338	0.0283
0.8	0.85	0.0339	0.0291
0.85	0.9	0.0298	0.0256
0.9	0.95	0.0293	0.0254
0.95	1	0.0253	0.0216

λ_A	λ_B	sA	sB
0	0.1	0.0402	0.0406
0.1	0.2	0.0416	0.0428
0.2	0.3	0.0497	0.0519
0.3	0.4	0.0468	0.0507
0.4	0.5	0.0706	0.0775
0.5	0.6	0.1064	0.1192
0.6	0.7	0.1502	0.1756
0.7	0.8	0.2271	0.2556
0.8	0.9	0.2935	0.2929
0.9	1	0.2593	0.2355

Two histograms from two different ensembles are presented in Figures 2.6 and 2.7 to show the relationship between the shape of the histograms and the relative entropy. Figure 2.6 shows the histogram of the Hamiltonian difference between $\lambda = 0$ and $\lambda = 0.1$ calculated from configurations in the trajectory of $\lambda = 0$ (red line) and $\lambda = 0.1$ (blue line). From Table 2.5, it can be seen that the relative entropies for these lambda values are 0.0402 and 0.0406, respectively. These values are very close to zero indicating there is a good overlap between the two-phase spaces, as is also suggested from the histogram. On the other hand, Figure 2.7 presents the histogram of the Hamiltonian difference between 0.6 and 0.7 calculated from configurations in the trajectory of $\lambda = 0.6$ (red line)

and λ= 0.7 (blue line). The relative entropies for these stages are 0.1502 and 0.1756, respectively. These values are higher than those for λ= 0 and λ= 0.1, and therefore, a higher degree of non-overlap can be observed. Also, it can be noticed that the histogram in Figure 2.7 presents a much wider range than the one in Figure 2.6.

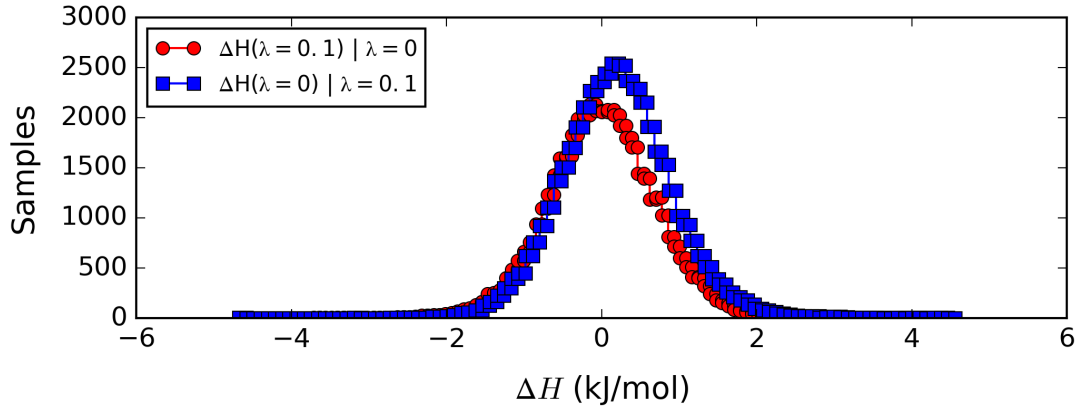

Figure 2.6: Histogram for the distribution of the Hamiltonian difference between 0 and 0.1. The red line shows the distribution of the values obtained from configurations in the λ= 0 trajectory and the blue line for the values calculated from configurations in the λ= 0.1 trajectory.

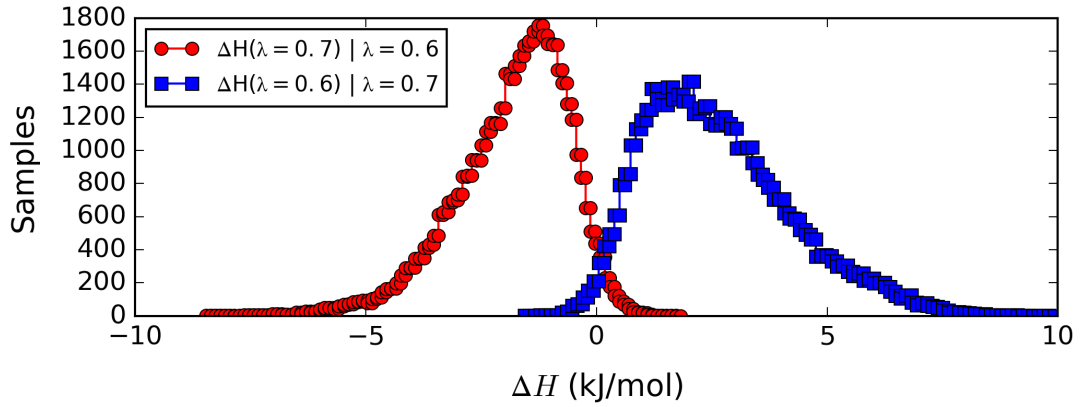

Figure 2.7: Histogram for the distribution of the Hamiltonian difference between 0.6 and 0.7. The red line shows the distribution of the values obtained from configurations in the λ= 0.6 trajectory and the blue line for the values calculated from configurations in the λ= 0.7 trajectory.

The number of λ-states and simulation length needed to reach convergence are correlated [161], nevertheless it seems that the minimal simulation length per λ-state varies a lot from system to system, and hence, it can only be determined on a case by case basis [163]. The number of λ-states for each particular system was obtained based on the relative entropies, and the simulation length was chosen by plotting free energy of self-solvation versus simulation time to ensure convergence. These values can be found

in the result chapters.

2.8.3 Simulation details

Free energy simulations were run using a leap-frog stochastic dynamics integrator [165] and, therefore, the temperature was kept constant using the Langevin method. A soft-core function was used to avoid instabilities close to the non-interacting state [166]. The soft-core parameters *sc-power* and *sc-sigma* were 1 and 0.3, respectively. *Sc-alpha* was 0.5 for the LJ term since this seems to be the ideal value for transformations of charge-less molecules [156], while a *sc-alpha* = 0 was used for the electrostatic component. All other parameters were the same as the ones used to calculate bulk properties except for the Fourier spacing, which was 0.12, and the time constant for pressure coupling, which in this case was 5 ps. These parameters were changed to improve efficiency. Long-range van der Waals interactions were calculated using a cut-off method with long-range corrections. The first calculations were run using PME for van der Waals interactions since this could be necessary for certain systems [167]. However, as can be seen from Figure 2.8, using a cut-off method with long-range corrections gives values that are consistent with those obtained using the more computationally expensive approach.

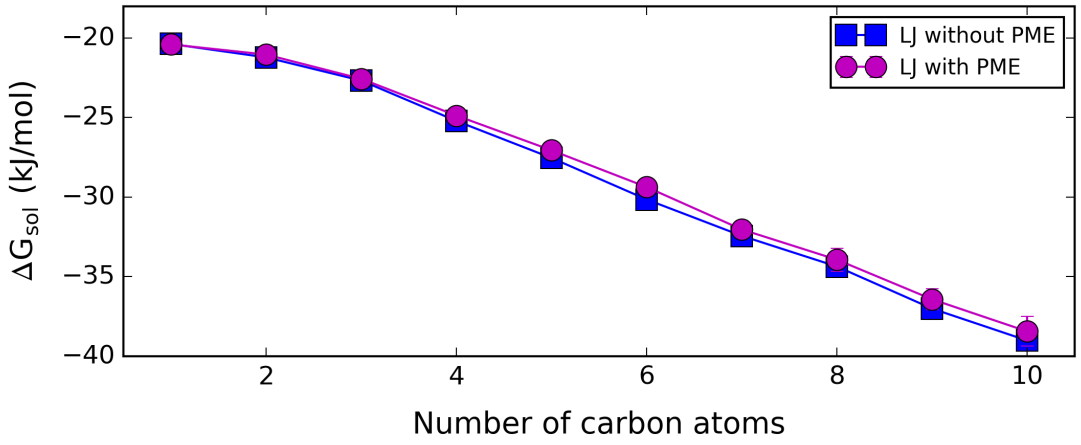

Figure 2.8: Free energies of self-solvation of primary alcohols calculated using TraPPE-UA. The circles in magenta were calculated using 15 λ-states for the LJ component and 7 λ-states for the electrostatics component, and van der Waals interactions were calculated using a cut-off with long-range corrections. The blue squares were obtained using PME to treat long-range van der Waals interactions, and the number of λ-states was 21 and 11 for the LJ and electrostatic components, respectively.

Polarisation corrections were added to the solvation free energies calculated from MD simulations. These corrections were calculated from Equations 2.47 and 2.49, but C_{pol} has the opposite sign since the solvation process moves a molecule from the vapour to the condensed phase (i.e. the direction is the opposite of the vaporisation process). Note that when applying those equations to solvation, the experimental values of μ_g and α correspond to the solute molecule, while values of ε_{sol} and ε_{el} correspond to the surrounding solvent, hence μ_l is the estimated dipole moment of the solute in the solvent of interest.

2.8.4 Experimental values

Experimental values for the free energies of solvation were taken from the Minnesota Solvation Database [168] and the Katritzky database [169]. Additionally, some free energies of self-solvation (i.e. when the solute and solvent are the same molecule) were calculated from the experimental vapour pressure and density at 298 K, using Equation 2.66 [168]:

$$\Delta G_s = -2.303\, RT \log\left(\frac{\rho/M_w}{P_v/24.45}\right) \tag{2.66}$$

where 24.45 atm is the pressure of an ideal gas at 1 molar concentration and 298 K, P_v is the vapour pressure in atm, ρ is the density in g/L and M_w is the molecular weight.

2.9 Force Field Parameters Optimisation

This section explains the procedure used for the development of the new PolCA force field.

First, a learning grid was created by running simulations for a few selected molecules using different non-bonded parameters and calculating certain target properties. The specific details about the calculated properties, the parameters' range and the chosen molecules are described in the result chapters since they varied for each functional group. Meta-models that predict molecular simulation results for a given set of pa-

rameters were used to decrease computational cost and allow for a more extensive exploration of the force field parameters. This technique has already been tested by Cailliez et al. [170] and proven to be successful. They also found that it was much more convenient and efficient to model each property used for the calibration process instead of directly modelling the full objective function [170]. Consequently, the learning set for each property was fitted to a second-degree model with cross-interaction terms using Equation 2.67 [171]:

$$f(X) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum \sum_{i<j} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 \tag{2.67}$$

where k is the number of parameters and $X = (x_1, x_2, .., x_k)$. To normalise the force field parameters used in Equation 2.67 and avoid working with different units, coded values of these parameters were used when creating the meta-models:

$$x_{coded} = \frac{2(x - x_c)}{x_{max} - x_{min}} \tag{2.68}$$

where x_{min} and x_{max} are the lowest and highest values for each control variable, respectively, and x_c is the initial guess used in the optimisation routine.

It is crucial to have a good correlation between the simulated values and the meta-models' values to obtain a reasonable estimation of the objective function's minimum. This was checked by plotting each calibration property's predicted versus simulated values and calculating the meta-model's predictivity coefficient Q_2 with Equation 2.69 [170]:

$$Q_2 = 1 - \frac{\sum_{i=1}^{N} [y_{sim}(X_i) - f_k(X_i)]^2}{\sum_{i=1}^{N} [y_{sim}(X_i) - \overline{y}]^2} \tag{2.69}$$

where, N is the number of points in the sample, $y_{sim}(X_i)$ is the simulated value, $f_k(X_i)$ is the value predicted by the meta-model and $\overline{y}$ is the mean value of $y_{sim}(X_i)$ on the whole sample. Also, once an optimum was found, new simulations were run using these parameters to corroborate the meta-model's predictions. As will be shown in later

chapters, the simple and easy to understand second-degree meta-models used in this work have predictivity coefficients close to 1 and thus, more complex machine learning algorithms were not tested. Nonetheless, in some cases the input parameters' space had to be divided into smaller regions as using a single set of fitted coefficients for the whole parameter domain did not lead to accurate predictions of the target properties. This problem could be solved using unsupervised machine learning algorithms that are able to cluster the input parameters regions.

An ideal force field is one where the difference between experimental and simulated values is zero. Consequently, an objective function was created that takes into account these differences:

$$F(X) = \sum_{j=1}^{N_{mol}} \sum_{k=1}^{N_{prop}} (f_k(X) - y_{exp_k})^2 \tag{2.70}$$

where N_{mol} corresponds to the number of molecules used in the optimisation and N_{prop} the target properties. Additionally, y_{exp_k} is the experimental value for each property.

Equation 2.70 was minimised using an optimised steepest descent algorithm. Starting from the initial guess (X_0), a new point (X_1) was found by searching along the opposite direction of the objective function's gradient ($\nabla F(X_0)$):

$$X_1 = X_0 - t\,\nabla F(X_0) \tag{2.71}$$

where t is the step length and determines the distance the algorithm needs to move along the specified direction. This value was found using 100 equidistant trial steps and selecting the one that returned the lowest function value. For each iteration, the minimum step length (t_{min}) was a tunable parameter and the maximum step length (t_{max}) was obtained from Equation 2.72 [172].

$$t_{max} = \frac{F(X_i)}{\nabla F(X_i) \nabla F(X_i)} \tag{2.72}$$

where X_i is the current iterate and $\nabla F(X_i)$ is the gradient of the function evaluated at X_i.

This procedure was repeated until either convergence was achieved with a tolerance of 10^{-8}, $X_{i+1} = X_i$ or the maximum number of iterations was reached. In each iteration, the new point X_{i+1} was obtained using the point found by the previous iteration (X_i) and the objective function's gradient evaluated at X_i:

$$X_{i+1} = X_i - t\,\nabla F(X_i) \tag{2.73}$$

The values of X_i and $F(X_i)$ obtained in each iteration were stored, and if convergence was not achieved, the algorithm returned the point X_i that produced the minimum $F(X_i)$ value. The optimisation routine was done in two parts. First, an optimum was found using larger trial steps ($t_{min} = 0.01\ t_{max}$), and then this optimum was chosen as the starting point for a new optimisation run with smaller steps ($t_{min} = 0.0001\ t_{max}$). The maximum number of iterations was 4000 for the first part (unless stated otherwise) and 100 for the second part since the second optimisation's initial point tends to be close to the optimum.

Furthermore, in some cases the optimum found by the algorithm was forced to stay inside some boundary conditions. In these situations, the coordinates of X_i were checked during each iteration and if one of these values was outside the boundaries, it was replaced by its closest bound to bring the point back into the desired region. For example, if X_i was (2, 0, 0) and the boundary conditions for the algorithm were X_{bmax} = (1,1,1) and X_{bmin} = (-1,-1,-1) in coded values, then X_i would become (1,0,0) before moving to the next iteration.

2.9.1 Sensitivity analysis

A variance-based global sensitivity analysis was performed to determine the most influential parameters for each target property. Global sensitivity analysis can be used to identify which factors make no significant contribution to the variance of the output and, therefore, can be fixed to any given value within their range of variation [173].

Sobol's indices [174] can be used to rank the input variables based on their importance. First-order indices (S_i) give the influence of each parameter taken alone while total sensitivity indices (S_T) consider the total effect of an input parameter [175]. The importance of interaction effects for a specific parameter depends on the difference between the first-order and total indices. If these two indices are close to each other, interaction effects are not important for that parameter. Furthermore, the sum of all S_i is equal to 1 for additive models (no interaction between parameters) and less than one for non-additive models. For this sensitivity analysis, only first-order indices were used.

First-order indices can be evaluated using Monte Carlo simulations to estimate the mean value, the total variance (D) and the partial variance due to variable x_i (D_i) (Equations 2.74 to 2.77 [175]).

$$f_0 = \frac{1}{N_{sim}} \sum_{m=1}^{N_{sim}} f(x_m) \tag{2.74}$$

$$D = \frac{1}{N_{sim}} \sum_{m=1}^{N_{sim}} f(x_m)^2 - f_0^2 \tag{2.75}$$

$$D_i = \frac{1}{N_{sim}} \sum_{m=1}^{N_{sim}} f(x^{(1)}_{(\sim i)m}, x^{(1)}_{im}) f(x^{(2)}_{(\sim i)m}, x^{(1)}_{im}) - f_0^2 \tag{2.76}$$

$$S_i = \frac{D_i}{D} \tag{2.77}$$

where N_{sim} is the number of samples, x_m denotes the mth sample point, and $x_{(\sim i)m}$ is the mth sample point without the variable i. Two matrices of random numbers, called (1) and (2), and order N_{sim} x k (k is the number of input variables) need to be generated first. The superscripts (1) and (2) indicate which matrix the sample point needs to be taken from.

The optimisation routine described here was implemented in python (code available at the University of Strathclyde KnowledgeBase 10.15129/eb1e582f-2891-4002-874b-

c76e12f88afb) and applied to develop the new molecular models for alcohols (Chapter 3), amines (Chapter 4) and ketones (Chapter 5).

Chapter 3

Polarisation-Consistent United Atom Force Field for Alcohols

3.1 Introduction

Alcohols are very interesting from a fundamental point of view [176–178], as they are the simplest molecules that combine a hydrophobic moiety with a hydrogen-bonding functional group. Additionally, these compounds are widely used in industry. For example, ethanol is used as a cosolvent to enhance the aqueous solubility of a drug by several orders of magnitude, and a drug's ability to diffuse through lipids can be described based on the octanol/water partition coefficient [7].

Some very well known force fields for alcohols are OPLS-AA [100], GROMOS [101], TraPPE-UA [69] and GAFF [103]. There are some modifications of these force fields like, for example, OPLS/2016 [99], which was reparameterised including solid-fluid experimental data. Several properties were used to validate the model, including the static dielectric constant, which was calculated using polarisation scaling corrections [87, 110]. However, the Berendsen correction was only applied to the enthalpy of vaporisation and was not considered when obtaining vapour-liquid coexistence properties, such as phase diagrams and vapour pressure. Another modification of OPLS was the one proposed by Kulschewski and Pleiss [143]. They modified the hydroxyl group's

partial charges in the OPLS-AA force field for different alcohols (including methanol to octanol) to consider the influence of the alkyl tail on the electronic structure of the hydroxyl group. Their simulated densities, self-diffusion coefficients and dielectric constants are closer to the experimental values than the values obtained using the original OPLS-AA force field. However, these modifications increase the model's complexity and reduce its transferability [143]. Another two relevant models are the improved GROMOS force field [77], obtained by fitting to the density, enthalpy of vaporisation and free energy of solvation in water and cyclohexane, and the GAFF-DC force field [179] which was parameterised to match methanol's dielectric constant. These last three models do not include polarisation corrections. In conclusion, although there are several force fields for alcohols available in the literature, none of them has been parameterised using consistent polarisation corrections to the best of our knowledge. For this reason, a new model was created, and a brief description of its development is presented in the following paragraphs.

First, TraPPE-UA was tested for ten linear alcohols with increasing chain length (from methanol to decanol), and it was noticed that the simulated density of longer alcohols was systematically overpredicted. After this, simulations were run using the alkyl parameters proposed by Jorge [1] combined with TraPPE original parameters for the hydroxyl group, and this model will be called "modified TraPPE" from now on. The densities of octanol, nonanol and decanol obtained using the modified TraPPE were closer to experimental data than those predicted by TraPPE. However, smaller alcohols' densities were considerably under-predicted, suggesting that the hydroxyl group's parameters needed to be re-optimised. Out of the six parameters in the hydroxyl group, σ, ε and partial charges of the oxygen and hydrogen atoms, only four are relevant for the parameterisation. The LJ parameters of the hydroxyl hydrogen can be set to zero because the electronic distribution of a hydrogen bond is approximately spherical around the centre of the more electronegative atom [180, 181].

The density, diffusion constant and enthalpy of vaporisation of methanol, 1-propanol, 1-pentanol and 1-heptanol were used for the parameterisation. The parameters for the alkyl chain groups were taken from the model proposed by Jorge for alkanes, alkenes

and alkynes [1] (Table 3.5). Additionally, the LJ parameters of the alkane methyl and methylene groups were used for the alpha carbons without modification. Each property used in the optimisation routine was fitted to a polynomial with cross-interaction terms, and then the score function was minimised using the steepest descent algorithm (see section 2.9).

3.2 Methodology

The optimisation procedure and the methods used to obtain bulk properties and free energies of solvation are described in detail in Chapter 2. Specific details for this chapter are presented below.

Bulk properties

Table 3.1 shows the number of molecules in each simulation box, which was selected to maintain an approximately constant box size. First, an alcohol molecule was inserted in a box of 27 nm^3 and then this box was solvated using 1000 iterations. In each iteration, a solvent molecule was inserted in a random position inside the box and depending on its distance to the existing atoms, the insertion of the molecule was accepted or rejected. [116]

Table 3.1: Number of molecules in the simulation box for each single-component system.

Compound	Number of molecules
Methanol	399
Ethanol	473
Propanol	361
Butanol	281
Pentanol	216
Hexanol	180
Heptanol	147
Octanol	128
Nonanol	111
Decanol	103
2-Propanol	367
2-Butanol	290
2-Pentanol	233
Tert-butanol	289
Tert-Amyl alcohol	228

Additionally, as explained in section 2.7.3, diffusion coefficients depend significantly on

the system size, and thus, corrections were calculated using TraPPE-UA to account for this. It was assumed that these corrections would be similar for other parameter sets, and therefore, the same correction values were used during force field optimisation and validation stages. New correction terms were later calculated using the final optimised force field to corroborate this hypothesis, and the values obtained were consistent with those obtained with TraPPE-UA. Table 3.2 presents the finite-size correction values to the diffusion coefficient for all alcohol molecules and two models: TraPPE-UA and the new PolCA.

Table 3.2: Correction terms for the diffusion constant in 10^{-5} cm^2/s.

	PolCA model	TraPPE-UA
Methanol	0.354	0.376
Ethanol	0.153	0.245
Propanol	0.152	0.116
Butanol	0.098	0.133
Pentanol	0.072	0.107
Hexanol	0.058	0.063
Heptanol	0.047	0.06
Octanol	0.054	0.051
Nonanol	0.043	0.038
Decanol	0.038	0.034

Solvation free energy calculations

For the first approach (Model 1), the Lennard Jones and electrostatic components were calculated using 21 and 11 equidistant λ-states, respectively. Additionally, long-range van der Waals interactions were calculated using PME [167]. For the other two cases, free energy simulations were run using 15 λ-states for the LJ component (0, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.8, 0.9 and 1) and 7 λ-states for the electrostatic component (0, 0.3, 0.6, 0.7, 0.8, 0.9 and 1). In these cases, van der Waals interactions were calculated using a cut-off method with long-range corrections. The number of intermediate λ-states and the way the Lennard Jones component was treated were changed to save computational time without losing accuracy, as discussed in section 2.8.

The length of the simulation for each system can be found in Table 3.3 and was chosen by plotting free energy of solvation versus simulation time to ensure convergence (see

Figures 3.1 and 3.2). A system was considered to have reached convergence when simulated values remained within +/- 0.15 kJ/mol of the final value and their error bars were not larger than 1 kJ/mol.

Table 3.3: Simulation time in ns for the free energy components of the systems studied here. Where N/A means not applicable.

Solute	**Solvent**	**LJ**	**Electrostatics**
Methanol	Methanol	5	5
Ethanol	Ethanol	5	15
Propanol	Propanol	5	15
Butanol	Butanol	7	35
Pentanol	Pentanol	7	35
Hexanol	Hexanol	7	50
Heptanol	Heptanol	7	50
Octanol	Octanol	10	50
Nonanol	Nonanol	10	50
Decanol	Decanol	10	50
2-Butanol	2-Butanol	7	35
Tert-butanol	Tert-butanol	7	35
Alcohols	Hexadecane	10	N/A

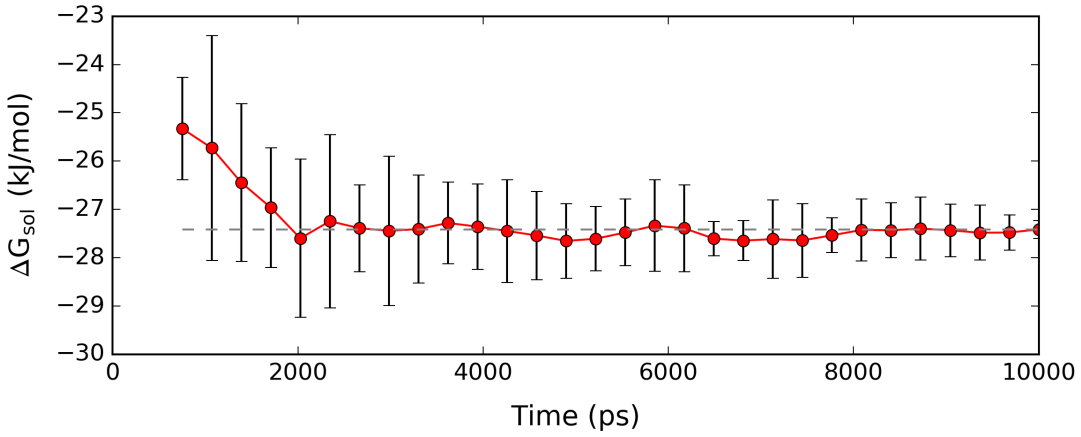

Figure 3.1: LJ component of the free energy of self-solvation of decanol at 298.15 K and 1 bar obtained using the PolCA model as a function of simulation time. Error bars were estimated as described in section 2.8, and the horizontal dashed line corresponds to the solvation free energy value of the last point.

The structure of liquid linear alcohols, which consists of hydrogen-bonded chains of molecules that sometimes link together to form branched aggregates [182], could explain why the Coulomb component of the free energy of solvation takes a long time to converge. To accurately simulate the solvation free energy, all possible scenarios need to be sampled. Between 15 and 25 % of alcohol molecules only have one hydrogen

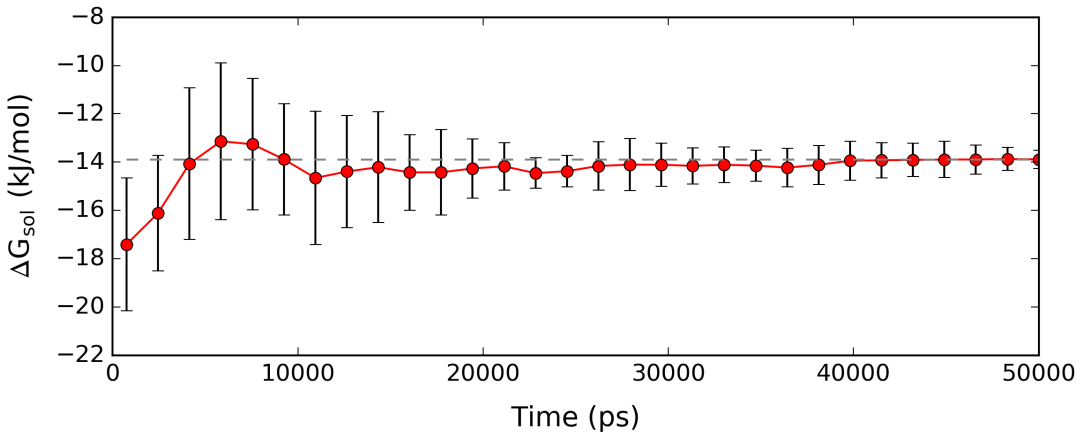

Figure 3.2: Electrostatic component of the free energy of self-solvation of decanol at 298.15 K and 1 bar obtained using the PolCA model as a function of simulation time. Error bars were estimated as described in section 2.8, and the horizontal dashed line corresponds to the solvation free energy value of the last point.

bond, while the percentage of molecules with two hydrogen bonds is around 60 and 75 %. Also, between 0 and 5 % of the molecules are not forming any hydrogen bonds, while 5 to 10 % have three hydrogen bonds [182]. Methanol and ethanol have the largest number of hydrogen bonds (approximately 1.87), while propanol has the lowest (around 1.82). Primary alcohols from butanol to octanol have all practically the same hydrogen bond distributions [182].

3.2.1 Polarisation corrections

As explained in section 2.7.2, there is no experimental estimate of the liquid phase dipole moment for most molecules, and Equation 2.49, proposed by Leontyev and Stuchebrukhov [92], was used to approximate the liquid dipole. The value obtained for methanol using this equation was 2.7 D, which is in agreement with values obtained from ab initio MD simulations. Pagliai et al. [183] reported a value of 2.64 D from simulations on a rather small box of 26 methanol molecules using the BLYP exchange-correlation functional. The same functional was used by Handgraaf et al. [184] with a larger box of 64 molecules, yielding a value of 2.59 D. In a study of the methanol vapour/liquid interface with a box of 120 molecules, Kuo et al. [185] report average bulk liquid dipole moments around 2.7 D, using both BLYP and PBE functionals. Finally, a more recent study by Sieffert et al. [186] report a range of values obtained with several

functionals in a box of 64 molecules, ranging from 2.58 to 2.84 D. Importantly, some of the calculations were carried out with dispersion corrections, which are known to play an important role in determining the structure of liquids [187].

Furthermore, all *post-facto* polarisation corrections used in this chapter are presented in Table 3.4. These corrections were used to calculate properties that involve a phase transition, and, as explained in section 2.7, they can be divided into two contributions: the negative distortion term and the positive electronic polarisation term. The values presented here correspond to a liquid-gas transition to be consistent with the definition of C_{pol}, and the corrections for the solvation free energy have the same magnitude but opposite sign. As well, this table contains the distortion correction proposed by Berendsen as a comparison. This correction uses the dipole moment of the model, and thus, it has different values for TraPPE and the new PolCA model, unlike C_{pol}, which is model-independent.

We can see from Table 3.4 that the distortion and electronic contributions nearly cancel each other out for pure alcohols, and thus, the net corrections are close to zero (ranging from -0.41 to 0.46 kJ/mol). This effect is not unexpected since it has previously been observed for water [92, 108]. In contrast, the Berendsen correction is significantly larger in magnitude, particularly for the TraPPE model, and always negative since it only considers distortion effects.

3.2.2 Optimisation

The density, diffusion constant and enthalpy of vaporisation of methanol, 1-propanol, 1-pentanol and 1-heptanol were simulated using different LJ parameters for the oxygen atom and different partial charges for the hydroxyl group. The partial charge of the α-carbon was obtained as the sum of the other two partial charges to assure the neutrality of the molecule. The parameters for the alkyl chain groups were taken from the model proposed by Jorge for alkanes, alkenes and alkynes [1] (Table 3.5). Additionally, the LJ parameters of the alkane methyl and methylene groups were used for the α-carbons without modification.

The dielectric constant was not considered as one of the fitting properties since it was

Table 3.4: Comparison between the polarisation corrections used in this work (C_{pol}) and the correction previously proposed by Berendsen et al. ($C_{Berendsen}$) for the TraPPE and PolCA models. The corrections are for transfer from liquid to gas and are expressed in kJ/mol.

Solute	**Solvent**	$C_{Distortion}$	$C_{Electronic}$	C_{pol}	$C_{Berendsen}$	
					TraPPE	**PolCA**
methanol	methanol	-9.251	8.842	-0.409	-2.896	-1.264
ethanol	ethanol	-7.297	7.074	-0.224	-1.907	-0.848
propanol	propanol	-6.169	6.073	-0.096	-1.455	-0.658
butanol	butanol	-5.079	5.101	0.022	-1.233	-0.576
pentanol	pentanol	-4.568	4.702	0.134	-0.890	-0.389
hexanol	hexanol	-3.611	3.860	0.249	-0.899	-0.426
heptanol	heptanol	-3.425	3.731	0.306	-0.637	-0.273
octanol	octanol	-2.783	3.175	0.393	-0.628	-0.284
nonanol	nonanol	-2.155	2.568	0.413	-0.730	-0.370
decanol	decanol	-1.848	2.307	0.459	-0.661	-0.335
2-propanol	2-propanol	-4.948	4.938	-0.010	-1.942	-1.019
2-butanol	2-butanol	-4.560	4.659	0.098	-1.455	-0.746
2-pentanol	2-pentanol	-4.042	4.251	0.209	-0.995	-0.470
Tert-butanol	Tert-butanol	-4.002	4.347	0.344	-1.160	-0.496
methanol	hexadecane	-1.860	8.730	6.870	-2.896	-1.264
ethanol	hexadecane	-1.168	5.483	4.314	-1.907	-0.848
propanol	hexadecane	-0.851	3.993	3.142	-1.455	-0.658
butanol	hexadecane	-0.658	3.087	2.429	-1.233	-0.576
pentanol	hexadecane	-0.572	2.682	2.111	-0.890	-0.389
hexanol	hexadecane	-0.458	2.152	1.693	-0.899	-0.426
heptanol	hexadecane	-0.429	2.014	1.585	-0.637	-0.273
octanol	hexadecane	-0.367	1.722	1.355	-0.628	-0.284
nonanol	hexadecane	-0.299	1.403	1.104	-0.730	-0.370
decanol	hexadecane	-0.271	1.271	1.000	-0.661	-0.335

Table 3.5: Lennard-Jones parameters and partial charges for the new PolCA united-atom force field for alcohols.

non-bonded [a] (Eq. 2.37)	σ [nm]	ϵ [kJ/mol]	Partial charge (q)
CH_x-(**O**)-H	0.2853	0.7733	-0.646
O-(**H**)	0	0	+0.406
(**CH_3**)-OH	0.379	0.833	+0.240
(**CH_3**)-CH_x	0.379	0.833	0
CH_x-(**CH_2**)-OH	0.399	0.392	+0.240
$(CH_x)_2$-(**CH_2**)	0.399	0.392	0
$(CH_x)_2$-(**CH**)-OH	0.438	0.085	+0.240
$(CH_x)_3$-(**CH**)	0.473	0.085	0
$(CH_x)_3$-(**C**)-OH	0.585	0.00426	+0.240
$(CH_x)_4$-(**C**)	0.646	0.00426	0

a) The non-bonded parameters correspond to the sites in bold.

estimated, using a bootstrapping technique, that at least six runs were needed to get an accurate value. This can be seen from Figure 3.3, which shows the average obtained using different combinations of 10 independent runs of methanol as a function of the number of runs. As such, including this property in the parameterisation stage would have significantly increased the computational cost by a factor of 6 since only one run was needed for each point in the learning grid to accurately predict the density, enthalpy of vaporisation and diffusion constant of alcohols. Furthermore, as will be shown in section 3.3, corrected dielectric constants predicted using the models obtained in this chapter are relatively close to experimental values (RMSD lower than 3.1), and thus, including meta-models for the dielectric constant created using simulated points with an error of approximately 1.25 units would probably lead to force fields with a similar or worse overall performance.

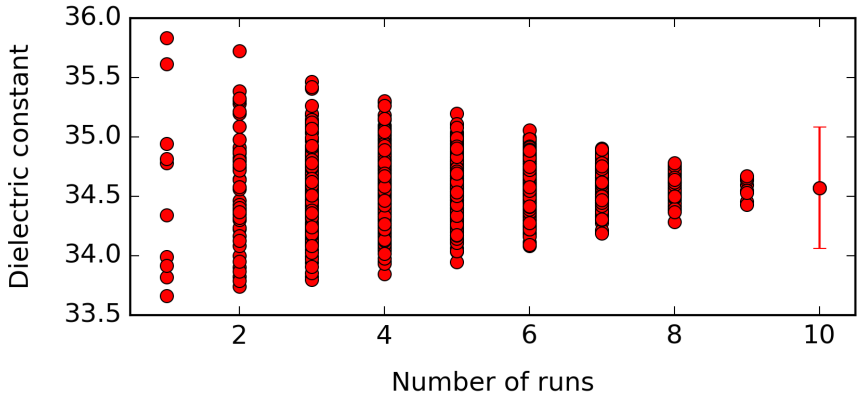

Figure 3.3: Average obtained for different combinations of 10 independent runs of methanol as a function of the number of runs, using the TraPPE-UA force field.

The simulations used to construct the meta-models were different for each approach. Nevertheless, in all cases, a full factorial design was used unless stated otherwise. In a full factorial design, different levels are assigned to each factor, and then all possible combinations of these levels across all factors are tested.

The correlation between simulated and predicted values was checked for each calibration property. In all cases, there was an almost perfect match between simulated values in the learning set and values predicted by the meta-models, as can be seen in Figures A7, A8 and A9 in Appendix A2. Additionally, simulations were run using the optimised parameters, and the values predicted by the meta-models were found to be

in agreement with the simulated values.

Equation 3.1 defines the objective function that was minimised using an optimised steepest descent algorithm (section 2.9). Here, j =1 corresponds to methanol, j =2 to propanol, j =3 to pentanol and j =4 to heptanol. Additionally, $f_k(X)$ is the value predicted using the meta-model at X, y_{exp} is the experimental value and ρ, D_A and ΔH represent the density, diffusion constant and enthalpy of vaporisation, respectively. The function ΔH_{error} was created to take into consideration the experimental error of the enthalpy of vaporisation ($error_{exp}$). This was only done for the enthalpy of vaporisation since its experimental error is higher than the one for the other two properties.

$$F(X) = \sum_{j=1}^{4} \left((f_k(X) - y_{exp})^2_{\rho_j} + (f_k(X) - y_{exp})^2_{D_{A_j}} + (f_k(X) - y_{exp})^2_{\Delta H_j} \, \Delta H_{error_j}(X) \right) \tag{3.1}$$

$$\Delta H_{error}(X) = min \left(1, \frac{|f(X) - y_{exp}|}{error_{exp}} \right) \tag{3.2}$$

The point that describes the original TraPPE force field was the initial guess. From this point, a new point was found by searching along the opposite direction of the objective function's gradient. This procedure was repeated until the maximum number of iterations was reached. In each iteration, the initial point was the point found by the previous iteration, and the objective function's gradient was evaluated at this point. As explained in section 2.9, the optimisation routine was done in two parts. During the first part, the maximum number of iterations was 3000 for the first approach and 4000 for all other cases. The number of iterations in the second approach was reduced because the initial point for the second optimisation routine tends to be close to the optimum, and thus, not many extra iterations are needed.

Model 1

The partial charges of the α-carbon and the hydrogen and oxygen of the hydroxyl group were changed simultaneously and in a proportional way. This was done by multiplying the partial charges taken from the TraPPE-UA force field by a scaling factor α. Consequently, a value of $\alpha = 1$ means that the partial charges of the α-carbon, the oxygen and the hydrogen are 0.265, -0.7 and 0.435; respectively.

The parameters range used to build the learning set for pentanol and heptanol was different from the one used for methanol and propanol. This was because methanol and propanol's properties were fitted first, and therefore, the learning set for pentanol and heptanol was created with a better idea about the optimum location. To construct an objective function based on the individual properties, coded values had to be consistent therefore, in all cases, the ranges for σ, ε and α were 0.032 nm, 0.34 kJ/mol and 0.25, respectively. It is important to mention than these ranges were not always symmetrical with respect to the central point (i.e. they are not +/- ranges). The set of parameters used for each molecule is presented in Table 3.6, and in every case, a full factorial design was applied. The total number of simulation runs was 360 for methanol and propanol, and 125 for pentanol and heptanol.

Table 3.6: Values used to obtain Model 1's meta-models. In all cases, a 3^k factorial was chosen.

Methanol Propanol	σ(nm)	0.278, 0.282, 0.286, 0.290, 0.294, 0.296, 0.298, 0.300, 0.302, 0.310
	ε(kJ/mol)	0.7, 0.77, 0.84, 0.91, 0.97, 1.04
	α	0.85, 0.90, 0.95, 1, 1.05, 1.1
Pentanol Heptanol	σ(nm)	0.278, 0.285, 0.292, 0.298, 0.305
	ε(kJ/mol)	0.7, 0.77, 0.84, 0.91, 0.97, 1.04
	α	0.85, 0.90, 0.95, 1, 1.05

Before creating the meta-models, every property was plotted with respect to each control variable keeping the other two variables constant, and it was noted that some design points did not follow the same trend as the rest (Figure 3.4). Thus, those points were not used to build the meta-models. In all cases, simulations with $\sigma \leq 0.286$ nm and $\alpha = 1.1$ were discarded. Additionally, when building the meta-models for propanol, the points obtained with $\sigma = 0.278$ nm and $\alpha = 1.05$ were not included in the learning set.

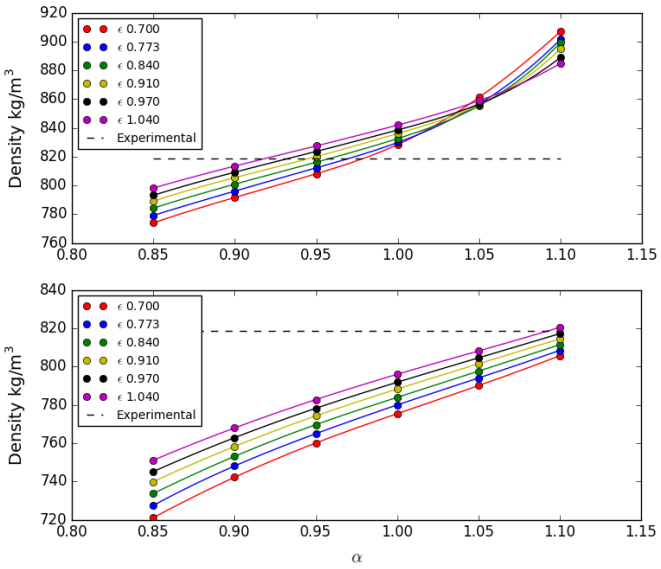

Figure 3.4: Simulated density of methanol at 298.15 K as a function of α and ϵ (kJ/mol) for the oxygen atom when σ = 0.278 nm (top) and σ = 0.298 nm (bottom). These points were created during the development of Model 1's learning set and the LJ parameters used for the carbon and hydrogen atoms were the same as the ones in Table 3.5.

Also, each molecule's learning set was broken into two parts to get a more accurate prediction. For methanol, two different sets of fitting parameters were generated, one for points with $\sigma \geq 0.294$ nm and one for points with a lower value of σ. In the case of propanol, the meta-model that predicts properties when $\sigma \geq 0.282$ nm was created using all points, except those with $\sigma = 0.278$ nm, while only points obtained with $\sigma \leq 0.290$ were used to calculate the fitting parameters needed to predict the properties obtained when $\sigma < 0.282$ nm. Finally, all points were considered when fitting the curve used to estimate the properties of pentanol and heptanol when $\sigma > 0.2915$ nm and only points with $\sigma \leq 0.2915$ nm were used to predict the value that each property would have when using a σ lower or equal to this value. The choice of what points to consider for each set was made by trial and error. This choice was based on the graph obtained when plotting predicted versus simulated values (the performance of the meta-model obtained using Model 1's final learning set can be seen in Figure A7 in Appendix A2).

Model 2

During Model 2's development, the partial charges of the hydroxyl group were not longer dependent on each other and instead they were allowed to vary independently.

Therefore, a total of four parameters were used during this model optimisation: σ, ε, the partial charge of the oxygen atom (q_O) and the partial charge of the hydrogen atom (q_H). The α-carbon's partial charge was calculated based on the hydroxyl group partial charges to maintain the neutrality of the molecule.

Initially, a learning set or grid was created with only 3 levels for each variable, with boundaries: $0.278 < \sigma < 0.326$, $0.700 < \varepsilon < 0.846$, $-0.8 < q_O < -0.6$ and $0.370 < q_H < 0.5$. After an optimum was found, a new grid was generated near those coordinates to increase the accuracy of the meta-models. The final levels for each variable were: σ = 0.302, 0.290, 0.294 and 0.298; ε = 0.846, 1.06 and 1.274; q_O = -0.6, -0.7 and -0.8 and q_H = 0.37, 0.435 and 0.5. Thus, the total number of simulations used in the final learning set was 108 for each molecule.

Furthermore, a meta-model for methanol's solvation free energy was created, hoping to get an optimum that can predict free energies of self-solvation of primary alcohols as well or better than TraPPE-UA. This meta-model was able to accurately predict the LJ component of methanol's solvation free energy, however, the predicted electrostatic component had an error of approximately 1 kJ/mol compared to the simulated value. Therefore, including this meta-model in the objective function did not lead to better results. The performance of this meta-model is shown in Figure 3.5, and the levels for each variable were: σ= 0.278, 0.290 and 0.302; ε= 0.7, 0.846 and 1.274; q_O = -0.6, -0.7 and -0.8 and q_H = 0.37, 0.435 and 0.5.

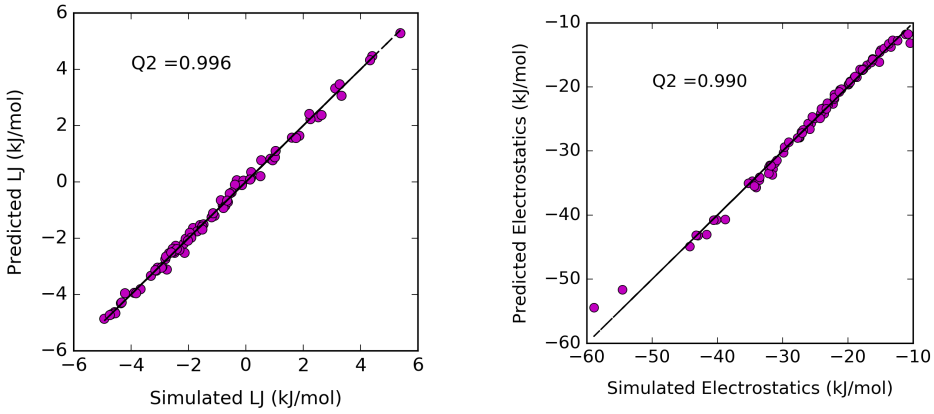

Figure 3.5: Plot of the simulated values versus the values predicted using the meta-models for methanol's self-solvation free energy at 298.15 K and 1 bar. The meta-models for the LJ and electrostatic components are shown on the left and right, respectively.

Model 3

The free energies predicted using the optimum obtained by changing σ, ε, q_O, and q_H were further away from the experimental data than the values obtained with TraPPE-UA (see results section later in this chapter). This discrepancy between experimental and simulated data was perhaps due to the LJ component's magnitude (the LJ component estimated with TraPPE was approximately 2 or 3 kJ/mol lower in magnitude than the one obtained with the new model). For this reason, a variance-based global sensitivity analysis was performed to determine the most influential factors for each property. From Figure 3.6, we can see that ε is not a relevant factor for the density, enthalpy of vaporisation and diffusion constant of alcohols, but it is quite important for the LJ component of the free energy of self-solvation. As a consequence, it was decided to keep ε constant and only change the other three parameters.

During Model 3's development, 1-propanol was not used as part of the parameterisation because including it produced a model that could predict the density of propanol slightly better at the expense of the diffusion constant of methanol. Furthermore, the function ΔH_{error} was not included in this model's objective function since its inclusion did not affect the optimised parameters. The levels for each variable were: σ= 0.290, 0.294, 0.298 and 0.302; ε= 0.7, 0.773 and 0.846; q_O = -0.6, -0.7 and -0.8 and q_H = 0.37, 0.435 and 0.5. To clarify, even though ε was kept constant during the optimisation, the meta-model was still a function of ε. Also, Model 3's learning set was constructed using the simulations ran during Model 2's development and therefore, it was not a full factorial. In particular, there were no combinations of σ= 0.294 or 0.298 with ε= 0.7 or 0.773.

The meta-models' performance for the three approaches can be found in Appendix A2 (Figures A7, A8, 3.5).

3.3 Results and Discussion

Several force fields are discussed in this section, and thus, a summary table showing all these models (Table 3.7) is included here to improve the readability of the chap-

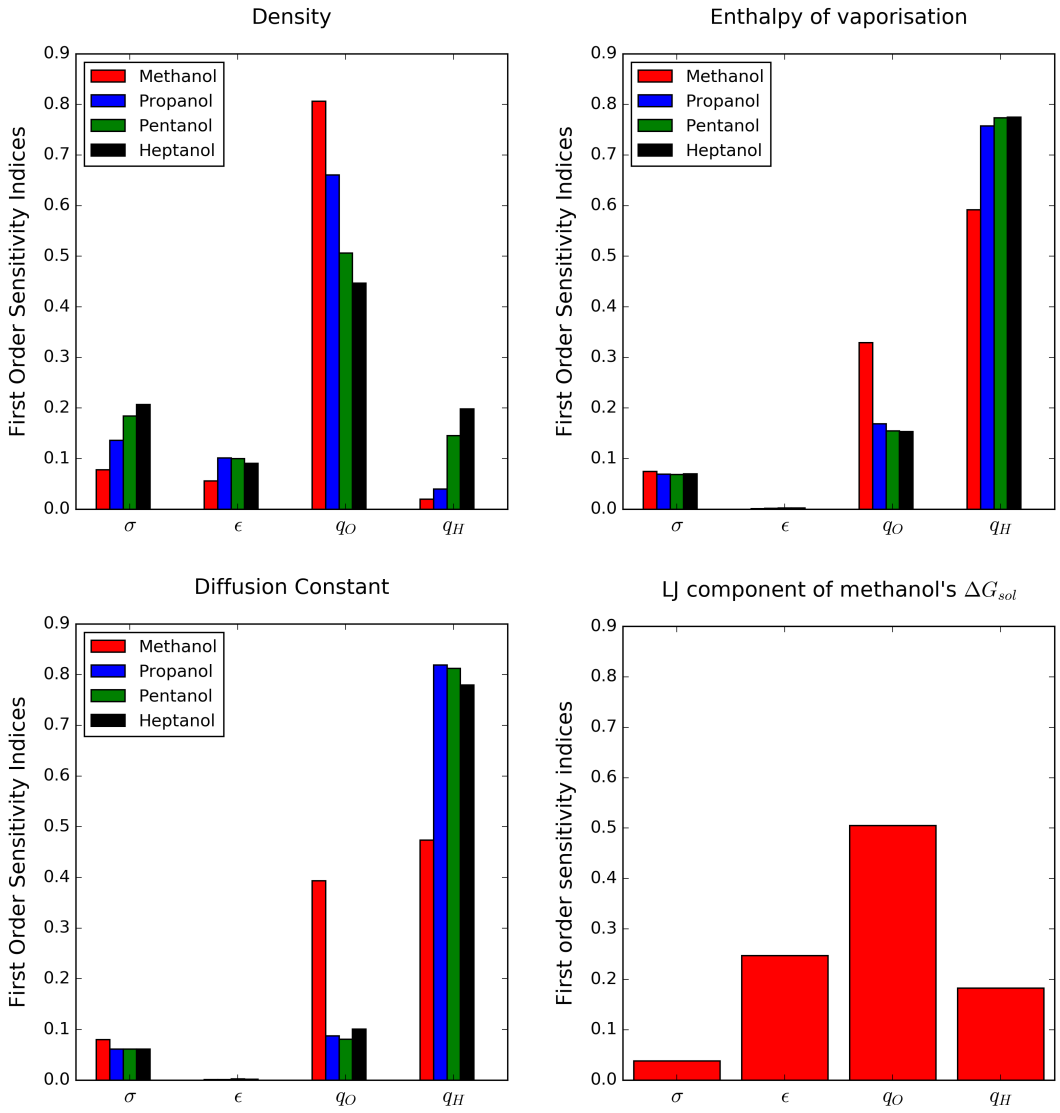

Figure 3.6: First order sensitivity indices for different properties obtained using Model 2's meta-models. Top: density (left) and enthalpy of vaporisation (right). Bottom: diffusion constant (left) and LJ component of the free energy of self-solvation of methanol (right). The different colours represent the different molecules: methanol (red), 1-propanol (blue), 1-pentanol (green), 1-heptanol (black).

ter.

3.3.1 TraPPE with polarisation corrections

The original TraPPE force field has been parameterised by fitting to the single-component vapour-liquid phase equilibria of a few alcohols, without applying polarisation corrections [69]. In this work, we test TraPPE's performance for several bulk properties (density, diffusion, enthalpy of vaporisation and dielectric constant)

Table 3.7: Force fields simulated in this chapter. Model 1, Model 2 and Model 3 use the LJ parameters proposed by Jorge [1] for the alkyl chain and the α-carbon, and have been parameterised taking into account post-facto polarisation corrections. Unless stated otherwise their results always include these corrections.

TraPPE	TraPPE-UA force field for alcohols [69]. This model has been fitted to the vapour-liquid coexistence curve of methanol, ethanol, 2-propanol and 2-methylpropan-2-ol.
TraPPE (C)	TraPPE-UA force field for alcohols with *post-facto* polarisation corrections.
Modified TraPPE	Combines TraPPE-UA's parameters for the hydroxyl group with the set of parameters for alkane groups proposed by Jorge [1]
Model 1	• Fitted to the density, enthalpy of vaporisation and diffusion constant of methanol, 1-propanol, 1-pentanol and 1-heptanol. • The hydrogen and oxygen's partial charges were taken from TraPPE and scaled using an scaling factor α. • Parameters optimised: σ and ε of the oxygen atom and the scaling factor α.
Model 2	• Fitted to the density, enthalpy of vaporisation and diffusion constant of methanol, 1-propanol, 1-pentanol and 1-heptanol. • The hydrogen and oxygen's partial charges were allowed to vary independently. • Parameters optimised: q_O, q_H and the LJ parameters of the oxygen atom (σ and ε).
Model 3/ PolCA	• Fitted to the density, enthalpy of vaporisation and diffusion constant of methanol, 1-pentanol and 1-heptanol. • The hydrogen and oxygen's partial charges were allowed to vary independently but ε was taken from TraPPE (ε = 0.773 kJ/mol) and kept constant during the optimisation. • Parameters optimised: σ of the oxygen atom, q_O and q_H.

and for solvation free energy calculations. For the dielectric constant and properties that involve a phase change, the model's performance without polarisation corrections was compared to the one obtained once these corrections were applied.

This force field can accurately predict the density of small primary alcohols (from methanol to pentanol), but from hexanol to decanol, it overpredicts this property (Figure 3.7). The difference between simulated and experimental values seems to increase with the alkyl chain's length, suggesting that the predictions will get worse as we move towards even heavier alcohols. Additionally, TraPPE overpredicts the self-diffusion constant of primary alcohols, except for methanol's diffusion, which is very close to the experimental value (Figure 3.7).

Figure 3.8 compares TraPPE's predictions for the enthalpy of vaporisation and self-

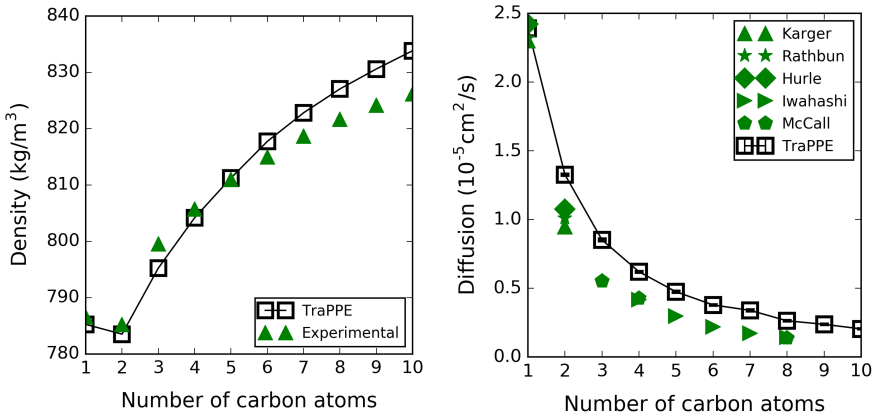

Figure 3.7: Density (left) and diffusion constant (right) of primary alcohols at 298.15 K and 1 bar obtained with the TraPPE model, represented with open black symbols. The green symbols are experimental values. Densities were taken from reference [188] and diffusion constants from references [189–193].

solvation free energy of alcohols, with and without polarisation corrections. First of all, the value for pentanol reported in the Minnesota Solvation Database [168] is most likely in error, and thus, it was excluded from the analysis. This value deviates significantly from the trend of the remaining alcohols and disagrees with the value reported by Katritzky [169] and with estimates based on the vapour pressure. As expected, the original TraPPE model performs very well for both properties since it was parameterised against vapour-liquid coexistence curves of pure alcohols. However, a small improvement is observed once the corrections from Table 3.4 are applied. Table 3.8 gives a quantitative comparison of each approach's root mean square deviation (RMSD) which was calculated using the mean of the experimental values when more than one source was available. The RMSD of the enthalpy of vaporisation decreases by 0.24 kJ/mol (from 3.20 kJ/mol without corrections to 2.96 kJ/mol with corrections), and the RMSD of the self-solvation free energy is reduced by 0.15 kJ/mol (from 1.74 kJ/mol without corrections to 1.59 kJ/mol with corrections). Furthermore, from this table we can see that applying the Berendsen correction worsens agreement with experimental data, particularly for the smaller alcohols, (RMSDs increase by 0.88 kJ/mol and 0.79 kJ/mol for the enthalpy of vaporisation and self-solvation free energy, respectively) confirming the conclusions of Leontyev and Stuchebrukhov [92].

The original TraPPE model significantly underpredicts all the alcohol molecules' experi-

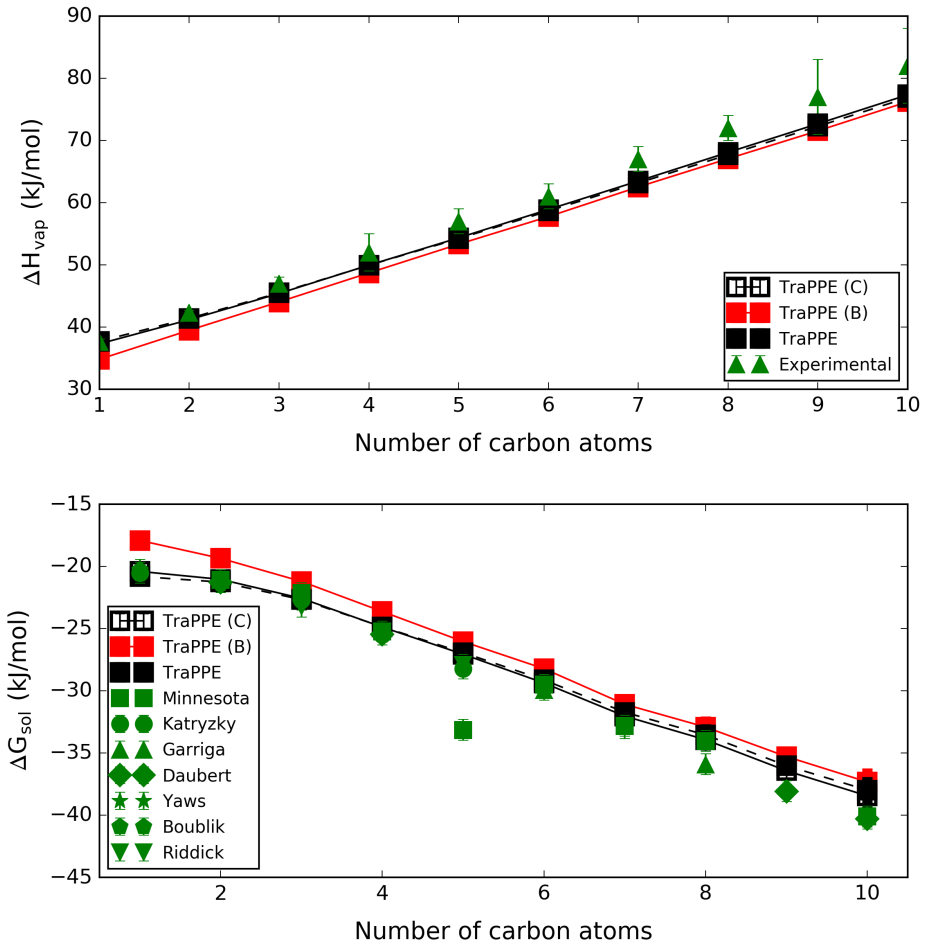

Figure 3.8: Enthalpy of vaporisation (top) and free energy of self-solvation (bottom) of primary alcohols at 298.15 K and 1 bar obtained with the TraPPE model: without any polarisation corrections ("TraPPE"; full black symbols); with the corrections proposed here ("TraPPE(C)"; open black symbols); and with the Berendsen corrections ("TraPPE(B)"; red symbols). The green symbols are experimental values, taken from NIST [194] for the enthalpy of vaporisation and from references [168, 169, 195–199] for the self-solvation free energy. The estimated average uncertainty for the experimental free energy was taken as 0.84 kJ/mol [168].

mental dielectric constant (RMSD = 5.24), as shown in Figure 3.9. However, agreement with experiment is highly improved across the entire range once simulated values are corrected to account for the difference between the liquid dipole and the model's dipole (RMSD = 2.06). This result reinforces the conclusions of a study performed by Jorge and Lue [87], where the corrections were seen to practically eliminate systematic deviations between simulated and experimental dielectric constants over a wide range of compounds and molecular models.

As explained in Chapter 1, most fixed-charge force fields struggle to predict solvation

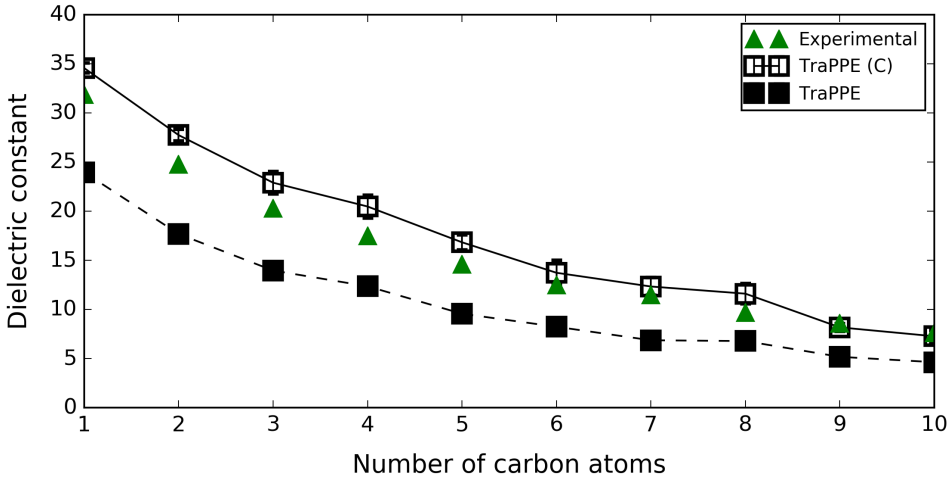

Figure 3.9: Dielectric constant of primary alcohols at 298.15 K and 1 bar obtained using the TraPPE model with and without polarisation corrections (open black squares and full black squares, respectively). Experimental values (green triangles) were taken from the supporting information of reference [87].

free energies of polar molecules in non-polar systems. Therefore, the solvation free energy in hexadecane of primary alcohols was simulated using TraPPE to assess its performance. The results are plotted in Figure 3.10, and it is clear that the original TraPPE systematically underestimates the solvation free energy (i.e. predicts less favourable solvation than observed experimentally) with an RMSD of 4.29 kJ/mol. However, once polarisation corrections (Table 3.4) are applied to take into account the solvent's dielectric constant, TraPPE can accurately predict ethanol to butanol's solvation free energies in hexadecane. These corrections also improve predicted solvation free energies of longer alcohols in hexadecane, however, deviation between experimental and simulated values increases with the alkyl chain (the solvation free energy of decanol in hexadecane is underpredicted by 4.26 kJ/mol). These results are not unexpected since TraPPE systematically underpredicts the magnitude of the solvation free energy of linear alkanes in hexadecane [1]. Furthermore, the *post-facto* polarisation corrections used in this work were compared to those obtained using the Berendsen approach, and it was noticed that Berendsen corrections shifted the free energies to even less negative values than those obtained with the original TraPPE, worsening the model's performance.

The TraPPE force field performs very well for properties with significant *post-facto*

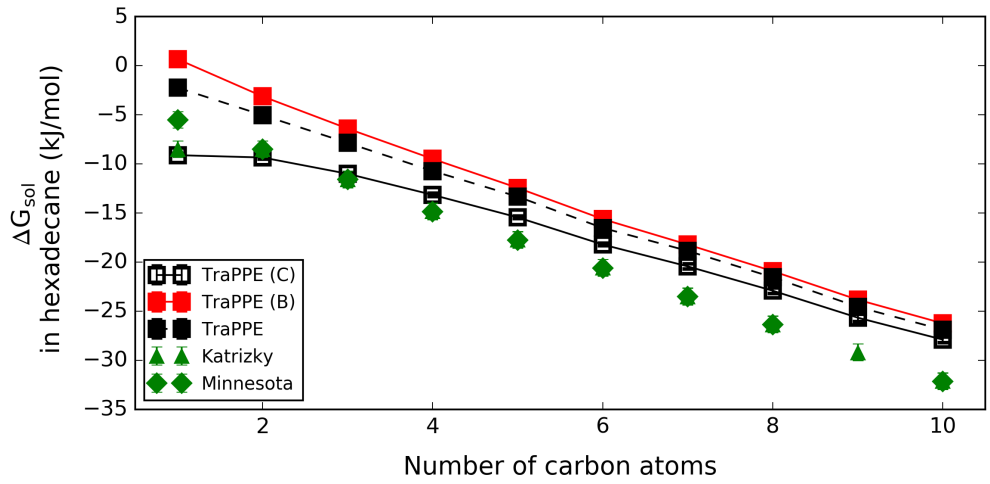

Figure 3.10: Free energy of solvation of primary alcohols in hexadecane at 298.15 K and 1 bar obtained with the TraPPE model: without any polarisation corrections ("TraPPE"; full black symbols and dashed line); with the corrections proposed here ("TraPPE(C)"; open black symbols); and with the Berendsen corrections ("TraPPE(B)"; red symbols). The green symbols are experimental values [168, 169].

Table 3.8: RMSD of TraPPE, TraPPE with Berendsen corrections (TraPPE (B)) and TraPPE with polarisation corrections (TraPPE (C)) for the dielectric constant, enthalpy of vaporisation and solvation free energies of primary alcohols at 298.15 K and 1 bar.

	TraPPE	**TraPPE (B)**	**TraPPE (C)**
ΔH_{vap} (kJ/mol)	3.20	4.08	2.96
Dielectric constant	5.24	N/A	2.06
ΔG_{sol} (kJ/mol)	1.74	2.53	1.59
ΔG_{sol} in hexadecane (kJ/mol)	4.29	5.45	2.83

polarisation corrections once these are added (i.e. dielectric constant and free energy of solvation in hexadecane), even though polarisation corrections were not considered during the parameterisation of the model. These corrections work well for this model because the net correction for pure alcohols is close to zero for the properties used in its parameterisation. Therefore, the original TraPPE parameters are, perhaps fortuitously, already close to providing optimal performance. In this chapter, a new polarisation-consistent model that includes polarisation corrections from its inception is developed.

3.3.2 Modified TraPPE

The set of alkane parameters proposed by Jorge [1] were initially combined with the parameters for the hydroxyl group taken from TraPPE-UA to decide whether these parameters needed to be modified or not. This force field was called "modified TraPPE" and it highly underpredicts small alcohols' densities as shown in Figure 3.11. However, the predicted densities of octanol to decanol are more accurate than those from TraPPE-UA. This result can be explained by the fact that the original TraPPE parameters were obtained from calculations of the vapour-liquid equilibria (including saturated liquid densities) of methanol and ethanol [69]. In contrast, Jorge's parameters were designed to fit the density, enthalpy of vaporisation and free energy of solvation of alkanes [1]. As the number of carbon atoms in an alcohol increases, non-polar interactions become more important, and the alcohol's behaviour becomes more similar to that of the alkane with the same amount of carbon atoms. Consequently, a new set of LJ parameters for the oxygen and new partial charges for the hydroxyl group and the α-carbon had to be found.

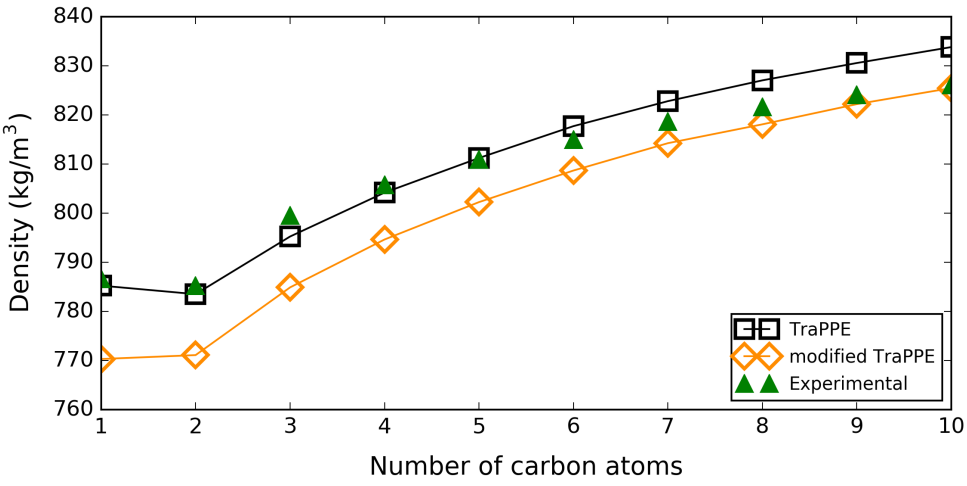

Figure 3.11: Densities of primary alcohols at 298.15 K and 1 bar obtained using the modified TraPPE model (orange diamonds) and the original TraPPE force field [69] (black squares). The green symbols are experimental values and were obtained from reference [188].

3.3.3 Model 1

Model 1 performs better than TraPPE at predicting the density, enthalpy of vaporisation and diffusion constant of alcohols with more than two carbon atoms, while still

giving good predictions for methanol and ethanol (Figures 3.12 to 3.15). Furthermore, simulations at different temperatures were run for methanol, ethanol, butanol, hexanol, and decanol to validate this model over a wider temperature range (from 283 K to 333 K). These results are plotted in Figure 3.13, where experimental values were obtained from references [200–204].

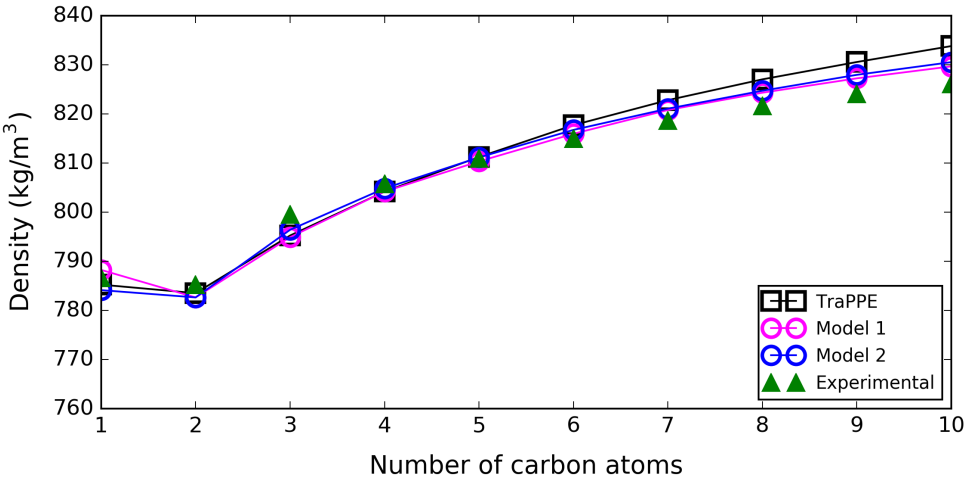

Figure 3.12: Densities of primary alcohols at 298.15 K and 1 bar obtained using Model 1 (magenta circles), Model 2 (blue circles) and the original TraPPE force field [69] (black squares). The green symbols are experimental values obtained from reference [188].

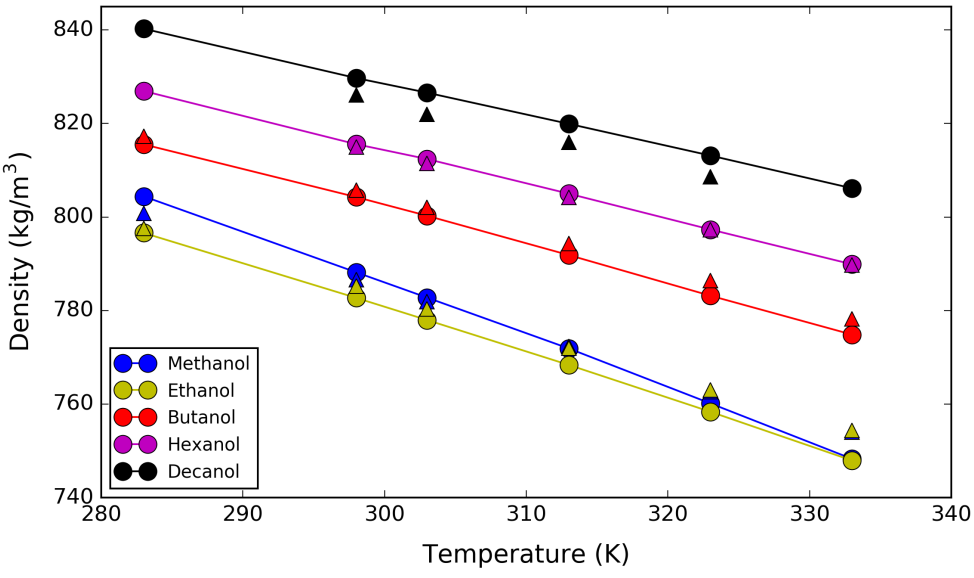

Figure 3.13: Simulated density of methanol, ethanol, butanol, hexanol, and decanol at different temperatures obtained using Model 1 versus their experimental values. Simulated values are represented with circles and experimental values with triangles. Experimental values were obtained from references [200–204].

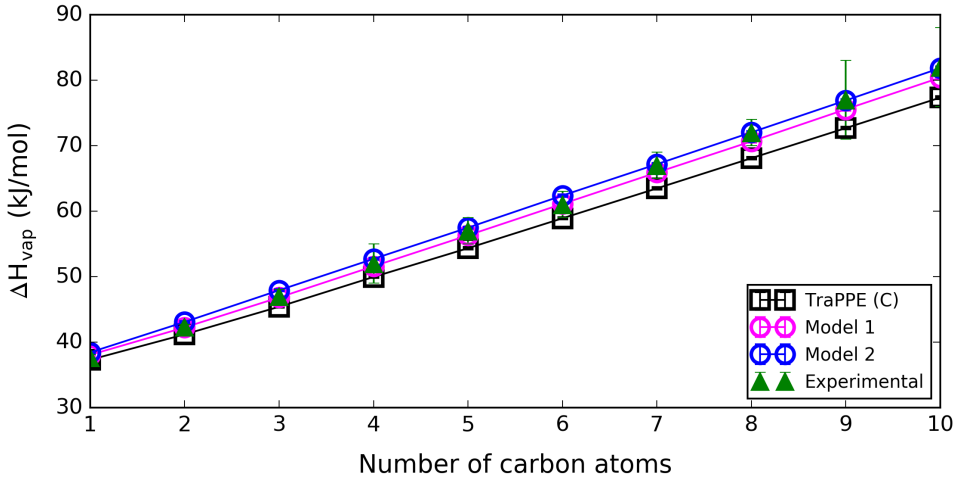

Figure 3.14: Enthalpies of vaporisation of primary alcohols at 298.15 K and 1 bar obtained using Model 1 (magenta circles), Model 2 (blue circles) and the original TraPPE force field [69] (black squares). The green symbols are experimental values taken from the NIST website [194].

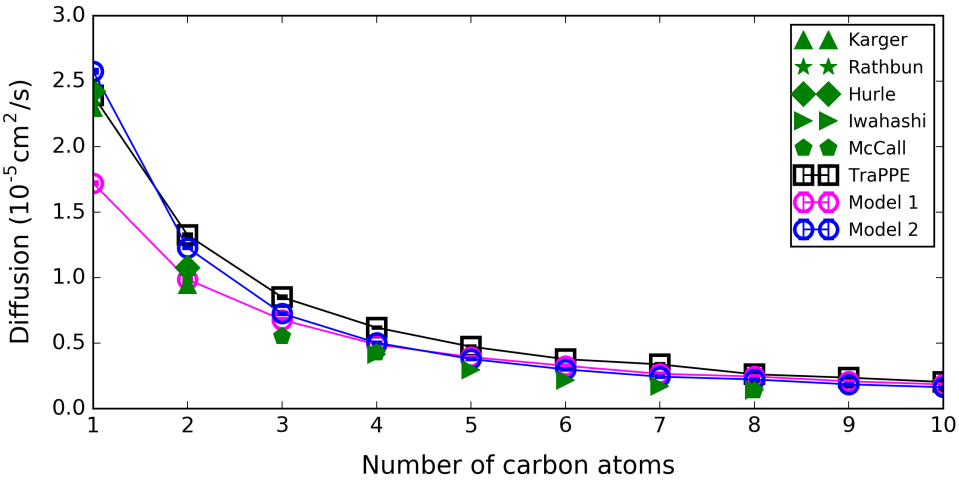

Figure 3.15: Diffusion constants of primary alcohols at 298.15 K and 1 bar obtained using Model 1 (magenta circles), Model 2 (blue circles) and the original TraPPE force field [69] (black squares). The green symbols are experimental values taken from references [189–193].

Additionally, the free energy of self-solvation of each alcohol was calculated and compared to the values obtained using TraPPE-UA (Figure 3.16). Both force fields perform very well at predicting the free energy of self-solvation of small alcohols. However, Model 1 performs better than TraPPE-UA as the alkyl chain increases because TraPPE-UA was parameterised to match small alcohols' properties while the new parameters were fitted taking into consideration larger molecules. Based on the RMSD, Model 1 has an overall better performance with an RMSD of 1.357 kJ/mol compared to 1.586

kJ/mol for TraPPE. Unfortunately, Model 1 systematically underpredicts the static dielectric constant for alcohols larger than butanol (Figure 3.17), and thus, different approaches were tried.

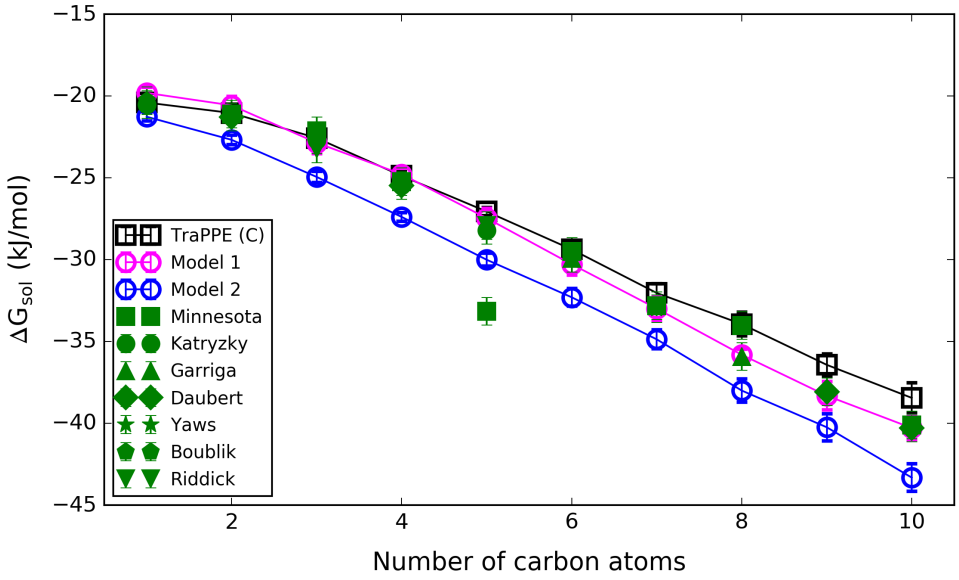

Figure 3.16: Free energy of self-solvation of primary alcohols at 298.15 K and 1 bar as a function of the number of carbons obtained using Model 1 (magenta circles), Model 2 (blue circles) and the TraPPE force field (black squares). Experimental values for the free energy were extracted from the Minnesota Solvation Database [168], the Katritzky database [169] and from references [195–199]. The estimated average uncertainty for the values extracted from the Minnesota Database is approximately 0.84 kJ/mol [168], the same uncertainty was used for the other experimental values.

3.3.4 Model 2

Model 2 performs very well at predicting the density of linear primary alcohols (Figure 3.12) with an RMSD of 2.752 kg/m^3 which is 1.481 kg/m^3 smaller than TraPPE's RMSD. Additionally, Model 2 performs significantly better than Model 1 at predicting the diffusion constant of methanol with an error of 9.62 % compared to an error of -26.72 % for Model 1 (Figure 3.15). Concerning the enthalpy of vaporisation, Model 2 has an RMSD of 0.697 kJ/mol, which is 2.26 kJ/mol smaller than the value for TraPPE, however, it slightly overpredicts (by less than 0.83 kJ/mol) the enthalpy of vaporisation of small alcohols. As the alkyl chain increases, Model 2's performance improves and thus, pentanol to decanol's simulated values are in almost perfect agreement with experimental data (Figure 3.14). Furthermore, this model was tested over a wider tem-

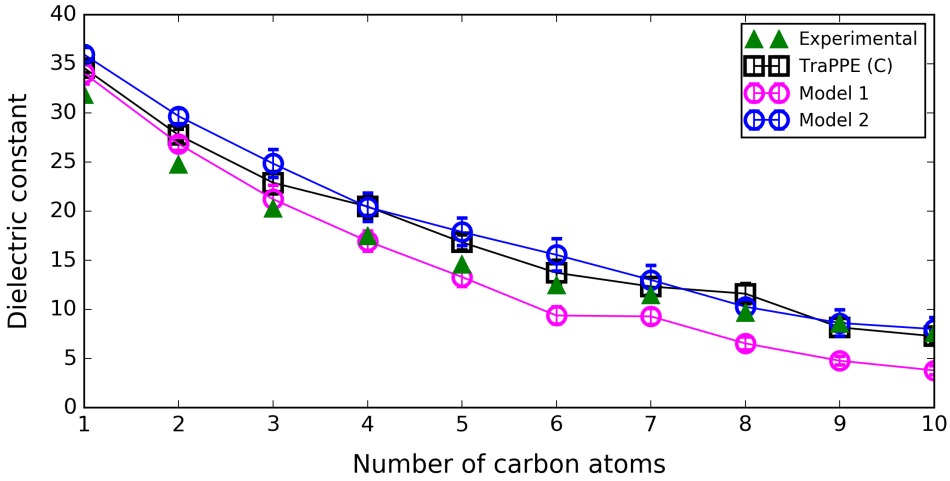

Figure 3.17: Dielectric constant at 298.15 K and 1 bar as a function of the number of carbons, obtained using Model 1 (magenta circles), Model 2 (blue circles) and the TraPPE force field (black squares). Experimental values (green triangles) were taken from the supporting information of reference [87].

perature range (from 283 K to 333 K) and simulated densities of methanol, ethanol, butanol, hexanol, and decanol at different temperature were in good agreement with experimental data as can be seen in Figure 3.18.

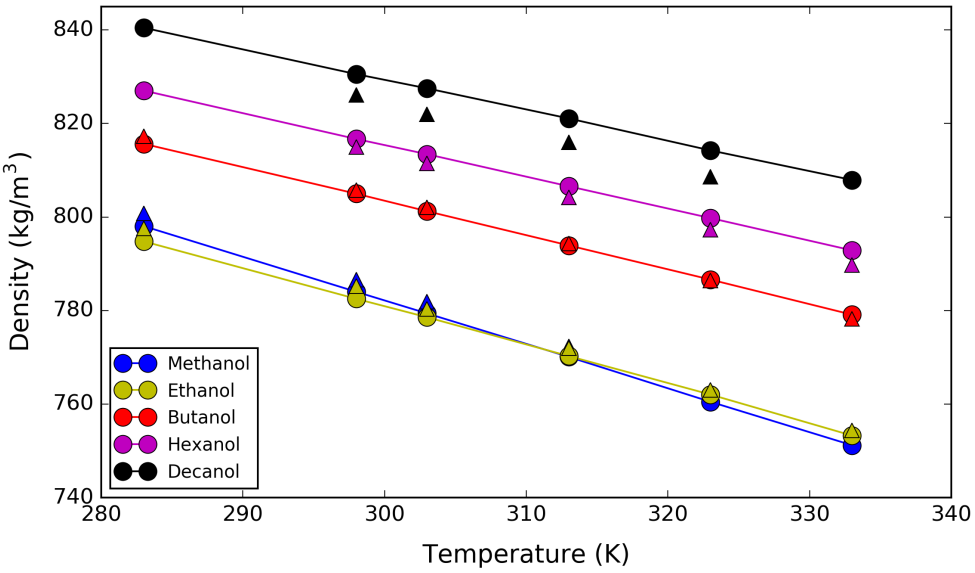

Figure 3.18: Simulated density of methanol, ethanol, butanol, hexanol, and decanol at different temperatures obtained using Model 2 versus their experimental values. Triangles were used for the experimental values, while circles were chosen for the simulated values. Experimental values were obtained from references [200–204].

As shown in Figure 3.17, Model 2 overpredicts the dielectric constant of methanol to

heptanol but can accurately predict this property from octanol onwards. This model performs overall worse than Model 1 with an RMSD of 3.034 compared to 2.562 for Model 1. However, it is important to point out that Model 1 underpredicts the dielectric constant of butanol onwards and agreement with experimental data systematically deteriorates as the alkyl chain increases (it underpredicts decanol's dielectric constant by 3.8), meanwhile the opposite effect is observed with Model 2 where performance improves for larger alcohols. Concerning the self-solvation free energy, Model 2 significantly overpredicts its magnitude with an RMSD of 2.305 kJ/mol (Figure 3.16), and thus, this model is not suitable for our purposes. Consequently, a third approach was tried, as described in section 3.2.2.

3.3.5 New PolCA model (Model 3)

The PolCA model performs better than TraPPE at predicting the density of alcohols from hexanol to decanol while still giving good predictions for smaller alcohols, and it has a similar performance to the first model. The root mean square deviation of the new model to experimental data is 2.79 kg/m^3, while for TraPPE it is 4.23 kg/m^3.

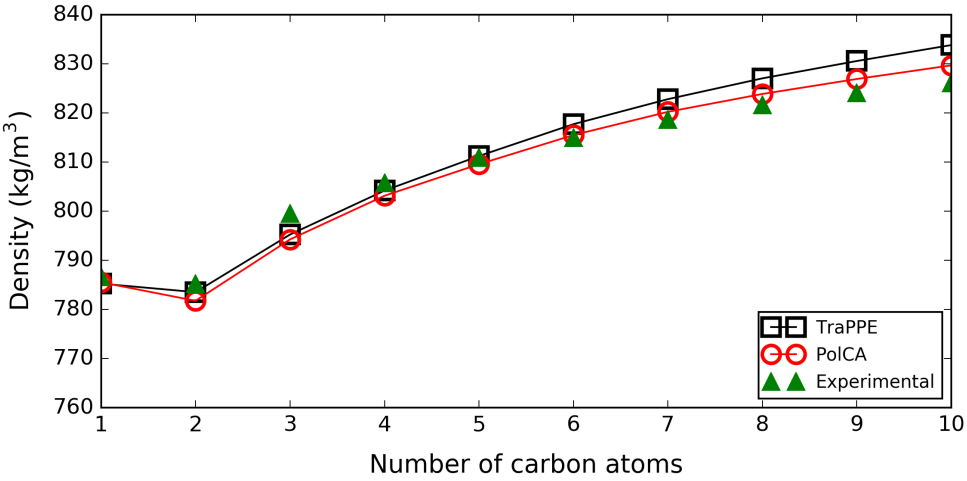

Figure 3.19: Densities of primary alcohols at 298.15 K and 1 bar obtained using the PolCA force field (red circles) and the original TraPPE force field [69] (black squares). The green symbols are experimental values [188].

Additionally, simulations were run at five other temperatures (283, 303, 313, 323 and 333 K) for methanol, ethanol, butanol, hexanol and decanol to validate the new PolCA model over this range of temperatures. Figure 3.20 shows that PolCA accurately cap-

tures the density's temperature dependence for all the alcohols.

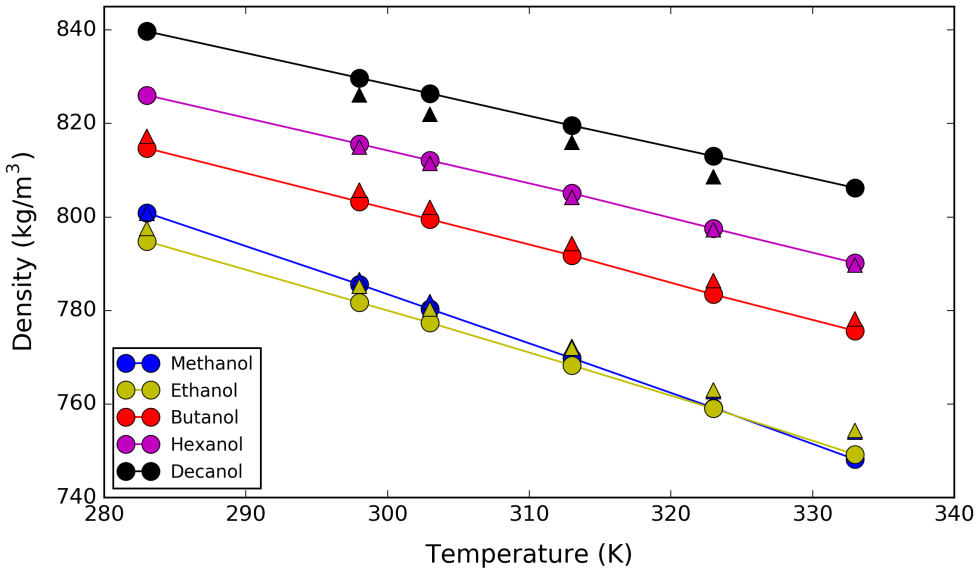

Figure 3.20: Simulated density of methanol, ethanol, butanol, hexanol, and decanol at different temperatures obtained using the PolCA force field versus their experimental values. Triangles were used for the experimental values, while circles were chosen for the simulated values. Experimental values were obtained from references [200–204].

As discussed previously, TraPPE overestimates the self-diffusion constant of primary alcohols, except for methanol's diffusion, which is very close to the experimental value. On the other hand, PolCA underestimates methanol's self-diffusion constant but accurately predicts this property for the other primary alcohols (see Figure 3.21). The root mean square deviation (RMSD) of the new PolCA model with respect to experimental data is 0.201 10^{-5} cm^2/s, while for TraPPE it is 0.181 10^{-5} cm^2/s. Concerning the diffusion, PolCA is similar to Model 1 and worse than Model 2. However, PolCA still gives a good description of the diffusion constant and is overall better than Model 2. The diffusion constant of methanol could be improved using different parameters for methanol but at the expense of a less general model.

Both TraPPE and PolCA do a great job predicting the enthalpy of vaporisation. PolCA slightly overpredicts this property for methanol, but it matches the experimental data almost exactly from ethanol onwards (Figure 3.22). It is important to mention that experimental values were taken from the NIST website [194] and the error bars for nonanol and decanol are quite large (6 kJ/mol). Therefore, the values obtained using

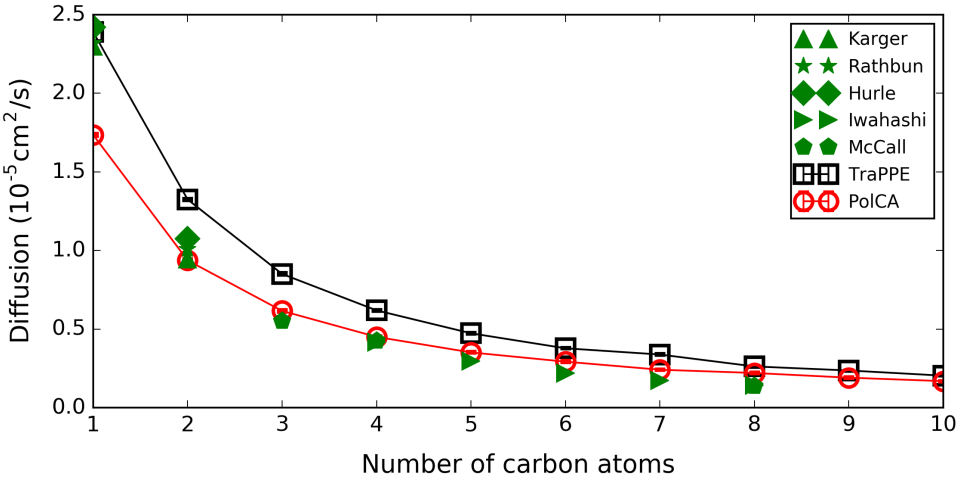

Figure 3.21: Self-diffusion constant of primary alcohols at 298.15 K and 1 bar obtained using the PolCA model (red circles) and the original TraPPE force field [69] (black squares). The green symbols are experimental values obtained from references [189–193].

TraPPE fall within the error bars but are a lot lower than the average. The RMSD of PolCA is 0.71 kJ/mol, and the RMSD of TraPPE (C) is 2.96 kJ/mol.

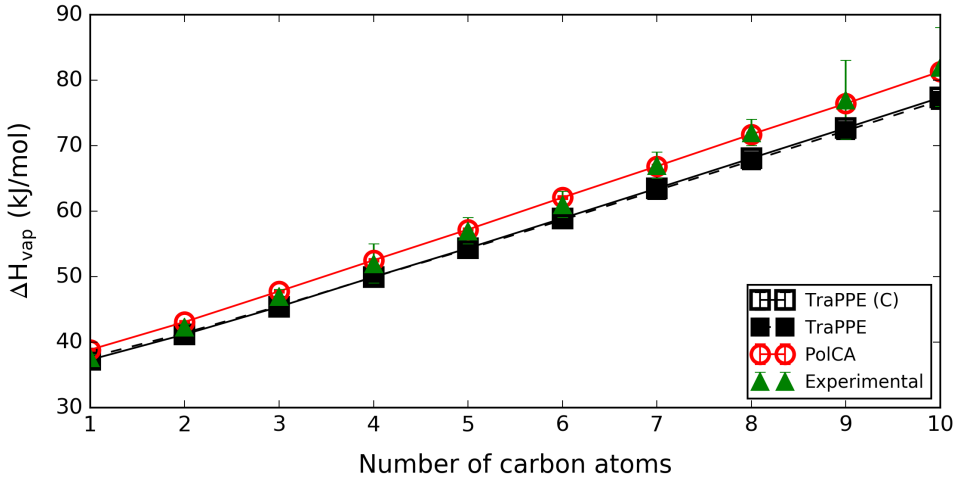

Figure 3.22: Enthalpy of vaporisation of primary alcohols at 298.15 K and 1 bar obtained using the new PolCA model (red circles) and the TraPPE force field with and without polarisation corrections (open black squares and filled black symbols, respectively). The green symbols are experimental values obtained from NIST website [194].

TraPPE and PolCA are both able to accurately predict the dielectric constant once polarisation corrections are applied (see Figure 3.23). The root mean square deviations for PolCA and TraPPE (C) are 1.84 and 2.06, respectively. This outcome is remarkable since neither TraPPE nor PolCA used this property in its parameterisa-

tion, and it emphasises the need to consistently account for polarisation effects when comparing predictions of non-polarisable force fields against experimental dielectric constants.

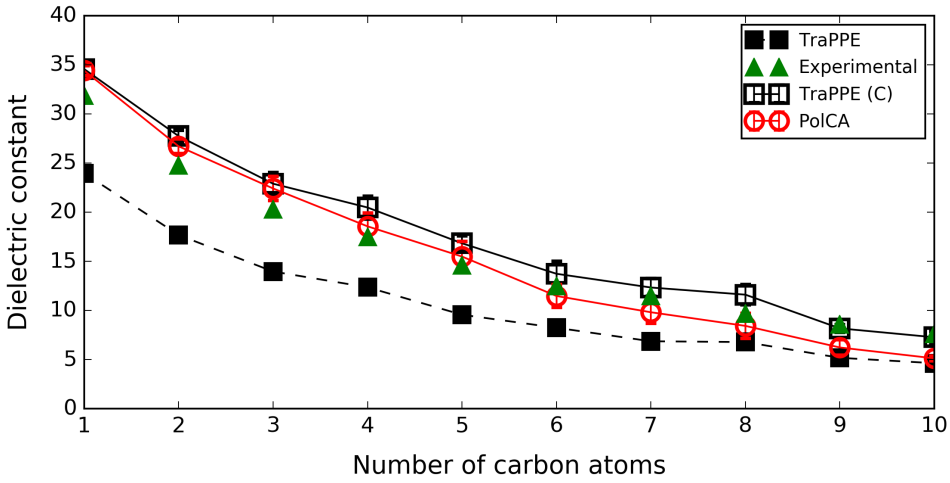

Figure 3.23: Dielectric constant of primary alcohols at 298.15 K and 1 bar as a function of the number of carbons, obtained using the PolCA model (red circles) and the TraPPE force field with corrections (open squares) and without corrections (dashed lines). Experimental values (green triangles) were taken from the supporting information of reference [87].

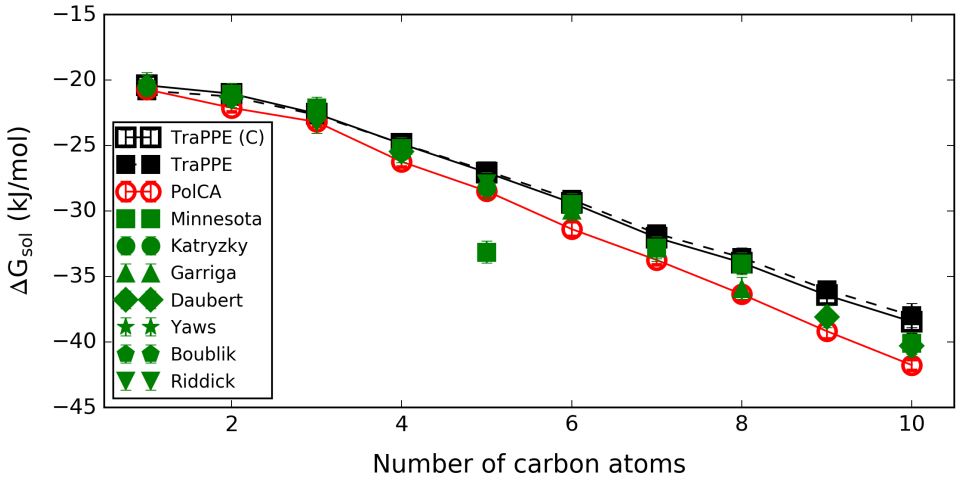

Figure 3.24: Free energy of self-solvation of primary alcohols at 298.15 K and 1 bar obtained using the new PolCA model (red circles), TraPPE and TraPPE with corrections (filled squares and open black squares, respectively), as a function of the number of carbons. Experimental values for the free energy were extracted from the Minnesota Solvation Database [168], the Katritzky database [169] and from references [195–199]. The estimated average uncertainty for the values extracted from the Minnesota Database is approximately 0.84 kJ/mol [168]; the same uncertainty was used for the other experimental values.

Figure 3.24 compares experimental data for the free energy of self-solvation of primary alcohols against predictions from TraPPE and PolCA. The RMSD for TraPPE (C) is

1.59 kJ/mol, while it is 1.45 kJ/mol for PolCA. This shows that both models do a good job at predicting this property, although the TraPPE (C) values are on the upper (i.e. less favourable) end of the range of experimental data, while those of our new model are on the lower end (i.e. more favourable). The polarisation corrections for the self-solvation free energy are the same as for the enthalpy of vaporisation, but with opposite sign; as such, their overall magnitude is also quite small.

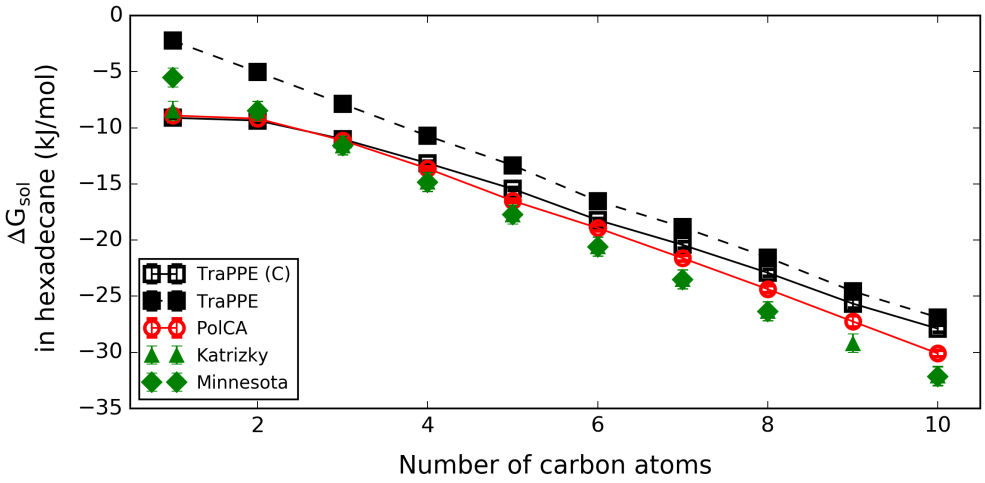

Figure 3.25: Free energy of solvation of primary alcohols in hexadecane at 298.15 K and 1 bar as a function of the number of carbons in the solute obtained using the new PolCA model (red circles), TraPPE and TraPPE with corrections (filled squares and open black squares, respectively). Experimental values were taken from references [168] and [169]. The estimated average uncertainty is approximately 0.84 kJ/mol [168].

Lastly, Figure 3.25 shows the free energy of solvation of primary linear alcohols in hexadecane, obtained using TraPPE and PolCA. Both models' performance is identical for small alcohols, but the PolCA model yields improved predictions for larger alcohols relative to TraPPE due to the reparameterised alkane force field parameters. In both cases, the simulation predictions for methanol are in agreement with the experimental data of Katritzky [169] but deviate from that of the Minnesota Solvation Database [168]. Overall, Figure 3.25 shows that PolCA leads to consistent predictions of solvation of alcohols in alkanes and that polarisation corrections are the key to achieve this good performance. The RMSD for PolCA is 1.833 kJ/mol, while for TraPPE (C) the RMSD is 2.828 kJ/mol.

3.3.6 Secondary and tertiary alcohols

Furthermore, the model was tested for secondary and tertiary alcohols. As a first attempt, the LJ parameters of the alkane CH and C pseudo-atoms were used for the α-carbons without modification. However, doing this resulted in a model that significantly underpredicted the density. This result could be explained by the fact that a C-O bond is shorter than a C-C bond and that the hydroxyl oxygen has an electron-withdrawing effect over the carbon [69]. The same effect was noticed by the TraPPE-UA force field developers, who found they had to reduce σ of the α-CH and α-C by 7.48 % and 9.38%, respectively, to reach satisfactory agreement with experiment [69]. Consequently, our second attempt was based on sigma values for α-CH and α-C pseudo-atoms that were 7.48 % and 9.38% smaller than their respective alkane pseudo-atoms. The PolCA model's performance for a few non-primary alcohols is presented in Table 3.9, and it can be seen that this model also performs very well for secondary alcohols. For tertiary alcohols, while the performance is acceptable for the density, enthalpy of vaporisation and self-solvation free energy, it fails to predict the dielectric constant of those molecules. D' Aprano et al. studied the dielectric constant of several isomeric pentanols and noticed that this property decreased with chain branching and the proximity of the hydroxyl group to the branch due to steric effects that reduce the association of the molecules through hydrogen bonding (co-association causes dipoles to reinforce each other) [205]. Therefore, the values predicted using PolCA suggest that our force field could be overestimating the steric effects possibly due to an overestimation of σ for the alkyl carbons attached to the α-carbon. Another explanation for the discrepancy between our values and the experimental dielectric constants could be an underestimation of the tertiary alcohols' hydrogen bond strength due to the use of the same partial charges as those parameterised for primary amines. Further work is needed to establish the exact cause.

3.3.7 Comparison between all models

The root mean square deviation (RMSD) for each model based on the different properties can be found in Table 3.10. These values were calculated based on the experimental

Table 3.9: Simulated properties for non-primary alcohols using the PolCA model. Experimental values are in bold. The values for ΔH were taken from the NIST website [194], and in the case of 2-butanol, the average was taken over 8 of the 9 data points because one was deemed to be an outlier.

	2-propanol	**2-butanol**	**2-pentanol**	**Tert-butanol**	**Tert-amyl**
Density (kg/m^3)	779.0 +/- 0.1 **781.31** [206]	798.8 +/- 0.2 **802.43** [206]	804.6 +/- 0.3 **805.9** [205]	775.4 +/- 0.1 **780.79** [207]	808.3 +/- 0.1 **806.9** [205]
ΔH_{vap} (kJ/mol)	45.68 +/- 0.03 **45 +/- 3**	50.88 +/- 0.04 **49 +/- 1**	55.5 +/- 0.1 **53 +/- 1**	44.00 +/- 0.03 **46 +/- 1**	49.6 +/- 0.2 **50 +/- 1**
Dielectric constant	19 +/- 1 **19.43** [208]	15 +/- 2 **16.60** [208]	12 +/- 2 **13.71** [205]	3.08 +/- 0.01 **12.49** [208]	2.80 +/- 0.01 **5.78** [205]
$\Delta G_{self\text{-}sol}$ (kJ/mol)		-24.2 +/- 0.4 **-23.1 +/- 0.8** [196]		-19.6 +/- 0.3 **-21.0 +/- 0.8** [196]	

and simulated values of the primary alcohols studied here. Also, this table contains the optimised parameter set for each approach.

Table 3.10: Parameters set and RMSD of the different models obtained in this chapter and TraPPE-UA. The RMSD was obtained considering only primary alcohols, and all properties were simulated at 298.5 K and 1 bar.

Force Field	**σ** [nm]	**ε** [kJ/mol]	**q_O** [C]	**q_H** [C]	**Density** [kg/m^3]	**Diffusion** [10^{-5} cm^2/s]	**ΔH_{vap}** [kJ/mol]	**Dielectric constant**	**ΔG** [kJ/mol]
Model 1	0.278	0.734	-0.618	0.384	**2.615**	0.213	0.926	2.562	**1.357**
Model 2	0.2938	1.155	-0.633	0.449	2.752	**0.127**	**0.697**	3.034	2.305
PolCA	0.2853	0.773	-0.646	0.406	2.79	0.201	0.710	**1.839**	1.454
TraPPE (C)	0.302	0.773	-0.7	0.435	4.233	0.181	2.958	2.062	1.586

All the models developed in this chapter have a similar performance at predicting the density of linear primary alcohols (RMSDs range from 2.62 to 2.79 kg/m^3), and they perform better than TraPPE by $\approx$ 1.4 kg/m^3) since pentanol and heptanol's densities were included in our parameterisation while TraPPE was fitted to only small alcohols' densities. Concerning the enthalpy of vaporisation, our models RMSDs are smaller than TraPPE (C)'s value by more than 2 kJ/mol likely due to the fact that this property was included in our optimisation routine but not in TraPPE's. From our models, Model 2 and Model 3 have almost the same performance (RMSDs = 0.7 kJ/mol), and they perform better than Model 1 by 0.2 kJ/mol. Model 1 also performs worse than the other three models at predicting the diffusion constant (RMSD = 0.21 10^{-5} cm^2/s), probably due to the restrictions imposed on the partial charges during its development. On the other hand, Model 2 is the best at predicting the diffusion constant with an RMSD of 0.13 10^{-5} cm^2/s but has the worst performance for the dielectric constant (RMSD =

3.03) and the free energy of self-solvation (RMSD = 2.31 kJ/mol). The models that can better predict the dielectric constant and the free energy of self-solvation are PolCA (RMSD = 1.84) and Model 1 (RMSD = 1.36 kJ/mol), respectively. PolCA has the second best performance for the self-solvation free energy (RMSD = 1.45 kJ/mol), and except for the diffusion constant, it performs better or almost the same as the other models for all the properties studied here. Consequently, it was selected as the best overall force field.

3.4 Conclusions

In this work, we have proposed a new non-polarisable force field for alcohols that was parameterised taking into consideration polarisation effects. The new parameters for the alcohol functional group are consistent with the UA model for aliphatic hydrocarbons developed by Jorge [1] that eliminates systematic deviations in solvation free energy predictions from existing UA force fields. When compared to the benchmark TraPPE-UA force field for alcohols, our force field performs better at predicting the density, enthalpy of vaporisation, dielectric constant, free energy of self-solvation, and free energy of solvation in hexadecane. The RMSD of the diffusion coefficient is slightly higher for PolCA than for TraPPE, but this is likely due to an inherent limitation of the UA approach since our model sacrifices methanol's diffusion to predict the diffusion of ethanol to decanol with more accuracy than TraPPE. Indeed, the improved performance of PolCA is seen most significantly for larger alcohol molecules, precisely due to the elimination of systematic errors in the description of the alkane moieties. Nevertheless, it is important to highlight that TraPPE-UA already performs quite well, especially for smaller alcohols, once *post-facto* polarisation corrections are added. Even though this model was not parameterised considering polarisation corrections, it performs well because the corrections for properties involving a change of phase in pure alcohols are quite small for alcohols with a short alkyl chain.

Chapter 4

Polarisation-Consistent United Atom Force Field for Amines

4.1 Introduction

Amines are chemical compounds that contain an sp^3 nitrogen atom bonded to other elements through single bonds, and ammonia is the simplest compound in this family. Ammonia consists of a nitrogen atom united by single bonds to three hydrogen atoms. If an alkyl group replaces one or more of these hydrogen atoms, the resulting amine is an organic compound called alkyl amine. Primary alkyl amines are obtained when only one hydrogen atom is replaced by an alkyl group, while secondary and tertiary amines are obtained when two and all three hydrogen atoms are replaced, respectively.

An important property of amines is their ability to act as Lewis bases thanks to the lone electron pair on the central nitrogen atom, and thus, they are found in several industrial applications [209]. For example, amines are used to remove acidic gases from natural gas in a process called gas sweetening. This process prevents corrosion problems on the pipelines, increases the calorific value of the final gas stream and reduces atmospheric pollution [210]. Additionally, aqueous amine solutions are commonly used in CO_2 adsorption/desorption processes designed to decrease CO_2 emissions, and thus, help mitigate climate change [211]. Amine groups are also widely prevalent in pharma-

ceutical compounds, and they play an important role in solvation and protein-binding mechanisms.

Molecular dynamic (MD) simulations can be used to predict thermodynamic and transport properties of amine-containing molecules with relevance to the above industrial applications. The accuracy of these simulations depends on the chosen force field, and thus, several force fields have been developed over the years using different approaches and target properties. Some well known amine force fields are described below.

The OPLS all-atom (AA) force field for many common organic groups was developed in 1996 by Jorgensen et al. [139]. Bond stretching and angle bending parameters were mostly taken from the AMBER-AA force field [212], while non-bonded and torsional parameters were determined by fitting to densities and enthalpies of vaporisation of 34 pure organic liquids and rotational energy profiles of over 50 organic molecules and ions, respectively. This force field only included parameters for primary amines, therefore, in 1999, Rizzo et al. [213] extended the OPLS-AA force field to ammonia and non-primary amines, and slightly adjusted the parameters for primary amines. Molecular structures, conformational energetics, hydrogen-bond strengths, densities, enthalpies of vaporisation and relative free energies of hydration of ammonia, methylamine, dimethylamine and trimethylamine were considered during the model's development. The partial charges and nitrogen's Lennard Jones (LJ) parameters were adjusted to match experimental pure liquid properties better. The LJ parameters for the nitrogen atom were the same for all amines. On the other hand, the partial charges of the nitrogen and its bonded hydrogen atoms differed for each amine type, and these charges were allowed to vary independently from each other during the optimisation (α-carbons' partial charges were obtained from the neutrality constraint). Missing bonded parameters were developed using optimised geometries and rotational energy profiles from *ab initio* calculations. Additionally, 13 pure liquid simulations of aliphatic, cyclic and aromatic amines and several relative free energy calculations of amines in water and chloroform were run, and good agreement with experimental data was obtained. Free energies of hydration were specifically important during this model's development since previous classical force fields had failed to reproduce the observed experimental trends for the

amine series, ammonia, methylamine, dimethylamine and trimethylamine [214]. Experimentally, ammonia and methylamine present an unexpected behaviour since the replacement of an amino hydrogen by a methyl group leads to a favourable (-1.09 kJ/mol) contribution to the free energy of hydration [215].

OPLS-AA is a good amine model and it captures the counterintuitive experimental trend for the free energies of hydration of ammonia, methylamine, dimethylamine and trimethylamine. However, it underpredicts methylamine's critical temperature by 10% and fails to predict this molecule's density at elevated temperatures that are far from its boiling point [216].

In 2005, the TraPPE explicit-hydrogen (EH) force field was extended to include amines [216]. This force field was parametrised using the vapour-liquid coexistence curves of methylamine, dimethylamine and trimethylamine, and it includes new specific LJ parameters and partial charges for the different types of amine nitrogens and a common partial charge for all α-carbon types. The parameters for tertiary amines were parametrised first to determine the α-carbon's partial charge, which was then transferred to primary and secondary amines. The neutrality requirement determined the partial charges of the hydrogens attached to the nitrogen atom, and consequently, only three new parameters (nitrogen's σ, ε and partial charge) were required for each amine type since all bonded parameters were taken from the OPLS-AA force field [213].

TraPPE-EH does a very good job at reproducing the liquid-vapour equilibrium of small amines. However, it treats all hydrogens explicitly, and thus, it has a higher computational cost than the united-atom (UA) approach that we have chosen to adopt here. Furthermore, this model lacks torsion parameters to extend its transferability to more complex amines. For these reasons, a new anisotropic UA force field for amines with extra torsion parameters, called AUA4 [217], was proposed by Orozco et al. in 2011. This model is an extension of the AUA4 model proposed by Ungerer [218] with three extra parameters for the nitrogen atom: σ, ε and δ (anisotropic UA displacement). These parameters were obtained by fitting to the density, enthalpy of vaporisation and vapour pressure of methylamine, ethylamine and propylamine at different temper-

atures, while the amine group's partial charges were taken from TraPPE-EH. Additionally, critical points, normal boiling points and viscosity coefficients of amines were calculated during the model's validation. In 2013, Orozco et al. studied the densities and excess volumes of amines in water finding an excellent agreement for the mixture densities and a good qualitative description of the excess volumes. The excess volumes reproduce the overall physical trends, however they are overestimated for all the studied systems (methylamine, ethylenediamine, diethylenetriamine, ethanolamine, diethanolamine, methyldiethanolamine). [219]

Lastly, GROMOS-53A6 [101] is another UA model that includes amines relevant to this work. This model was developed in 2004 to reproduce free energies of solvation in water and non-polar solvents. In 2016, Horta et al. [80] developed an improved GROMOS-compatible parameter set using the liquid density, enthalpy of vaporisation, hydration free energy and solvation free energy in cyclohexane of small organic molecules as the target properties.

None of the force fields described above include polarisation corrections, and thus, the PolCA model is extended here to include amines. The methodology specific for this chapter is presented in section 4.2. This section contains information about simulation times, polarisation corrections, force field parameters and details about the optimisation procedure. Three different methods were investigated during the optimisation routine, resulting in three potential models for amines. These models are compared in section 4.3.2 and the model with the best performance overall is compared to another united-atom force field for amines (GROMOS-2016H66) in section 4.3.3. Lastly, parameters for branched primary amines and non-primary amines are presented.

4.2 Methodology

The optimisation procedure, as well as the methods used to obtain bulk properties and free energies of solvation, are described in detail in Chapter 2. Specific details for the present chapter are described below.

Bulk properties

Table 4.1 shows the number of molecules in each simulation box, which was selected to maintain an approximately constant box size of 27 nm^3.

Table 4.1: Number of molecules in the simulation box for each single-component system.

Compound	Number of molecules
Methylamine	654
Ethylamine	447
Propylamine	327
Butylamine	250
Pentylamine	218
Hexylamine	177
Heptylamine	147
Octylamine	129
Nonylamine	113
Decylamine	103
Isopropylamine	319
Tertbutylamine	257
1,1-Dimethylpropylamine	216
2-amino-3-methylpentane	183
2-butylamine	260
Dimethylamine	506
Diethylamine	284
Trimethylamine	404
Triethylamine	201

Solvation free Energy calculations

As for the other compounds examined in this thesis, the sampling length was determined by carefully analysing the convergence of the free energy over time. Table 4.2 contains simulation lengths for each system and decylamine's convergence plots can be found in Figures A1 and A2 in Appendix A1. Longer simulation times were used to assess the performance of the Gromos 2016 force field since these simulations were run first.

4.2.1 Polarisation corrections

Polarisation corrections were used to calculate properties that involve a phase transition, as explained in Chapter 2. One of the input parameters needed for their calculation is the solute's liquid dipole moment. Unfortunately, we are not aware of any experimental or quantum mechanical estimates of the liquid dipole moment of amine molecules. Therefore, and following on from successful estimates for water [108] and

Table 4.2: Simulation time in ns for the free energy components of the systems studied in this chapter. N/A means not applicable.

Solute	**Solvent**	**LJ**	**Electrostatics**
Methylamine	Methylamine	5	5
Ethylamine	Ethylamine	5	10
Propylamine	Propylamine	5	10
Butylamine	Butylamine	5	20
Pentylamine	Pentylamine	5	20
Hexylamine	Hexylamine	7	20
Heptylamine	Heptylamine	7	20
Octylamine	Octylamine	10	30
Nonylamine	Nonylamine	10	30
Decylamine	Decylamine	10	30
Isopropylamine	Isopropylamine	5	10
Tertbutylamine	Tertbutylamine	5	20
1,1 Dimethylpropylamine	1,1 Dimethylpropylamine	5	20
2-amino-3-methylpentane	2-amino-3-methylpentane	7	20
2-butylamine	2-butylamine	5	20
Dimethylamine	Dimethylamine	5	10
Trimethylamine	Trimethylamine	5	10
All amines	Hexadecane	10	N/A

alcohols (see Chapter 3), the liquid amine dipole moments were calculated from Equation 2.49, proposed by Leontyev and Stuchebrukhov [92]. Furthermore, we were only able to find experimental gas-phase dipole moments for the smallest amine molecules (see below for details). Therefore, unless stated otherwise, the gas-phase dipole for each molecule was estimated from a quantum mechanical calculation in vacuum using Gaussian 09W [220]. A geometry optimisation was performed first until a stationary point on the potential energy surface was found. After this, the electron density was calculated using the B3LYP functional and the basis set cc-pvtz with the AUG- prefix to add diffuse functions. Table 4.3 compares experimental and estimated values for the gas-phase dipole of several molecules, and there is a good agreement overall.

The gas phase dipoles for all the molecules used here are presented in Table 4.4. Additionally, this table contains all *post-facto* polarisation corrections used in this chapter. These corrections are divided into two contributions: the negative distortion term (C_{dist}) and the positive electronic polarisation term (C_{elec}), see section 2.7.2 for de-

Table 4.3: Experimental gas-phase dipoles versus QM estimates obtained using Gaussian 09W for several amines.

Molecule	**Experimental dipole (D)**	**QM dipole (D)**
Methylamine	1.31 [221]	1.29
Ethylamine	1.22 [221]	1.32
Propylamine	1.17 [221]	1.20
Butylamine	1.33 [222]	1.28
Dimethylamine	1.02 [223]	0.96
Trimethylamine	0.64 [223]	0.51

tails. The values presented here correspond to a liquid-gas transition to be consistent with the definition of C_{pol} in section 2.7.2 , and the corrections for the solvation free energy have the same magnitude but opposite sign. The polarisation corrections for amine self-solvation, like those for alcohols, are relatively small in magnitude (below 1 kJ/mol), even though the individual distortion and electronic contributions are somewhat larger. However, as observed for alcohols, the corrections are much larger in magnitude for solvation in hexadecane, since the electronic component dominates over the distortion for solvation in a non-polar solvent. Interestingly, the amine corrections take on a positive value in all systems, whereas for alcohols there were both positive and negative corrections (cf. Table 3.4 in Chapter 3).

4.2.2 PolCA's parameters

In Chapter 3, a new model for alcohols that includes *post-facto* polarisation corrections was developed using TraPPE as a starting point. As explained previously, TraPPE was chosen for this task because it performed better than OPLS-UA and GROMOS at predicting solvation free energies of 52 solute-solvent pairs of linear, branched and cyclic alkanes [81]. Unfortunately, TraPPE uses an explicit-hydrogen model, called TraPPE-EH, to describe amines. TraPPE-EH treats all hydrogens explicitly and has extra interaction sites located at the centres of carbon-hydrogen bonds to represent the molecular shape of alkanes better [216]. This model was created to improve the performance of TraPPE-UA when predicting liquid and vapour phases of alkanes, however, this improvement comes at the expense of a significantly higher computational

*Table 4.4: Polarisation corrections for a liquid-gas phase transition (C_{pol}), expressed in kJ/mol. C_{dist} and C_{elec} are the distortion and electronic polarisation terms, respectively, α [224] is the polarizability of the solute in the gas phase and ϵ_{el} [224] and ϵ_{sol} are the high-frequency dielectric permittivity and the static dielectric constant, respectively, of the medium. *Due to lack of space, the acronyms "1-1-DMPA" and "2A3MP" were used to represent 1-1 dimethylpropylamine and 2-amino-3-methyl-pentane, respectively. No experimental data was found for their dielectric constants, and thus, they were estimated using PolCA.*

Solute	**Solvent**	μ_g **(D)**	α ($\AA^3$)	ϵ_{sol}	ϵ_{el}	μ_l **(D)**	C_{dist}	C_{elec}	C_{pol}
methylamine	methylamine	1.31 [221]	4.03	9.40	1.82	2.01	-3.65	4.37	0.72
ethylamine	ethylamine	1.22 [221]	5.86	6.61	1.86	1.83	-1.94	2.67	0.73
propylamine	propylamine	1.17 [221]	7.69	4.99	1.92	1.73	-1.25	2.00	0.76
butylamine	butylamine	1.33 [222]	9.36	4.62	1.94	1.97	-1.30	2.20	0.90
pentylamine	pentylamine	1.19	11.38	4.20	1.99	1.75	-0.84	1.53	0.68
hexylamine	hexylamine	1.27	13.25	4.03	2.01	1.87	-0.82	1.53	0.71
heptylamine	heptylamine	1.18	15.06	3.77	2.02	1.72	-0.59	1.17	0.58
octylamine	octylamine	1.27	16.91	3.55	2.04	1.84	-0.58	1.21	0.63
nonylamine	nonylamine	1.18	18.76	3.38	2.05	1.69	-0.43	0.94	0.51
decylamine	decylamine	1.27	20.58	3.28	2.06	1.82	-0.44	1.00	0.56
isopropylamine	isopropylamine	1.26	7.77	4.85	1.88	1.83	-1.26	2.08	0.82
tertbutylamine	tertbutylamine	1.18	9.67	4.01	1.89	1.67	-0.74	1.42	0.68
*1-1-DMPA	1-1-DMPA	1.13	11.43	4.75	1.95	1.68	-0.80	1.33	0.53
*2A3MP	2A3MP	1.20	13.25	4.16	1.99	1.77	-0.73	1.33	0.60
2 butylamine	2 butylamine	1.20	9.67	4.51	1.94	1.76	-0.99	1.71	0.72
dimethylamine	dimethylamine	1.02 [223]	6.02	5.26	1.84	1.48	-1.04	1.64	0.59
diethylamine	diethylamine	0.90	9.62	3.90	1.91	1.28	-0.44	0.86	0.42
trimethylamine	trimethylamine	0.64 [223]	7.90	2.44	1.81	0.80	-0.10	0.34	0.25
triethylamine	triethylamine	0.56	13.48	2.40	1.96	0.72	-0.06	0.21	0.15
methylamine	hexadecane	1.31 [221]	4.03	2.09	2.05	1.66	-0.89	4.19	3.30
ethylamine	hexadecane	1.22 [221]	5.86	2.09	2.05	1.54	-0.53	2.50	1.97
propylamine	hexadecane	1.17 [221]	7.69	2.09	2.05	1.48	-0.37	1.75	1.38
butylamine	hexadecane	1.33 [222]	9.36	2.09	2.05	1.68	-0.40	1.86	1.46
pentylamine	hexadecane	1.19	11.38	2.09	2.05	1.50	-0.26	1.23	0.96
hexylamine	hexadecane	1.27	13.25	2.09	2.05	1.61	-0.26	1.20	0.94
heptylamine	hexadecane	1.18	15.06	2.09	2.05	1.49	-0.19	0.91	0.72
octylamine	hexadecane	1.27	16.91	2.09	2.05	1.61	-0.20	0.94	0.74
nonylamine	hexadecane	1.18	18.76	2.09	2.05	1.49	-0.16	0.73	0.58
decylamine	hexadecane	1.27	20.58	2.09	2.05	1.61	-0.16	0.77	0.61
dimethylamine	hexadecane	1.02 [223]	6.02	2.09	2.05	1.29	-0.36	1.70	1.34
diethylamine	hexadecane	0.90	9.62	2.09	2.05	1.14	-0.18	0.83	0.66
trimethylamine	hexadecane	0.64 [223]	7.90	2.09	2.05	0.81	-0.11	0.51	0.40
triethylamine	hexadecane	0.56	13.48	2.09	2.05	0.71	-0.05	0.23	0.18

cost and parameterisation complexity [225]. Consequently, bond and angle parameters were taken from TraPPE-EH (which uses bonded parameters from OPLS-AA [213]) and dihedral's parameters from AUA4 [217]. The latter force field does not use 1-4 LJ or electrostatic interactions, except for ethylenediamine which was not considered here, and therefore its dihedral parameters do not depend on the non-bonded parameters. The PolCA model developed in this work also does not use 1-4 non-bonded interactions, and thus, AUA4's dihedral parameters are consistent with our force field. For these

reasons, AUA4 was chosen as the source for the missing dihedrals, and a curve fitting technique was used to convert its parameters into the Ryckaert-Bellemans convention (Equation 2.38) (see Figure 4.1). Table 4.5 contains the dihedral parameters extracted from the AUA4 model [217] which need to be used with Equation 4.1:

$$U_{tor}^{AUA4} = \sum_{i=0}^{8} a_i \cos(\phi + 180°)^i \tag{4.1}$$

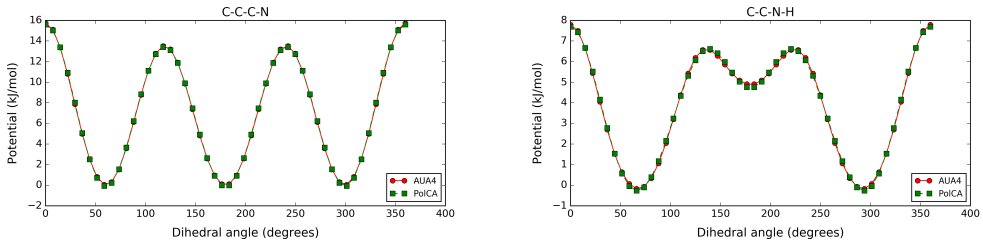

Figure 4.1: Dihedrals obtained using PolCA's parameters with the Ryckaert-Bellemans convention (dashed green line) and AUA4's parameters with Equation 4.1 (full red line). The plot on the left corresponds to the dihedral C-C-C-N, and the one on the right to C-C-N-H.

Table 4.5: Torsion parameters for the AUA4 model [217] to be used with Equation 4.1. The parameters a_i are divided by k_B, and thus, their units are K.

Torsion	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
CH_x-CH_2-CH_2-N	816.65	2509.94	9.01	-3609.00	-54.51	286.01	-104.22	-133.18	279.10
CH_x-CH_2-N-H	154.98	869.66	902.48	-586.25	-651.38	-607.86	253.22	149.74	102.90

The PolCA non-bonded and bonded parameters for amines can be found in Tables 4.6 and 4.7, respectively.

Table 4.6: Non-bonded parameters for the PolCA force field.

non-bonded [a] (Eq 2.37)	σ [nm]	ε [kJ/mol]	Partial charge (q)
CH_x-**N**-H_2	0.3401	0.814	-0.871
$(CH_x)_2$-**N**-H	0.389	0.723	-0.605
$(CH_x)_3$-**N**	0.4918	0.151	-0.339
N-**H**	0	0	0.379
CH_3-CH_x	0.379	0.833	0
CH_3-N	0.379	0.833	0.113
$(CH_x)_2$-**CH_2**	0.399	0.392	0
CH_2-N	0.399	0.392	0.113
$(CH_x)_3$-**CH**	0.473	0.085	0
CH-N	0.4645	0.085	0.113
$(CH_x)_4$-**C**	0.646	0.00426	0
C-N	0.630	0.00426	0.113

a) The non-bonded parameters correspond to the sites in bold.

Table 4.7: Bonded parameters for the PolCA Force Field.

Bonds

	Ideal bond length b_0 [nm]
CH_x-CH_y	0.154
N-H	0.101
CH_n-N	0.1448

Angles

	Force constant k_θ [kJ $mol^{-1}rad^{-2}$]	Ideal bond angle θ_0 [degree]
CH_x-CH_2-CH_y	519.655	114
H-N-H	364.84	106.4
CH_n-N-H	292.88	109.5
CH_x-CH_x-N	470.281	109.47

Dihedrals

	C_0	C_1	C_2	C_3	C_4	C_5
CH_x-CH_2-CH_2-CH_y	8.397	16.785	1.134	-26.316		
CH_x-CH_y-CH_n-N	6.892	20.732	-1.499	-28.956	2.349	0.358
CH_x-CH_y-N-H	1.403	7.356	5.549	-5.907	-0.736	-2.932

Force constants for the angles are twice the value found in the OPLS-AA paper [213] since their equation includes the coefficient 1/2 in k_θ.

4.2.3 GROMOS-2016H66 Force Field

As explained in section 4.2.2, TraPPE does not have UA parameters for amines, and thus, the GROMOS-compatible 2016H66 parameter set developed by Horta et al. [80] was selected as a reference case to assess the PolCA model's performance. This set was obtained by fitting to the pure-liquid density, the enthalpy of vaporisation, the hydration free energy and the solvation free energy in cyclohexane of small organic molecules. The molecules used in the optimisation were at atmospheric pressure and at a temperature close or equal to 298.15 K, and they belonged to different families including alcohols, ethers, aldehydes, ketones, carboxylic acids, esters, amines, amides, thiols, sulfides, disulfides, aromatic compounds and nucleic-acid bases. Ethylamine, diethylamine and triethylamine were used in both the calibration and validation, while propylamine, butylamine and ethylenediamine were only part of the validation process. Additionally, seven other properties were studied during this model's validation: the molar isobaric heat capacity, the isothermal compressibility, the isobaric thermal-expansion coefficient, the surface tension coefficient, the self-diffusion coefficient, the

shear viscosity and the static relative dielectric permittivity.

Interaction Parameters

The LJ and bonded parameters of the GROMOS-2016H66 set are the same as the ones from the GROMOS-53A6 force field [101], except for a few exceptions that do not affect the molecules studied here [80]. This model represents aliphatic carbon groups as one pseudo-atom and uses a Lennard-Jones potential and Coulomb interactions with a reaction field to calculate van der Waals and electrostatic interactions, respectively.

$$U^{Gromos} = \sum_{pairs\ i,j} \left(\frac{C12_{ij}}{r_{ij}^{12}} - \frac{C6_{ij}}{r_{ij}^{6}} + \frac{q_i\ q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \right) + u^{RF} \tag{4.2}$$

$$u^{RF}(r_{ij}) = \sum_{pairs\ i,j} \frac{q_i\ q_j}{4\pi\epsilon_0\epsilon_1} \left[\frac{-0.5\ C_{rf}\ r_{ij}^2}{r_c^3} + \frac{-(1 - 0.5\ C_r f)}{r_c} \right] \tag{4.3}$$

where,

$$C_{rf} = \frac{(2\epsilon_1 - 2\epsilon_2)\ (1 + k\ r_c) - \epsilon_2\ (k\ r_c)^2}{(\epsilon_1 + 2\epsilon_2)\ (1 + k\ r_c) + \epsilon_2\ (k\ r_c)^2} \tag{4.4}$$

Here, r_{ij}, q, ϵ_0 and ϵ_1 are the distance between atoms i and j, the partial charges, the vacuum permittivity and the relative permittivity of the medium, respectively. Additionally, ϵ_2 and k are the relative permittivity and inverse Debye screening length of the medium outside the cut-off sphere, respectively. The reaction-field contribution term (u^{RF}) only appears outside the cut-off and represents the interaction between the atoms and the dielectric medium. Furthermore, the mixed parameters $C12_{ij}$ and $C6_{ij}$ are obtained from the parameters defined for each atom type using geometric combination rules (Equation 4.6) and taking into consideration the type of interaction.

$$C12_{ij} = \sqrt{C12_{ii}\ C12_{jj}} \tag{4.5}$$

$$C6_{ij} = \sqrt{C6_{ii}\ C6_{jj}} \tag{4.6}$$

Each atom type can have up to three different $C12$ parameters ($C12[I]$, $C12[II]$ and $C12[III]$), being type I the default. Type II is slightly larger than type I and is normally used when the atoms i and j can form hydrogen bonds. Lastly, type III is used for ions. Furthermore, 1-4 interactions have a special set of parameters labelled $C6_{nei}^{1/2}$ and $C12_{nei}^{1/2}$. The parameters $C12$ and $C6$ can easily be related to σ and ε using Equations 2.27 and 2.28, as is explained in the methodology (section 2.5).

The GROMOS force field calculates bonded interactions using different equations than TraPPE. GROMOS uses Equations 4.7, 4.8 and 4.9 to obtain the potential energy due to bonds, angles and dihedrals, respectively:

$$u_{bonds}^{G} = \frac{K_b}{4} \left[r_{ij}^2 - b_0^2\right]^2 \tag{4.7}$$

$$u_{angles}^{G} = \frac{K_\theta}{2} \left[\cos(\theta) - \cos(\theta_0)\right]^2 \tag{4.8}$$

$$u_{torsion}^{G} = K_\phi \ \left\{1 + \cos[m(\phi - 180^\circ)]\cos(\delta)\right\} \tag{4.9}$$

where, r_{ij}, θ, ϕ are the actual distance, angle and dihedral angle between particles, respectively. All other parameters are defined according to the type of atoms involved in the interaction. Tables 4.8 and 4.9 present the GROMOS bonded and non-bonded parameters for primary amines.

4.2.4 Optimisation

The density, enthalpy of vaporisation and dielectric constant of methylamine, 1-propylamine, 1-pentylamine and 1-heptylamine were simulated using different LJ parameters for the nitrogen atom (LJ parameters for amine hydrogen atoms were set to zero by construction) and different partial charges for the amino group (namely, on the nitrogen and hydrogen atoms), while the α-carbon's charge was adjusted accordingly to keep the neutrality of the molecule. This led to a total of 4 fitting parameters for primary amines: ε, σ, nitrogen's partial charge (q_N) and hydrogen's

Table 4.8: GROMOS-2016H66's bonded parameters for primary amines [101]. The second column shows the code used in the GROMOS force field to identify each bond, angle and dihedral type.

Bonds				
	Code	Ideal bond length b_0 [nm]		
NT-H	2	0.1		
CH_n-NT	21	0.147		
CH_x-CH_y	27	0.153		
Angles				
	Code	Force constant K_θ [kJmol^{-1}]	Ideal bond angle θ_0 [degree]	
H-NT-H	10	380	109.5	
CH_n-NT-H	11	425	109.5	
CH_x-CH_y-CH_n	15	530	111	
CH_x-CH_y-NT	15	530	111	
Dihedrals				
	Code	Force constant K_ϕ [kJmol^{-1}]	Phase shift cos(δ)	Multiplicity (m)
X-CH_x-CH_y-X	34	5.92	1	3
X-CH_n-NT-X	41	3.77	1	6

Table 4.9: GROMOS-2016H66's non-bonded parameters for primary amines [80, 101]. The C12(II)$^{1/2}$ for two nitrogen atoms that interact with each other is 1.6 (10^{-3} kJmol^{-1} nm^{12})$^{1/2}$. The units for C6$^{1/2}$ and C12(I)$^{1/2}$ are (kJmol^{-1} nm^6)$^{1/2}$ and (10^{-3} kJmol^{-1} nm^{12})$^{1/2}$, respectively. The same units apply for the 1-4 parameters.

Code	**Atom type**	**C6$^{1/2}$**	**C6$_{nei}$$^{1/2}$**	**C12 (I)$^{1/2}$**	**C12$_{nei}$** $^{1/2}$	q
7	**NT**	0.04975	0.04936	1.600	1.301	-0.98
15	**CH2**	0.08642	0.06873	5.828	2.178	0
15	**CH2**-NT	0.08642	0.06873	5.828	2.178	0.25
16	**CH3**	0.09805	0.08278	5.162	2.456	0
16	**CH3**-NT	0.09805	0.08278	5.162	2.456	0.25
21	**H**	0	0	0	0	0.365

Non-bonded parameters correspond to the atoms in bold.

partial charge (q_H). In Chapter 3, the diffusion constant was used during the alcohol model calibration; however, no experimental data was found for the self-diffusion of alkyl amines, and thus, the dielectric constant was chosen instead. Amines' dielectric constant converged a lot faster than alcohols', and consequently, the meta-models for this property have a good performance (see below).

The alkyl chain's parameters were the same as the ones used for alcohols (Table 4.6),

and they were developed by Jorge [1]. Also, the LJ parameters of the alkane methyl and methylene groups were used for the α-carbons without modification.

As explained in Chapter 2, meta-models were created to predict how the target properties vary with the amino group's non-bonded parameters. These meta-models included polarisation corrections for the enthalpy of vaporisation and the dielectric constant. Additionally, each meta-model's performance was checked by plotting simulated versus predicted values, and once an optimum was found, the values predicted by the meta-models were always in good agreement with the results from the simulations.

Once the meta-models were obtained and their performance checked, they were used to create an objective function which was minimised using the procedure explained in section 2.9. The objective function was:

$$F(X) = \sum_{j=1}^{4} \left((f_k(X) - y_{exp})^2_{\rho_j} + (f_k(X) - y_{exp})^2_{\Delta H_j} + (f_k(X) - y_{exp})^2_{\epsilon_{sol\,j}} \right) \qquad (4.10)$$

where j=1, j=2, j=3 and j=4 correspond to methylamine, propylamine, pentylamine and heptylamine, respectively. Also, $f_k(X)$ is the value predicted using the meta-model at X, y_{exp} is the experimental value and ρ, ΔH and ϵ_{sol} represent the density, enthalpy of vaporisation and dielectric constant, respectively.

Different learning sets were investigated during the parameterisation, resulting in three potential models for amines (called Model 1, Model 2 and Model 3 from now on). The learning sets used to obtain each model are described below.

Construction of the learning set

An initial grid was created for methylamine, pentylamine and heptylamine around TraPPE-EH's non-bonded parameters (σ = 0.334 nm, ε = 0.923 kJ/mol, q_N = -0.892 and q_H = 0.356). The levels used for each variable are shown in Table 4.10, and the total number of simulations for each molecule was 342.

The meta-models obtained using this initial grid were not accurate, and the predicted

Table 4.10: Parameter levels used in the initial grid.

	Levels
σ [nm]	0.325, 0.331, 0.334, 0.340, 0.346 and 0.352
ε [kJ/mol]	0.523, 0.623, 0.723, 0.823, 0.923, 1.023, 1.123 and 1.223
q_N	-0.792, -0.892 and -0.992
q_H	0.256, 0.356 and 0.456

and simulated properties of the possible optimums did not match. For example, the experimental density of pentylamine is 751 kg/m^3, and looking at Figure 4.2 we can see that an optimum with a predicted density of 751 kg/m^3 can actually give simulated values that range from 745 to 757 kg/m^3 (an error of 6 kg/m^3). Consequently, a sensitivity analysis was carried out to detect each property's most influential parameter. In all cases, the hydrogen's partial charge was the predominant variable, as can be seen in Figure 4.2.

After uncovering the problems with the initial grid, a more accurate learning set was created using a smaller range for q_H ($0.336 \leq q_H \leq 0.396$). The parameters' levels in this grid for all other variables were the same as the ones presented in Table 4.10, while the levels for the hydrogen's partial charge were 0.336, 0.356, 0.376 and 0.396. The optimum obtained using this grid ($\sigma = 0.324$, $\varepsilon = 0.253$, q_N = -0.970, $q_H = 0.356$) fell outside the grid's boundaries and thus, a smaller grid was created for higher accuracy around those values. This new grid was used to obtained Model 1, and it is explained below.

Model 1

The meta-models were created using a full factorial grid and three levels for each variable (see Table 4.11); however, 5 points that had low σ, low ε and high q_N were removed to improve the meta-models performance. These points were chosen based on the predictivity plots, and their removal increased the meta-models performance for all the properties (see Figure A10 in Appendix A2). Three of the removed points were combinations of $\sigma = 0.325$ nm with $\varepsilon = 0.223$ kJ/mol and q_N = -0.992, while the other two points were: $\sigma = 0.325$, $\varepsilon = 0.323$, q_N = -0.992, $q_H = 0.396$ and $\sigma = 0.331$, $\varepsilon = 0.223$, q_N = -0.992, $q_H = 0.396$. Therefore, Model 1's learning set consisted of

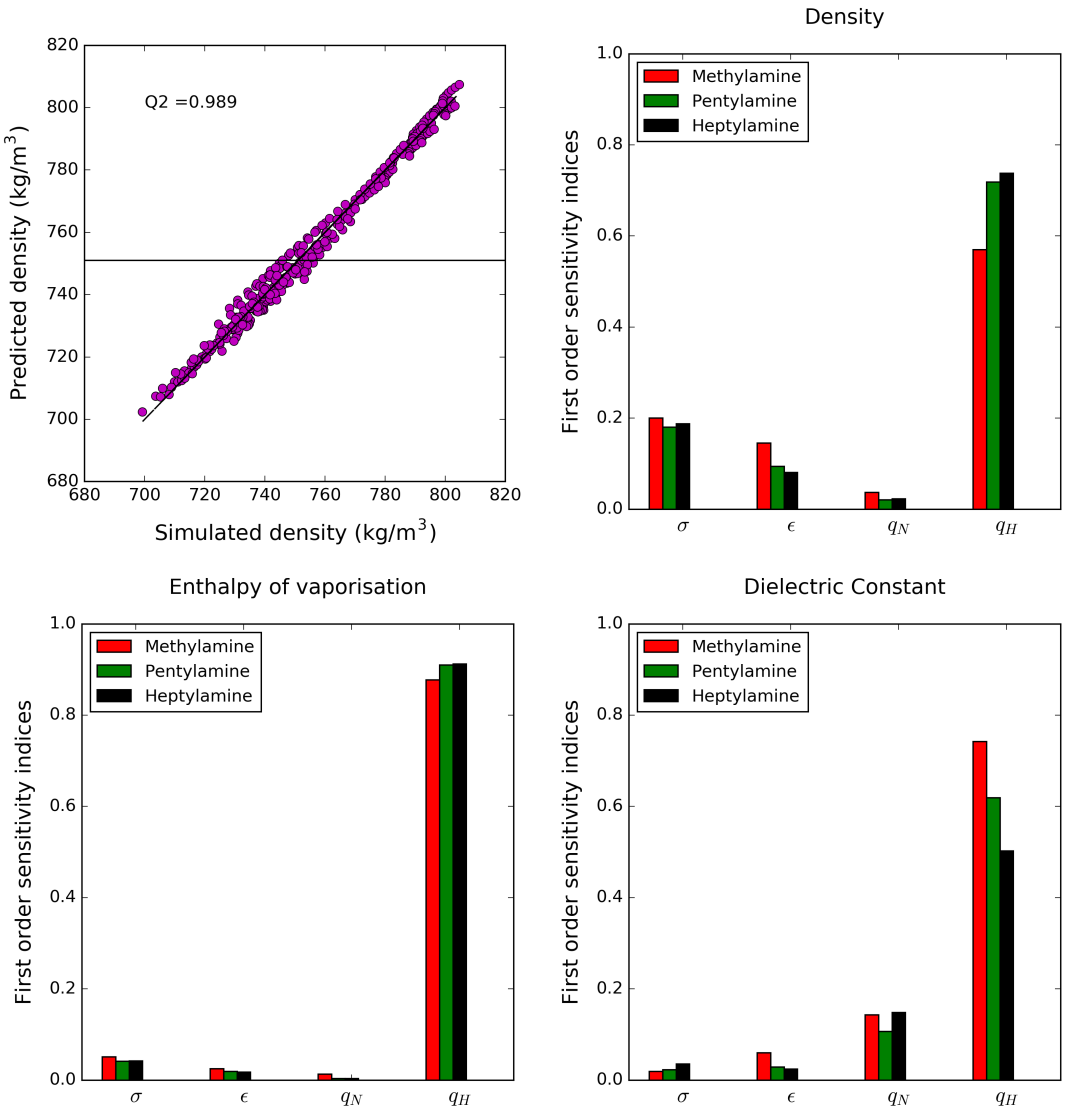

Figure 4.2: The top left figure shows the performance of the density's meta-model for pentylamine at 298.15 K and 1 bar obtained using the initial grid. This plot compares simulated values to the meta-model's predictions, and a horizontal black line represents pentylamine's experimental density. The three other plots are the first-order sensitivity indices of the initial grid's meta-models for each property. The studied property in the top-right plot is the density while the properties in the figures at the bottom are the enthalpy of vaporisation (left) and the dielectric constant (right). The different colours represent the different molecules: methylamine (red), 1-pentylamine (green) and 1-heptylamine (black).

76 simulations for each molecule, including propylamine which was added to improve the prediction of small amines. The initial point for the optimisation was $\sigma = 0.334$, $\varepsilon = 0.423$, q_N = -0.892, $q_H = 0.356$.

The objective function used to obtain Model 1 did not include methylamine's dielectric constant since removing this property from the parameterisation slightly improved the

Table 4.11: Levels for each parameter used in Model 1's learning set.

	Levels
σ [nm]	0.325, 0.331 and 0.334
ε [kJ/mol]	0.223, 0.323 and 0.423
q_N	-0.792, -0.892 and -0.992
q_H	0.356, 0.376 and 0.396

enthalpy of vaporisation and the dielectric constant of pentylamine and heptylamine at the expense of a very small deterioration in the dielectric constant of methylamine and propylamine. However, the models obtained including and excluding methylamine's dielectric constant from the objective function had very similar parameters, and thus, either approach could have been used. Furthermore, the optimisation procedure was forced to stay inside the parameters boundaries to assure the predictivity of the model since lifting this restriction generated optimums that were outside the region where the meta-models have a good performance. This outcome suggested that a grid expansion was needed for a more accurate model. The meta-models' performance within the grid's region can be found in Appendix A2 (Figure A11).

Model 2

As explained above, the method used to obtain Model 1 was restricted by the parameters boundaries, and thus, Model 1's grid was extended to incorporate more σ and ε values. The learning set used for methylamine, pentylamine and heptylamine was constructed using 165 simulations from previous grids without having to run extra points. This set consisted of combinations of σ = 0.325, 0.331, 0.334 and 0.340 nm with ε = 0.423, 0.523, 0.623, 0.723, 0.823, 0.923, 1.023, 1.123 and 1.223 kJ/mol, with the same levels for the partial charges used in Model 1's grid. Propylamine, unlike methylamine, pentylamine and heptylamine, was not included in the initial grid shown in Table 4.10, and thus, different parameter levels were used for σ and ε when constructing its meta-models to avoid running many extra simulations. Propylamine's learning set consisted of 108 simulations, out of which 81 simulations were new, and the levels used for σ and ε were 0.325, 0.331 and 0.334 nm and 0.423, 0.523, 0.623 and 0.723 kJ/mol, respectively. Additionally, the optimum model's predicted values

were compared with the simulated values to check the performance of propylamine's meta-model, and, since there was a good agreement, no extra simulations were run for this molecule. The levels used for each parameter are summarised in Table 4.12.

Table 4.12: Levels for each parameter used in Model 2's learning set.

	Levels
	For propylamine
σ [nm]	0.325, 0.331 and 0.334
ε [kJ/mol]	0.423, 0.523, 0.623, 0.723
q_N	-0.792, -0.892 and -0.992
q_H	0.356, 0.376 and 0.396
	For methyl, pentyl and heptyl amine
σ [nm]	0.325, 0.331, 0.334 and 0.340
ε [kJ/mol]	0.423, 0.523, 0.623, 0.723, 0.823, 0.923, 1.023, 1.123, 1.223
q_N	-0.792, -0.892 and -0.992
q_H	0.356, 0.376 and 0.396

Furthermore, the initial point for the optimisation was the same as the one used for Model 1; however, unlike in the previous approach, the search was no longer forced to stay inside the boundaries. The optimum's predicted properties were in good agreement with the simulations even though the optimum's σ and ε were outside the boundaries for propylamine's learning set. The meta-models performance inside their respective boundaries can be found in Appendix A2 (Figure A12).

Model 3

Model 3 was created since the previous two models were not able to predict all the properties studied here. Model 1 severely underpredicts the free energy of solvation of amines in hexadecane, while Model 2 is able to predict this property, but it overpredicts the dielectric constant of primary amines. Additionally, the dipoles obtained with Model 2 are higher than the estimated liquid dipoles. These results are explained in detail in section 4.3.2.

Model 3 was obtained using an extended grid for propylamine, pentylamine and heptylamine, while methylamine was not included in the objective function. A total of 174 simulations were used to create each molecule's learning set, and, unlike in the previous approach, the same number of simulations was used for all three molecules.

Additionally, the optimum was not forced to stay inside the learning grid's boundaries but an extra restriction was included in the objective function to assure that the optimum's dipole moments for the fitted molecules were lower than their estimated liquid dipoles from Table 4.4. Table 4.13 contains the levels used for each parameter, and, the meta-models performances can be found in Appendix A2 (Figure A13). Similar to the previous approach, one of the model's parameters was slightly outside the boundaries; however, simulated values were in agreement with the predictions.

Table 4.13: Levels for each parameter used in Model 3's learning set.

	Levels
σ [nm]	0.325, 0.331, 0.334, 0.340, 0.344, 0.348 and 0.352
ε [kJ/mol]	0.723, 0.823, 0.923, 1.023, 1.1, 1.123, 1.223, 1.3 and 1.5
q_N	-0.792, -0.892 and -0.992
q_H	0.322, 0.336, 0.352, 0.356, 0.376 and 0.382

4.3 Results and Discussion

Several force fields are discussed in this section, and thus, a summary table showing all these models (Table 4.14) is included here to improve the readability of the chapter.

4.3.1 Gromos 2016H66 with polarisation corrections

The simulation protocol used during the development of the GROMOS 2016H66 force field [80] differs from the procedure used in this work, and thus, values obtained using three different methods are presented here for easier comparison with the literature. The first method, called Gromos-2016-RF from now on, is the same as the one used during Gromos-2016H66's parameterisation. This approach uses a cut-off of 1.4 nm for non-bonded interactions and a reaction-field correction [226] to account for electrostatic interactions beyond the cut-off. However, long-range dispersion interactions beyond the cut-off are not included. The second method, called Gromos-2016-RF-Dispcorr, is the same as the first approach except for the inclusion of long-range dispersion corrections for energy and pressure. These corrections are essential since it has previously been shown that the bulk density depends strongly on the cut-off radius when long-range

Table 4.14: Force fields simulated in this chapter. Model 1, Model 2 and Model 3 use the LJ parameters proposed by Jorge [1] for the alkyl chain and the α-carbon, and have been parameterised taking into account post-facto polarisation corrections. Unless stated otherwise their results always include these corrections.

Gromos-2016-RF	• GROMOS 2016H66 force field for amines [80]. This model was fitted to the liquid density, enthalpy of vaporisation, hydration free energy and solvation free energy in cyclohexane of ethylamine, diethylamine and triethylamine. • Same simulation protocol as the one used during its development: cut-off equal to 1.4 nm without long-range dispersion corrections and with a reaction-field correction for electrostatic interactions beyond the cut-off.
Gromos-2016-RF-Dispcorr	• Same as Gromos-2016-RF except for the inclusion of long-range dispersion corrections for the energy and pressure.
Gromos-2016-PME	• GROMOS 2016H66 force field simulated using the same protocol as the one used for PolCA • Includes long-range dispersion corrections and uses PME to treat electrostatic interactions beyond the cut-off (1 nm).
Model 1	• Fitted to the density, enthalpy of vaporisation and dielectric constant of propylamine, pentylamine and heptylamine and to methylamine's density and enthalpy of vaporisation. • Optimisation algorithm was forced to stay inside the boundaries of the learning set.
Model 2	• Fitted to the density, enthalpy of vaporisation and dielectric constant of propylamine, pentylamine and heptylamine and to methylamine's density and enthalpy of vaporisation. • The meta-models used during its parameterisation were trained using a larger grid that the one used for Model 1. • No boundary restrictions were imposed during the optimisation routine.
Model 3	• Fitted to density, enthalpy of vaporisation and dielectric constant of propylamine, pentylamine and heptylamine. • No boundary restrictions were imposed during the optimisation routine but the dipole moments of the fitted molecules were forced to be lower than their estimated liquid dipoles.

corrections are not applied, and as the radius increases the density gets closer to the value obtained using long-range corrections [1]. The last protocol, called Gromos-2016-PME, is the same as the one used for PolCA and TraPPE, and it is described in detail in section 2.7. This approach uses PME to treat long-range electrostatic interactions and a cut-off of 1 nm for non-bonded interactions. Also, it applies long-range dispersion corrections for energy and pressure.

Figure 4.3 shows the predicted density of 10 linear primary amines obtained using

Gromos-2016H66 with the three approaches explained above. Gromos-2016-RF accurately predicts the density from propylamine to decylamine, and its trend suggests that longer linear primary amines will also be accurately predicted. However, methylamine's density is highly overpredicted, and ethylamine's density is slightly overpredicted. This result is not surprising since Gromos performs very well at predicting alkane densities when this protocol is used [142]. On the other hand, all densities are highly overpredicted once long-range dispersion corrections are added, and these values are similar to those obtained using Gromos-2016-PME (especially from pentylamine onwards). This outcome suggests that Gromos-2016-RF underpredicts long-range dispersion forces and that the difference between the different methods is mostly due to dispersion corrections and not to the way long-range electrostatic interactions are treated.

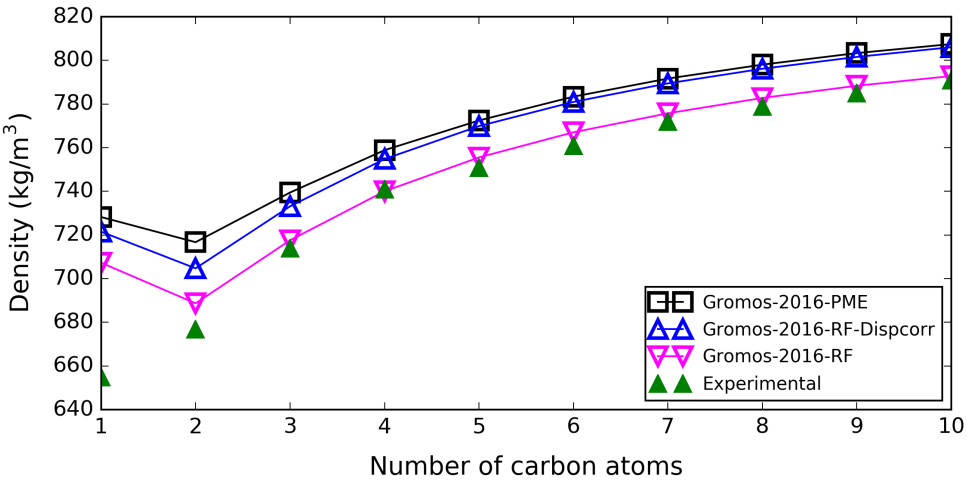

Figure 4.3: Density of linear primary amines at 298.15 K and 1 bar obtained using Gromos-2016H66 with three different methods. Simulations ran using a reaction field correction without and with dispersion corrections are represented with upside-down magenta triangles and blue triangles, respectively. Black squares correspond to values obtained using PME with dispersion corrections, and green symbols are experimental values [224].

Figure 4.4 shows the predicted enthalpy of vaporisation obtained with and without polarisation corrections. Gromos-2016-RF does a great job at predicting this property for all the primary amines studied here with an RMSD of 1.695 kJ/mol, and once polarisation corrections are applied its performance improves by 0.467 kJ/mol (RMSD = 1.228 kJ/mol). However, once dispersion corrections are added, Gromos-2016 overpredicts the enthalpy of vaporisation for molecules larger than ethylamine and the RMSD of Gromos-2016-RF-Dispcorr without polarisation corrections is 1.153 kJ/mol larger than

Gromos-2016-RF's RMSD. This outcome is not unexpected since, as explained above, Gromos-2016-RF underpredicts attractive long-range forces. Furthermore, adding polarisation corrections (which are small but positive) to Gromos-2016-RF-Dispcorr actually worsens agreement with experimental data by 0.561 kJ/mol since the uncorrected model already overpredicts this property. A similar effect is observed for Gromos-2016-PME which has an RMSD of 2.347 kJ/mol before applying polarisation corrections and an RMSD of 2.844 kJ/mol once these corrections are incorporated.

On the other hand, the dielectric constant does not seem to be affected by the long-range dispersion corrections or the way long-range electrostatic interactions are treated beyond the cut-off (Figure 4.5). In all cases, Gromos can accurately predict the dielectric constant of primary amines larger than methylamine, once simulated values are corrected using Equation 2.52, while methylamine's dielectric constant is overpredicted by a factor of 1.4. If corrections are not applied, Gromos 2016H66 does a better job at predicting methylamine's dielectric constant but it underpredicts the dielectric constant for ethylamine and beyond (the curves for all versions of Gromos nearly overlap, and so only one is shown).

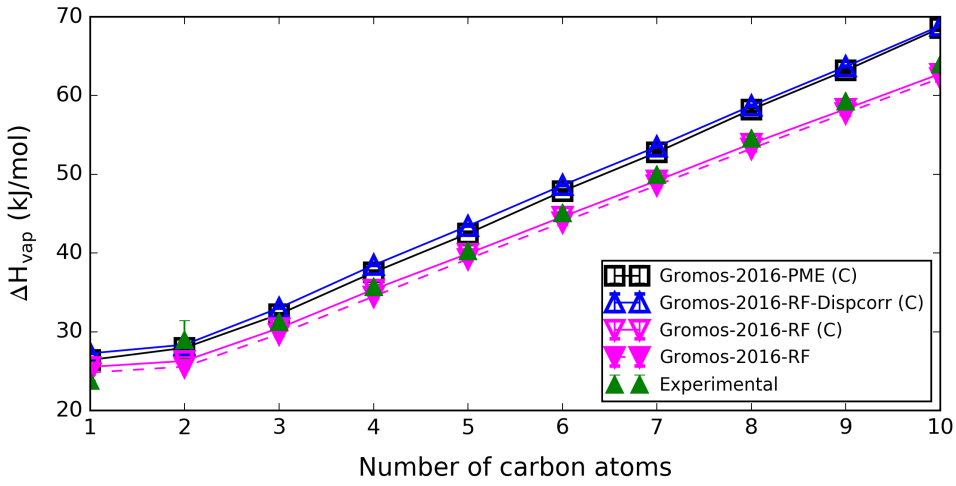

Figure 4.4: Enthalpy of vaporisation of primary amines at 298.15 K and 1 bar obtained with Gromos-2016H66. Here, the term (C) in the labels means that polarisation corrections were included (full lines and empty symbols). Magenta upside-down triangles represent simulations ran using a reaction field correction without dispersion corrections, while blue triangles correspond to values that include dispersion corrections, and black squares represent values obtained using PME with long-range dispersion corrections. Experimental values (green symbols) were taken from references [194] and [227].

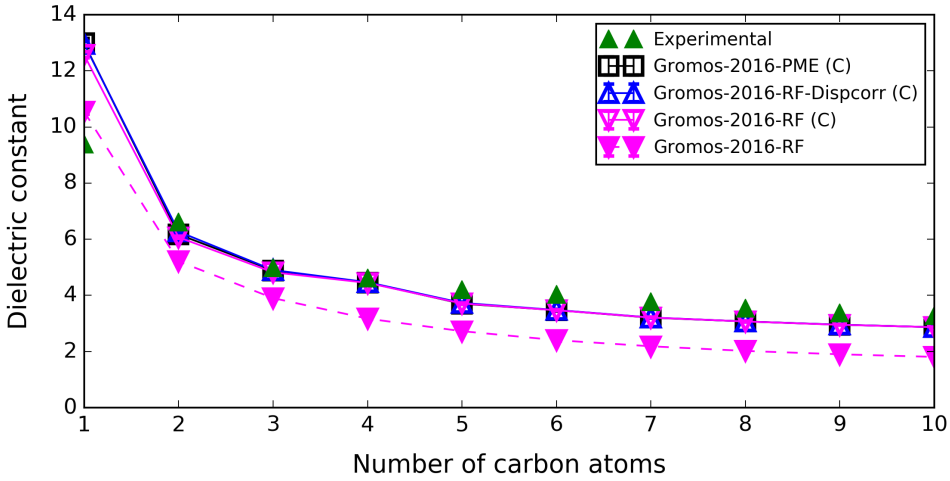

Figure 4.5: Dielectric constant of primary amines at 298.15 K and 1 bar obtained with Gromos-2016H66 at 298.15 K and 1 bar. Here, the term (C) in the labels means that polarisation corrections were included (full lines and empty symbols). Magenta upside-down triangles represent simulations ran using a reaction field correction without dispersion corrections, while blue triangles correspond to values that include dispersion corrections, and black squares represent values obtained using PME with long-range dispersion corrections. Experimental values (green symbols) were taken from references [223] and [221].

Figure 4.6 presents the model's performance at predicting free energies of self-solvation of primary amines with and without polarisation corrections. Free energy simulations are very time consuming, and thus, only simulations using the Gromos-2016-PME and Gromos-2016-RF were run. Gromos-2016-RF-Dispcorr was not part of the calculations since we can see from Figures 4.3, 4.4 and 4.5 that Gromos-2016 is not significantly affected by the way long-range electrostatic forces are treated. The free energy of self-solvation of primary amines is accurately predicted using Gromos-2016-PME, although adding polarisation corrections slightly increases the RMSD from 1.032 to 1.525 kJ/mol. As expected, the opposite behaviour is observed for Gromos-2016-RF since adding polarisation corrections decreases the RMSD from 1.664 to 1.170. Nonetheless, Gromos-2016-PME's comparatively poor performance at predicting the density and enthalpy of vaporisation suggests that this model could be improved.

The free energy of solvation of linear primary amines in hexadecane, calculated using Gromos-2016-PME and Gromos-2016-RF with and without polarisation corrections, is shown in Figure 4.7. The polarisation corrections for these systems are significant, as can be seen from Table 4.4, especially for smaller amines, and they highly improve

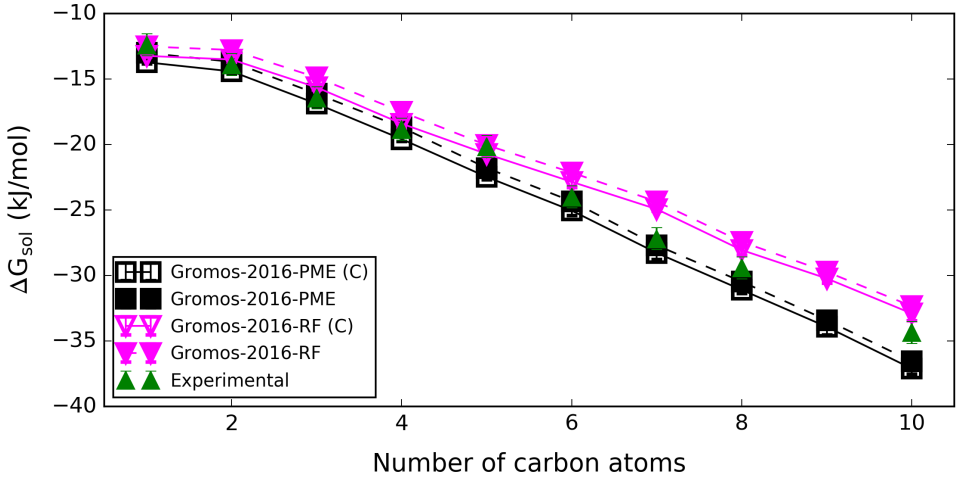

Figure 4.6: Free energy of self-solvation of primary amines at 298.15 K and 1 bar obtained using Gromos-2016H66 with RF (magenta upside-down triangles) and PME (black squares). Full lines with empty symbols contain polarisation corrections, while dashed lines with filled symbols represent values without corrections. The green symbols are experimental values obtained from vapour pressure data [228, 229].

the model's performance once added. The RMSD of Gromos-2016-PME decreases from 3.738 to 2.147 kJ/mol, while Gromos-2016-RF's RMSD decreases from 5.049 to 3.563. However, even with these corrections, Gromos-2016-PME still significantly underpredicts solvation free energies magnitudes from methylamine to propylamine, while Gromos-2016-RF significantly underpredicts this property's magnitude for all ten linear primary amines studied here. Gromos-2016-RF does not use dispersion corrections during the simulations and thus, its predicted values are lower than those obtained using PME, especially from pentylamine onwards.

The root-mean-square deviation of the three methods for the properties studied here can be found in Table 4.15. The Gromos-2016-RF's RMSD for the density is high due to methylamine's density which is greatly overpredicted. However, the RMSD decreases to 5.27 kg/m^3 if methylamine is not included. Furthermore, this table also contains the RMSD of the enthalpy of vaporisation and the dielectric constant without polarisation corrections. For the three approaches, agreement with experimental data improves for the dielectric constant when simulated values are corrected using Equation 2.52, while adding polarisation corrections to the enthalpy of vaporisation improves accuracy only for Gromos-2016-RF, which corresponds to the protocol used in the original parameteri-

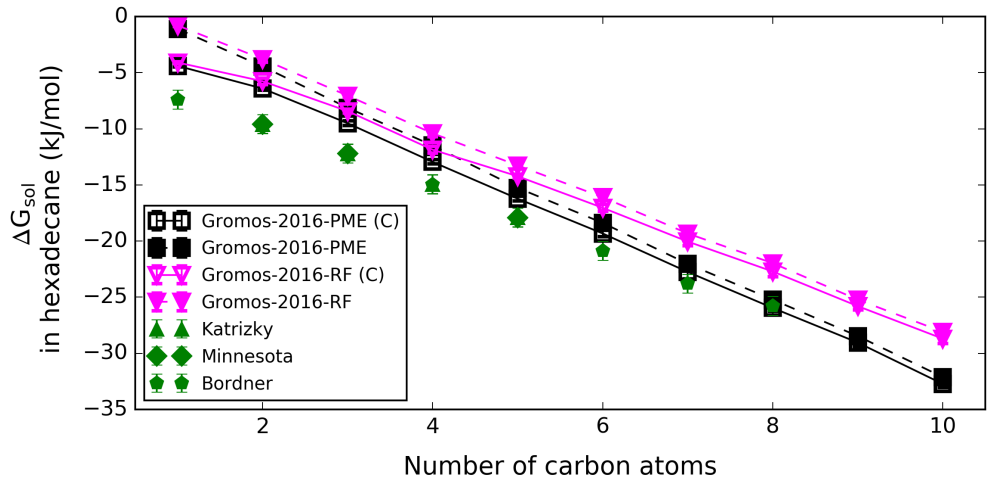

Figure 4.7: Free energy of solvation of linear primary amines in hexadecane at 298.15 K and 1 bar obtained using Gromos-2016H66 with RF (magenta upside-down triangles) and PME (black squares). Full lines with empty symbols contain polarisation corrections, while dashed lines with filled symbols represent values without corrections. The green symbols are experimental values obtained from the Katrizky and Minnesota databases [168, 169] and reference [230]'s supporting information.

sation. Therefore, it can be concluded that the Gromos 2016H66 amine force field would benefit from inclusion of *post-facto* polarisation corrections, as observed for TraPPE in the case of alcohols (see section 3.3.5). For the other protocols considered here, the results deviate more strongly from experimental data when the positive polarisation corrections are added, since they already overestimate the vaporisation enthalpy.

Table 4.15: RMSD of the density, dielectric constant, enthalpy of vaporisation and free energy of solvation of primary amines in themselves and hexadecane at 298.15 K and 1 bar, obtained using the Gromos-2016H66 parameter set and the three protocols explained above. The values for properties with a (C) next to their names were obtained including post-facto polarisation corrections. No experimental data was found for the diffusion constant of the simulated amines, and thus, this property was not included in the study.

	Gromos-2016-RF	**Gromos-2016-RF-Dispcorr**	**Gromos-2016-PME**
Density [kg/m^3]	17.278	27.572	31.975
ΔH_{vap} [kJ/mol]	1.695	2.848	2.347
ΔH_{vap} (C) [kJ/mol]	1.228	3.409	2.844
Dielectric	1.435	1.438	1.457
Dielectric (C)	1.088	1.194	1.207
ΔG_{sol} [kJ/mol]	1.664		1.032
ΔG_{sol} (C) [kJ/mol]	1.170		1.525
ΔG_{hex} [kJ/mol]	5.049		3.738
ΔG_{hex} (C) [kJ/mol]	3.563		2.147

4.3.2 Comparison between the new potential models

Table 4.16 contains the optimised parameters for each approach. Model 1 has the lowest σ and ε, while Model 3 has the highest values for these parameters. Model 2 fits in between the other two models, with σ = 0.34 nm and ε = 0.814 kJ/mol, and, as it will be shown in this section, this model has the best performance overall. The next paragraphs contain each model's predicted densities, enthalpies of vaporisation, dielectric constants, self-solvation free energies and solvation free energies in hexadecane of ten primary amines.

Table 4.16: Non-bonded parameters for each model.

Model	σ (nm)	ε (kJ/mol)	q_N	q_H
Gromos-2016H66	0.314	0.640	-0.980	0.365
Model 1	0.333	0.423	-0.935	0.373
Model 2	0.340	0.814	-0.871	0.379
Model 3	0.353	1.580	-0.902	0.350

Density

As can be seen from Figure 4.8, all three models do a good job at predicting the liquid density of primary linear amines. However, Model 1 systematically overpredicts the density of linear primary amines larger than butylamine, and Model 3 overpredicts the density of methylamine, ethylamine and propylamine. On the other hand, Model 2 accurately predicts the density of all the linear amines studied here, except for ethylamine which is overestimated by 1.03 %.

The RMSD of Model 1, Model 2 and Model 3 with respect to experimental values are 4.975, 3.625 and 5.21 kg/m^3,respectively.

Enthalpy of vaporisation and free energy of self-solvation

Figure 4.9 shows the enthalpy of vaporisation of ten linear primary amines obtained using the three models after inclusion of polarisation corrections. Concerning the enthalpy of vaporisation, all three models behave very similarly, which is somewhat expected since this property was included in the parameterisation of all models. However, Model 1 has the best performance with an RMSD of 0.883 kJ/mol, followed by Model 3

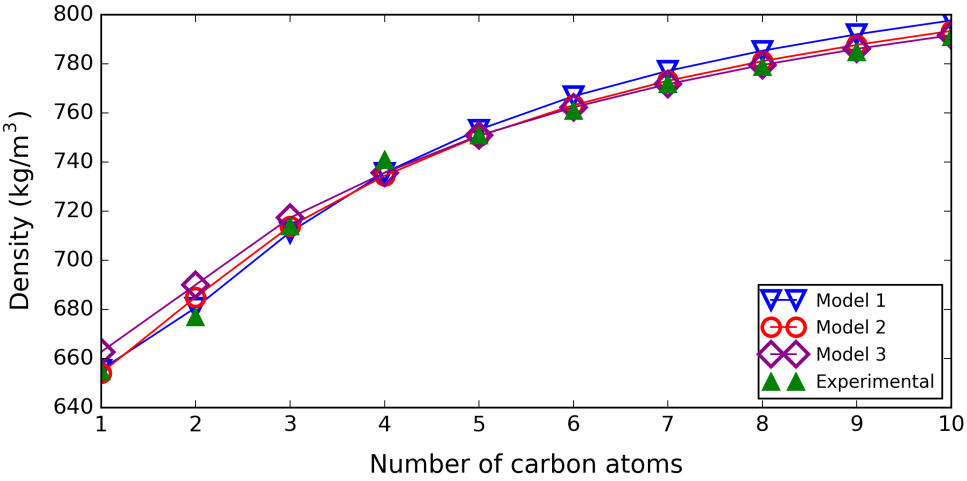

Figure 4.8: Density of linear primary amines at at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (red circles) and Model 3 (purple rhombus). The green symbols are experimental values [224].

with an RMSD of 0.899 kJ/mol and lastly by Model 2 with an RMSD of 1.437 kJ/mol. Model 2 does an excellent job at predicting the enthalpy of vaporisation of small amines (from methylamine to butylamine), but from pentylamine to decylamine it seems to systematically underpredict this property by a small amount (3 % for decylamine). It is also important to mention that no data was found for the experimental error bars of nonylamine and decylamine, and thus they were plotted without error bars. All other molecules have error bars, but some of them are too small to be visible in the plot.

From Figure 4.10 we can see that Model 2 is the best at predicting free energies of self-solvation of linear primary amines. This property is particularly relevant because it was not included in the training set (hence it represents a pure prediction) and because the PolCA model is aimed at accurately predicting solvation properties. This model has an RMSD of 0.730 kJ/mol, and its predicted free energies of solvation fall within the error bars of the experimental values for all ten linear primary amines studied here, except for methylamine and ethylamine. The free energies of self-solvation of these two molecules are overpredicted (more negative values) by 1.3 kJ/mol and 1.1 kJ/mol, respectively. On the other hand, Model 3 has the worst performance out of the three models with an RMSD of 2.942 kJ/mol. This model systematically predicts self-solvation free energies that are approximately 3 kJ/mol more negative than the

experimental values. Finally, Model 1 is better than Model 2 for methylamine and ethylamine, but it slightly underpredicts the magnitude of the self-solvation free energy from butylamine onwards. The RMSD of Model 1 is 1.336 kJ/mol.

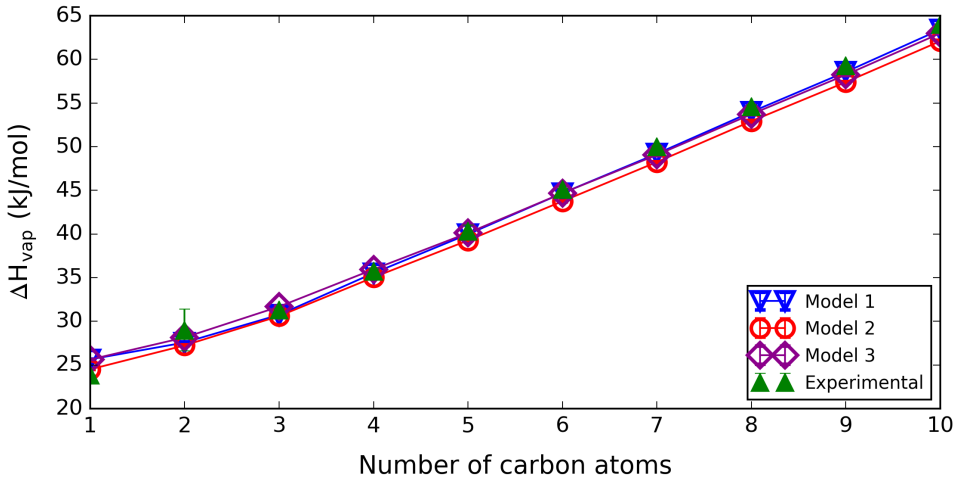

Figure 4.9: Enthalpy of vaporisation of linear primary amines at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (red circles) and Model 3 (purple rhombus). In all cases, polarisation corrections were included. Experimental values (green symbol) were taken from references [194] and [227].

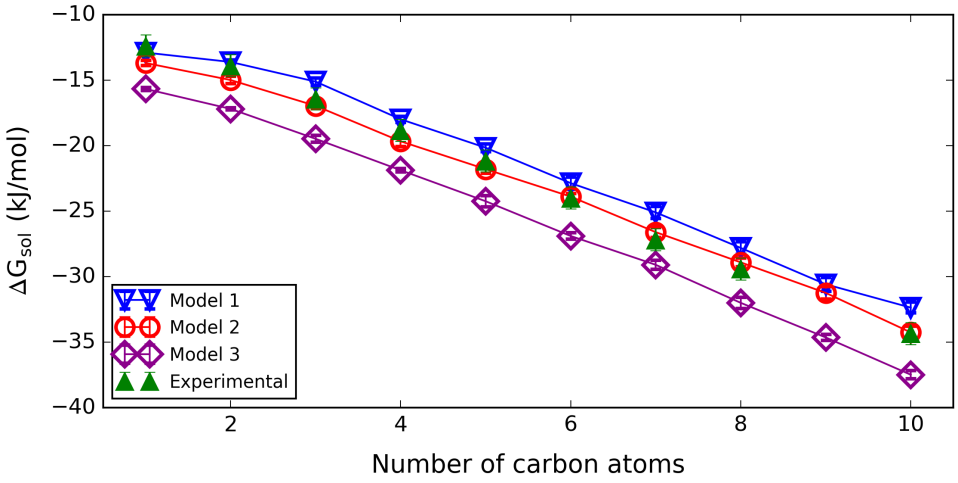

Figure 4.10: Free energy of self-solvation of linear primary amines at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (red circles) and Model 3 (purple rhombus). In all cases, polarisation corrections were included. Experimental values (green symbol) were obtained from vapour pressures data [228, 229].

Dielectric constant

The dielectric constant was one of the properties used to parameterise the three models. Figure 4.11 shows the dielectric constants of methylamine to decylamine obtained using

Model 1, Model 2 and Model 3. All models overpredict methylamine's dielectric constant by more than 60 %, but their performance improves as the alkyl chain increases. Concerning the dielectric constant, Model 1 and Model 3 are similar. However, Model 1 is better with an RMSD of 1.890 compared to an RMSD of 2.096 for Model 3. On the other hand, Model 2 is the worst, with an RMSD of 2.672. This model overpredicts the dielectric constant of all linear primary amines studied here, especially from methylamine to propylamine. From butylamine onwards, the discrepancy between simulated and experimental values is small, and thus, the model does a reasonable job at predicting this property for larger molecules.

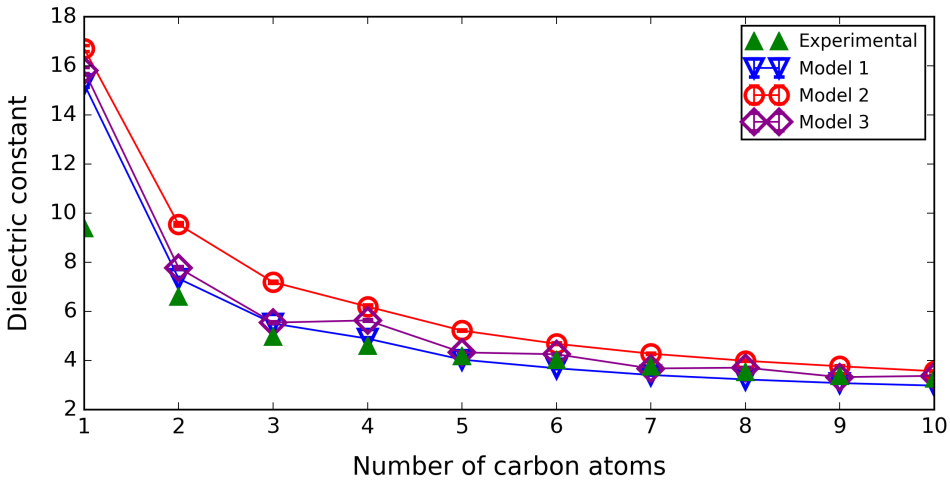

Figure 4.11: Dielectric constant of linear primary amines at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (red circles) and Model 3 (purple rhombus). In all cases, polarisation corrections were applied. Experimental dielectric constants (green symbol) were taken from references [223] and [221].

Another issue with Model 2 is its dipole as can be observed from Figure 4.12. Except for methylamine, simulated dipoles obtained using this model are higher than the estimated liquid dipoles from Table 4.4 (the lack of a clear trend in these values is due to the gas dipoles obtained with Gaussian as shown in Figure A17 in Appendix A5). Model 1 has a dipole moment that is lower than Model 2's but it is still higher than the estimated liquid dipole for most amines. Consequently, Model 3 was optimised forcing the algorithm to only return optimums with a dipole lower than the estimated liquid dipole. Unfortunately, this model fails to predict the free energy of self-solvation, one of the main properties of interest in this work. There are other amine models in the

literature with a dipole larger than the estimated liquid dipoles, like for example AUA4. The AUA4 force field uses the TraPPE-EH charges, reported by Wick et al. [216], and its dipole is approximately 1.8 D [217]. Interestingly, Jorge and Lue [87] analysed 31 alkylamines' dielectric constants obtained using GAFF [103] and OPLS-AA [213], and concluded that 0.88 was the optimal scaling factor between the liquid and model dipoles for that data set. This scaling factor represents the ratio between the dipole moment of the real liquid and that of the molecular model, so a value below 1 suggests that the model dipoles tend to be larger than the liquid dipoles. Taken together, these observations suggest that either Equation 2.49 does not give an accurate estimate of the liquid dipole moments of amines or that a model dipole moment larger than that of the liquid might be necessary to accurately represent the potential energy surface of amine molecules. Further investigation is needed to clarify this.

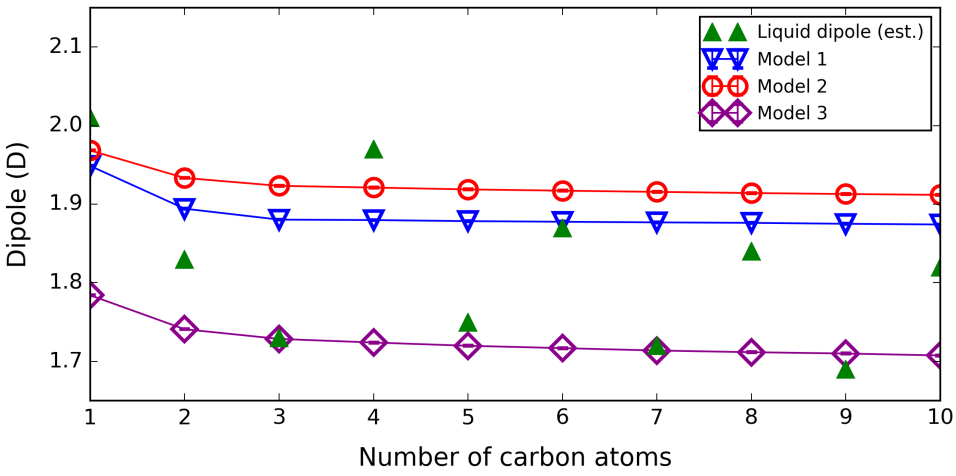

Figure 4.12: Liquid dipole of linear primary amines at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (red circles) and Model 3 (purple rhombus). The green symbols are the liquid dipoles obtained with the equation proposed by Leontyev and Stuchebrukhov (Equation 2.49) [92] (see section 4.2.1).

Free energy of solvation in hexadecane

The free energy of solvation in hexadecane was one of the properties considered before selecting the best model. Like the self-solvation free energy, it represents a pure prediction of the models, and assesses their transferability to a medium with significantly different polarity. Figure 4.13 shows the free energy of solvation of ten linear primary amines in hexadecane obtained using the three models. Looking at this plot,

it is clear that Model 2 can accurately predict this property for all the linear primary amines studied here, while Model 1 systematically underpredicts the magnitude of the solvation free energy in hexadecane by 4 to 5 kJ/mol and Model 3 consistently overpredicts this property's magnitude by 3 to 4 kJ/mol. We note also that the inclusion of polarisation corrections is essential to capture the correct trend of the experimental data (refer to Table 4.4 for values of the corrections).

The RMSDs with respect to experimental data for Model 1, Model 2 and Model 3 are 4.541 kJ/mol, 1.514 kJ/mol and 3.274 kJ/mol; respectively.

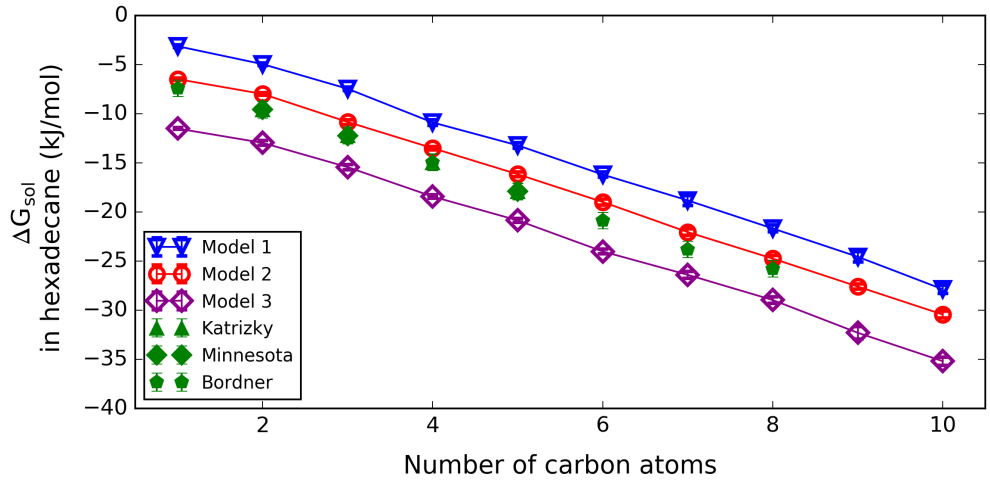

Figure 4.13: Free energy of solvation of linear primary amines in hexadecane at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (red circles) and Model 3 (purple rhombus). Polarisation corrections were applied in all cases. The green symbols are experimental values obtained from the Katrizky and Minnesota databases [168, 169] and reference [230]'s supporting information.

Root-mean-square deviation

To summarise this section, several properties' RMSDs obtained using the different models are shown in Table 4.17. These values were calculated based on the experimental and simulated values of the primary amines studied here.

Overall, Model 2 has the best performance, especially for free energies of solvation. As explained above, Model 1 has the lowest value of ε (0.423 kJ/mol) and Model 3 has the highest value (1.580 kJ/mol), while Model 2's ε falls in between these two values (0.814 kJ/mol). In section 3.2.2, it was shown that this parameter had a high influence in the LJ component of methanol's free energy of self-solvation, and thus, the variation

Table 4.17: RMSD with respect to experimental data for all the models in this chapter. Models 1, 2 and 3 include polarisation corrections in all cases, while Gromos-2016H66's RMSD is presented with and without polarisation corrections. The symbol (C) indicates that corrections have been added.

Force Field	**Density [kg/m^3]**	**ΔH_{vap} [kJ/mol]**	**Dielectric constant**	**ΔG_{sol} [kJ/mol]**	**ΔG_{hex} [kJ/mol]**
Gromos-RF	17.28	1.70	1.44		
Gromos-RF (C)	17.28	1.23	1.09		
Gromos-PME	31.98	2.35	1.46	1.03	3.74
Gromos-PME (C)	31.98	2.84	1.21	1.53	2.15
Model 1	4.98	0.88	1.89	1.34	4.54
Model 2	3.62	1.44	2.67	0.73	1.51
Model 3	5.21	0.90	2.10	2.94	3.27

of ε between the force fields could explain why Model 1 underpredicts the magnitude of the self-solvation free energy and the solvation free energy in hexadecane while Model 3 overpredicts it. Model 2 is better than the other two at predicting liquid densities, free energies of self-solvation and free energies of solvation in hexadecane. Additionally, it performs well when predicting the enthalpies of vaporisation, even though the other two models are slightly better. On the other hand, this model overpredicts dielectric constants of linear primary amines (especially for methylamine to propylamine). However, this is a reasonable price to pay, and thus, Model 2 represents a very good overall force field for primary amines. It is possible that a better performing model could in principle be obtained by lifting some of the constraints enforced above. For example, the models developed here use the same amino group's partial charges for all primary amines for simplicity and transferability, which ignores how the alkyl tail affects the electronic structure of the amino group. This could potentially explain the limitation of the models to accurately predict the dielectric constant of all amines at the same time. It is also possible that a better optimum exists outside the region explored here, although this appears unlikely since the training grids applied here encompass the majority of the corresponding parameters for existing models of amines.

4.3.3 PolCA

As explained in the previous section, Model 2 has the best performance overall, compared to the other two options, and from now on, it will be referred to simply as the

"PolCA model". In this section, PolCA's performance is compared against Gromos-PME for linear primary amines, taking into consideration bulk properties and free energies of solvation. Furthermore, PolCA is tested for several branched primary amines, and lastly, a set of parameters for secondary and tertiary amines is proposed.

Primary amines

PolCA's performance is compared to Gromos-2016H66 simulated using PME for long-range electrostatic interactions and dispersion corrections for LJ interactions (Gromos-2016-PME). See section 4.3.1 to analyse Gromos-2016H66's performance using the same methodology employed in its development.

PolCA does an excellent job overall at predicting liquid densities of linear primary amines at 298 K from methylamine to decylamine (RMSD = 3.625 kg/m^3), as can be seen in Figure 4.14. On the other hand, Gromos-2016-PME fails to predict this property (RMSD = 31.975 kg/m^3), and it highly overpredicts methylamine's density by 11 %. From ethylamine onwards, the densities are overpredicted by a significant amount; however, the relative error between experimental and simulated data decreases as the alkyl chain increases.

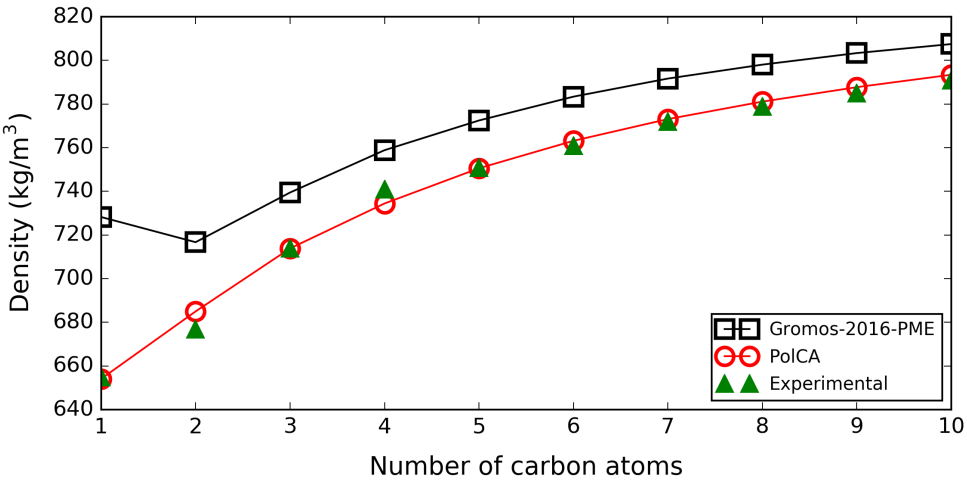

Figure 4.14: Density of linear primary amines at 298.15 K and 1 bar obtained using PolCA (red circles) and Gromos-2016-PME (black squares). The green symbols are experimental values [224].

The opposite behaviour is observed in Figure 4.15. This plot compares the dielectric constant of linear primary amines obtained using PolCA and Gromos-2016-PME

with and without polarisation corrections. Here, Gromos-2016-PME accurately predicts the dielectric constant of ethylamine to decylamine once polarisation corrections are included (RMSD = 1.21), while PolCA significantly overpredicts this property from methylamine to butylamine. PolCA's RMSD is 2.67, and thus, PolCA still is a reasonable model for most amines' dielectric properties, except for methylamine. Methylamine's dielectric constant is overpredicted by 7.3.

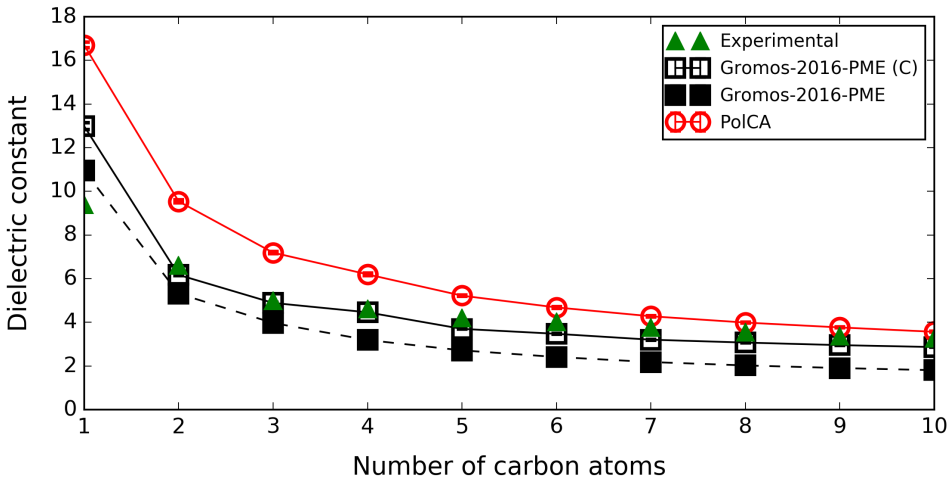

Figure 4.15: Dielectric constant of linear primary amines at 298.15 K and 1 bar obtained with PolCA and GROMOS-2016-PME. Here, the term (C) in the labels for Gromos means that polarisation corrections were included (full lines and empty symbols). Experimental values (green symbol) were obtained from references [223] and [221].

The enthalpy of vaporisation and free energy of self-solvation of linear primary amines obtained using PolCA and Gromos-2016-PME are shown in Figures 4.16 and 4.17, respectively. PolCA accurately predicts the enthalpy of vaporisation of linear primary amines smaller than pentylamine, while from pentylamine onwards this model slightly underpredicts this property. However, PolCA does an overall great job at predicting enthalpies of vaporisation with an RMSD equal to 1.437 kJ/mol. In contrast, Gromos-2016-PME overpredicts the enthalpy of vaporisation of all the amines presented in Figure 4.16, except for ethylamine, and, the discrepancy between experimental and calculated values systematically increases as we move towards larger amines. Adding polarisation corrections to Gromos-2106-PME worsens the model's performance (the RMSD increases from 2.347 kJ/mol to 2.844 kJ/mol). Concerning the free energy of self-solvation, both PolCA (RMSD = 0.730 kJ/mol) and Gromos-2016-PME (RMSD

= 1.032 kJ/mol) can accurately predict the experimental values; however PolCA is the only model that matches decylamine's free energy of self-solvation. The RMSD of Gromos-2016-PME with polarisation corrections is 1.525 kJ/mol.

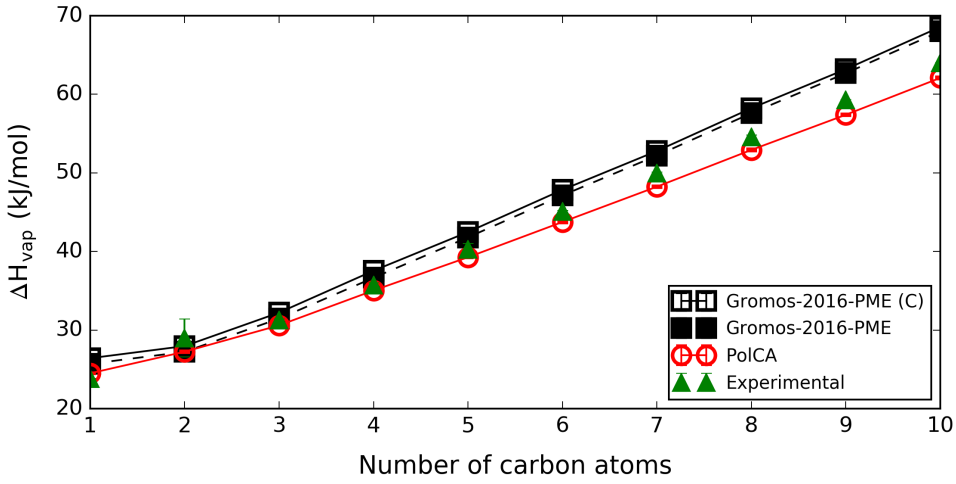

Figure 4.16: Enthalpy of vaporisation at 298.15 K and 1 bar of linear primary amines obtained with PolCA and GROMOS-2016-PME. Here, the term (C) in the labels for Gromos means that polarisation corrections were included (full lines and empty symbols). Experimental values (green symbol) for the enthalpy were taken from references [194] and [227].

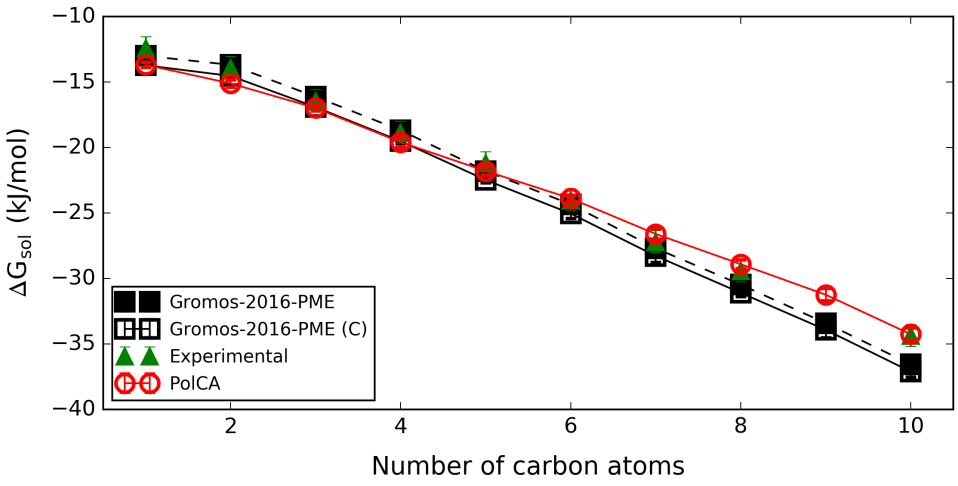

Figure 4.17: Free energy of self-solvation of linear primary amines at 298.15 K and 1 bar obtained using PolCA and GROMOS-2016-PME. Full lines with empty symbols contain polarisation corrections, while dashed lines with filled symbols represent values without corrections. The green symbols are experimental values obtained from vapour pressure data [228, 229].

Figure 4.18 shows the free energy of solvation of linear primary amines in hexadecane obtained using both force fields. PolCA accurately predicts free energies of solvation of primary amines in hexadecane with an RMSD equal to 1.514 kJ/mol, while Gromos-

2016-PME without polarisation corrections highly underpredicts the free energy of solvation in hexadecane of small linear primary amines (RMSD = 3.738 kJ/mol). Once polarisation corrections are added, the model's performance improves significantly, especially for smaller amines, however, its performance is still worse than PolCA's for the amines studied here (RMSD = 2.147).

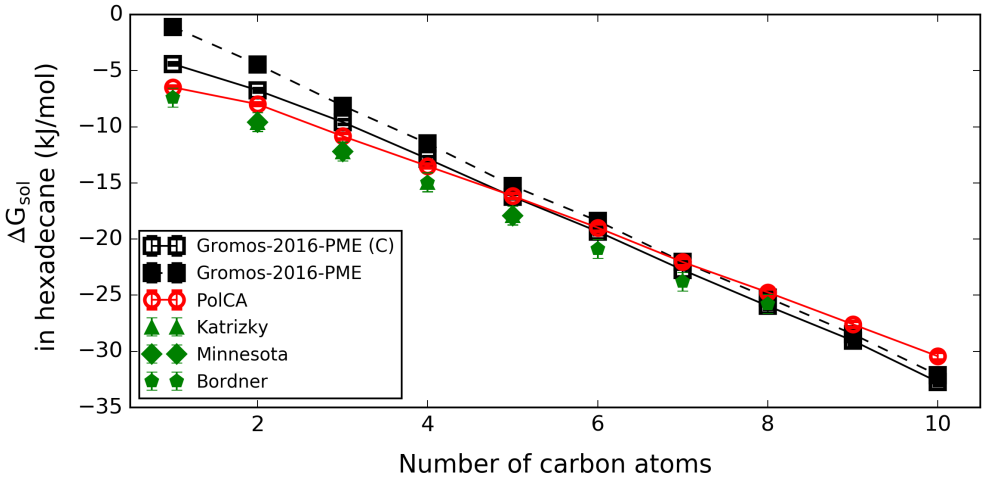

Figure 4.18: Free energy of solvation of linear primary amines in hexadecane at 298.15 K and 1 bar obtained using PolCA and GROMOS-2016-PME. Full lines with empty symbols contain polarisation corrections, while dashed lines with filled black squares represent values without corrections. The green symbols are experimental values obtained from the Katrizky and Minnesota databases [168, 169] and reference [230]'s supporting information.

Branched primary amines

The PolCA force field was also extended to branched primary amines to increase its transferability. Initially, the α-carbon pseudo-atoms' LJ parameters were identical to the corresponding parameters in pure alkanes, resulting in a model that underpredicted the liquid density at 298 K of isopropylamine and tertbutylamine by 6.3 and 14 kg/m^3, respectively. Consequently, in a second attempt, σ was slightly adjusted to account for the replacement of a C-C bond with a shorter C-N by fitting to isopropylamine and tertbutylamine's liquid density at 298 K. We note that there is a precedent for applying such adjustments in the TraPPE force field for alcohols, as we discussed in section 3.3.6. This is because the substitution of a carbon atom by a more electronegative atom (like O or N) leads to a change in the electron cloud of the α-carbon. A straight linear fit between two σ values was sufficient to obtain good predictions since all other parameters were

kept constant and the initial model was not far from the experimental densities.

The density, enthalpy of vaporisation, dielectric constant and free energy of self-solvation of isopropylamine, tertbutylamine, 1-1 dimethylpropylamine, 2-amino-3-methyl pentane and 2-butylamine were simulated using PolCA to validate the model for branched amines. The results are presented in Table 4.18, and agreement was quite satisfactory for all properties.

Table 4.18: Density, enthalpy of vaporisation, dielectric constant and free energy of self-solvation of several branched amines at 298.15 K and 1 bar, obtained using PolCA. Experimental values are in bold, and they were extracted from references [194, 224, 228, 231].

	Density [kg/m^3]		**ΔH_{vap} [kJ/mol]**		**Dielectric constant**		**ΔG_{sol} [kJ/mol]**	
Isopropylamine	685.1	**684**	27.94	**28.71**	6.63	**4.85**	-15.2 +/- 0.1	**-14.74**
Tertbutylamine	688.9	**688**	27.16	**29.80**	5.33	**4.01**	-15.1 +/- 0.4	**-15.30**
1-1-dimethylpropylamine	725.2	**727.6**	31.20	**32.50**	4.75		-17.7 +/- 0.2	**-18.68**
2-amino-3-methyl-pentane	762.2	**750**	38.66	**39.29**	4.16		-21.7 +/- 0.1	
2-butylamine	715.4	**720**	30.99	**32.7**	5.27	**4.51**	-17.6 +/- 0.2	**-17.55**

Secondary and tertiary amines

A new grid was created for dimethylamine and trimethylamine by changing the nitrogen atom's LJ parameters and keeping the partial charges constant. The partial charges of the hydrogen and α-carbon were the same as the ones used for primary amines ($q_C = 0.113$ and $q_H = 0.379$), while the nitrogen's charge was modified to assure the neutrality of the molecule. The nitrogen's partial charge was -0.871 for primary amines, -0.605 for secondary amines and -0.339 for tertiary amines. For comparison, Table 4.19 shows the CHELPG [232] charges for methylamine, dimethylamine and trimethylamine in vacuum, obtained using Gaussian. Here, we can see that nitrogen's partial charge varies the most when moving from primary to tertiary amines, while the hydrogen charge stays practically constant and the α-carbon is only slightly affected by the degree of substitution. This outcome suggests that changing nitrogen's charge while keeping the other charges constant is a physical realistic approximation.

Table 4.20 shows the σ and ε values used to create the meta-models for dimethylamine and trimethylamine, and the properties used in the optimisation were the liquid density, enthalpy of vaporisation and dielectric constant with equal weight. The centre points

Table 4.19: CHELPG charges obtained using Gaussian 9 for methylamine, dimethylamine and trimethylamine in vacuum.

	Methylamine	**Dimethylamine**	**Trimethylamine**
Carbon	0.26	0.18	0.119
Nitrogen	-0.9	-0.69	-0.357
Hydrogen	0.32	0.33	

of the grids were used as the initial guess for both molecules, and the optimisation was forced to stay inside the boundaries.

Table 4.20: Parameter levels used to create the meta-models for non-primary amines.

	For dimethylamine
σ [nm]	0.352, 0.356, 0.360, 0.364, 0.368, 0.372, 0.376, 0.386 and 0.396
ε [kJ/mol]	0.323, 0.423, 0.523, 0.623 and 0.723
	For trimethylamine
σ [nm]	0.356, 0.360, 0.364, 0.368, 0.372, 0.376, 0.475, 0.5 and 0.525
ε [kJ/mol]	0.023, 0.123 and 0.223

The optimised parameters for the nitrogen atom were σ = 0.389 nm and ε = 0.723 kJ/mol for dimethylamine and σ = 0.4918 and ε = 0.151 kJ/mol for trimethylamine (Table 4.6). These parameters were used to predict bulk properties and free energies of solvation of two other amines (diethylamine and triethylamine) to validate their transferability, and the calculated properties are presented in Table 4.21. PolCA does a great job at predicting the properties used in the fitting. However, it overpredicts the magnitude of free energies of self-solvation by 2.5 kJ/mol and free energies of solvation in hexadecane by 3.1 kJ/mol for both secondary and tertiary amines. Furthermore, this force field overpredicts the densities of diethylamine and trimethylamine by 3 % and 5.6 %, respectively, and it overestimates triethylamine's enthalpy of vaporisation by approximately 3 kJ/mol. To summarise, PolCA does a reasonable job at predicting bulk properties and free energies of solvation of secondary and tertiary amines, but it can potentially be improved if different partial charges for the amino group are used for each amine type. For example, TraPPE-EH uses a common partial charge for all types of α-carbons, but the partial charges for the amino group differ for each amine type. The TraPPE-EH's LJ parameters and partial charges for the amine nitrogens were fitted to reproduce the vapour-liquid coexistence curves of methylamine, dimethylamine and trimethylamine. Their fitting procedure started with trimethylamine to determine the

α-carbon's partial charge, which is equal to -1/3 of the nitrogen's charge due to the molecule's neutrality. [216]. Other options to improve PolCA's accuracy would be to extend the learning set or to use more molecules during the fitting; however, this was tried for secondary amines without success.

Table 4.21: Density, enthalpy of vaporisation, dielectric constant and free energy of solvation in themselves and hexadecane of dimethylamine, diethylamine, trimethylamine and triethylamine at 298.15 K and 1 bar, obtained using PolCA. Experimental values are in bold, and they were extracted from references [194, 224, 228], except for the liquid dipoles which are estimations obtained using Equation 2.49.

	Dimethylamine	**Diethylamine**	**Trimethylamine**	**Triethylamine**
Density	649.322	723.633	628.102	763.358
[kg/m^3]	***650***	***702***	***629***	***724***
ΔH_{vap}	24.079	32.379	22.256	37.795
[kJ/mol]	***25.400***	***31.470***	***22.180***	***34.810***
Dielectric	6.179	4.359	2.387	2.354
constant	***5.26***	***3.9***	***2.44***	***2.4***
ΔG_{sol}	-15.3 +/- 0.2		-14.4 +/- 0.2	
[kJ/mol]	**-12.9 +/- 0.8**		**-11.7 +/- 0.8**	
ΔG_{hex}	-12.2 +/- 0.3		-12.6 +/- 0.3	
[kJ/mol]	**-9.1 +/- 0.8**		**-9.3 +/- 0.8**	
Liq. dipole (D)	1.532	1.517	0.874	
	1.48	***1.28***	***0.798***	

4.4 Conclusions

In this chapter, the PolCA model has been extended to include primary amines. PolCA can accurately predict the density of linear and branched primary amines, which is expected since this was one of the properties used in the parameterisation. The largest relative error for the molecules studied in this work was 1.62 % which corresponds to 2-amino-3-methyl pentane. Furthermore, it does a very good job at predicting enthalpies of vaporisation and free energies of self-solvation of linear primary amines with RMSDs of 1.44 kJ/mol and 0.73 kJ/mol compared to experimental data, respectively, and a similar performance is observed for branched primary amines. It is important to mention that the enthalpy of vaporisation was included in the parameterisation of linear primary amines. Meanwhile, the free energy of self-solvation was not included in the optimisation algorithm, however, it was taken into consideration when choosing PolCA from the other two potential models (Model 1 and Model 3).

Furthermore, free energies of solvation in hexadecane were calculated for linear primary amines, and in all cases, there was a very good agreement between calculated and experimental values (RMSD = 1.51 kJ/mol). It is important to remember that PolCA uses *post-facto* polarisation corrections for the free energy of solvation in hexadecane and these corrections range from 3.3 kJ/mol (methylamine) to 0.58 kJ/mol (nonylamine). The positive results obtained here once again reinforce the importance of adding *post-facto* polarisation corrections to non-polarisable force fields.

On the other hand, PolCA greatly overpredicts the dielectric constant of methylamine and significantly overpredicts this property for molecules smaller than pentylamine. The dielectric constant was part of the parameterisation; however, it was decided that this property had to be compromised in order to achieve a better overall model performance. Concerning branched amines, the dielectric constant is also overpredicted for the three molecules with available experimental data.

Lastly, parameters for secondary and tertiary amines have been proposed as a starting point for a future parameterisation. These parameters do a reasonable job at predicting bulk properties and free energies of solvation of secondary and tertiary amines, but they can be improved if each amine type has its own independent partial charges for the amino group.

Chapter 5

Polarisation-Consistent United Atom Force Field for Ketones

5.1 Introduction

Ketones consist of a carbonyl group (C=O) bonded to two carbon atoms, and they can be obtained by oxidation of secondary alcohols [233]. One of the most well-known ketones is propanone, also known as acetone, which is widely used as a solvent for fats, oils, waxes, resins, rubber, plastics, lacquers, varnishes, and rubber cements. This compound dissolves in water in all proportions since the carbonyl oxygen's lone pair electrons can act as a hydrogen bond acceptor, and thus, it forms hydrogen bonds with water. Another excellent solvent for many organic compounds is 2-butanone, also known as methyl ethyl ketone (MEK). [233]. Some ketones, like 2-nonanone and 2-undecanone, are found in essential oils [234], and many compounds containing a carbonyl group are relevant to the pharmaceutical industry like the oral contraceptive norethindrone [233]. Thus, it is important to be able to accurately predict solvation properties of molecules containing ketone groups.

As explained in previous chapters, molecular dynamics simulations are extremely useful since they can predict many compounds' properties saving time and money when designing an industrial process. However, their accuracy depends on the set of parameters

and equations used during their calculation, and thus, several important force fields for ketones are mentioned below as an example.

The united-atom OPLS force field for acetone [235] was developed by Jorgensen et al. in 1989 to reproduce thermodynamic and structural properties of pure liquid acetone, and the authors showed that it does a good job at predicting the free energy change obtained when mutating acetic acid to acetone in water and chloroform. Later on, Jorgensen et al. developed the OPLS all-atom (AA) force field for many common organic groups, and they included propanone and butanone in its development [139]. Bonded parameters were taken from the AMBER-AA force field [212] or determined by fitting to rotational energy profiles. Meanwhile, non-bonded parameters were obtained by fitting to densities and enthalpies of vaporisation of 34 organic liquids at 298.15 K, except for a few molecules that were simulated at their boiling points or 398.15 K

In general, OPLS becomes less accurate for conditions further away from 298.15 K and 1 atm, as has been shown by Martin and Siepmann [102]. Consequently, those authors developed the TraPPE-UA force field for n-alkanes [102] to better reproduce critical temperatures and saturated liquid densities. Later on, Stubbs et al. extended the TraPPE-UA force field to ketones by fitting the carbonyl carbon's LJ parameters to the vapour-liquid coexistence curve of acetone [146]. To keep the number of fitting parameters at a manageable level, the carbonyl group's partial charges were taken from the OPLS-UA force field [235], while the LJ parameters for the carbonyl oxygen were taken from the TraPPE force field for carbon dioxide [236].

Another UA force field for ketones focused on vapour-liquid equilibria is the anisotropic united-atom force field developed by Kranias et al. [237] in 2003. This model is an extension of the AUA4 model by Ungerer et al. [218], with partial charges from ab initio calculations and new LJ parameters for the carbon and oxygen atom of the ketone group. The LJ parameters of these two atoms were fitted to reproduce the vapour pressures, enthalpies of vaporisation and liquid densities of propanone and butanone at ambient conditions and at a reduced temperature of 0.8. This force field yields good agreement between experimental and simulated data for the coexistence curves of acetone, butanone and 2-pentanone, and it does a very good job at predicting the

enthalpies of vaporisation of these molecules from 250 to 500 K.

Other important force fields for ketones are AMBER [212] and GROMOS-53A6 [101]. The latter was developed in 2004 to reproduce free energies of solvation in water and non-polar solvents, and in 2011, Horta et al. [77] proposed new interaction parameters for ketones by fitting to the liquid density, enthalpy of vaporisation, hydration free energy and solvation free energy in cyclohexane of propanone and butanone. Meanwhile, 2-pentanone, 3-pentanone, 2-hexanone and 3-hexanone were part of the validation stage.

In this chapter, the PolCA force field is extended to include ketones, and to the best of our knowledge, this is the first ketone force field developed using polarisation corrections. The first section contains details about the parameterisation and the polarisation corrections used in this chapter. Section 5.3.1 shows how using a force field with PolCA parameters for the alkyl chain and TraPPE's parameters for the ketone group results in liquid densities that are underpredicted. Consequently, the ketone group's non-bonded parameters were optimised using three different approaches and their performances are shown in section 5.3.2. Lastly, the PolCA model for ketones is compared to TraPPE-UA in section 5.3.3. Preliminary results for the ketone PolCA force field's parameterisation were carried out by a Masters student, Jordan Cree, under direct supervision from the author of this thesis. The precise contributions from the student are explicitly mentioned in the sections below.

5.2 Methodology

The optimisation procedure, as well as the methods used to obtain bulk properties and free energies of solvation, are described in detail in Chapter 2. Specific details for this chapter are described below.

Bulk properties

Table 5.1 shows the number of molecules in each simulation box, which was selected to maintain an approximately constant box size of 27 nm^3.

Furthermore, the finite-size correction terms for the diffusion coefficients of the ketones

Table 5.1: Number of molecules in the simulation box for each single-component system.

Compound	**Number of molecules**
Propanone	433
Butanone	328
Pentanone	261
Hexanone	213
Heptanone	168
Octanone	151
Nonanone	127
Decanone	121

studied here are shown in Table 5.2. These corrections take into account the relationship between system size and simulated diffusion (see section 2.7.3 for more details), and they were assumed to be the same as those for TraPPE-UA during the parameterisation and validation stages. This assumption was proven to be valid for linear alcohols, as shown in section 3.2

Table 5.2: Correction terms for the diffusion constant in 10^{-5} cm^2/s obtained using the TraPPE force field. These values were calculated by Jordan Cree during his Master's project.

	TraPPE-UA
Propanone	0.486
Butanone	0.466
Pentanone	0.360
Hexanone	0.328
Heptanone	0.303
Octanone	0.238
Nonanone	0.239
Decanone	0.170

Free energy calculations

The free energies of solvation calculated in this chapter were obtained using 14 λ-states for the LJ component (0, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.7, 0.85 and 1) and 5 λ-states for the electrostatics (0, 0.2, 0.4, 0.7 and 1). The LJ contribution to the solvation free energy was simulated for 5 ns while 10 ns were used for the electrostatic component. The λ-states were selected based on their relative entropies to assure a good degree of overlap between them, as described in section 2.8.2, and the simulation times were chosen by plotting free energies of solvation against simulation times, as has been done for alcohols (section 3.2) and amines (section 4.2). The plots for the LJ and

electrostatic components of decanone can be found in Figures A3 and A4 in Appendix A1.

5.2.1 Polarisation corrections

As explained in section 2.7.2 of the methodology, and following the same procedure used in Chapters 3 and 4, liquid phase dipole moments for the ketones studied in this chapter were estimated using Equation 2.49. Table 5.3 shows the estimated liquid dipoles and the *post-facto* polarization corrections (C_{pol}) used to obtain enthalpies of vaporisation. These corrections are meant to be used with enthalpies of vaporisation (liquid to gas transition), and they need to be multiplied by -1 before using them to correct free energies of solvation (gas to liquid transition).

Table 5.3: Polarisation corrections for a liquid-gas phase transition (C_{pol}), expressed in kJ/mol. C_{dist} and C_{elec} are the distortion and electronic polarisation terms, respectively, α [224] is the polarizability of the solute in the gas phase and ϵ_{el} [224] and ϵ_{sol} are the high-frequency dielectric permittivity and the static dielectric constant, respectively, of the medium.

Solute	**Solvent**	μ_g **(D)**	α ($\AA^3$)	ϵ_{sol}	ϵ_{el}	μ_l **(D)**	C_{dist}	C_{elec}	C_{pol}
propanone	propanone	2.88	6.40	20.49	1.85	4.72	-15.98	15.91	-0.07
butanone	butanone	2.78	8.22	18.50	1.91	4.70	-13.45	13.48	0.03
2-pentanone	2-pentanone	2.70	10.05	15.45	1.93	4.58	-10.56	10.90	0.33
2-hexanone	2-hexanone	2.66	11.89	14.60	1.96	4.57	-9.22	9.57	0.36
2-heptanone	2-heptanone	2.59	13.73	11.66	1.98	4.42	-7.33	8.01	0.68
2-octanone	2-octanone	2.70	15.56	10.30	2.00	4.59	-6.92	7.81	0.90
2-nonanone	2-nonanone	2.73	17.42	9.14	2.02	4.61	-6.12	7.18	1.05
2-decanone	2-decanone	2.55	19.25	8.30	2.03	4.29	-4.72	5.72	1.00
propanone	hexadecane	2.88	6.40	2.09	2.05	3.64	-2.72	12.76	10.04
butanone	hexadecane	2.78	8.22	2.09	2.05	3.51	-1.97	9.25	7.28
2-pentanone	hexadecane	2.70	10.05	2.09	2.05	3.41	-1.52	7.15	5.62
2-hexanone	hexadecane	2.66	11.89	2.09	2.05	3.36	-1.25	5.86	4.61
2-heptanone	hexadecane	2.59	13.73	2.09	2.05	3.27	-1.03	4.81	3.79
2-octanone	hexadecane	2.70	15.56	2.09	2.05	3.41	-0.98	4.61	3.63
2-nonanone	hexadecane	2.73	17.42	2.09	2.05	3.45	-0.90	4.21	3.32
2-decanone	hexadecane	2.55	19.25	2.09	2.05	3.22	-0.71	3.33	2.62

The polarisation corrections for pure ketones presented in Table 5.3 are small (lower or equal to 1.05 kJ/mol), and the distortion and electronic contributions nearly cancel each other out from propanone to hexanone. The same effect has been observed for water [92, 108], alcohols (Table 3.4) and amines (Table 4.4). In contrast, the net corrections for ketones in hexadecane are quite significant, ranging from 10.04 kJ/mol for propanone to 2.62 kJ/mol for 2-decanone. Nonetheless, it is important to remind

the reader that these corrections have been obtained using several assumptions and simplifications, and due to lack of experimental data, we cannot guarantee that the equation used to obtain the liquid dipoles (Equation 2.49) is accurate for ketones.

5.2.2 Optimisation

The procedure used to develop the PolCA force field is explained in this section. Several approaches were attempted before fine-tuning the optimisation protocol, and thus, three models, which from now on are called Model 1, Model 2 and Model 3, are presented here for comparison.

Following the same methodology used in Chapter 3 and Chapter 4, meta-models were created for propanone, 2-hexanone and 2-decanone using simulations with different non-bonded parameters for the oxygen atom (the partial charge of the α-carbon was modified accordingly to maintain the molecule's neutrality). The alkyl chain and α-carbon parameters [1] were kept constant in all the simulations, and they were the same as those used in our PolCA model (see Table 5.4). These meta-models include *post-facto* polarisation corrections, and they can predict how different target properties change when the oxygen's ε, σ and partial charge are modified. In addition, Model 2 and Model 3 use partial charges on the pseudo-atoms that are bonded to α-carbons, which were obtained by scaling partial charges from ab initio calculations (see below for details).

Table 5.4: Lennard-Jones parameters and partial charges for the new PolCA united-atom force field for ketones.

non-bonded [a] (eq 2.37)	σ [nm]	ε [kJ/mol]	Partial charge (q)
C(=O)	0.382	0.333	0.709
O (sp^2)	0.300	0.200	-0.609
(**CH_3**)-CH_x	0.379	0.833	0
(**CH_3**)-C=O	0.379	0.833	-0.05
$(CH_x)_2$-(**CH_2**)	0.399	0.392	0
(CH_x)-(**CH_2**)-C=O	0.399	0.392	-0.05
$(CH_x)_3$-(**CH**)	0.473	0.085	0
$(CH_x)_2$-(**CH**)-C=O	0.473	0.085	-0.05
$(CH_x)_4$-(**C**)	0.646	0.00426	0
$(CH_x)_3$-(**C**)-C=O	0.646	0.00426	-0.05

a) The non-bonded parameters correspond to the sites in bold.

The learning sets used to create the meta-models differed for each model, and they are

explained in detail below. Also, the fitting properties and, consequently, the objective functions were different for each approach. Equations 5.1, 5.2 and 5.3 are the objective functions for Model 1, Model 2 and Model 3, respectively:

$$F(X) = \sum_{j=1}^{3} \left((f_k(X) - y_{exp})^2_{\rho_j} + (f_k(X) - y_{exp})^2_{D_{A_j}} + (f_k(X) - y_{exp})^2_{\Delta H_j} \right) \quad (5.1)$$

$$F(X) = \sum_{j=1}^{3} \left((f_k(X) - y_{exp})^2_{\rho_j} + (f_k(X) - y_{exp})^2_{D_{A_j}} + (f_k(X) - y_{exp})^2_{\Delta H_j} \right) + (f_k(X) - y_{exp})^2_{\Delta G_P} \quad (5.2)$$

$$F(X) = \sum_{j=1}^{3} \left((f_k(X) - y_{exp})^2_{\rho_j} + (f_k(X) - y_{exp})^2_{D_{A_j}} + (f_k(X) - y_{exp})^2_{\Delta H_j} \right) + (f_k(X) - y_{exp})^2_{\Delta G_P} + (f_k(X) - y_{exp})^2_{\Delta G_D} \quad (5.3)$$

where j=1 corresponds to propanone, j=2 to hexanone and j=3 to decanone. Additionally, $f_k(X)$ is the value predicted using the meta-model at X, y_{exp} is the experimental value and ρ, D_A and ΔH represent the density, self-diffusion coefficient and enthalpy of vaporisation, respectively. ΔG_P and ΔG_D are the free energies of self-solvation of propanone and decanone, respectively. Therefore, in Model 1 only the three base properties were included in the optimisation; Model 2 also includes the solvation free energy of propanone, while Model 3 incorporates propanone and decanone's self-solvation free energies.

Furthermore, the meta-models accuracy was checked by plotting simulated versus predicted values, and the predicted properties for the models explained in this work were in perfect agreement with the simulations. The meta-models performances for each model are described in more detail below.

Sensitivity analysis

The sensitivity analysis was carried out by Jordan Cree as part of his Master's project. Propanone's σ, ε and partial charges of the oxygen and α-carbon were modified by +/- 10 %, while keeping all other parameters constant, to see their effect on the density, enthalpy of vaporisation and dielectric constant (Table 5.5). The partial charges for the oxygen and α-carbon have the same magnitude, but opposite sign, and thus, only oxygen's partial charge appears in Table 5.5. A modified TraPPE force field was used for this analysis (TraPPE-UA [146] with improved parameters for the alkyl chain, proposed by Jorge [1]).

Table 5.5: Sensitivity analysis of σ, ϵ and partial charges of the oxygen and α-carbon of propanone. TraPPE-UA [146] with improved parameters for the alkyl chain, proposed by Jorge [1], was used for this study. The symbols -ve and +ve indicate that the original non-bonded parameter has been decreased or increased by 10 %, respectively.

	σ_C	ε_C	σ_O	ε_O	q_O	**Density [kg/m^3]**	**ΔH [kJ/mol]**	**Dielectric Constant**
Original	*0.382*	*0.333*	*0.305*	*0.657*	*-0.424*	*767*	*26.68*	*8.9*
σ_C -ve	**0.343**	0.333	0.305	0.657	-0.424	-1.30 %	-7.81 %	-3.60 %
σ_C +ve	**0.420**	0.333	0.305	0.657	-0.424	-1.04 %	+5.85 %	+6.97 %
ε_C -ve	0.382	**0.299**	0.305	0.657	-0.424	-0.91 %	-2.89 %	+2.02 %
ε_C +ve	0.382	**0.367**	0.305	0.657	-0.424	+0.78 %	-2.51 %	-0.56 %
σ_O -ve	0.382	0.333	**0.275**	0.657	-0.424	+3.00 %	+0.75 %	+5.62 %
σ_O +ve	0.382	0.333	**0.336**	0.657	-0.424	-3.78 %	-0.75 %	-0.45 %
ε_O -ve	0.382	0.333	0.305	**0.591**	-0.424	-0.65 %	-2.36 %	+3.93 %
ε_O +ve	0.382	0.333	0.305	**0.723**	-0.424	+0.52 %	+2.10 %	-0.34 %
q_O -ve	0.382	0.333	0.305	0.657	**-0.466**	+1.3 %	+6.6 %	+28.65 %
q_O +ve	0.382	0.333	0.305	0.657	**-0.424**	-1.17 %	-5.77 %	-19.45 %

An unexpected behaviour is observed in Table 5.5 for the α-carbon's σ and ε. Both increasing and decreasing σ reduces the density by approximately 1 %, while the enthalpy of vaporisation is decreased by approximately 2.7 % when ε is modified in either direction. This behaviour is likely due to the α-carbon being almost buried by the surrounding atoms. On the other hand, altering the oxygen's parameters produced the expected behaviour making it a good choice as a fitting parameter. The density decreases when either σ increases or ε decreases. Increasing σ increases the excluded volume of each molecule while decreasing ε reduces the cohesive energy of the liquid. The enthalpy of vaporisation highly depends on the interaction between molecules, and thus, it significantly decreases when either ε or the absolute value of the partial charge

are reduced (these actions reduce the dispersion and electrostatic forces, respectively). Lastly, the dielectric constant is an electric property, therefore, it mainly depends on the partial charges. Consequently, only oxygen's non-bonded parameters were changed during the parameterisation.

Model 1

This model was developed by Jordan Cree and was obtained by fitting σ, ε and the partial charge of the oxygen atom to the densities, enthalpies of vaporisation and diffusion constants of propanone, 2-hexanone and 2-decanone (Equation 5.1). The parameter levels shown in Table 5.6 were used to create a full factorial grid, and thus, this model's learning set consisted of 27 simulations per molecule. The initial guess and the central point used to create the coded values was σ = 0.305 nm, ε = 0.721 kJ/mol and q_O = -0.462.

Table 5.6: Parameter levels used in the parameterisation of Model 1.

	Levels
σ [nm]	0.2745, 0.305 and 0.3355
ε [kJ/mol]	0.591, 0.657 and 0.850
q_O	-0.5, -0.466 and -0.424

The meta-models' performance can be found in Appendix A2 (Figure A14), and in all cases predicted and simulated values are in good agreement. Model 1's optimum is inside the learning set's boundaries, and the predicted values match those obtained with the simulations.

As shown below (section 5.3.2), Model 1 could not predict all thermodynamic properties of ketones to a good enough degree of accuracy. One possible reason is that the charge distribution on the molecule was overly simplified. To test this hypothesis, it was decided to add partial charges to pseudo-atoms that were bonded to α -carbons, as has been done for a molecule of propanone by Kamath et al. [238]. Consequently, Model 2 and Model 3 were developed using QM partial charges for propanone which were allowed to vary in a proportional way through a scaling factor α during the optimisation routine. A value of α = 1 corresponds to the partial charges depicted in Figure 5.1 calculated with Gaussian G09W [220] using the B3LYP functional [239] and the aug-

cc-pvtz basis set [240] which includes diffuse functions. During the QM calculation, propanone was solvated in itself using a polarisable continuum model (PCM) [36] and the atomic charges were produced by fitting to the electrostatic potential at points selected according to the CHelpG scheme [232].

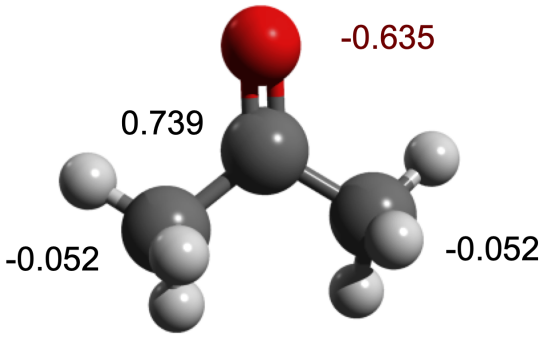

Figure 5.1: Partial charges for solvated propanone in itself obtained using a polarisable continuum model with Gaussian G09W. The charges were obtained using the CHelpG scheme.

Model 2 and Model 3

Model 2's parameters were fitted to the densities, enthalpies of vaporisation and diffusion constants of propanone, 2-hexanone and 2-decanone and propanone's self-solvation free energy (Equation 5.2). Meanwhile, Model 3 includes all these properties plus the free energy of self-solvation of decanone in its objective function (Equation 5.3). Both models used the same meta-models during their parameterisation, and their parameter levels are shown in Table 5.7. The meta-model for 2-decanone's free energy of self-solvation was first introduced in Model 3, and thus, it does not apply to Model 2. The initial guess and the central point used to create the coded values was $\sigma = 0.305$ nm $\varepsilon = 0.657$ kJ/mol and $\alpha = 1$, for both models.

Table 5.7: Parameter levels used in the parameterisation of Model 2 and Model 3.

	Bulk properties
	Levels
σ [nm]	0.290,0.295,0.3,0.305,0.31 and 0.315
ε [kJ/mol]	0.200, 0.400, 0.520, 0.656, 0.750 and 0.850
α	0.75, 0.95, 1 and 1.05
	Free energies of self-solvation
	Levels
σ [nm]	0.290, 0.3 and 0.305
ε [kJ/mol]	0.400, 0.657 and 0.85
α	0.95, 1 and 1.05

During Model 3's development, the optimisation algorithm was forced to stay inside the boundaries of the most extensive grid, while no boundary restrictions were applied to Model 2. Nevertheless, Model 2's optimum fell inside the grid's boundaries. The meta-models performances inside their respective boundaries can be found in Appendix A2 (Figures A15 and A16), and, in all cases, the meta-models can accurately predict the target properties. It is worth mentioning that Model 3's ε was outside the meta-models' boundaries for the free energies of self-solvation. Nonetheless, the predicted values were consistent with the simulated results obtained subsequently.

Other models

Other methods of optimisation were attempted without success. One of these approaches was to include the dielectric constant in the parameterisation using the same learning set as Model 2 and Model 3. However, the predicted optimum had an α value equal to 0.58, and thus, a model's dipole lower than the gas-phase dipoles. Jordan Cree also attempted to include the dielectric constant, as well as the other three bulk properties, in the parameterisation with similar outcomes. His dipole moment was significantly lower than the gas dipoles (1.79 compared to 2.66 for 2-hexanone), which does not make physical sense. Hence, these models were not considered further.

Another technique explored was fitting to the dielectric constant using variable polarisation corrections instead of those shown in Table 3.2. Liquid dipoles were calculated based on the best scaling factor for the dielectric constant, in an iterative manner, and then, these dipoles were substituted into Equation 2.47 to calculate the correc-

tions. The resulting model systematically overpredicted the enthalpies of vaporisation and self-solvation free energies, and thus, it was also discarded from further consideration.

5.3 Results and discussion

Bulk properties and free energies of self-solvation of TraPPE and modified-TraPPE were simulated by Jordan Cree under our supervision as part of his Master's project. Additionally, Model 1 was developed by him.

Several force fields are discussed in this section, and thus, a summary table showing all these models (Table 5.8) is included here to improve the readability of the chapter.

Table 5.8: Force fields simulated in this chapter. Model 1, Model 2 and Model 3 use the LJ parameters proposed by Jorge [1] for the alkyl chain and the α-carbon, and have been parameterised taking into account post-facto polarisation corrections. Unless stated otherwise their results always include these corrections.

TraPPE	TraPPE-UA force field for ketones [146]. This model has been fitted to the vapour-liquid coexistence curve of acetone.
TraPPE (C)	TraPPE-UA force field for ketones with *post-facto* polarisation corrections.
Modified TraPPE	Combines TraPPE-UA's parameters for the carbonyl group with the set of parameters for alkane groups proposed by Jorge [1]
Model 1	• Fitted to the density, enthalpy of vaporisation and diffusion constant of propanone, 2-hexanone and 2-decanone. • Parameters optimised: LJ parameters (σ and ε) and partial charge (q_O) of the oxygen atom.
Model 2	• Fitted to propanone's self-solvation free energy and to the density, enthalpy of vaporisation and diffusion constant of propanone, 2-hexanone and 2-decanone. • Partial charges were obtained using Gaussian 09 and then scaled using an scaling factor α. • Parameters optimised: σ and ε of the oxygen atom and the scaling factor α.
Model 3	• Fitted to propanone and decanone's self-solvation free energies and to the density, enthalpy of vaporisation and diffusion constant of propanone, 2-hexanone and 2-decanone. • Partial charges were obtained using Gaussian 09 and then scaled using an scaling factor α. • Parameters optimised: σ and ε of the oxygen atom and the scaling factor α.

5.3.1 Modified TraPPE

As a preliminary step, as done previously for alcohols, the densities of propanone to 2-decanone were calculated using a modified TraPPE force field that combines TraPPE-UA's parameters for the carbonyl group with the set of parameters for alkane groups proposed by Jorge [1]. The latter parameters were designed to fit the density, enthalpy of vaporisation and free energy of solvation of alkanes [1].

It can be seen from Figure 5.2 that this model considerably underpredicts the density of these molecules. However, non-polar interactions become more important as the alkyl chain increases, and thus, the deviation between simulated and experimental values decreases when the number of carbon atoms increases since the molecules become more similar to alkanes. Consequently, an optimisation procedure was carried out to find a new set of LJ parameters and partial charges for the oxygen atom, as described in section 5.2.2. The LJ parameters of the α-carbon were kept unchanged based on the sensitivity analysis results carried out using this force field (see section 5.2.2).

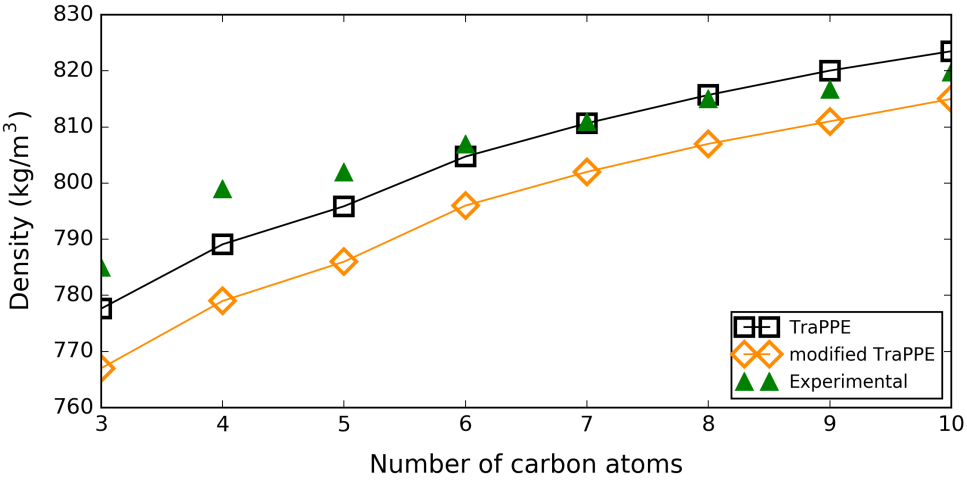

Figure 5.2: Densities of linear ketones at 298.15 K and 1 bar obtained using the modified TraPPE model (orange diamonds) and the original TraPPE force field [146] (black squares). The green symbols are experimental values and were obtained from references [224, 241].

5.3.2 Comparison between all the new models

This section analyses the performance of the three potential models described in section 5.2.2. The properties studied here are the density, enthalpy of vaporisation, di-

electric constant, and free energy of self-solvation.

The models' parameters are shown in Table 5.9. All models have similar values of σ, including TraPPE. On the other hand, ε and the partial charges significantly differ for all the force fields. Model 3 has the lowest ε (0.2 kJ/mol), and thus, the largest magnitude of the charges to compensate for the lower dispersion forces. Meanwhile, Model 1 and Model 2 have the largest values of ε (0.805 and 0.861 kJ/mol, respectively), and their charges are closer to TraPPE's. TraPPE's LJ parameters are in between Model 2 and Model 3's parameters, however, its partial charges are the lowest and hence, also its dipole moment.

Table 5.9: Force field parameters for TraPPE-UA and the different models obtained in this chapter.

Model	**σ [nm]**	**ε [kJ/mol]**	$\mathbf{q_O}$	$\mathbf{q_C}$
Model 1	0.303	0.805	-0.461	0.461
Model 2	0.311	0.861	-0.498	0.580
Model 3	0.300	0.200	-0.609	0.709
TraPPE	0.305	0.657	-0.424	0.424

Density and diffusion

Figure 5.3 shows the simulated densities of propanone to 2-decanone obtained using Model 1, Model 2 and Model 3. All three models behave very similarly, and they can accurately predict the densities of the ketones studied here, except for butanone and 2-pentanone, which are slightly underpredicted (less than 1 %). From butanone to 2-octanone, Model 1 is the best compared to the other two models, while Model 2 and Model 3 are better for 2-nonanone and 2-decanone. The RMSD of Model 1, Model 2 and Model 3 with respect to experimental values are 2.425, 3.052 and 3.932 kg/m^3, respectively.

Concerning the diffusion constants (Figure 5.4), all models do a very good job predicting this property for the eight ketones tested here. Model 3 has the lowest RMSD with a value of 0.096 10^{-5} cm^2/s compared to 0.200 for Model 2 and 0.229 for Model 1. Nonetheless, it is important to remember that these models are using the finite-

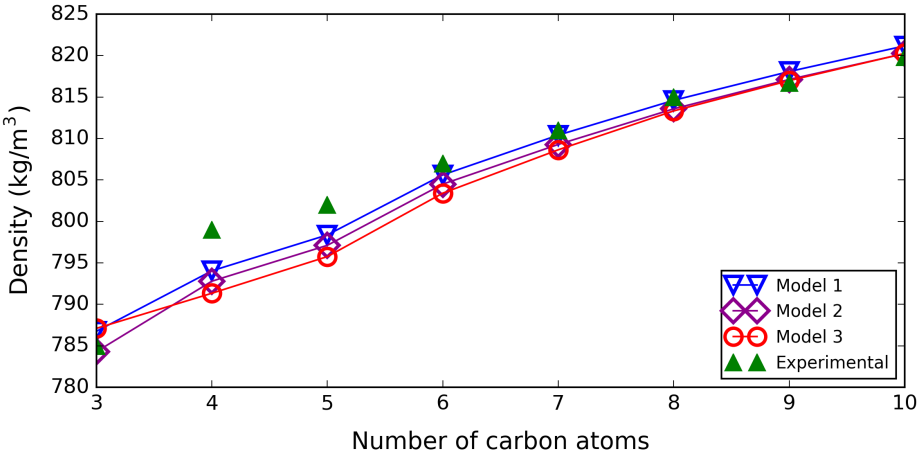

Figure 5.3: Density of linear ketones at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (purple rhombus) and Model 3 (red circles). The green symbols are experimental values [224, 241].

size corrections obtained for TraPPE, and their specific finite-size corrections could be slightly different.

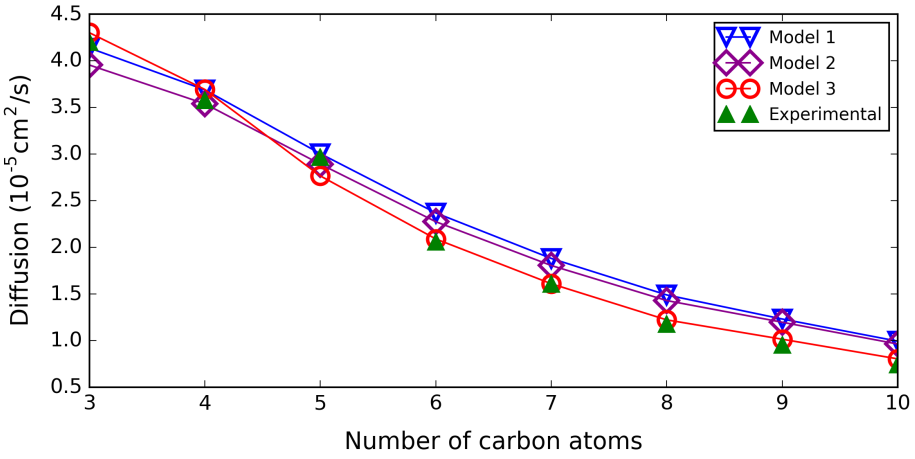

Figure 5.4: Diffusion of linear ketones at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (purple rhombus) and Model 3 (red circles). The green symbols are experimental values [242].

Enthalpy of vaporisation and free energy of self-solvation

The enthalpies of vaporisation of propanone to 2-decanone, obtained using the three potential models, are depicted in Figure 5.5. All models behave almost identically, and their plots overlap, especially for Model 2 and Model 3. This is not surprising since the enthalpy of vaporisation was one of the properties used during the parameterisation, and Model 2 and Model 3 were created using the same meta-models but different

objective functions. The models are in very good agreement with experimental data, with RMSDs equal to 1.108, 0.816 and 0.776 kJ/mol for Model 1, Model 2 and Model 3, respectively.

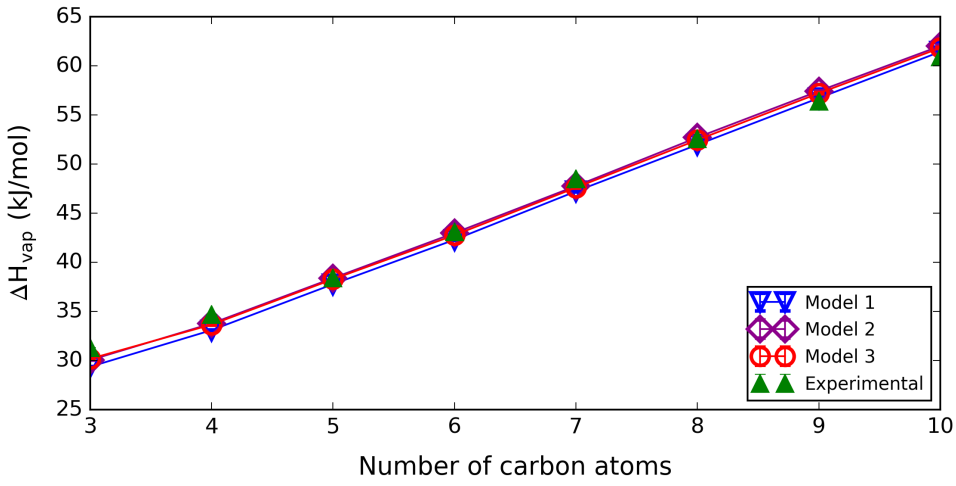

Figure 5.5: Enthalpy of vaporisation of linear ketones at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (purple rhombus) and Model 3 (red circles). In all cases, polarisation corrections were included. Experimental values (green triangles) were taken from reference [243].

Figure 5.6 shows the free energies of self-solvation obtained using these three models. As expected, Model 3 has the best performance (RMSD = 1.397 kJ/mol) due to propanone and 2-decanone's free energies of self-solvation being included in its parameterisation. On the other hand, Model 2 is slightly worse than Model 1 (RMSD equal to 2.485 kJ/mol compared to 2.209 kJ/mol for Model 1) even though propanone's free energy of self-solvation was one of the fitted properties used during Model 2's parameterisation, while Model 1 was obtained by fitting to only bulk properties. Model 3 can accurately predict free energies of self-solvation from propanone to 2-hexanone, except for pentanone. Pentanone's experimental value does not seem to follow the same trend as the rest of the ketones plotted here, however, experimental free energies of self-solvation from 2-hexanone onwards were calculated using extrapolated values for the vapour pressures since their temperature ranges for the Antoine coefficients did not include 298.15 K [228] (section 2.8.4 explains the relationship between the vapour pressure and the self-solvation free energy). Therefore, we expect a greater degree of uncertainty in those experimental values.

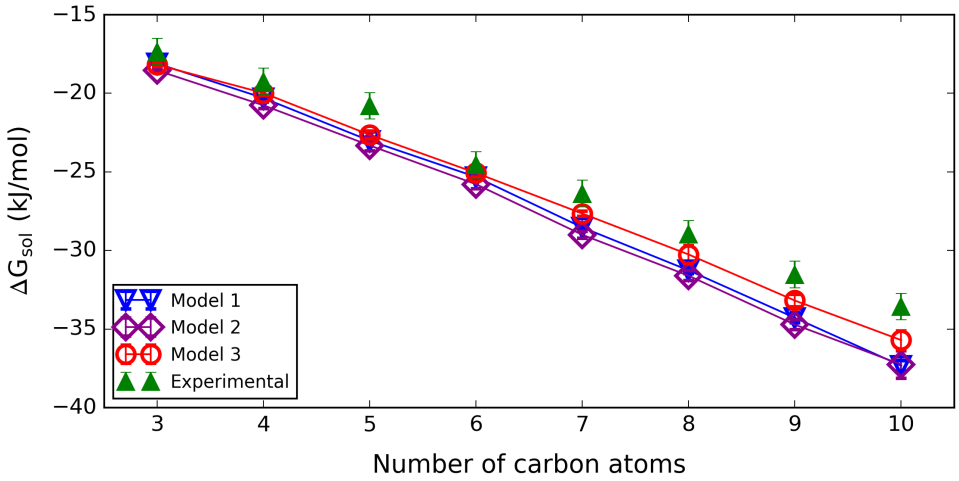

Figure 5.6: Free energy of self-solvation of linear ketones at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (purple rhombus) and Model 3 (red circles). In all cases, polarisation corrections were included (full lines and open symbols). Experimental values (green triangles) were calculated using the vapour pressures found in references [228]. The vapour pressures of hexanone to decanone at 298.15 K are extrapolations since this temperature was not included in their Antoine coefficients' temperature range.

Dielectric constant and dipole moment

Figure 5.7 shows the dielectric constants of propanone to 2-decanone, obtained using Model 1, Model 2 and Model 3. All models considerably overpredict this property, especially for smaller ketones. Model 2 has the best performance out of the three, however, it is still quite poor with an RMSD of 5.343. Meanwhile, Model 3 has the highest RMSD (RMSD = 6.920), followed by Model 1 with an RMSD of 6.629.

As explained in section 2.7.4, the equation proposed by Jorge and Lue [87] (Equation 2.52) was used to correct the simulated dielectric constants. This equation uses a simple dipole moment scaling factor and accounts for polarisation effects that cannot be captured using non-polarisable models. The scaling factor is the ratio between the liquid dipole, estimated using Equation 2.49 [92], and the dipole moment of the model. From Figure 5.8, we can see that the estimated liquid dipoles for linear ketones range from approximately 4.3 D to 4.7 D, while the dipoles for Model 1, Model 2 and Model 3 are 2.7, 2.6 and 3.2 D, respectively, and thus, this scaling factor is relatively high for all the models. However, as mentioned above, Equation 2.49 may be overestimating the liquid phase dipole of ketones, which would lead to an overestimation of the corrected

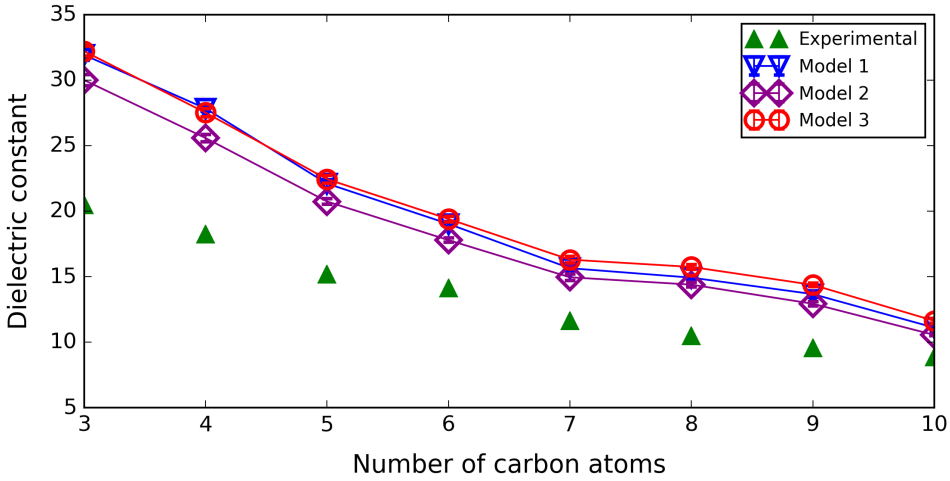

Figure 5.7: Dielectric constant of linear ketones at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (purple rhombus) and Model 3 (red circles). In all cases, polarisation corrections were applied. Experimental dielectric constants (green triangles) were taken from references [242, 244].

dielectric constant. This equation is an approximation, and due to a lack of experimental data, we do not know how well it performs for ketones. The Kirkwood-Onsager model [149–151] was used during its development, and the radius of the spherical cavity, where the polarisable dipole is placed, was estimated using a simple model. Furthermore, the polarisability of the molecule was assumed to be always the same as the value in the gas phase. [92]. Consequently, more accurate methods, like the one proposed by Jorge et al. [245], should be used in future work to estimate ketones' liquid dipole moments and assess Equation 2.52's performance.

Concerning the dipole moment (Figure 5.8), Model 1 and Model 2 have dipoles that are very close to or lower than the experimental dipoles in the gas phase. It has been argued that the dipole moment of a fixed-charge force field should fall somewhere between the experimental dipoles in the liquid and gas phases [92], as is the case for Model 3.

Root-mean-square deviation

Table 5.10 shows each model's root-mean-square deviation (RMSD) for the density, diffusion constant, enthalpy of vaporisation, dielectric constant and free energy of self-solvation. These values were calculated based on the experimental and simulated values of the ketones studied here. TraPPE is analysed in section 5.3.3, however, its RMSDs

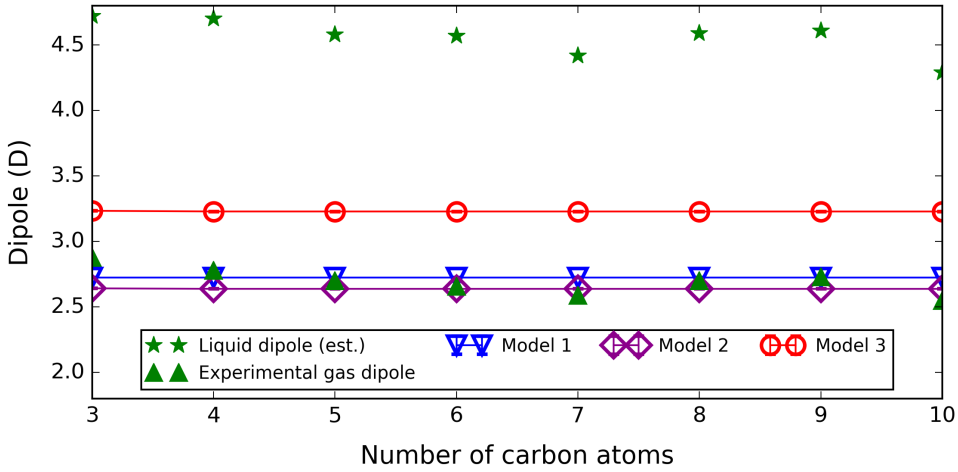

Figure 5.8: Simulated dipoles at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (purple rhombus) and Model 3 (red circles). The green asteriks are the liquid dipoles obtained with the equation proposed by Leontyev and Stuchebrukhov (Equation 2.49) [92] (see section 5.2.1), while the green triangles are the experimental dipoles in the gas-phase [246].

are presented here for completeness.

Table 5.10: RMSD for TraPPE-UA and the different models obtained in this chapter.

Force Field	**Density** [kg/m^3]	**Diffusion** [10^{-5} cm^2/s]	**ΔH_{vap}** [kJ/mol]	**Dielectric constant**	**ΔG** [kJ/mol]
Model 1	2.43	0.23	1.11	6.63	2.21
Model 2	3.05	0.20	0.82	5.34	2.49
Model 3	3.93	0.10	0.78	6.92	1.40
TraPPE	5.29	0.65	5.06	7.82	1.18
TraPPE (C)	5.29	0.65	4.53	6.13	0.82

All models perform better than TraPPE for the density, enthalpy of vaporisation, and self-diffusion constant, however, TraPPE is the best at predicting the free energy of self-solvation. TraPPE's RMSD for the free energy of self-solvation is 1.18 kJ/mol and once polarisation corrections are included this value decreases to 0.82 kJ/mol. Out of the models developed in this work, Model 3 is the best at predicting the free energy of self-solvation, a key aim of this work, however its RMSD is 0.58 kJ/mol larger that the value for TraPPE (C). Model 3 also has the best performance for the diffusion constant and enthalpy of vaporisation, and it leads to a model dipole moment that is more in line with the current knowledge about implicitly accounting for polarisation effects in fixed-

charge force fields [92, 108]. On the other hand, Model 1 and Model 2 are better than Model 3 at predicting the density, especially for butanone and 2-pentanone. Lastly, none of the models can accurately predict the dielectric constant; however, Model 3' dipole moment makes more physical sense than the dipole moments of Model 1 and Model 2. In conclusion, Model 3 is the best model overall, and thus, it was chosen for the PolCA force field.

5.3.3 PolCA

In this section, PolCA's performance is compared against TraPPE-UA (with and without polarisation corrections) for linear ketones, considering bulk properties and free energies of solvation. TraPPE-UA for ketones has been parameterised to reproduce the vapour-liquid coexistence curve of acetone [146].

As shown in Figure 5.9, both PolCA and TraPPE do a good job predicting the densities of the ketones studied here. However, neither can fully recreate the shape of the plotted experimental values. TraPPE accurately predicts the densities of 2-hexanone to 2-octanone, but from 2-nonanone onwards, it overpredicts this property. This is likely due to the shortcomings of the alkane parameters identified by Jorge [81]. On the other hand, PolCA accurately predicts 2-nonanone and 2-decanone's densities while slightly underpredicting this property for 2-hexanone to 2-octanone. Both models significantly underpredict the densities of butanone and 2-pentanone; nonetheless, PolCA accurately predicts propanone's density, unlike TraPPE. The RMSDs of PolCA and TraPPE are 3.932 and 5.291 kg/m^3, respectively.

Concerning the diffusion constant (Figure 5.10), PolCA does an excellent job predicting this property with an RMSD of 0.096 10^{-5} cm^2/s, while TraPPE systematically overpredicts this property for all the ketones studied here (RMSD = 0.647 10^{-5} cm^2/s). The self-diffusion constants obtained using TraPPE follow a similar trend to the experimental results but are shifted towards higher values by approximately 0.6 x 10^{-5} cm^2/s. A similar effect was observed when testing alcohols. TraPPE-UA overpredicts the self-diffusion constant of linear primary alcohols, except for methanol's (Figure 3.7) and the density of linear primary alcohols larger than pentanol (see section 3.3.1).

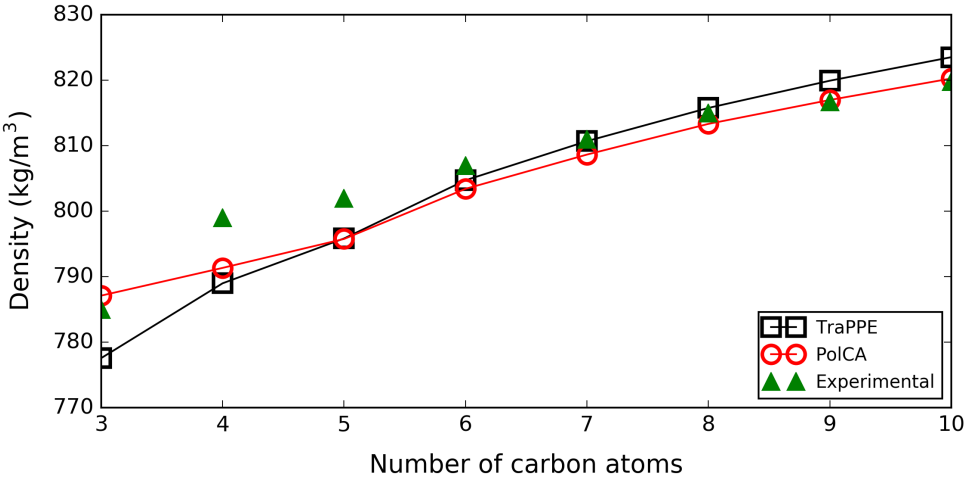

Figure 5.9: Densities of linear ketones at 298.15 K and 1 bar obtained using PolCA (red circles) and the original TraPPE force field (black squares). The green triangles are experimental values [224, 241].

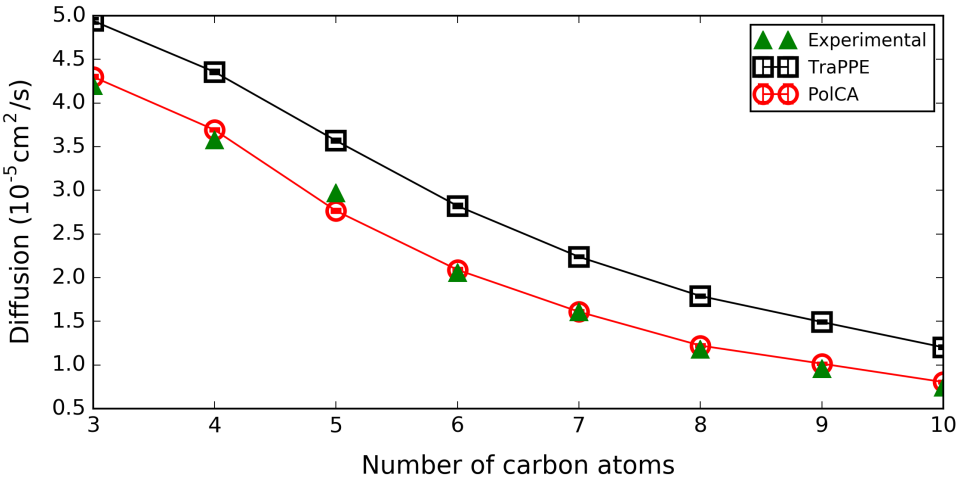

Figure 5.10: Self-diffusion constant of linear ketones at 298.15 K and 1 bar obtained using PolCA (red circles) and the original TraPPE force field (black squares). The green triangles are experimental values obtained from reference [242].

Figures 5.11 and 5.12 show the enthalpy of vaporisation and the free energy of self-solvation of ketones obtained using TraPPE and PolCA. TraPPE systematically underpredicts the enthalpy of vaporisation of all the ketones studied here by approximately 4 to 5 kJ/mol. Once polarisation corrections are added, TraPPE's performance slightly improves, and its RMSD decreases from 5.056 to 4.529 kJ/mol. On the other hand, PolCA included this property in its parameterisation, and thus, it can accurately predict the enthalpy of vaporisation with an RMSD of 0.776 kJ/mol. The opposite effect

is observed when looking at the free energy of self-solvation, PolCA slightly overpredicts the free energy of self-solvation (more negative values) with an RMSD of 1.397 kJ/mol while TraPPE accurately predicts this property, especially when polarisation corrections are added. The magnitude of these corrections increases from near zero to approximately 1 kJ/mol when we move from propanone to decanone, and thus, they have a bigger impact on larger ketones. TraPPE's RMSD for the free energy of solvation decreases from 1.178 to 0.817 kJ/mol once *post-facto* polarisation corrections are included. These results suggest a possible imbalance between TraPPE's ability to describe the energetic and entropic contributions to the solvation free energy since $G = H - TS$. For the solvation process, the enthalpy change ($\Delta H_{v\rightarrow l}$) is equal to minus the enthalpy of vaporisation, and thus, $\Delta G_{sol} = -\Delta H_{vap} - T\Delta S_{v\rightarrow l}$. TraPPE can accurately predict the self-solvation free energy but underpredicts the enthalpy of vaporisation, therefore, it also underpredicts the magnitude of $\Delta S_{v\rightarrow l}$ (i.e. the entropy of the liquid is likely being overpredicted reducing the magnitude of the difference between the entropy of the liquid and the gas). Nonetheless, as explained above, no experimental values were found for 2-hexanone to 2-decanone's vapour pressures at 298.15 K, and thus, extrapolated values had to be used when calculating experimental free energies. Consequently, more accurate experimental data is needed for a better comparative assessment of these two models.

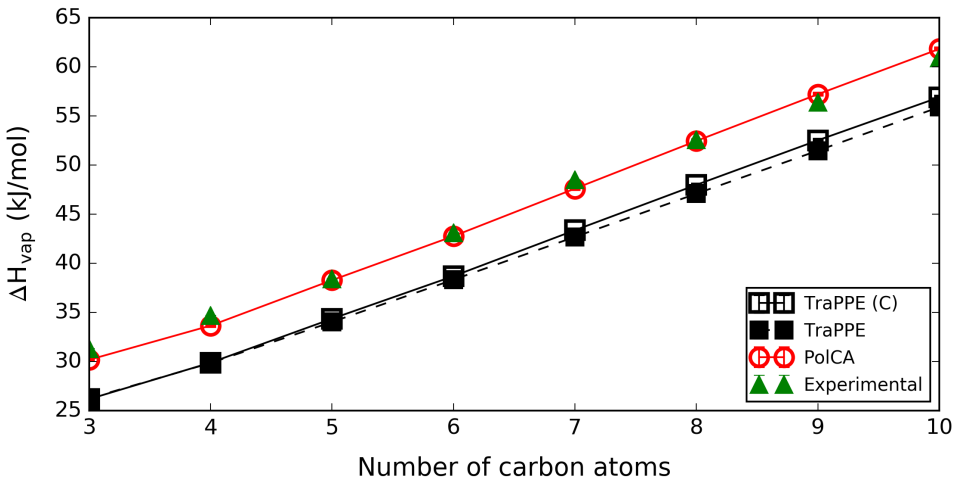

Figure 5.11: Enthalpy of vaporisation of linear ketones at 298.15 K and 1 bar obtained using PolCA (red circles) and the TraPPE force field with and without polarisation corrections (open black squares and filled black symbols, respectively). The green symbols are experimental values obtained from reference [243].

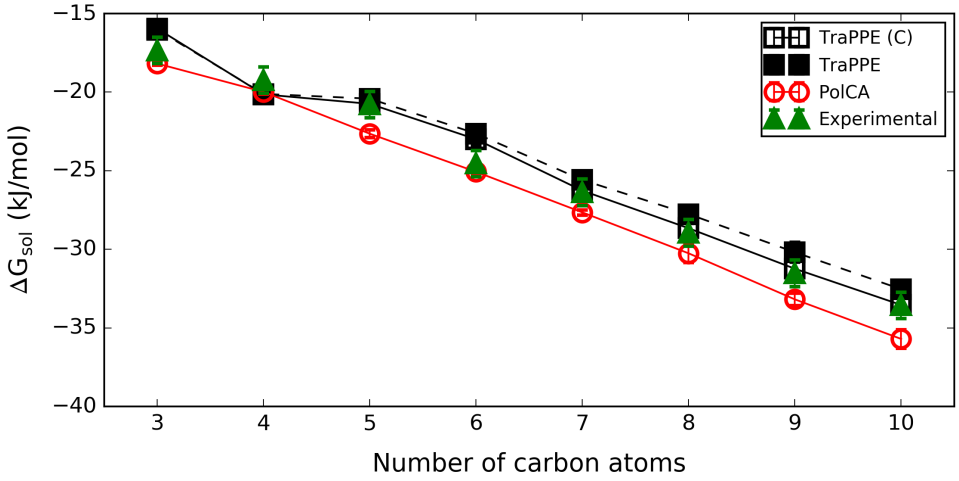

Figure 5.12: Free energy of self-solvation of linear ketones at 298.15 K and 1 bar obtained using PolCA (red circles), TraPPE and TraPPE with corrections (filled squares and open black squares, respectively), as a function of the number of carbons. Experimental values for the free energy were calculated using the vapour pressures found in references [228]. The vapour pressures of hexanone to decanone at 298.15 K are extrapolations since this temperature was not included in their Antoine coefficients' temperature range. An estimated average uncertainty of 0.84 kJ/mol was used for all the values. This value was taken from the Minnesota database [168].

Both PolCA and TraPPE (C) greatly overpredict the dielectric constant of the ketones studied here with RMSDs of 6.920 and 6.125, respectively (Figure 5.13). On the other hand, TraPPE without polarisation corrections substantially underpredicts this property and its RMSD, equal to 7.821, is higher than that for TraPPE (C). Propanone's dielectric constant obtained using TraPPE with and without corrections differs by 21.7 due to the high scaling factor used in Equation 2.52. This scaling factor is the ratio between the high estimated liquid dipole (4.7 Debye) and the model's low dipole moment (2.5 Debye), however, as explained in section 5.3.2, the estimated liquid dipoles obtained using Equation 2.49 are probably being overestimated causing the discrepancy between calculated and experimental values for both models. Nonetheless, it is important to emphasize that TraPPE's dipole moment is lower than the gas phase while PolCA's dipole (3.2 Debye) lies in between the gas and liquid dipole moments.

Lastly, Figure 5.14 shows the free energy of solvation of linear ketones in hexadecane obtained using both force fields, with and without polarisation corrections. These corrections are quite significant and range from 10.04 kJ/mol for propanone to 2.62 kJ/mol for 2-decanone (Table 5.3). Perhaps surprisingly, TraPPE without polarisation correc-

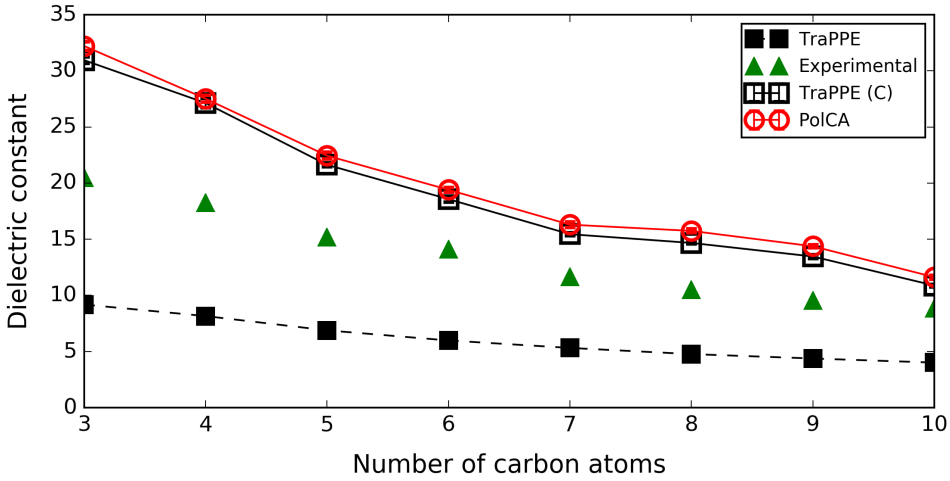

Figure 5.13: Dielectric constant of linear ketones at 298.15 K and 1 bar obtained using PolCA (red circles) and the TraPPE force field with corrections (open squares) and without corrections (dashed lines), as a function of the number of carbons. The green symbols are experimental values [242, 244].

tions does an excellent job at predicting this property with an RMSD of 0.475 kJ/mol, and thus, the model's performance significantly deteriorates once these corrections are added (RMSD = 6.654 kJ/mol). PolCA also overpredicts this property with an RMSD of 4.380 kJ/mol, and if polarisation corrections are not included, PolCA underpredicts the solvation free energy in hexadecane by 2.003 kJ/mol. Nonetheless, as explained above, it is possible that the equations used to obtain these corrections do not perform well for ketones, and more accurate methods should be used instead.

5.4 Conclusions

The PolCA model has been extended to ketones, and its performance was compared to TraPPE's. Concerning the density, both PolCA and TraPPE do a good job; however, both underpredict the densities of butanone and 2-pentanone. PolCA has a lower RMSD for the density than TraPPE, and it accurately predicts the enthalpies of vaporisation and diffusion constants of propanone to 2-decanone. Meanwhile, TraPPE underpredicts and overpredicts these two properties, respectively. Furthermore, both models do a good job predicting free energies of self-solvation.

PolCA works well for the properties included in the optimisation. However, both PolCA

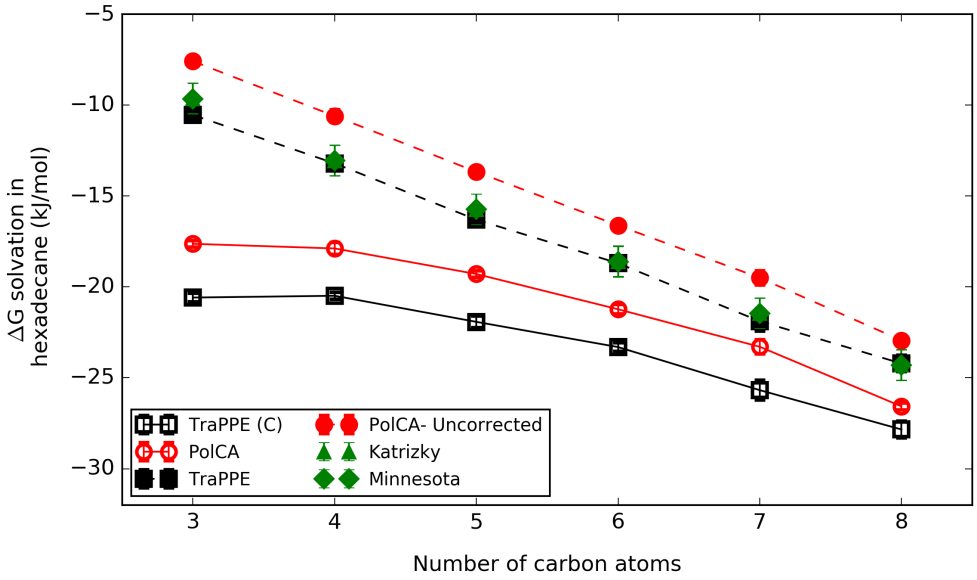

Figure 5.14: Free energy of solvation of primary ketones in hexadecane at 298.15 K and 1 bar as a function of the number of carbons in the solute, obtained using PolCA (red circles) and TraPPE (black squares), with and without corrections. Experimental values were taken from references [168] and [169]. The estimated average uncertainty is approximately 0.84 kJ/mol [168].

and TraPPE (C) fail to reproduce ketones' dielectric constants and free energies of solvation in hexadecane, although PolCA performs significantly better for the latter. Likely, the polarisation corrections and the liquid dipoles used in this chapter are not accurately predicted by the equations proposed by Leontyev and Stuchebrukhov [92], and more accurate methods should be pursued before extending the meta-models' learning set. Based on the previous chapters, these equations seem to work well for water [92], alcohols and amines but not for ketones. Nonetheless, out of the two models, PolCA performs overall better than TraPPE.

Chapter 6

PolCA's Transferability

6.1 Introduction

In this chapter, we study PolCA's performance at predicting solvation free energies in octanol as well as multifunctional compounds' properties to assess its transferability. Alcohols, amines and ketones' solvation free energies in hexadecane were presented in their corresponding chapters since only the alkyl parameters proposed by Jorge [1] and each chapter's new parameters were needed for this property's calculation. The interactions between the alkyl parameters and, separately, the hydroxyl, amino or carbonyl groups were indirectly accounted for during PolCA's development since molecules of different alkyl chain lengths were included in the parameterisation. In contrast, this chapter focuses on systems that contain several independently parameterised functional groups (e.g. hydroxyl group's parameters were not considered during the amino and carbonyl's groups optimisation). One exception is the solvation free energy of alkanes in octanol which is included here for a more complete comparison.

The octanol/water partition coefficient (P_{ow}) measures the ratio between a neutral solute's concentration in octanol and its concentration in water ($P_{ow} = [solute]_{octanol} \,/\, [solute]_{aqueous}$) [65], and it is widely used in the pharmaceutical, metallurgic and agrochemical industries to predict a solute's behaviour in both hydrophilic and hydrophobic media (e.g. drug partitioning, hydrophobicity in environmental problems, etc.). This important parameter can be calculated from the solute's

hydration free energy and solvation free energy in octanol [64], which is why solvation free energies in octanol were chosen for PolCA's validation. As explained in Chapter 1, several researchers have calculated octanol/water partition coefficients using molecular simulations. Garrido et al. simulated $\log P_{ow}$ of linear alkanes from methane to octane using TraPPE, OPLS-AA and Gromos with the MSPC/E water model, and they found that simulating the solute with the OPLS-AA force field and octanol with TraPPE-UA produced the most accurate results for solvation free energies in octanol (TraPPE-UA was used instead of OPLS-AA to decrease computational time) while Gromos was the best at predicting hydration free energies. The partition coefficients obtained using this combination had an absolute deviation of 0.1 $\log P$ units. [64] Bannan et al. calculated $\log P_{ow}$ using GAFF and GAFF-DC, which was parameterised to improve methanol's dielectric constant, and noticed that both force fields had very similar accuracy with RMSDs of 1.1 +/- 0.2 and 0.9 +/- 0.1 log units compared to the experimental $\log P_{ow}$, respectively [65]. Zhang et al. studied neutral amino acid side chain analogues' $\log P_{ow}$ using Amber ff9x, Gromos 536A, OPLS-AA/L, Amber ff03 and CHARMM 27, and concluded that Amber ff9x had the best performance with an RMSD equal to 0.4 +/- 0.1 log units. However, all these force fields overpredicted $\log P_{ow}$ due to an underestimation of the hydration free energies' magnitudes, except for Gromos which predicts more negative values for the solvation free energies in water and octanol, with RMSDs equal to $\approx$ 1 kcal/mol and 1.9 kcal/mol, respectively [66].

The second part of this chapter is dedicated to the study of multifunctional compounds to further test the interactions between PolCA's parameters. Due to limited experimental data availability, eight out of the ten compounds studied here are alkanolamines. These molecules are widely used in the pharmaceutical, petroleum, textile and detergent industries since they are inexpensive and can absorb acid gases like H_2S and CO_2. [247, 248] However, short alkanolamines' simulations obtained using a simple combination of the amino and hydroxyl groups parameters do not usually produce accurate results [248]. This outcome is expected due to the proximity between the functional groups, and thus, several force fields specific for alkanolamines have been

proposed. Alejandre et al. developed a monoethanolamine (MEA) UA model using charges and bonded parameters from ab initio calculations with OPLS LJ parameters. This model accurately predicts monoethanolamine's liquid density at ambient conditions and underpredicts its critical temperature and surface tension at 323 K by 6% and 4%, respectively [247], and it was later extended to other alkanolamines [249]. Another force field specific for this molecule is the all-atom model proposed by da Silva et al. [250] which uses GAFF parameters and torsional terms obtained from ab initio calculations with scaled QM charges to better reproduce MEA's density and enthalpy of vaporisation at 333 K. Orozco et al. [248] simulated MEA at four different temperatures using AUA4's hydroxyl and amino group parameters and noticed that the predicted liquid densities, enthalpies of vaporisation and vapour pressures deviated from experimental data by 5%, 7% and 45%, respectively. Consequently, they proposed using an independent set of partial charges for each molecule and fitted the oxygen atom's σ value to monoethanolamine, diethanolamine and methyldiethanolamine's liquid densities, vapour pressures and enthalpies of vaporisation. Furthermore, they calculated MEA's OCH_2CH_2N torsional potential from QM simulations and added an explicit 1-4 LJ intramolecular interaction between the amino and hydroxyl groups due to their strong attraction.

In the previous paragraph, we discussed the transferability issues that arise when modelling short alkanolamines, however, this problem also applies to other molecules with polar functional groups in close proximity. For example, TraPPE-UA uses a special short-range repulsive potential between the positively charged hydrogen atoms and the negatively charged heteroatoms in conjunction with Coulomb 1-4 interactions scaled by 0.5 to simulate glycols and ethers [146].

All these considerations mean that it is essential to test the transferability of PolCA parameters in a wide range of environments, and this is what we aim to achieve in this chapter.

6.2 Methodology

The methods used to obtain bulk properties and free energies of solvation are described in detail in Chapter 2. Specific details for this chapter are described below.

Compounds studied in this chapter

The structure of the multifunctional molecules studied in this chapter can be found in Figure 6.1, while the number of molecules used for each simulation is shown in Table 6.1. The simulation box size for each system was approximately 27 nm^3.

Table 6.1: Number of molecules in the simulation box for each system.

System	Number of molecules
4-amino-2-butanol	214
2-aminoethanol	340
1-amino-2-propanol	255
3-amino-1-propanol	257
2-amino-2-methyl-1-propanol	223
4-amino-4-methyl-2-pentanone	164
6-amino-2-methyl-2-heptanol	123
N-methyldiethanolamine	161
Diethanolamine	176
Diacetone alcohol	173
Alkanes, amines and ketones in octanol	$\approx$ 134
Alcohols in butanone	$\approx$ 328

Solvation free energy calculations

Free energies of solvation in octanol were run using fifteen λ-states (0, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.8, 0.9 and 1) and seven λ-states (0, 0.3, 0.6, 0.7, 0.8, 0.9 and 1) for the LJ and electrostatic components, respectively. The LJ and electrostatic components were run for 10 ns and 50 ns, respectively. The same methodology was followed when calculating 2-aminoethanol's free energy of self-solvation.

On the other hand, free energy simulations in 2-butanone converged much faster and thus, the LJ and Coulomb components were run for 5 ns and 10 ns, respectively. Additionally, the intermediate λ-states used to run these simulations were the same as

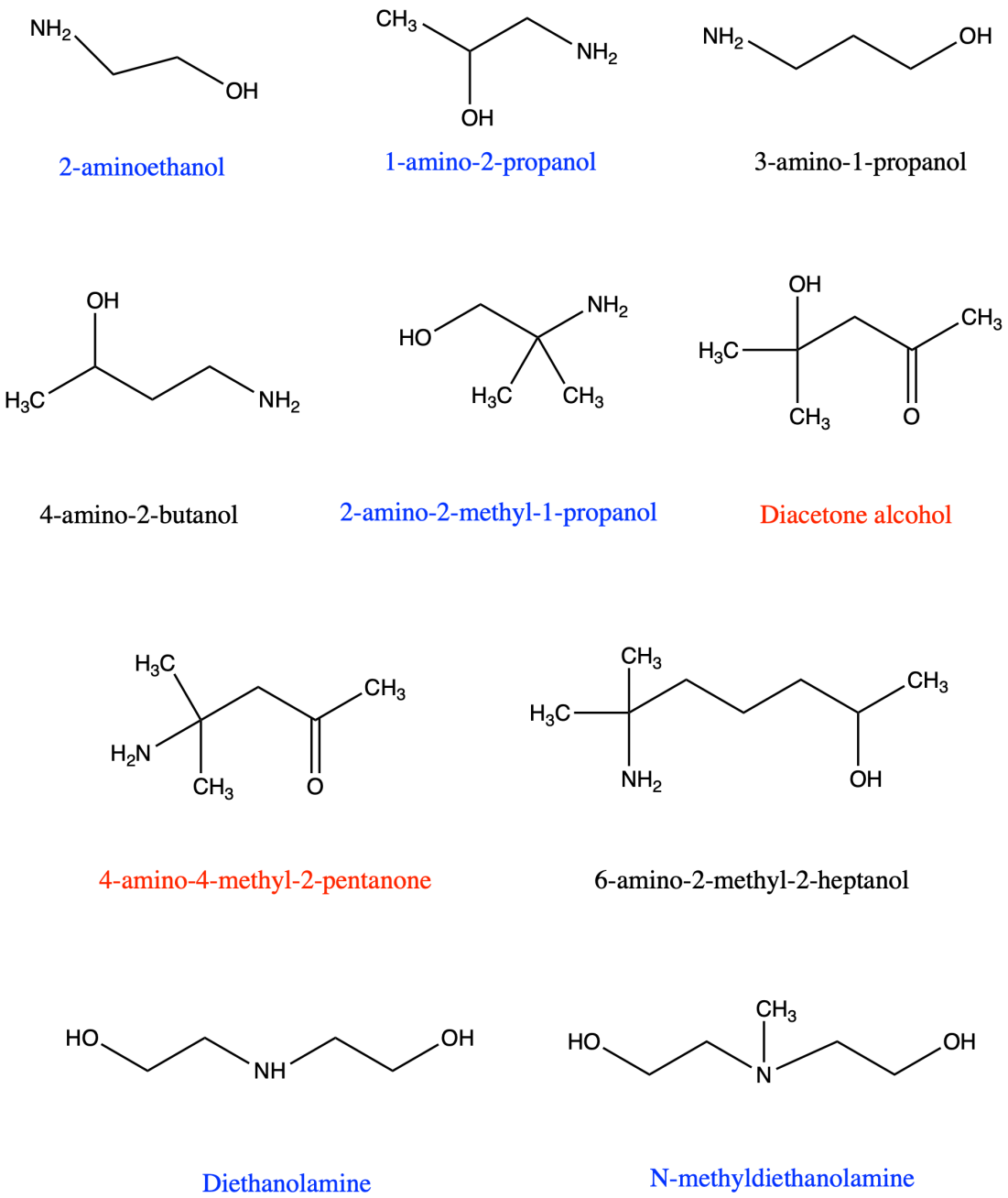

Figure 6.1: Structure of the multifunctional molecules studied in this chapter, drawn using the software ChemDraw 20.0. Out of the ten molecules, eight are alkanolamines. The other two molecules, represented with a red label, contain a carbonyl group and either a hydroxyl or amino group. Molecules with blue labels contain the dihedral N-CH_x-CH_y-OH, and thus, explicit 1-4 interactions had to be added during their simulation (see text for details).

those used when running ketones free energies of self-solvation (Chapter 5). Fourteen λ-states were used for the LJ component (0, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45,

0.5, 0.55, 0.7, 0.85 and 1) and five λ-states for the electrostatics (0, 0.2, 0.4, 0.7 and 1).

Simulation times were chosen by plotting free energy of solvation against simulation time to ensure convergence. The LJ and electrostatic components of the free energy of solvation of octylamine in octanol as a function of simulation time can be found in Figures A5 and A6 in Appendix A1.

6.2.1 Polarisation corrections

Table 6.2 shows *post-facto* polarisation corrections (C_{pol}) for the systems studied in this chapter. These corrections correspond to a liquid-gas transition, and thus, they need to be multiplied by -1 before applying them to free energies of solvation. The equations used to obtain these corrections and the assumptions made during their development are explained in section 2.7.2 of the methodology. Furthermore, Table 6.2 presents the estimated liquid dipole moments for the molecules studied here obtained using Equation 2.49.

The polarisation corrections for amines in octanol are relatively small in magnitude (below 1 kJ/mol), even though the individual distortion and electronic contributions are somewhat larger. On the other hand, the corrections for ketones in octanol are more significant, and they range from 2.91 kJ/mol for propanone to 1.05 kJ/mol for 2-octanone. Additionally, the distortion and electronic polarisation terms for ketones in octanol are quite large, getting as high as 23.53 kJ/mol. Furthermore, polarisation corrections for alkanes in octanol are equal to zero and corrections for alcohols in butanone are almost zero. Alkanes are non-polar molecules, and hence the electrostatic contribution to the solvation free energy is effectively zero [76], therefore, their corrections were expected to be negligible. On the other hand, the polarisation corrections for alcohols in butanone are a surprising result.

The last three molecules in Table 6.2 are alkanolamines with very high distortion and electronic polarisation terms. The distortion and electronic terms for diethanolamine are as high as -23.22 kJ/mol and 21.39 kJ/mol, respectively. However, these molecules' net polarisation corrections range from -2.06 kJ/mol to -1.16 kJ/mol. Nonetheless, their

Table 6.2: Polarisation corrections for a liquid-gas phase transition (C_{pol}), expressed in kJ/mol. C_{dist} and C_{elec} are the distortion and electronic polarisation terms, respectively, α [224] is the polarisability of the solute in the gas phase and ϵ_{el} [224] and ϵ_{sol} are the high-frequency dielectric permittivity and the static dielectric constant, respectively, of the medium. N-MDEA stands for N-methyldiethanolamine.

Solute	**Solvent**	μ_g **(D)**	α ($\AA^3$)	ϵ_{sol}	ϵ_{el}	μ_l **(D)**	C_{dist}	C_{elec}	C_{pol}
methane	octanol	0.00	2.60	9.78	2.04	0.00	0.00	0.00	0.00
ethane	octanol	0.00	4.47	9.78	2.04	0.00	0.00	0.00	0.00
propane	octanol	0.08	6.29	9.78	2.04	0.14	-0.02	0.02	0.00
butane	octanol	0.05	8.20	9.78	2.04	0.09	0.00	0.01	0.00
pentane	octanol	0.10	10.11	9.78	2.04	0.17	-0.02	0.02	0.00
hexane	octanol	0.10	11.94	9.78	2.04	0.17	-0.01	0.02	0.00
heptane	octanol	0.10	13.81	9.78	2.04	0.17	-0.01	0.01	0.00
octane	octanol	0.10	15.60	9.78	2.04	0.17	-0.01	0.01	0.00
nonane	octanol	0.10	17.45	9.78	2.04	0.17	-0.01	0.01	0.00
decane	octanol	0.10	19.33	9.78	2.04	0.17	-0.01	0.01	0.00
methylamine	octanol	1.31 [221]	4.03	9.78	2.04	2.26	-6.78	7.73	0.96
ethylamine	octanol	1.22 [221]	5.86	9.78	2.04	2.11	-4.04	4.61	0.57
propylamine	octanol	1.17 [221]	7.69	9.78	2.04	2.02	-2.83	3.23	0.40
butylamine	octanol	1.33 [222]	9.36	9.78	2.04	2.30	-3.01	3.43	0.42
pentylamine	octanol	1.19	11.38	9.78	2.04	2.06	-1.98	2.26	0.28
hexylamine	octanol	1.27	13.25	9.78	2.04	2.19	-1.94	2.21	0.27
heptylamine	octanol	1.18	15.06	9.78	2.04	2.04	-1.47	1.68	0.21
octylamine	octanol	1.27	16.91	9.78	2.04	2.19	-1.52	1.73	0.21
propanone	octanol	2.88	6.4	9.78	2.04	4.97	-20.62	23.53	2.91
butanone	octanol	2.78	8.22	9.78	2.04	4.80	-14.95	17.06	2.11
2-pentanone	octanol	2.7	10.05	9.78	2.04	4.66	-11.55	13.18	1.63
2-hexanone	octanol	2.66	11.89	9.78	2.04	4.59	-9.47	10.81	1.34
2-heptanone	octanol	2.59	13.73	9.78	2.04	4.47	-7.77	8.87	1.10
2-octanone	octanol	2.7	15.56	9.78	2.04	4.66	-7.46	8.51	1.05
methanol	butanone	1.70	3.26	18.50	1.91	2.87	-12.69	12.72	0.02
ethanol	butanone	1.69	5.13	18.50	1.91	2.86	-7.97	7.99	0.02
propanol	butanone	1.68	6.96	18.50	1.91	2.84	-5.81	5.82	0.01
butanol	butanone	1.66	8.79	18.50	1.91	2.80	-4.49	4.50	0.01
pentanol	butanone	1.70	10.61	18.50	1.91	2.87	-3.90	3.91	0.01
hexanol	butanone	1.65	12.46	18.50	1.91	2.79	-3.13	3.13	0.01
heptanol	butanone	1.71	14.30	18.50	1.91	2.89	-2.93	2.93	0.01
octanol	butanone	1.68	16.14	18.50	1.91	2.84	-2.50	2.51	0.00
2-aminoethanol	2-aminoethanol	2.27	6.44	31.40	2.11	4.46	-22.43	20.36	-2.06
N-MDEA	N-MDEA	2.86	12.77	22.40	2.16	5.68	-18.76	17.60	-1.16
Diethanolamine	Diethanolamine	2.81	10.76	25.26	2.17	5.69	-23.22	21.39	-1.82

estimated liquid dipoles could be significantly overestimated since these values range from 4.46 D to 5.69 D. The same effect was observed when studying ketones, where estimated liquid dipoles were as high as 4.72 D (see section 5.2.1). The alkanolamines from Table 6.2 and the ketones studied in this work have gas dipoles that range from 2.27 D to 2.88 D, and these values are significantly larger than those for alcohols (1.6 D to 1.7 D) and amines (1.18 D to 1.31 D).

6.2.2 Force Field parameters

Section 6.3.1 analyses PolCA's performance at predicting solvation free energies in octanol and butanone against TraPPE-UA [69, 146] and Gromos2016H66 [80]. TraPPE-UA was used for alkanes and ketones in octanol, while Gromos-2016 was used to calculate amines in octanol due to the lack of TraPPE-UA parameters for amines.

PolCA's non-bonded parameters for each functional group were first introduced in their corresponding results chapter, but the full set of PolCA's non-bonded parameters are presented here in Table 6.3 for clarity. Additionally, TraPPE-UA and Gromos2016H66's parameters can be found in sections 2.6 and 4.2.2, respectively.

Table 6.3: Lennard-Jones parameters and partial charges for the PolCA united-atom force field for alcohols, ketones and amines.

non-bonded [a] (Eq. 2.37)	σ [nm]	ε [kJ/mol]	Partial charge (q)
		Alkanes [1]	
(**CH_3**)-CH_x	0.379	0.833	0
$(CH_x)_2$-(**CH_2**)	0.399	0.392	0
$(CH_x)_3$-(**CH**)	0.473	0.085	0
$(CH_x)_4$-(**C**)	0.646	0.00426	0
		Alcohols	
CH_x-(**O**)-H	0.2853	0.7733	-0.646
O-(**H**)	0	0	+0.406
(**CH_3**)-OH	0.379	0.833	+0.240
CH_x-(**CH_2**)-OH	0.399	0.392	+0.240
$(CH_x)_2$-(**CH**)-OH	0.438	0.085	+0.240
$(CH_x)_3$-(**C**)-OH	0.585	0.00426	+0.240
		Amines	
CH_x-**N**-H_2	0.3401	0.814	-0.871
$(CH_x)_2$-**N**-H	0.389	0.723	-0.605
$(CH_x)_3$-**N**	0.4918	0.151	-0.339
N-**H**	0	0	0.379
CH_3-N	0.379	0.833	0.113
CH_2-N	0.399	0.392	0.113
CH-N	0.4645	0.085	0.113
C-N	0.630	0.00426	0.113
		Ketones	
C(=O)	0.382	0.333	0.709
O (sp^2)	0.300	0.200	-0.609
(**CH_3**)-C=O	0.379	0.833	-0.05
(CH_x)-(**CH_2**)-C=O	0.399	0.392	-0.05
$(CH_x)_2$-(**CH**)-C=O	0.473	0.085	-0.05
$(CH_x)_3$-(**C**)-C=O	0.646	0.00426	-0.05

a) The non-bonded parameters correspond to the sites in bold.

Furthermore, Table 6.4 contains extra dihedral parameters needed to simulate the molecules studied here. These parameters were taken from either other force fields

or from similar dihedrals, and their source is specified in the table. Molecules with two adjacent α-carbon atoms bonded to either an amino or hydroxyl group had to be treated differently due to the strong attraction between the negatively charged oxygen or nitrogen atoms and the positively charged hydrogen atoms unprotected by a LJ potential. These molecules were simulated using an explicit 1-4 LJ intramolecular interaction between the oxygen and nitrogen atoms and the dihedral O-CH_2-CH_2-N's parameters were taken from the anisotropic AUA4 force field for alkanolamines [248]. This dihedral was used as an approximation since its parameters were obtained using different 1-4 LJ intramolecular interactions between the oxygen and nitrogen atoms, which cannot be easily included in PolCA due to the LJ parameters' displacement from the atom centres in the AUA4 force field. The possible implications of this assumption will be discussed later in this chapter.

Table 6.4: Extra dihedrals parameters for the PolCA force field. These dihedrals appear in some of the molecules studied in this chapter and have not been defined previously. The parameters were converted to the Ryckaert-Bellemans convention and should be used with Equation 2.38.

New dihedral parameters taken from the AUA4 force field [217, 248]						
	C0	**C1**	**C2**	**C3**	**C4**	**C5**
N-CH_2-CH_2-OH	3.412	38.189	-39.902	6.627	65.536	-46.443
CH_x-CH-N-H	0.548	2.069	0.970	-1.663	-0.335	-0.793
CH_x-CH-CH_2-N	4.698	10.788	-3.226	-14.850	2.807	-17.393
CH_2-CH_2-CH-N	3.616	9.497	1.217	-13.796	2.086	-1.697
Dihedrals used to approximate missing parameters						
			Used to estimate			
CH_x-CH_2-CH_2-OH			CH_x-CH-CH-OH, C(=O)-CH_2-C-OH			
CH_2-CH_2-CH-N			C(=O)-CH_2-C-N			
C(=O)-CH_2-CH_2-CH_x			C(=O)-CH_2-C-CH_x			
CH_x-C(=O)-CH_2-CH_y			CH_x-C(=O)-C-CH_y			
N-CH_2-CH_2-OH			N-CH_x-CH_y-OH			
CH_x-CH-N-H			CH_x-C-N-H			
CH_x-CH-CH_2-N			CH_x-CH-CH-N			

6.3 Results and discussion

6.3.1 Free energy of solvation in octanol

This section analyses PolCA's performance at predicting solvation free energies of alkanes, amines and ketones in octanol. TraPPE-UA's performance was also tested for comparison. However, TraPPE does not have a united-atom model for amines, and thus, Gromos-2016 was used for these molecules instead.

Figure 6.2 shows the free energy of solvation of linear alkanes in octanol obtained using PolCA and TraPPE. The PolCA model yields excellent agreement with experimental data (RMSD of 0.68 kJ/mol), although TraPPE also performs quite well (RMSD of 1.10 kJ/mol). As explained above, no polarisation corrections need to be applied because alkanes are non-polar molecules, and hence the electrostatic contribution to the solvation free energy is effectively zero [76].

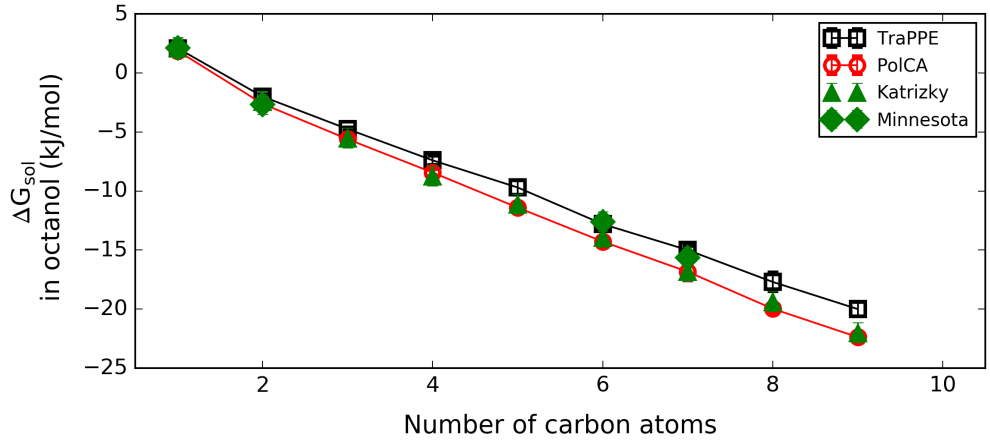

Figure 6.2: Free energy of solvation of alkanes in octanol at 298.15 K and 1 bar as a function of the number of carbons in the solute. Experimental values were taken from references [168] and [169]. The estimated average uncertainty is approximately 0.84 kJ/mol [168].

Free energies of solvation of linear primary amines in octanol, obtained using PolCA and Gromos-2016-PME, are shown in Figure 6.3. PolCA greatly underpredicts (less negative values) free energies of solvation in octanol with an RMSD of 7.184 kJ/mol, compared to the experimental values from the Minnesota database [168] and references [230, 251]. Similarly, Gromos-2016-PME greatly underpredicts the free energy of solvation in octanol with an RMSD of 7.208 kJ/mol, however from pentylamine to octylamine, this model is better than PolCA, and its trend suggests that the same effect will

be observed for larger amines. Once polarisation corrections are added, Gromos-2016-PME's performance slightly improves with an RMSD equal to 6.737 kJ/mol. From pentylamine onwards, polarisation corrections are smaller than -0.3 kJ/mol, and thus, the plots for Gromos-2016-PME with and without polarisation corrections overlap for these molecules. Methylamine and ethylamine are the most affected by the addition of polarisation corrections since their corrections are -0.96 kJ/mol and -0.57 kJ/mol, respectively (see Table 6.2). The values from the Katrizky database [169] were not taken into account when calculating the RMSDs since they are not consistent with the other two sources. Nonetheless, it is important to highlight the uncertainty associated to the experimental solvation free energies in octanol. Experimental values for this property are normally obtained from the solute's experimental octanol/water partition coefficient and hydration free energy, and thus, errors in these values will propagate to the solvation free energy in octanol. Simulated solvation free energies in water and octanol are run at infinite dilution, and hence, experimental $\log P_{ow}$ obtained using concentrations that significantly deviate from infinite solutions might not be appropriate for the calculation of ΔG_{sol}. Also, the equation used to correlate solvation free energies with $\log P_{ow}$ assumes a complete immiscibility between the solvents, however, water and octanol are partially miscible. [65] Furthermore, hydration free energy databases contain some inconsistencies and their values are often over 50 years old [82].

Lastly, solvation free energies of ketones in octanol obtained using PolCA and TraPPE can be found in Figure 6.4. TraPPE systematically underpredicts free energies of solvation with an RMSD of 1.445 kJ/mol; however, its RMSD decreases to 0.846 kJ/mol once polarisation corrections are added. TraPPE (C) overpredicts propanone's free energy of solvation in octanol, but it accurately predicts this property from butanone onwards. On the contrary, PolCA significantly and systematically overpredicts (more negative values) free energies of solvation of ketones in octanol with an RMSD of 7.183 kJ/mol. Furthermore, free energies of solvation were run for alcohols in butanone using both force fields, and the results are plotted in Figure 6.5. Even though TraPPE (C) can accurately predict free energies of solvation of ketones in octanol, it underpredicts free energies of solvation of alcohols in butanone by 3.445 kJ/mol. As can be seen

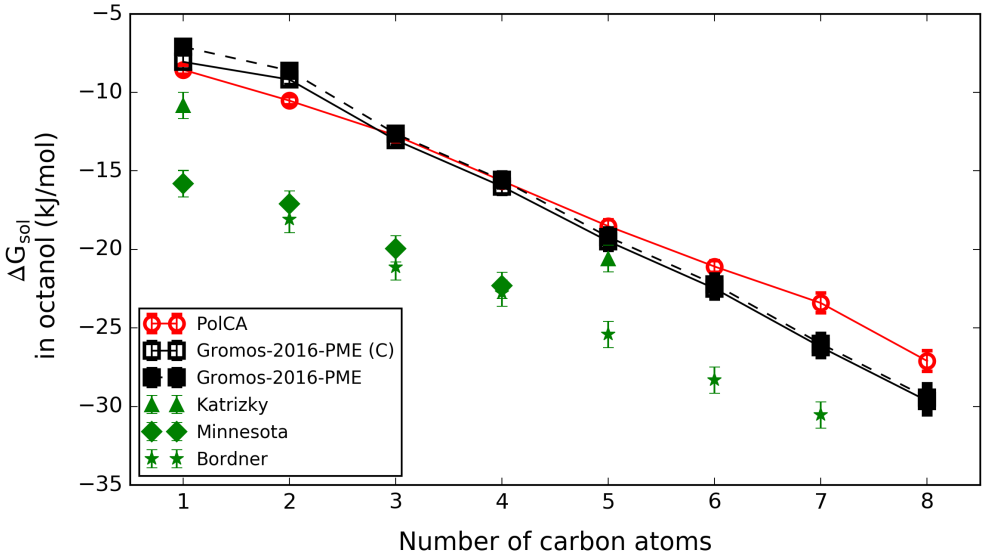

Figure 6.3: Free energy of solvation of linear primary amines in octanol at 298.15 K and 1 bar obtained using PolCA and GROMOS-2016-PME. Full lines with empty symbols contain polarisation corrections, while dashed lines with filled black squares represent values without corrections. The green symbols are experimental values obtained from the Katrizky and Minnesota databases [168, 169] and references [230, 251].

from Table 6.2, polarisation corrections are practically zero for these systems and thus, TraPPE and TraPPE (C) overlap in Figure 6.5. Meanwhile, PolCA systematically overpredicts free energies of solvation of alcohols in butanone by 5.684 kJ/mol. This result and ketones solvation free energies in alcohols suggest that PolCA overpredicts the interaction between the hydroxyl and carbonyl groups. We will return to this point later in this chapter.

6.3.2 Multifunctional compounds

Several molecules containing more than one functional group were simulated using PolCA (see Figure 6.1) to test the model's transferability. The density was the main property of interest due to the lack of experimental data for other properties. Nonetheless, enough experimental data was found for the calculation and analysis of three of these molecules' dielectric constants, and the results are presented at the end of this section.

Table 6.5 shows the simulated density of ten multifunctional compounds compared to their experimental values. Most of these molecules were simulated at 298.15 K since

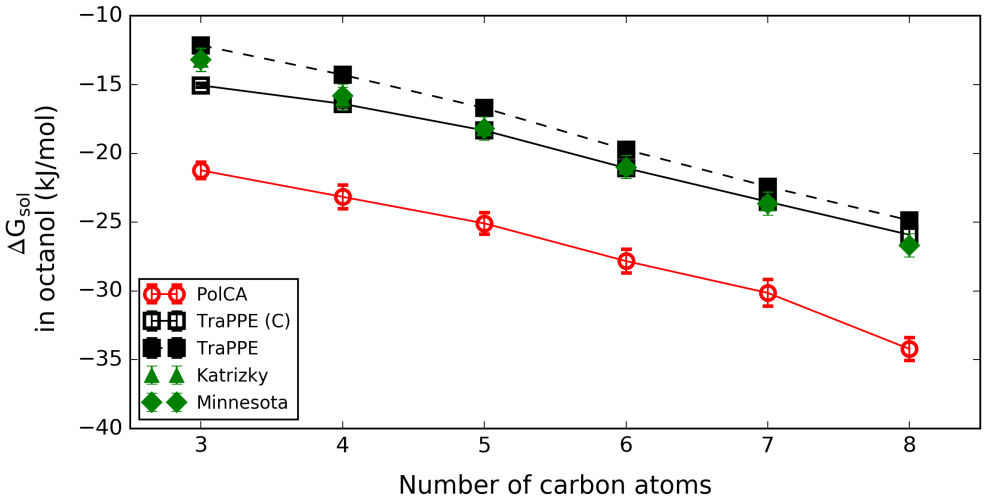

Figure 6.4: Free energy of solvation of linear ketones in octanol at 298.15 K and 1 bar as a function of the number of carbons in the solute. Experimental values were taken from references [168] and [169]. The estimated average uncertainty is approximately 0.84 kJ/mol [168].

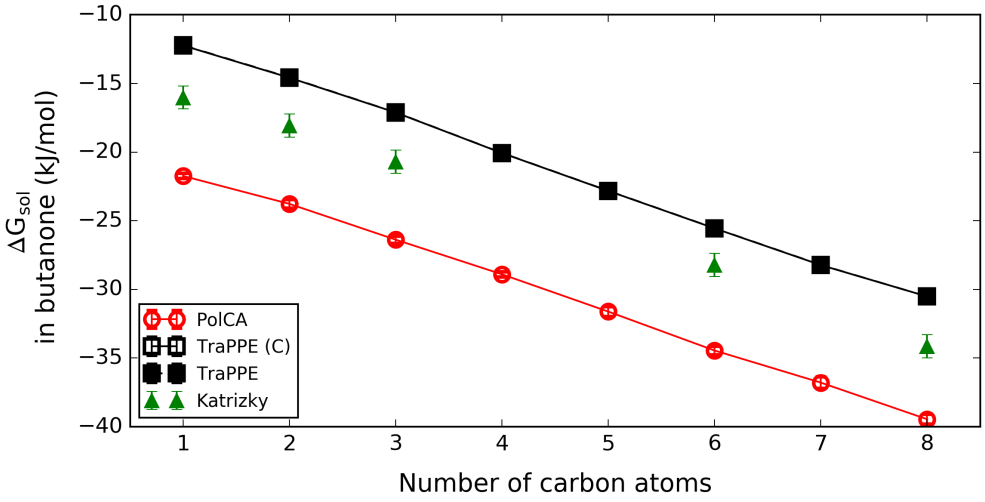

Figure 6.5: Free energy of solvation of linear alcohols in butanone at 298.15 K and 1 bar as a function of the number of carbons in the solute. Experimental values were taken from reference [169]. The estimated average uncertainty is approximately 0.84 kJ/mol [168].

this was the temperature used during PolCA's development. Unfortunately, two of these compounds are solids at room temperature: 2-amino-2-methyl-1-propanol and diethanolamine, with freezing points of 298.65 K and 301.15 [224], respectively. Therefore, 2-amino-2-methyl-1-propanol and diethanolamine were simulated at 313.15 K and 303.15 K, respectively.

Table 6.5: Densities, in kg/m^3, of several multifunctional compounds obtained using PolCA. Simulated and experimental temperatures were 298.15 K except for 2-amino-2-methyl-1-propanol (313.15 K) and diethanolamine (303.15 K). Experimental values were obtained from reference [224], unless stated otherwise. 4-amino-4-methyl-2-pentanone's experimental value taken from reference [224] was labelled as a "rough estimation" (code 3 in the Handbook, and no extra information was given).

	PolCA	**Experimental**	**Error (%)**
4-amino-2-butanol	950.7 +/- 0.4	930	2.228
2-aminoethanol	982.2 +/- 0.2	1014	-3.132
1-amino-2-propanol	1013.5 +/- 0.6	957	5.900
3-amino-1-propanol	975.3 +/- 0.2	972	0.339
2-amino-2-methyl-1-propanol	866.3 +/- 0.1	917.2 [252]	-5.547
4-amino-4-methyl-2-pentanone	894.1 +/- 0.1	983.7	-9.112
6-amino-2-methyl-2-heptanol	935.8 +/- 1.1	895	4.556
N-methyldiethanolamine	1034.722 +/- 1.4	1029	0.556
Diethanolamine	1053.2 +/- 0.8	1090.4 [253]	-3.414
Diacetone alcohol	961.9 +/- 0.7	934	2.982

The density of 3-amino-1-propanol is accurately predicted using PolCA with an error of only 0.3 %. However, 4-amino-2-butanol and 6-amino-2-methyl-2-heptanol's densities are overpredicted by 2 % and 5 %, respectively. These results are surprising since these molecules have a similar structure, and PolCA can accurately predict 2-butanol's bulk properties and only slightly underpredicts the density of tert-butanol by 0.7 % (see section 3.3.6). These three molecules contain two non-adjacent α-CH_2 pseudo-atoms that are bonded to either a hydroxyl or an amino group. However, 3-amino-1-propanol has a linear structure, while the other two molecules are branched alkanolamines.

Concerning molecules with two adjacent α-carbon atoms, PolCA overpredicts 1-amino-2-propanol's density by 6 % and underpredicts 2-aminoethanol, 2-amino-2-methyl-1-propanol and diethanolamine's densities by 3 %, 5 % and 3 %, respectively. As explained in section 6.2.2, these molecules were simulated using an explicit 1-4 LJ intramolecular interaction due to the strong attraction between the amino and hydroxyl groups, and the parameters used for the dihedral O-CH_2-CH_2-N are an approximation. Orozco et al. [248] and Alejandre et al. [247] found that 2-aminoethanol's intermolecular hydrogen bonds are mainly formed between the hydroxyl hydrogen atoms and the oxygen or nitrogen atoms. Hence, an underestimation of the hydroxyl and amino groups' interactions would explain the alkanolamines' underpredicted densities and the less negative solvation free energies of amines in octanol (figure 6.3). Meanwhile, N-

methyldiethanolamine's density is only slightly overpredicted by 0.6 %, possibly due to the lack of hydrogen bonds between hydroxyl hydrogen atoms and the central nitrogen. Furthermore, a re-parametrisation of the amino or hydroxyl groups and the O-CH_2-CH_2-N dihedral could be needed to improve PolCA's performance. It is important to mention that 2-aminoethanol's enthalpy of vaporisation and self-solvation free energy were also calculated using PolCA and the predicted values differed from experimental data by -11.6 % and -18.6 % (e.i. the value predicted by PolCA is less negative than the experimental one), respectively.

The density of 4-amino-4-methyl-2-pentanone is significantly underpredicted, however, the density value used as a reference was described as a rough estimation in the Yaws Handbook [224], and no data was found elsewhere. Also, several dihedral parameters were missing for this molecule, and thus, approximations were used instead. Furthermore, we could not find experimental solvation free energies of ketones in amines or vice versa to assess PolCA's performance at modelling carbonyl and amino groups' interactions (aromatics amines and ketones were not an option due to the lack of PolCA parameters for aromatics).

Diacetone alcohol's density is overpredicted by 3 %. From Figure 6.4, we can see that PolCA overpredicts free energies of solvation of ketones in octanol (more negative values), and thus, interactions between ketone and alcohol groups are possibly being overestimated. Consequently, an overestimation of these functional groups' interaction would explain the higher density.

Lastly, the dielectric constants of 2-aminoethanol, N-methyldiethanolamine and diethanolamine obtained using PolCA are shown in Table 6.6. Once simulated values are corrected with Equation 2.52 to account for polarisation effects, PolCA overpredicts 2-aminoethanol and diethanolamine's dielectric constants by 47 % and 84 %, respectively. As seen in Table 6.2, these molecules have high estimated liquid dipoles, and thus, the calculated dipole moment scaling factors (μ_{liq}/μ_{model}) are high. Furthermore, 2-aminoethanol's low simulated dipole (lower than the experimental gas dipole) contributes to this problem. Due to the approximations made during Equation 2.49's development, it is possible that these molecules' liquid dipoles are being overestimated.

Since the dipole moment scaling factor is elevated to the power of two in Equation 2.52, an error in this value would lead to higher errors on the corrected dielectric constant. The same effect was observed when simulating ketones as explained in section 5.3.2. On the other hand, N-methyldiethanolamine's experimental dielectric constant falls inside the 95 % confidence interval of the predicted value, even though its estimated liquid dipole is 5.68 D. As explained above, this is probably due to the molecule interacting more like an alcohol molecule due to the buried nitrogen atom. It is also important to mention that the statistical error associated with N-methyldiethanolamine and diethanolamine's simulated dielectric constants obtained with ten simulations of 25 ns each is high, and longer simulations should be run for more precise results.

Table 6.6: Dielectric constants of 2-aminoethanol, N-methyldiethanolamine and diethanolamine at 298.15 K and 1 bar obtained using PolCA. The second column is the simulated dipole, and the last column is the percentual error between the corrected dielectric constant and experimental values.

	μ **(D)**	ϵ_{sim}	$\epsilon_{corrected}$	**Exp.**	**Error (%)**
2-aminoethanol	2.50	14.8 +/- 0.4	46.0 +/- 1.3	31.40 [254]	46.60
N-methyldiethanolamine	3.19	7.8 +/- 0.9	23.8 +/- 3.0	22.4 [87]	6.41
Diethanolamine	3.10	14.1 +/- 1.9	46.4 +/- 6.5	25.26	83.7

6.3.3 Other force fields' performance

In this section, PolCA's performance at predicting 2-aminoethanol's properties is compared to other well-known force fields using simulated data found in the literature. This molecule was chosen based on data availability for both experimental and simulated values. Cardona et al. [255] simulated 2-aminoethanol at 293 K using three different all-atom force fields: OPLS [139], GAFF [103] and MEAa [250], and the density, enthalpy of vaporisation and dielectric constant are among their calculated properties. Therefore, new simulations were run for 2-aminoethanol at 293 K using PolCA for a fair and clear comparison. Table 6.7 compares predicted densities and enthalpies of vaporisation obtained using the different models, and Table 6.8 shows the simulated and corrected dielectric constants. The latter property was obtained using Equation 2.52 and the reported simulated dipoles and dielectric constants.

Table 6.7: Density and enthalpy of vaporisation of 2-aminoethanol at 293 K without and with polarisation corrections obtained using different force fields. Simulated values for OPLS, GAFF and MEAa were taken from reference [255], and their corrected enthalpies of vaporisation (ΔH_{vap} (C)) were obtained by adding 2-aminoethanol's polarisation correction from Table 6.2 to the reported simulated values. The enthalpy of vaporisation experimental value was interpolated from data at 281.1 K and 294.0 K taken from reference [256].

	Density kg/m^3	**Error %**	**ΔH_{vap} kJ/mol**	**Error %**	**ΔH_{vap} (C) kJ/mol**	**Error %**
PolCA	986.7 +/- 0.2	-3.08	52.87 +/- 0.03	-11.59	50.81	-15.03
OPLS-AA	1024.12 +/- 0.06	0.601	59.222 +/- 0.002	-0.97	57.16	-4.41
GAFF	1130.6 +/- 0.1	11.06	85.765 +/- 0.008	43.42	83.71	39.97
MEAa	1053.56 +/- 0.02	3.49	54.14 +/- 0.004	-9.46	52.08	-12.91
Experimental	**1018** [257]		**59.8**		**59.8**	

Table 6.8: Simulated dipole moment and dielectric constant of 2-aminoethanol at 293 K obtained using different force fields. Simulated values for OPLS, GAFF and MEAa were taken from reference [255] and they were used to calculate the corrected dielectric constant ($\epsilon_{corrected}$). The experimental dielectric constant was estimated as the average of the dielectric constants at 288 K and 298 K taken from reference [258].

	μ (D)	ϵ_{sim}	$\epsilon_{corrected}$	**Error (%)**
PolCA	2.491 +/- 0.002	14.7 +/- 0.5	46.0 +/- 1.5	34.54
OPLS-AA	3.462 +/- 0.008	50.7 +/- 0.9	84.6 +/- 1.5	145.25
GAFF	3.002 +/- 0.004	40.9 +/- 0.9	90.2 +/- 2.0	161.44
MEAa	1.994 +/- 0.004	10.5 +/- 0.2	49.65 +/- 1.0	43.9
Experimental			**34.5**	

OPLS-AA can accurately predict the density and enthalpy of vaporisation of 2-aminoethanol, but it significantly overpredicts the dielectric constant (47 %). The addition of polarisation corrections worsens OPLS-AA's performance, with percentual errors of - 4 % and 145 % for the enthalpy of vaporisation and dielectric constant, respectively. A similar effect is observed when polarisation corrections are added to the GAFF model's dielectric constant (percentual error increases from 19 % to 161 %), however, the enthalpy of vaporisation predicted by GAFF with corrections is slightly better than that without corrections but still far of the experimental value with a percentual error of 40 %. GAFF also significantly overpredicts the density by 11 %, and thus, this model has the worst performance out of the other four force fields.

PolCA does the best job at estimating the dielectric constant; however, all the models fail to reproduce this property with reasonable accuracy. Also, PolCA's density is slightly better than the one obtained with MEAa. On the other hand, MEAa is better than PolCA for the enthalpy of vaporisation, however, the difference between the two models' enthalpies of vaporisation is only 1.27 kJ/mol. Also, the addition of polarisation corrections worsens both models' performances. MEAa is a model specific for 2-aminoethanol that uses GAFF's bonds, angles and LJ parameters with new values for the charges and dihedrals. The charges were first obtained from gas-phase QM simulations and then scaled by 0.96 to reproduce the density and enthalpy of vaporisation at 333 K. However, the model's developers could not accurately predict both properties. [250] The use of reduced gas-phase charges explains MEAa's low dipole moment (1.994 D), which is lower than 2-aminoethanol's experimental gas dipole (2.27 D).

6.4 Conclusions

Free energies of solvation of alkanes in octanol are accurately predicted using PolCA; however, this model greatly underpredicts the absolute value of free energies of solvation of amines in octanol. Also, PolCA predicts free energies of solvation of ketones in octanol and alcohols in butanone that are significantly more negative that the corresponding experimental values.

Concerning multifunctional molecules, PolCA can accurately predict 3-amino-1-propanol and N-methyldiethanolamine's densities. However, it fails to predict other alkanolamines' densities studied here with percentual errors that range from 2.2 to 5.9 % in magnitude. Also, 4-amino-4-methyl-2-pentanone's density is greatly underpredicted by this model.

Nonetheless, it is still important to emphasise that for 2-aminoethanol, PolCA performs significantly better than GAFF, which is an all-atom force field, and has a similar performance to that of MEAa, a model developed specifically for this molecule.

Based on the results described above, it seems like PolCA cannot correctly capture the

interactions between amino and hydroxyl groups, and a re-parameterisation is needed to improve its performance. However, the liquid dipoles estimated here for alkanolamines are suspiciously high, and thus, they should be validated using more accurate methods before re-parametrising PolCA. If the new polarisation corrections differ from the ones used in this thesis, the properties presented in this chapter should be re-calculated. Additionally, if the new polarisation corrections are not in agreement with those used during PolCA's development, a re-parameterisation of our model will be needed. Lastly, a new alkanolamine dihedral might be needed to improve PolCA's performance.

Chapter 7

Conclusions and Future Work

The main goal of this thesis was to develop a non-polarisable force field able to accurately predict solvation free energies in solvents of different polarity through the use of *post-facto* polarisation corrections during the model's parameterisation and validation. We hypothesised that the inclusion of these corrections would allow us to create a transferable model capable of predicting bulk liquid properties (density and diffusion), dielectric properties and properties that involve a phase change (enthalpies of vaporisation and solvation free energies) with a high degree of accuracy. Being able to accurately predict all these properties is not a trivial task since, as explained in Chapter 1, classical force fields struggle to simultaneously predict bulk liquid and dielectric properties and do not usually perform well in environments that significantly differ from those used during their development.

In this thesis, we have studied the effect of adding *post-facto* polarisation corrections to existing non-polarisable force fields, proposed a simple and semi-automated process for force field development and created a new force field for alcohols, amines and ketones called PolCA. These outcomes and their potential impact on different fields are discussed below.

Adding polarisation corrections to existing non-polarisable force fields

The impact of including *post-facto* polarisation corrections to the alcohol and ketone TraPPE-UA force field was analysed in Chapter 3 and Chapter 5. This force field was

fitted to pure vapour-liquid equilibrium curves, and polarisation corrections were not included in its development. However, phase change properties' polarisation corrections for the alcohol and ketone molecules used during TraPPE's parameterisation are quite small, and thus, it is likely that their omission did not significantly affect the optimum parameters. On the contrary, polarisation corrections for the dielectric constant and solvation free energy in hexadecane are significant, and thus, their inclusion has a high impact on the model's performance. Once polarisation corrections are added, TraPPE's performance at predicting alcohols' dielectric properties and solvation free energies in hexadecane significantly improves, and its predicted values are in very good agreement with experimental data. On the other hand, TraPPE with polarisation corrections fails to reproduce ketones' solvation free energies in hexadecane and dielectric constants. The solvation free energy of ketones in hexadecane is the only property studied in this thesis for which TraPPE performs worse with polarisation corrections than without them. TraPPE with polarisation corrections significantly overpredicts ketones' dielectric constants with an RMSD equal to 6.13, however, simulated values greatly underpredict this property (RMSD = 7.82) if no polarisation corrections are added. Furthermore, TraPPE's inability to predict ketones' dielectric constant could be due to an overestimation of the dipole moment scaling factor used to correct simulated dielectric constants since their estimated liquid dipoles are suspiciously high (4.72 D for propanone). Consequently, the polarisation corrections used in this thesis should first be validated using more accurate methods as they were obtained using several approximations and have been mostly proposed and tested for water.

The performance of the Gromos-2016H66 force field for amines was tested in chapter 4. This model does a very good job at predicting primary amines' dielectric constants, and its performance improves once the simulated values are corrected to account for polarisation effects. Gromos-2016H66 with polarisation corrections accurately predicts the dielectric constant of ethylamine to decylamine and overpredicts methylamine's dielectric constant by a factor of 1.4. Also, the model's performance at predicting solvation free energies in hexadecane improves when polarisation corrections are applied. However, from methylamine to propylamine, Gromos-2016H66 still predicts values that are

significantly less negative than the experimental ones. On the other hand, polarisation corrections for the enthalpy of vaporisation and self-solvation free energy are relatively small, and thus, they do not have a big impact on the model's performance.

As explained in Chapter 1, Jorge and Lue [87] calculated dielectric constants for many compounds using several non-polarisable force fields' simulated data with *post-facto* polarisation corrections and noticed that agreement with experiment significantly improves once polarisation corrections are applied. However, they used a fitted scaling factor equal to 1.26 instead of the ratio between the liquid and gas dipoles. This scaling factor, obtained by fitting to methanol's dielectric constant, is not optimal for all the compounds' classes (e.g., a scaling factor of 1 is better for predicting hydrocarbons, halogenated hydrocarbons, ethers, and alkylamines' dielectric constants), and it can vary for different force fields. In addition, Cardona et al. [259] studied the static dielectric constant of benzene, water, ethanol and three binary mixtures at different temperatures, and they found that simulated values matched experimental data once polarisation corrections were applied. They used a temperature independent charge scaling factor which varied for each force field and system. On the other hand, the scaling factor used in this thesis is independent of the force field, however, our liquid dipoles are estimated using an approximated equation that needs experimental dielectric constants as input parameters. Hence, dielectric constants of compounds for which there is no available experimental data need to be corrected using liquid dipoles of similar molecules. Also, each molecule's scaling factor should be obtained from first-principles calculations to increase the accuracy of the correction scheme used in this thesis.

Force field optimisation routine

The optimisation routine used in this thesis is straightforward and easily transferable to other molecules. The use of meta-models that are able to predict how the target properties change with the input variables allow the user to explore a larger parameter space and to partially automatise the process. Additionally, multiple force fields can be obtained from the same learning set if different objective functions or *post-facto* polarisation corrections are used during their development. However, the meta-models in this project were obtained using a second-order equation with cross interaction terms,

which worked well only for a small region of input parameters, and thus, the learning sets used for each meta-model had to be chosen manually to assure the desired level of accuracy. In future work, machine learning can be used to fully automate the optimisation routine. One option could be to use several independent second-order models for the different input space's regions, and automatically identify these models and their corresponding regions using machine learning. As mentioned in Chapter 1, a similar method has been applied by Chinta and Rengaswamy [23] to create a QSPR method for the prediction of drug solubility in binary solvent systems.

Further work needs to be done to find the optimal objective function for force field development. During PolCA's parameterisation, three potential amine models were fitted to the density, enthalpy of vaporisation and dielectric constant using different training sets. Even though all the models performed similarly for the fitted properties, the chosen model was the only one able to accurately predict free energies of self-solvation and solvation free energies in hexadecane. This outcome highlights that there is more than one set of parameters able to fit the target properties, however, many of these parameter sets likely have an inappropriate balance between the LJ and Coulomb interactions. A model's performance at predicting bulk liquid properties and solvation free energies in non-polar solvents once polarisation corrections are added is likely a good indicator of the balance between these two types of interactions. If all the particles in the non-polar solvent are zero (like is the case for hexadecane modelled with PolCA), only the LJ component of the solvation free energy needs to be simulated and thus, the solute's partial charges do not affect the calculated values. However, to successfully assess the model's description of the LJ interactions, the repolarisation of the solute and its interaction with the electronic continuum around it need to be accurately accounted for by the polarisation corrections. Consequently, it might be possible to fit σ and ε to solvation free energies in non-polar solvents, and then use these parameters to obtain the partial charges by fitting to liquid bulk properties.

There is a need for additional experimental data for the development and validation of new force fields, especially solvation free energies in solvents of different polarity. A lot of time and effort was spent searching for experimental properties and thus, force fields'

development could be sped up if there was a standardised database with bulk liquid properties, dielectric properties and solvation free energies for each functional group. Also, it is important to encourage researchers to measure experimental properties for compounds that might not be relevant for industrial purposes but that are interesting from a structural point of view, like diffusion constants of amines at 298 K and 1 bar. In the meantime, it might be possible to use properties predicted using other computational methods, like statistical methods or group-contribution equations of state, as "experimental properties".

PolCA

The force field developed in this thesis is called PolCA, standing for "Polarisation-Consistent Approach", and it is an extension of the UA model for aliphatic hydrocarbons proposed by Jorge [1]. The PolCA force field for alcohols was developed and validated in Chapter 3, and it can accurately predict methanol to decanol's densities, diffusion constants (except for methanol), enthalpies of vaporisation, free energies of self-solvation, dielectric constants and solvation free energies in hexadecane. Both TraPPE with polarisation corrections and PolCA do a very good job at predicting the studied properties, however, PolCA had a better performance overall for the alcohol molecules simulated in this thesis.

PolCA was extended to amines in Chapter 4. Our model does a very good job at predicting the densities, enthalpies of vaporisation and free energies of self-solvation of linear and branched primary amines, and its predicted solvation free energies of linear primary amines in hexadecane are in very good agreement with experimental data. PolCA performs better than Gromos-2016H66 for all the primary amine properties studied in that chapter, except for the dielectric constant. Our model greatly overpredicts the dielectric constant of methylamine and significantly overpredicts other linear and branched amines' dielectric constants even though this property was included in the parameterisation routine. Solvation free energies were not included as target properties in the optimisation routine, however, they were taken into account when choosing the PolCA's parameters.

In Chapter 5, PolCA was extended to ketones. PolCA can accurately predict the densities, enthalpies of vaporisation and diffusion constants of propanone to 2-decanone (except for butanone and 2-pentanone's densities which are underpredicted), however, these properties were included in the optimisation. On the other hand, TraPPE underpredicts the enthalpy of vaporisation and overpredicts the diffusion constant of these ketones. Both PolCA and TraPPE perform well at predicting free energies of self-solvation. This outcome was expected since TraPPE-UA was parameterised using acetone' vapour-liquid coexistence curve and PolCA was fitted to propanone and 2-decanone's self-solvation free energies. Also, PolCA greatly overpredicts ketones' dielectric constants and solvation free energies in hexadecane (more negative values), and thus, PolCA's parameters for ketones should be re-optimised. The same results were observed for TraPPE once polarisation corrections were added, as explained above.

Lastly, PolCA's transferability was tested in Chapter 6, and the results obtained there also suggest that some of the polarisation corrections used in this work are perhaps not entirely accurate, given the approximations involved. PolCA can accurately predict solvation free energies of alkanes in octanol, however, it fails to reproduce solvation free energies of amines and ketones in octanol. Consequently, the polarisation corrections used in this thesis need to be tested using more accurate methods, like the QM/MM method proposed by Jorge et al. [245], and if they differ from those used here, PolCA should be re-parameterised using these corrections. Depending on the value of the new corrections, it might be possible to construct the new meta-models using the same simulations performed in this thesis. If this is the case or if only a few extra simulations are needed, PolCA re-parameterisation should not take a significant amount of time or effort. However, if the meta-models' learning sets need to be extended significantly, the computational time required for the meta-models' construction could be considerably decreased by using fewer fitted parameters. Partial charges could be obtained from QM calculations and then scaled using a fitted parameter. In Chapter 3, a similar idea was implemented during the development of one of the potential alcohol models, however, this model used TraPPE-UA's partial charges instead of high-level QM charges and its polarisation corrections were obtained using several approximations.

In conclusion, the approach used in this thesis is very promising, however, we could not obtain a force field able to simultaneously predict solvation free energies in hexadecane and octanol, and thus a re-parameterisation of PolCA should be carried out in future work using polarisation corrections obtained with more accurate methods. Nonetheless, this work provides a semi-automated and efficient approach for force field development and highlights the benefits of including *post-facto* polarisation corrections to non-polarisable force fields. Agreement with experiment significantly improves when these corrections are applied to simulated dielectric constants of a wide range of compounds [87, 259], and solvation free energies of alcohols and amines in hexadecane are significantly benefited by the use of these corrections, as demonstrated in this thesis. Transfer free energies from water to hexadecane can be used to estimate water permeabilities through lipid bilayers [75], hence, it is vital for biological force fields to accurately predict solvation free energies in solvents of different polarity. Many non-polarisable force fields struggle to predict water/hexadecane transfer free energies [75], and thus, the inclusion of *post-facto* polarisation corrections could have a key impact on the study of biological systems and drug development. In future work, hydration free energies should be analysed using polarisation corrections. Polarisation corrections for water's enthalpy of vaporisation and self-solvation free energy, obtained using QM calculations, are close to zero [108]. Therefore, non-polarisable water models fitted to a few structural and thermodynamic properties without including the Berendsen correction (e.g. SPC, TIP3P, TIP4P) could be compatible with PolCA, however, these models do not properly account for long-range electrostatic interactions. Newer water models use Ewald summation to treat long-range electrostatic interactions, but they were parameterised using the Berendsen correction for the enthalpy of vaporisation, which was close to 5 kJ/mol [108]. Consequently, a re-parameterisation of these water models might be necessary. Furthermore, in this thesis we only focus on polarisation corrections for the dielectric constant, the solvation free energy in different solvents and the enthalpy of vaporisation. However, other properties like vapour-liquid and liquid-liquid equilibria, surface tension and electrical conductivity would also benefit from the addition of *post-facto* polarisation corrections, and thus, future work needs to focus on how to obtain and implement their respective corrections.

Bibliography

[1] Miguel Jorge. Predicting hydrophobic solvation by molecular simulation: 2. New united-atom model for alkanes, alkenes, and alkynes. *Journal of Computational Chemistry*, 38(6):359–369, 2017.

[2] John D Chodera, David L Mobley, Michael R Shirts, Richard W Dixon, Kim Branson, and Vijay S Pande. Alchemical free energy methods for drug discovery: progress and challenges. *Current Opinion in Structural Biology*, 21(2):150–160, 2011.

[3] Thorsteinn Loftsson and Marcus E Brewster. Pharmaceutical applications of cyclodextrins: basic science and product development. *Journal of Pharmacy and Pharmacology*, 62(11):1607–1621, 2010.

[4] Yung-Chi Lee, Philip D Zocharski, and Brian Samas. An intravenous formulation decision tree for discovery compound formulation development. *International Journal of Pharmaceutics*, 253(1-2):111–119, 2003.

[5] Kai Lüder, Lennart Lindfors, Jan Westergren, Sture Nordholm, and Roland Kjellander. In silico prediction of drug solubility. 3. Free energy of solvation in pure amorphous matter. *Journal of Physical Chemistry B*, 111(25):7303–7311, 2007.

[6] Ketan T Savjani, Anuradha K Gajjar, and Jignasa K Savjani. Drug solubility: importance and enhancement techniques. *ISRN Pharmaceutics*, 2012, 2012.

[7] Albert Leo and DH Hoekman. *Exploring QSAR: Fundamentals and applications in chemistry and biology*, volume 1. An American Chemical Society Publication, 1995.

[8] Kai Lüder, Lennart Lindfors, Jan Westergren, Sture Nordholm, and Roland Kjellander. In silico prediction of drug solubility: 2. Free energy of solvation in pure melts. *Journal of Physical Chemistry B*, 111(7):1883–1892, 2007.

[9] Antonio Llinas, Robert C Glen, and Jonathan M Goodman. Solubility challenge: can you predict solubilities of 32 molecules using a database of 100 reliable measurements? *Journal of Chemical Information and Modeling*, 48(7):1289–1303, 2008.

[10] Cameron J Brown, Thomas McGlone, Stephanie Yerdelen, Vijay Srirambhatla, Fraser Mabbott, Rajesh Gurung, Maria L Briuglia, Bilal Ahmed, Hector Polyzois, John McGinty, et al. Enabling precision manufacturing of active pharmaceutical ingredients: Workflow for seeded cooling continuous crystallisations. *Molecular Systems Design and Engineering*, 3(3):518–549, 2018.

[11] Jan Westergren, Lennart Lindfors, Tobias Höglund, Kai Lüder, Sture Nordholm, and Roland Kjellander. In silico prediction of drug solubility: 1. Free energy of hydration. *Journal of Physical Chemistry B*, 111(7):1872–1882, 2007.

[12] Son Tung Ngo, Hung Minh Nguyen, Pham Minh Quan, Vi Khanh Truong, Nguyen Thanh Tung, Van V Vu, et al. Assessing potential inhibitors of SARS-CoV-2 main protease from available drugs using free energy perturbation simulations. *RSC Advances*, 10(66):40284–40290, 2020.

[13] Christel AS Bergström and Per Larsson. Computational prediction of drug solubility in water-based systems: Qualitative and quantitative approaches used in the current drug discovery and development setting. *International Journal of Pharmaceutics*, 540(1-2):185–193, 2018.

[14] RE Skyner, JL McDonagh, CR Groom, T Van Mourik, and JBO Mitchell. A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Physical Chemistry Chemical Physics*, 17(9):6174–6191, 2015.

[15] Christian L Silveira, Alessandro C Galvão, Weber S Robazza, and João Vic-

tor T Feyh. Modeling and parameters estimation for the solubility calculations of nicotinamide using UNIFAC and COSMO-based models. *Fluid Phase Equilibria*, 535:112970, 2021.

[16] Simon Dufal, Vasileios Papaioannou, Majid Sadeqzadeh, Thomas Pogiatzis, Alexandros Chremos, Claire S Adjiman, George Jackson, and Amparo Galindo. Prediction of thermodynamic properties and phase behavior of fluids and mixtures with the SAFT-γ Mie group-contribution equation of state. *Journal of Chemical and Engineering Data*, 59(10):3272–3288, 2014.

[17] Shuai Liu, Shannon Cao, Kevin Hoang, Kayla L Young, Andrew S Paluch, and David L Mobley. Using MD simulations to calculate how solvents modulate solubility. *Journal of Chemical Theory and Computation*, 12(4):1930–1941, 2016.

[18] Jürgen Gmehling, Michael Kleiber, Bärbel Kolbe, and Jürgen Rarey. *Chemical thermodynamics for process simulation*. Wiley Online Library, 2019.

[19] Neera Jain and Samuel H Yalkowsky. Estimation of the aqueous solubility I: Application to organic nonelectrolytes. *Journal of Pharmaceutical Sciences*, 90(2):234–252, 2001.

[20] Garrett B Goh, Nathan O Hodas, and Abhinav Vishnu. Deep learning for computational chemistry. *Journal of Computational Chemistry*, 38(16):1291–1307, 2017.

[21] PR Duchowicz. QSPR studies on water solubility, octanol-water partition coefficient and vapour pressure of pesticides. *SAR and QSAR in Environmental Research*, 31(2):135–148, 2020.

[22] Soumaya Kherouf, Nabil Bouarra, Amel Bouakkadia, and Djelloul Messadi. Modeling of linear and nonlinear quantitative structure property relationships of the aqueous solubility of phenol derivatives. *Journal of the Serbian Chemical Society*, 84(6):575–590, 2019.

[23] Sivadurgaprasad Chinta and Raghunathan Rengaswamy. Machine learning derived quantitative structure property relationship (QSPR) to predict drug solu-

bility in binary solvent systems. *Industrial and Engineering Chemistry Research*, 58(8):3082–3092, 2019.

[24] Aage Fredenslund. *Vapor-liquid equilibria using UNIFAC: a group-contribution method*. Elsevier, 1986.

[25] Aage Fredenslund, Russell L Jones, and John M Prausnitz. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal*, 21(6):1086–1099, 1975.

[26] Jürgen Lohmann, Ralph Joh, and Jürgen Gmehling. From UNIFAC to modified UNIFAC (dortmund). *Industrial and Engineering Chemistry Research*, 40(3):957–964, 2001.

[27] Arnold Aaron Bondi et al. Physical properties of molecular crystals liquids, and glasses. 1968.

[28] Ulrich Weidlich and Juergen Gmehling. A modified UNIFAC model. 1. Prediction of VLE, he, and gamma. infin. *Industrial and Engineering Chemistry Research*, 26(7):1372–1381, 1987.

[29] Zhigang Lei, Chengna Dai, Xing Liu, Li Xiao, and Biaohua Chen. Extension of the UNIFAC model for ionic liquids. *Industrial and Engineering Chemistry Research*, 51(37):12135–12144, 2012.

[30] Brock C Roughton, Brianna Christian, John White, Kyle V Camarda, and Rafiqul Gani. Simultaneous design of ionic liquid entrainers and energy efficient azeotropic separation processes. *Computers and Chemical Engineering*, 42:248–262, 2012.

[31] Yuqiu Chen, Xinyan Liu, John M Woodley, and Georgios M Kontogeorgis. Gas solubility in ionic liquids: UNIFAC-IL model extension. *Industrial and Engineering Chemistry Research*, 59(38):16805–16821, 2020.

[32] Hans Grensemann and Jürgen Gmehling. Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods. *Industrial and Engineering Chemistry Research*, 44(5):1610–1624, 2005.

[33] Jaber Yousefi Seyf and Abolfazl Shojaeian. Vapor-liquid (azeotropic systems) and liquid-liquid equilibrium calculations using UNIFAC and NRTL-SAC activity coefficient models. *Fluid Phase Equilibria*, 494:33–44, 2019.

[34] Ericsem Pereira, Fernando Torres Junqueira, Antonio José de Almeida Meirelles, and Guilherme José Maximo. Prediction of the melting behavior of edible fats using UNIFAC and UNIQUAC models. *Fluid Phase Equilibria*, 493:58–66, 2019.

[35] Miquel Garcia-Ratés and Frank Neese. Effect of the solute cavity on the solvation energy and its derivatives within the framework of the gaussian charge scheme. *Journal of Computational Chemistry*, 41(9):922–939, 2020.

[36] J Tomasi, Benedetta Mennucci, and E Cancès. The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM ab initio level. *Journal of Molecular Structure: THEOCHEM*, 464(1-3):211–226, 1999.

[37] Errol Lewars. *Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Springer, 2016.

[38] Jacopo Tomasi, Benedetta Mennucci, and Roberto Cammi. Quantum mechanical continuum solvation models. *Chemical Reviews*, 105(8):2999–3094, 2005.

[39] Aleksandr V Marenich, Christopher J Cramer, and Donald G Truhlar. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *Journal of Physical Chemistry B*, 113(18):6378–6396, 2009.

[40] Andreas Klamt. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *Journal of Physical Chemistry*, 99(7):2224–2235, 1995.

[41] Ryo Kato and Jürgen Gmehling. Systems with ionic liquids: Measurement of vle and γ^{∞} data and prediction of their thermodynamic behavior using original UNIFAC, mod. UNIFAC (do) and COSMO-RS (ol). *Journal of Chemical Thermodynamics*, 37(6):603–619, 2005.

[42] Andreas Klamt, Frank Eckert, Martin Hornig, Michael E Beck, and Thorsten Bürger. Prediction of aqueous solubility of drugs and pesticides with COSMO-RS. *Journal of Computational Chemistry*, 23(2):275–281, 2002.

[43] Oualid Alioui, Yacine Benguerba, and Inas M Alnashef. Investigation of the CO2-solubility in deep eutectic solvents using COSMO-RS and molecular dynamics methods. *Journal of Molecular Liquids*, 307:113005, 2020.

[44] Weiguo Hu, Zeren Shang, Ning Wei, Baohong Hou, Junbo Gong, and Yan Wang. Solubility of benorilate in twelve monosolvents: Determination, correlation and COSMO-RS analysis. *Journal of Chemical Thermodynamics*, 152:106272, 2021.

[45] Thomas Holderbaum and J Gmehling. PSRK: A group contribution equation of state based on UNIFAC. *Fluid Phase Equilibria*, 70(2-3):251–265, 1991.

[46] Giorgio Soave. Equilibrium constants from a modified Redlich-Kwong equation of state. *Chemical Engineering Science*, 27(6):1197–1203, 1972.

[47] Jiding Li, Magnus Topphoff, Kai Fischer, and Jürgen Gmehling. Prediction of gas solubilities in aqueous electrolyte systems using the predictive Soave- Redlich-Kwong model. *Industrial and Engineering Chemistry Research*, 40(16):3703–3710, 2001.

[48] Vasileios Papaioannou, Thomas Lafitte, Carlos Avendano, Claire S Adjiman, George Jackson, Erich A Müller, and Amparo Galindo. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments. *Journal of Chemical Physics*, 140(5):054107, 2014.

[49] Walter G Chapman, Keith E Gubbins, George Jackson, and Maciej Radosz. SAFT: Equation-of-state solution model for associating fluids. *Fluid Phase Equilibria*, 52:31–38, 1989.

[50] Vasileios Papaioannou. *A Molecular-based Group Contribution Equation of State for the Description of Fluid Phase Behaviour and Thermodynamic Derivative*

Properties of Mixtures (SAFT-γ Mie). PhD thesis, Imperial College London, 2013.

[51] Sadia Rahman, Olga Lobanova, Guadalupe Jiménez-Serratos, Carlos Braga, Vasilios Raptis, Erich A Müller, George Jackson, Carlos Avendaño, and Amparo Galindo. SAFT-γ force field for the simulation of molecular fluids. 5. Hetero-group coarse-grained models of linear alkanes and the importance of intramolecular interactions. *Journal of Physical Chemistry B*, 122(39):9161–9177, 2018.

[52] Panatpong Hutacharoen, Simon Dufal, Vasileios Papaioannou, Ravi M Shanker, Claire S Adjiman, George Jackson, and Amparo Galindo. Predicting the solvation of organic compounds in aqueous environments: From alkanes and alcohols to pharmaceuticals. *Industrial and Engineering Chemistry Research*, 56(38):10856–10876, 2017.

[53] Fritz London. The general theory of molecular forces. *Transactions of the Faraday Society*, 33:8b–26, 1937.

[54] Carlos Avendano, Thomas Lafitte, Amparo Galindo, Claire S Adjiman, George Jackson, and Erich A Müller. SAFT-γ force field for the simulation of molecular fluids. 1. A single-site coarse grained model of carbon dioxide. *The Journal of Physical Chemistry B*, 115(38):11154–11169, 2011.

[55] Oliver L Watson, Amparo Galindo, George Jackson, and Claire S Adjiman. Computer-aided design of solvent blends for the cooling and anti-solvent crystallisation of ibuprofen. In *Computer Aided Chemical Engineering*, volume 46, pages 949–954. Elsevier, 2019.

[56] Amulya K Pervaje, Christopher C Walker, and Erik E Santiso. Molecular simulation of polymers with a SAFT-γ Mie approach. *Molecular Simulation*, 45(14-15):1223–1241, 2019.

[57] Jos Tasche, Elise FD Sabattié, Richard L Thompson, Mario Campana, and Mark R Wilson. Oligomer/polymer blend phase diagram and surface concentration profiles for squalane/polybutadiene: Experimental measurements and pre-

dictions from SAFT-γ Mie and molecular dynamics simulations. *Macromolecules*, 53(7):2299–2309, 2020.

[58] Christopher C Walker, Jan Genzer, and Erik E Santiso. Extending the fused-sphere SAFT-γ Mie force field parameterization approach to poly (vinyl butyral) copolymers. *Journal of Chemical Physics*, 152(4):044903, 2020.

[59] Christopher C Walker, Jan Genzer, and Erik E Santiso. Development of a fused-sphere SAFT-γ Mie force field for poly (vinyl alcohol) and poly (ethylene). *Journal of Chemical Physics*, 150(3):034901, 2019.

[60] Michael P Allen and Dominic J Tildesley. *Computer simulation of liquids*. Oxford university press, 2017.

[61] Kai Lüder, Lennart Lindfors, Jan Westergren, Sture Nordholm, Rasmus Persson, and Mikaela Pedersen. In silico prediction of drug solubility: 4. Will simple potentials suffice? *Journal of Computational Chemistry*, 30(12):1859–1871, 2009.

[62] Michael J Schnieders, Jonas Baltrusaitis, Yue Shi, Gaurav Chattree, Lianqing Zheng, Wei Yang, and Pengyu Ren. The structure, thermodynamics, and solubility of organic crystals from simulation with a polarizable force field. *Journal of Chemical Theory and Computation*, 8(5):1721–1736, 2012.

[63] Fanfan Fan, Haikuan Yuan, Yonghao Feng, Fengjiao Liu, Lijuan Zhang, Xijian Liu, Xueyan Zhu, Wei An, Sohrab Rohani, and Jie Lu. Molecular simulation approaches for the prediction of unknown crystal structures and solubilities of (R)-and (R, S)-crizotinib in organic solvents. *Crystal Growth and Design*, 19(10):5882–5895, 2019.

[64] Nuno M Garrido, António J Queimada, Miguel Jorge, Eugénia A Macedo, and Ioannis G Economou. 1-Octanol/water partition coefficients of n-alkanes from molecular simulations of absolute solvation free energies. *Journal of Chemical Theory and Computation*, 5(9):2436–2446, 2009.

[65] Caitlin C Bannan, Gaetano Calabró, Daisy Y Kyu, and David L Mobley. Cal-

culating partition coefficients of small molecules in octanol/water and cyclohexane/water. *Journal of Chemical Theory and Computation*, 12(8):4015–4024, 2016.

[66] Haiyang Zhang, Yang Jiang, Ziheng Cui, and Chunhua Yin. Force field benchmark of amino acids. 2. Partition coefficients between water and organic solvents. *Journal of Chemical Information and Modeling*, 58(8):1669–1681, 2018.

[67] Andrew S Paluch, Sreeja Parameswaran, Shuai Liu, Anasuya Kolavennu, and David L Mobley. Predicting the excess solubility of acetanilide, acetaminophen, phenacetin, benzocaine, and caffeine in binary water/ethanol mixtures via molecular simulation. *Journal of Chemical Physics*, 142(4):044508, 2015.

[68] David L Mobley, Christopher I Bayly, Matthew D Cooper, Michael R Shirts, and Ken A Dill. Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations. *Journal of Chemical Theory and Computation*, 5(2):350–358, 2009.

[69] Bin Chen, Jeffrey J Potoff, and J Ilja Siepmann. Monte Carlo calculations for alcohols and their mixtures with alkanes. Transferable potentials for phase equilibria. 5. United-atom description of primary, secondary, and tertiary alcohols. *Journal of Physical Chemistry B*, 105(15):3093–3104, 2001.

[70] Noor Asidah Mohamed, Richard T Bradshaw, and Jonathan W Essex. Evaluation of solvation free energies for small molecules with the AMOEBA polarizable force field. *Journal of Computational Chemistry*, 37(32):2749–2758, 2016.

[71] Devleena Shivakumar, Joshua Williams, Yujie Wu, Wolfgang Damm, John Shelley, and Woody Sherman. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *Journal of Chemical Theory and Computation*, 6(5):1509–1519, 2010.

[72] Wenkun Wu and John Kieffer. New hybrid method for the calculation of the solvation free energy of small molecules in aqueous solutions. *Journal of Chemical Theory and Computation*, 15(1):371–381, 2018.

[73] Sathish Dasari and Bhabani S Mallik. Solubility and solvation free energy of

a cardiovascular drug, LASSBio-294, in ionic liquids: a computational study. *Journal of Molecular Liquids*, 301:112449, 2020.

[74] Sadra Kashefolgheta, Marina P Oliveira, Salomé R Rieder, Bruno AC Horta, William E Acree Jr, and Philippe H Hünenberger. Evaluating classical force fields against experimental cross-solvation free energies. *Journal of Chemical Theory and Computation*, 16(12):7556–7580, 2020.

[75] Andreas Krämer, Frank C Pickard IV, Jing Huang, Richard M Venable, Andrew C Simmonett, Dirk Reith, Karl N Kirschner, Richard W Pastor, and Bernard R Brooks. Interactions of water and alkanes: Modifying additive force fields to account for polarization effects. *Journal of Chemical Theory and Computation*, 15(6):3854–3867, 2019.

[76] Nuno M Garrido, Miguel Jorge, António J Queimada, Eugénia A Macedo, and Ioannis G Economou. Using molecular simulation to predict solute solvation and partition coefficients in solvents of different polarity. *Physical Chemistry Chemical Physics*, 13(20):9155–9164, 2011.

[77] Bruno AC Horta, Patrick FJ Fuchs, Wilfred F van Gunsteren, and Philippe H Hünenberger. New interaction parameters for oxygen compounds in the GROMOS force field: Improved pure-liquid and solvation properties for alcohols, ethers, aldehydes, ketones, carboxylic acids, and esters. *Journal of Chemical Theory and Computation*, 7(4):1016–1031, 2011.

[78] Patrick FJ Fuchs, Halvor S Hansen, Philippe H Hünenberger, and Bruno AC Horta. A GROMOS parameter set for vicinal diether functions: properties of polyethyleneoxide and polyethyleneglycol. *Journal of Chemical Theory and Computation*, 8(10):3943–3963, 2012.

[79] Bruno AC Horta, Zhixiong Lin, Wei Huang, Sereina Riniker, Wilfred F Van Gunsteren, and Philippe H Hünenberger. Reoptimized interaction parameters for the peptide-backbone model compound N-methylacetamide in the GROMOS force field: Influence on the folding properties of two beta-peptides in methanol. *Journal of Computational Chemistry*, 33(24):1907–1917, 2012.

[80] Bruno AC Horta, Pascal T Merz, Patrick FJ Fuchs, Jozica Dolenc, Sereina Riniker, and Philippe H Hünenberger. A GROMOS-compatible force field for small organic molecules in the condensed phase: The 2016h66 parameter set. *Journal of Chemical Theory and Computation*, 12(8):3825–3850, 2016.

[81] Miguel Jorge, Nuno M Garrido, Carlos JV Simões, Cândida G Silva, and Rui MM Brito. Predicting hydrophobic solvation by molecular simulation: 1. Testing united-atom alkane models. *Journal of Computational Chemistry*, 38(6):346–358, 2017.

[82] Martin Stroet, Bertrand Caron, Koen M Visscher, Daan P Geerke, Alpeshkumar K Malde, and Alan E Mark. Automated Topology Builder version 3.0: Prediction of solvation free enthalpies in water and hexane. *Journal of Chemical Theory and Computation*, 14(11):5834–5845, 2018.

[83] David L Mobley, Karisa L Wymer, Nathan M Lim, and J Peter Guthrie. Blind prediction of solvation free energies from the SAMPL4 challenge. *Journal of Computer-aided Molecular Design*, 28(3):135–150, 2014.

[84] Igor Leontyev and Alexei Stuchebrukhov. Electronic continuum model for molecular dynamics simulations. *Journal of Chemical Physics*, 130(8):085102, 2009.

[85] Christopher M Baker. Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 5(2):241–254, 2015.

[86] Igor Leontyev and Alexei Stuchebrukhov. Accounting for electronic polarization in non-polarizable force fields. *Physical Chemistry Chemical Physics*, 13(7):2613–2626, 2011.

[87] Miguel Jorge and Leo Lue. The dielectric constant: Reconciling simulation and experiment. *Journal of Chemical Physics*, 150(8):084108, 2019.

[88] Piotr Cieplak, James Caldwell, and Peter Kollman. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl

acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *Journal of Computational Chemistry*, 22(10):1048–1057, 2001.

[89] Pengyu Ren and Jay W Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *Journal of Physical Chemistry B*, 107(24):5933–5947, 2003.

[90] Jay W Ponder, Chuanjie Wu, Pengyu Ren, Vijay S Pande, John D Chodera, Michael J Schnieders, Imran Haque, David L Mobley, Daniel S Lambrecht, Robert A DiStasio Jr, et al. Current status of the AMOEBA polarizable force field. *Journal of Physical Chemistry B*, 114(8):2549–2564, 2010.

[91] Thomas A Halgren and Wolfgang Damm. Polarizable force fields. *Current Opinion in Structural Biology*, 11(2):236–242, 2001.

[92] Igor Leontyev and Alexei Stuchebrukhov. Electronic polarizability and the effective pair potentials of water. *Journal of Chemical Theory and Computation*, 6(10):3153–3161, 2010.

[93] HJC Berendsen, JR Grigera, and TP Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91(24):6269–6271, 1987.

[94] Jose LF Abascal and Carlos Vega. A general purpose model for the condensed phases of water: TIP4P/2005. *Journal of Chemical Physics*, 123(23):234505, 2005.

[95] Hans W Horn, William C Swope, Jed W Pitera, Jeffry D Madura, Thomas J Dick, Greg L Hura, and Teresa Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *Journal of Chemical Physics*, 120(20):9665–9678, 2004.

[96] Saeed Izadi, Ramu Anandakrishnan, and Alexey V Onufriev. Building water models: A different approach. *Journal of Physical Chemistry Letters*, 5(21):3863–3871, 2014.

[97] Lee-Ping Wang, Todd J Martinez, and Vijay S Pande. Building force fields: An

automatic, systematic, and reproducible approach. *Journal of Physical Chemistry Letters*, 5(11):1885–1891, 2014.

[98] Samantha Weerasinghe and Paul E Smith. A Kirkwood- Buff derived force field for methanol and aqueous methanol solutions. *Journal of Physical Chemistry B*, 109(31):15080–15086, 2005.

[99] D Gonzalez-Salgado and C Vega. A new intermolecular potential for simulations of methanol: The OPLS/2016 model. *Journal of Chemical Physics*, 145(3):034508, 2016.

[100] William L Jorgensen. Optimized intermolecular potential functions for liquid alcohols. *Journal of Physical Chemistry*, 90(7):1276–1284, 1986.

[101] Chris Oostenbrink, Alessandra Villa, Alan E Mark, and Wilfred F Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry*, 25(13):1656–1676, 2004.

[102] Marcus G Martin and J Ilja Siepmann. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *Journal of Physical Chemistry B*, 102(14):2569–2577, 1998.

[103] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.

[104] Igor Leontyev and Alexei Stuchebrukhov. Polarizable molecular interactions in condensed phase and their equivalent nonpolarizable models. *Journal of Chemical Physics*, 141(1):06B621_1, 2014.

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

[106] YS Badyal, M-L Saboungi, DL Price, SD Shastri, DR Haeffner, and AK Soper.

Electron distribution in water. *Journal of Chemical Physics*, 112(21):9206–9208, 2000.

[107] Cui Zhang, Jun Wu, Giulia Galli, and François Gygi. Structural and vibrational properties of liquid water from van der Waals density functionals. *Journal of Chemical Theory and Computation*, 7(10):3054–3061, 2011.

[108] Andrew W Milne and Miguel Jorge. Polarization corrections and the hydration free energy of water. *Journal of Chemical Theory and Computation*, 15(2):1065–1078, 2018.

[109] Ardavan Farahvash, Igor Leontyev, and Alexei Stuchebrukhov. Dynamic and electronic polarization corrections to the dielectric constant of water. *Journal of Physical Chemistry A*, 122(48):9243–9250, 2018.

[110] Carlos Vega. Water: one molecule, two surfaces, one mistake. *Molecular Physics*, 113(9-10):1145–1163, 2015.

[111] Jeffery B Klauda, Richard M Venable, J Alfredo Freites, Joseph W O'Connor, Douglas J Tobias, Carlos Mondragon-Ramirez, Igor Vorobyov, Alexander D MacKerell Jr, and Richard W Pastor. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *Journal of Physical Chemistry B*, 114(23):7830–7843, 2010.

[112] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier, 2001.

[113] Robert Swendsen. *An introduction to statistical mechanics and thermodynamics*. Oxford University Press, USA, 2012.

[114] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[115] William L Jorgensen and Julian Tirado-Rives. Monte Carlo vs molecular dynamics for conformational sampling. *Journal of Physical Chemistry*, 100(34):14508–14513, 1996.

[116] M.J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team. *GROMACS User Manual version 5.0.7, www.gromacs.org*, 2015.

[117] William C Swope, Hans C Andersen, Peter H Berens, and Kent R Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *Journal of Chemical Physics*, 76(1):637–649, 1982.

[118] RW Hockney, SP Goel, and JW Eastwood. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics*, 14(2):148–158, 1974.

[119] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977.

[120] Hans C Andersen. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52(1):24–34, 1983.

[121] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[122] Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R Hutchison. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4(1):17, 2012.

[123] H Bekker, HJC Berendsen, EJ Dijkstra, S Achterop, R Van Drunen, D Van der Spoel, A Sijbers, H Keegstra, B Reitsma, and MKR Renardus. Gromacs: A parallel computer for molecular dynamics simulations. *Physics Computing*, 92:252–256, 1993.

[124] M Born and Th von Karman. Vibrations in space gratings (molecular frequencies). *Zeitschrift für Physik*, 13:297–309, 1912.

[125] Herman JC Berendsen, JPM van Postma, Wilfred F van Gunsteren, ARHJ Di-

Nola, and Jan R Haak. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81(8):3684–3690, 1984.

[126] Hans C Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *Journal of Chemical Physics*, 72(4):2384–2393, 1980.

[127] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[128] William G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

[129] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *Journal of Chemical Physics*, 126(1):014101, 2007.

[130] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.

[131] Raymond Chang et al. *Chemistry*. McGraw-Hill, 2010.

[132] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[133] AJ Stone. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chemical Physics Letters*, 83(2):233–239, 1981.

[134] Jon Applequist, James R Carl, and Kwok-Kueng Fung. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *Journal of the American Chemical Society*, 94(9):2952–2960, 1972.

[135] Michiel Sprik. Computer simulation of the dynamics of induced polarization fluctuations in water. *Journal of Physical Chemistry*, 95(6):2283–2291, 1991.

[136] Hanne S Antila and Emppu Salonen. Polarizable force fields. *Biomolecular Simulations*, pages 215–241, 2013.

[137] Siewert J Marrink, H Jelger Risselada, Serge Yefimov, D Peter Tieleman, and Alex H De Vries. The MARTINI force field: coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B*, 111(27):7812–7824, 2007.

[138] Lijiang Yang, Chun-hu Tan, Meng-Juei Hsieh, Junmei Wang, Yong Duan, Piotr Cieplak, James Caldwell, Peter A Kollman, and Ray Luo. New-generation AMBER united-atom force field. *Journal of Physical Chemistry B*, 110(26):13166–13176, 2006.

[139] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[140] Alex D MacKerell Jr, Donald Bashford, MLDR Bellott, Roland Leslie Dunbrack Jr, Jeffrey D Evanseck, Martin J Field, Stefan Fischer, Jiali Gao, H Guo, Sookhee Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*, 102(18):3586–3616, 1998.

[141] Junmei Wang, Piotr Cieplak, and Peter A Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21(12):1049–1074, 2000.

[142] Lukas D Schuler, Xavier Daura, and Wilfred F Van Gunsteren. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *Journal of Computational Chemistry*, 22(11):1205–1218, 2001.

[143] Tobias Kulschewski and Jürgen Pleiss. A molecular dynamics study of liquid aliphatic alcohols: simulation of density and self-diffusion coefficient using a modified OPLS force field. *Molecular Simulation*, 39(9):754–767, 2013.

[144] Robert J Good and Christopher J Hope. New combining rule for intermolecular

distances in intermolecular potential functions. *Journal of Chemical Physics*, 53(2):540–543, 1970.

[145] Huai Sun. COMPASS: An ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. *Journal of Physical Chemistry B*, 102(38):7338–7364, 1998.

[146] John M Stubbs, Jeffrey J Potoff, and J Ilja Siepmann. Transferable potentials for phase equilibria. 6. United-atom description for ethers, glycols, ketones, and aldehydes. *Journal of Physical Chemistry B*, 108(45):17596–17605, 2004.

[147] William L Jorgensen, Jeffry D Madura, and Carol J Swenson. Optimized intermolecular potential functions for liquid hydrocarbons. *Journal of the American Chemical Society*, 106(22):6638–6646, 1984.

[148] William Mendenhall, Robert J Beaver, and Barbara M Beaver. *Introduction to probability and statistics*. Cengage Learning, 2012.

[149] John G Kirkwood. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *Journal of Chemical Physics*, 2(7):351–361, 1934.

[150] John G Kirkwood. The dielectric polarization of polar liquids. *Journal of Chemical Physics*, 7(10):911–919, 1939.

[151] Lars Onsager. Electric moments of molecules in liquids. *Journal of the American Chemical Society*, 58(8):1486–1493, 1936.

[152] In-Chul Yeh and Gerhard Hummer. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *Journal of Physical Chemistry B*, 108(40):15873–15879, 2004.

[153] Martin Neumann. Dipole moment fluctuation formulas in computer simulations of polar systems. *Molecular Physics*, 50(4):841–858, 1983.

[154] Niels Hansen and Wilfred F Van Gunsteren. Practical aspects of free-energy calculations: A review. *Journal of Chemical Theory and Computation*, 10(7):2632–2647, 2014.

[155] Andrew Pohorille, Christopher Jarzynski, and Christophe Chipot. Good practices in free-energy calculations. *Journal of Physical Chemistry B*, 114(32):10235–10253, 2010.

[156] Thomas Steinbrecher, InSuk Joung, and David A Case. Soft-core potentials in thermodynamic integration: Comparing one-and two-step transformations. *Journal of Computational Chemistry*, 32(15):3253–3263, 2011.

[157] Thomas Steinbrecher, David L Mobley, and David A Case. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *Journal of Chemical Physics*, 127(21):214108, 2007.

[158] Miguel Jorge, Nuno M Garrido, António J Queimada, Ioannis G Economou, and Eugénia A Macedo. Effect of the integration method on the accuracy and computational efficiency of free energy calculations using thermodynamic integration. *Journal of Chemical Theory and Computation*, 6(4):1018–1027, 2010.

[159] Conrad Shyu and F Marty Ytreberg. Accurate estimation of solvation free energy using polynomial fitting techniques. *Journal of Computational Chemistry*, 32(1):134–141, 2011.

[160] Charles H Bennett. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, 22(2):245–268, 1976.

[161] Stefan Bruckner and Stefan Boresch. Efficiency of alchemical free energy simulations. I. A practical comparison of the exponential formula, thermodynamic integration, and Bennett's acceptance ratio method. *Journal of Computational Chemistry*, 32(7):1303–1319, 2011.

[162] Himanshu Paliwal and Michael R Shirts. A benchmark test set for alchemical free energy transformations and its use to quantify error in common free energy methods. *Journal of Chemical Theory and Computation*, 7(12):4115–4134, 2011.

[163] Stefan Bruckner and Stefan Boresch. Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *Journal of Computational Chemistry*, 32(7):1320–1333, 2011.

[164] Di Wu and David A Kofke. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. *Journal of Chemical Physics*, 123(5):054103, 2005.

[165] N Goga, AJ Rzepiela, AH De Vries, SJ Marrink, and HJC Berendsen. Efficient algorithms for langevin and dpd dynamics. *Journal of Chemical Theory and Computation*, 8(10):3637–3649, 2012.

[166] Thomas C Beutler, Alan E Mark, René C van Schaik, Paul R Gerber, and Wilfred F Van Gunsteren. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical Physics Letters*, 222(6):529–539, 1994.

[167] Christian L Wennberg, Teemu Murtola, Berk Hess, and Erik Lindahl. Lennard-Jones lattice summation in bilayer simulations has critical effects on surface tension and lipid properties. *Journal of Chemical Theory and Computation*, 9(8):3527–3537, 2013.

[168] Aleksandr V Marenich, Casey P Kelly, Jason D Thompson, Gregory D Hawkins, Candee C Chambers, David J Giesen, Paul Winget, Christopher J Cramer, and Donald G Truhlar. Minnesota solvation database. 2012.

[169] Alan R Katritzky, Alexander A Oliferenko, Polina V Oliferenko, Ruslan Petrukhin, Douglas B Tatham, Uko Maran, Andre Lomaka, and William E Acree. A general treatment of solubility. 1. The QSPR correlation of solvation free energies of single solutes in series of solvents. *Journal of Chemical Information and Computer sciences*, 43(6):1794–1805, 2003.

[170] Fabien Cailliez, Arnaud Bourasseau, and Pascal Pernot. Calibration of forcefields for molecular simulation: Sequential design of computer experiments for building cost-efficient kriging metamodels. *Journal of Computational Chemistry*, 35(2):130–149, 2014.

[171] André I Khuri and Siuli Mukhopadhyay. Response surface methodology. 2(2):128–149, 2010.

[172] Michele Di Pierro and Ron Elber. Automated optimization of potential parameters. *Journal of Chemical Theory and Computation*, 9(8):3311–3320, 2013.

[173] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[174] Ilya M Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[175] Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, 93(7):964–979, 2008.

[176] Aleksander Vrhovsek, Orsolya Gereben, Andrej Jamnik, and László Pusztai. Hydrogen bonding and molecular aggregates in liquid methanol, ethanol, and 1-propanol. *Journal of Physical Chemistry B*, 115(46):13473–13488, 2011.

[177] DL Wertz and RK Kruh. Reinvestigation of the structures of ethanol and methanol at room temperature. *Journal of Chemical Physics*, 47(2):388–390, 1967.

[178] JK Vij, CJ Reid, and MW Evans. Molecular dynamics of methanol: Far infra-red laser spectroscopy, interferometry and computer simulation. *Molecular Physics*, 50(5):935–947, 1983.

[179] Christopher J Fennell, Karisa L Wymer, and David L Mobley. A fixed-charge model for alcohol polarization in the condensed phase, and its role in small molecule hydration. *Journal of Physical Chemistry B*, 118(24):6438–6446, 2014.

[180] RFW Bader, Ian Keaveny, and Paul E Cade. Molecular charge distributions and chemical binding. II. First-row diatomic hydrides, AH. *Journal of Chemical Physics*, 47(9):3381–3402, 1967.

[181] AT Hagler, E Huler, and Shneior Lifson. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond

from amide crystals. *Journal of the American Chemical Society*, 96(17):5319–5327, 1974.

[182] J Lehtola, M Hakala, and K Hamalainen. Structure of liquid linear alcohols. *Journal of Physical Chemistry B*, 114(19):6426–6436, 2010.

[183] Marco Pagliai, Gianni Cardini, Roberto Righini, and Vincenzo Schettino. Hydrogen bond dynamics in liquid methanol. *Journal of Chemical Physics*, 119(13):6655–6662, 2003.

[184] Jan-Willem Handgraaf, Evert Jan Meijer, and Marie-Pierre Gaigeot. Density-functional theory-based molecular simulation study of liquid methanol. *Journal of Chemical Physics*, 121(20):10111–10119, 2004.

[185] I-F William Kuo, Christopher J Mundy, Matthew J McGrath, and J Ilja Siepmann. Structure of the methanol liquid-vapor interface: A comprehensive particle-based simulation study. *Journal of Physical Chemistry C*, 112(39):15412–15418, 2008.

[186] Nicolas Sieffert, Michael Bühl, Marie-Pierre Gaigeot, and Carole A Morrison. Liquid methanol from DFT and DFT/MM molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 9(1):106–118, 2012.

[187] Philipp Schienbein and Dominik Marx. Liquid-vapor phase diagram of RPBE-D3 water: Electronic properties along the coexistence curve and in the supercritical phase. *Journal of Physical Chemistry B*, 122(13):3318–3329, 2017.

[188] Begoña García, Santiago Aparicio, Ana M Navarro, Rafael Alcalde, and José M Leal. Measurements and modeling of thermophysical behavior of (C1- C4) Alkylbenzoate/(C1- C11) alkan-1-ol mixed solvents. *Journal of Physical Chemistry B*, 108(40):15841–15850, 2004.

[189] N Karger, T Vardag, and H-D Lüdemann. Temperature dependence of self-diffusion in compressed monohydric alcohols. 93(5):3437–3444, 1990.

[190] RE Rathbun and AL Babb. Self-diffusion in liquids. III. temperature dependence in pure liquids. *Journal of Physical Chemistry*, 65(6):1072–1074, 1961.

[191] Robert L Hurle, Allan J Easteal, and Lawrence A Woolf. Self-diffusion in monohydric alcohols under pressure. Methanol, methan (2 h) ol and ethanol. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 81(3):769–779, 1985.

[192] Makio Iwahashi, Yoko Ohbu, Tadashi Kato, Yumi Suzuki, Kazuhiro Yamauchi, Yoshimi Yamaguchi, and Mitsuo Muramatsu. The dynamical structure of normal alcohols in their liquids as determined by the viscosity and self-diffusion measurements. *Bulletin of the Chemical Society of Japan*, 59(12):3771–3774, 1986.

[193] David W McCall and Dean C Douglass. Self-diffusion in the primary alcohols. *Journal of Chemical Physics*, 32(6):1876–1877, 1960.

[194] NIST standard reference simulation website. `https://webbook.nist.gov/`, 2018. [Online; accessed 1-December-2018].

[195] Rosa Garriga, Santiago Martinez, Pascual Pérez, and Mariano Gracia. Vapour pressures at several temperatures between 288.15 K and 323.15 K of di-n-propylether with 1-hexanol or 1-octanol. application of the ERAS model. *Fluid Phase Equilibria*, 147(1-2):195–206, 1998.

[196] R.P. Danner T.E. Daubert. *Physical and thermodynamic properties of pure chemicals: data compilation*. Hemisphere Pub. Corp, New York, 1989.

[197] CL Yaws. *Handbook of Vapor pressure; Gulf Pub.* Co., Houston, 1994.

[198] Tomáš Boublík, Vojtěch Fried, and Eduard Hála. *The vapour pressures of pure substances*. Elsevier Science Pub. Co., Inc., New York, NY, 1984.

[199] JA Riddick, WB Bunger, and TK Sakano. Techniques of chemistry 4th ed., volume ii. Organic solvents, 1985.

[200] SZ Mikhail and WR Kimel. Densities and viscosities of methanol-water mixtures. *Journal of Chemical and Engineering Data*, 6(4):533–537, 1961.

[201] Fred Kurata, Thomas W Yergovich, and George W Swift. Density and viscosity of aqueous solutions of methanol and acetone from the freezing point to 10. deg. *Journal of Chemical and Engineering Data*, 16(2):222–226, 1971.

[202] Babak Mokhtarani, Ali Sharifi, Hamid Reza Mortaheb, Mojtaba Mirzaei, Morteza Mafi, and Fatemeh Sadeghian. Density and viscosity of 1-butyl-3-methylimidazolium nitrate with ethanol, 1-propanol, or 1-butanol at several temperatures. *Journal of Chemical Thermodynamics*, 41(12):1432–1438, 2009.

[203] Fatima Alaoui, E Montero, Jean-Patrick Bazile, MJP Comuñas, and Christian Boned. Liquid density of oxygenated additives to bio-fuels: 1-Hexanol at pressures up to 140 MPa and from 293.15 k to 403.15 K. *Fluid Phase Equilibria*, 320:43–48, 2012.

[204] Muhammad A Saleh, Shamim Akhtar, Shahanara Begum, M Shamsuddin Ahmed, and Syeda K Begum. Density and viscosity of 1-alkanols. *Physics and Chemistry of Liquids*, 42(6):615–623, 2004.

[205] Alessandro D'Aprano, Dorina I Donato, and Valeria Agrigento. Static dielectric constant, viscosity, and structure of pure isomeric pentanols. *Journal of Solution Chemistry*, 10(9):673–680, 1981.

[206] Akl M Awwad, Hatem M Alsyouri, Malyuba A Abu-Daabes, and Kifah A Jbara. Densities and volumetric properties of (N-(2-hydroxyethyl) morpholine+ ethanol,+ 1-propanol,+ 2-propanol,+ 1-butanol, and+ 2-butanol) at (293.15, 298.15, 303.15, 313.15, and 323.15) K. *Journal of Chemical Thermodynamics*, 40(4):592–598, 2008.

[207] Santiago Martinez, Rosa Garriga, Pascual Pérez, and Mariano Gracia. Densities and viscosities of binary mixtures of butanone with butanol isomers at several temperatures. *Fluid Phase Equilibria*, 168(2):267–279, 2000.

[208] Agnieszka Chmielewska, Malgorzata Zurada, Krzysztof Klimaszewski, and Adam Bald. Dielectric properties of methanol mixtures with ethanol, isomers of propanol, and butanol. *Journal of Chemical and Engineering Data*, 54(3):801–806, 2008.

[209] Stephen A Lawrence. *Amines: synthesis, properties and applications*. Cambridge University Press, 2004.

[210] Lars Peters, A Hussain, M Follmann, T Melin, and M-B Hägg. CO2 removal from natural gas by employing amine absorption and membrane technology- A technical and economical analysis. *Chemical Engineering Journal*, 172(2-3):952–960, 2011.

[211] Graeme Puxty, Robert Rowland, Andrew Allport, Qi Yang, Mark Bown, Robert Burns, Marcel Maeder, and Moetaz Attalla. Carbon dioxide postcombustion capture: A novel screening study of the carbon dioxide absorption performance of 76 amines. *Environmental Science and Technology*, 43(16):6427–6433, 2009.

[212] Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[213] Robert C Rizzo and William L Jorgensen. OPLS all-atom model for amines: Resolution of the amine hydration problem. *Journal of the American Chemical Society*, 121(20):4827–4836, 1999.

[214] Yanbo Ding, Dan N Bernardo, Karsten Krogh-Jespersen, and Ronald M Levy. Solvation free energies of small amides and amines from molecular dynamics/free energy perturbation simulations using pairwise additive and many-body polarizable potentials. *Journal of Physical Chemistry*, 99(29):11575–11583, 1995.

[215] A Ben-Naim and Y Marcus. Solvation thermodynamics of nonionic solutes. *Journal of Chemical Physics*, 81(4):2016–2027, 1984.

[216] Collin D Wick, John M Stubbs, Neeraj Rai, and J Ilja Siepmann. Transferable potentials for phase equilibria. 7. Primary, secondary, and tertiary amines, nitroalkanes and nitrobenzene, nitriles, amides, pyridine, and pyrimidine. *Journal of Physical Chemistry B*, 109(40):18974–18982, 2005.

[217] Gustavo A Orozco, Carlos Nieto-Draghi, Allan D Mackie, and Véronique Lachet. Transferable force field for equilibrium and transport properties in linear,

branched, and bifunctional amines I. Primary amines. *Journal of Physical Chemistry B*, 115(49):14617–14625, 2011.

[218] Philippe Ungerer, Christele Beauvais, Jérôme Delhommelle, Anne Boutin, Bernard Rousseau, and Alain H Fuchs. Optimization of the anisotropic united atoms intermolecular potential for n-alkanes. *Journal of Chemical Physics*, 112(12):5499–5510, 2000.

[219] Gustavo A Orozco, Véronique Lachet, and Allan D Mackie. A molecular simulation study of aqueous solutions of amines and alkanolamines: mixture properties and structural analysis. *Molecular Simulation*, 40(1-3):123–133, 2014.

[220] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09. Gaussian Inc. Wallingford CT 2016.

[221] William M Haynes. *CRC handbook of chemistry and physics*. CRC press, 2014.

[222] S Profeta Jr and NL Allinger. Molecular mechanics calculations on aliphatic amines. *Journal of the American Chemical Society*, 107(7):1907–1918, 1985.

[223] RJW Le Fèvre and P Russell. The dependence on state of the apparent dipole moments of ammonia, methylamine, dimethylamine, and trimethylamine. *Transactions of the Faraday Society*, 43:374–393, 1947.

[224] Carl L Yaws. *The Yaws handbook of physical properties for hydrocarbons and chemicals: Physical properties for more than 54,000 organic and inorganic chemical compounds, Coverage for C1 to C100 Organics and Ac to Zr Inorganics*. Gulf Professional Publishing, 2015.

[225] Bin Chen and J Ilja Siepmann. Transferable potentials for phase equilibria. 3. Explicit-hydrogen description of normal alkanes. *Journal of Physical Chemistry B*, 103(25):5370–5379, 1999.

[226] Ilario G Tironi, René Sperb, Paul E Smith, and Wilfred F van Gunsteren. A generalized reaction field method for molecular dynamics simulations. *Journal of Chemical Physics*, 102(13):5451–5459, 1995.

[227] Alexei K Baev. *Specific intermolecular interactions of nitrogenated and bioorganic compounds*. Springer Science and Business Media, 2013.

[228] Carl L Yaws. *The Yaws handbook of vapor pressure: Antoine coefficients*. Gulf Professional Publishing, 2015.

[229] Chase Gobble, Nigam Rath, and James Chickos. The vaporization enthalpies and vapor pressures of some primary amines of pharmaceutical importance by correlation gas chromatography. *Journal of Chemical and Engineering Data*, 58(9):2600–2609, 2013.

[230] AJ Bordner, CN Cavasotto, and RA Abagyan. Accurate transferable model for water, n-octanol, and n-hexadecane solvation free energies. *Journal of Physical Chemistry B*, 106(42):11009–11015, 2002.

[231] Carl L Yaws. *Thermophysical Properties of Chemicals and Hydrocarbons (Second Edition)*. Gulf Professional Publishing, 2014.

[232] Curt M Breneman and Kenneth B Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *Journal of Computational Chemistry*, 11(3):361–373, 1990.

[233] J David Rawn and Robert J Ouellette. *Organic Chemistry: Structure, Mechanism, Synthesis*. Academic Press, 2018.

[234] Robert Tisserand and Rodney Young. *2 - Essential oil composition*. Churchill Livingstone, second edition edition, 2014.

[235] William L Jorgensen, James M Briggs, and M Leonor Contreras. Relative partition coefficients for organic solutes from fluid simulations. *Journal of Physical Chemistry*, 94(4):1683–1686, 1990.

[236] Jeffrey J Potoff and J Ilja Siepmann. Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE journal*, 47(7):1676–1682, 2001.

[237] Spyridon Kranias, Denis Pattou, Bernard Lévy, and Anne Boutin. An optimized potential for phase equilibria calculation for ketone and aldehyde molecular fluids. *Physical Chemistry Chemical Physics*, 5(19):4175–4179, 2003.

[238] Ganesh Kamath, Grigor Georgiev, and Jeffrey J Potoff. Molecular modeling of phase behavior and microstructure of acetone- chloroform- methanol binary mixtures. *Journal of Physical Chemistry B*, 109(41):19463–19473, 2005.

[239] Axel D Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A*, 38(6):3098, 1988.

[240] Rick A Kendall, Thom H Dunning Jr, and Robert J Harrison. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *Journal of Chemical Physics*, 96(9):6796–6806, 1992.

[241] Fabio Comelli and Romolo Francesconi. Densities and excess molar volumes of propylene carbonate+ linear and cyclic ketones at 298.15 K. *Journal of Chemical and Engineering Data*, 40(4):808–810, 1995.

[242] Dharshani N Bopege, Matt Petrowsky, Matthew B Johnson, and Roger Frech. Mass and ion transport in ketones and ketone electrolytes: comparison with acetate systems. *Journal of Solution Chemistry*, 42(3):584–591, 2013.

[243] James S Chickos and William E Acree Jr. Enthalpies of vaporization of organic

and organometallic compounds, 1880–2002. *Journal of Physical and Chemical Reference Data*, 32(2):519–878, 2003.

[244] David R Lide. *CRC handbook of chemistry and physics*, volume 84. CRC press, 2003.

[245] Miguel Jorge, José RB Gomes, and Andrew W Milne. Self-consistent electrostatic embedding for liquid phase polarization. *Journal of Molecular Liquids*, 322:114550, 2021.

[246] Ralph D Nelson Jr, David R Lide Jr, and Arthur A Maryott. Selected values of electric dipole moments for molecules in the gas phase. Technical report, National Standard Reference Data System, 1967.

[247] Jose Alejandre, Jose Luis Rivera, Marco Antonio Mora, and Virgínia de la Garza. Force field of monoethanolamine. *Journal of Physical Chemistry B*, 104(6):1332–1337, 2000.

[248] Gustavo A Orozco, Véronique Lachet, Carlos Nieto-Draghi, and Allan D Mackie. A transferable force field for primary, secondary, and tertiary alkanolamines. *Journal of Chemical Theory and Computation*, 9(4):2097–2103, 2013.

[249] Roberto López-Rendón, Marco A Mora, José Alejandre, and Mark E Tuckerman. Molecular dynamics simulations of aqueous solutions of ethanolamines. *Journal of Physical Chemistry B*, 110(30):14652–14658, 2006.

[250] Eirik F da Silva, Tatyana Kuznetsova, Bjørn Kvamme, and Kenneth M Merz. Molecular dynamics study of ethanolamine as a pure liquid and in aqueous solution. *Journal of Physical Chemistry B*, 111(14):3695–3703, 2007.

[251] Chieh-Ming Hsieh and Shiang-Tai Lin. Prediction of 1-octanol-water partition coefficient and infinite dilution activity coefficient in water from the PR+ COSMOSAC model. *Fluid Phase Equilibria*, 285(1-2):8–14, 2009.

[252] Ma Esther Rebolledo-Libreros and Arturo Trejo. Density and viscosity of aqueous blends of three alkanolamines: N-methyldiethanolamine, diethanolamine, and 2-

amino-2-methyl-1-propanol in the range of 303 to 343 K. *Journal of Chemical and Engineering Data*, 51(2):702–707, 2006.

[253] Jingyi Han, Jing Jin, Dag A Eimer, and Morten C Melaaen. Density of water (1)+ diethanolamine (2)+ co2 (3) and water (1)+ N-methyldiethanolamine (2)+ co2 (3) from (298.15 to 423.15) K. *Journal of Chemical and Engineering Data*, 57(6):1843–1850, 2012.

[254] NK Balabaev, DK Belashchenko, MN Rodnikova, SV Kraevskii, and IA Solonina. Modeling the structure of liquid monoethanolamine by molecular dynamics. *Russian Journal of Physical Chemistry A*, 89(3):398–405, 2015.

[255] Javier Cardona, Rui Fartaria, Martin B Sweatman, and Leo Lue. Molecular dynamics simulations for the prediction of the dielectric spectra of alcohols, glycols and monoethanolamine. *Molecular Simulation*, 42(5):370–390, 2016.

[256] Simon Kapteina, Krzysztof Slowik, Sergey P Verevkin, and Andreas Heintz. Vapor pressures and vaporization enthalpies of a series of ethanolamines. *Journal of Chemical and Engineering Data*, 50(2):398–402, 2005.

[257] David R Lide. *CRC handbook of chemistry and physics*, volume 90. CRC press, 2009.

[258] Prabhakar Undre, SN Helambe, SB Jagdale, PW Khirade, and SC Mehrotra. Study of solute–solvent interaction through dielectrics properties of N,N-dimethylacetamide in ethanolamine. *Journal of Molecular Liquids*, 137(1-3):147–151, 2008.

[259] Javier Cardona, Miguel Jorge, and Leo Lue. Simple corrections for the static dielectric constant of liquid mixtures from model force fields. *Physical Chemistry Chemical Physics*, 22(38):21741–21749, 2020.

[260] David L Mobley and J Peter Guthrie. Freesolv: A database of experimental and calculated hydration free energies, with input files. *Journal of Computer-aided Molecular Design*, 28(7):711–720, 2014.

[261] Ramón Bosque and Joaquim Sales. Polarizabilities of solvents from the chem-

ical composition. *Journal of Chemical Information and Computer sciences*, 42(5):1154–1163, 2002.

Appendix

A1. Converge plots for solvation free energies calculations

Amines

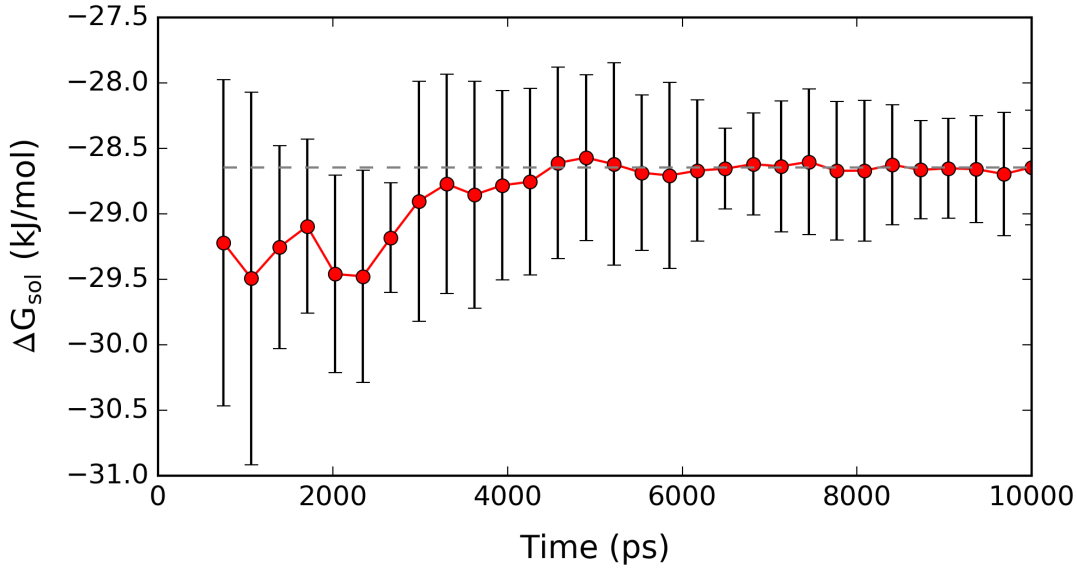

Figure A1: LJ component of the free energy of self-solvation of decylamine at 298.15 K and 1 bar, obtained using PolCA, as a function of simulation time. Error bars were estimated as described in section 2.8, and the horizontal dashed line corresponds to the solvation free energy value of the last point.

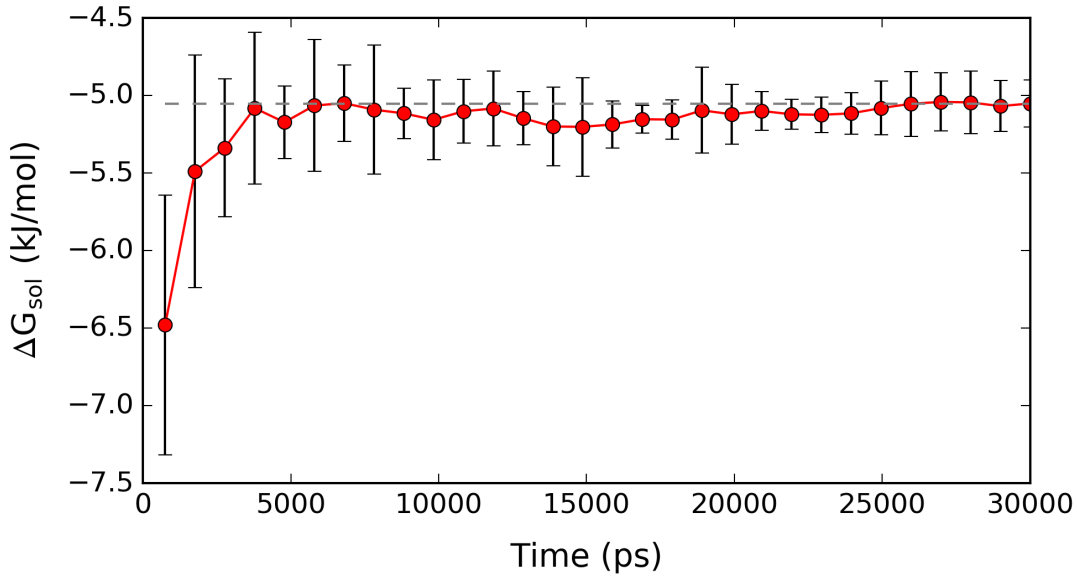

Figure A2: Electrostatic component of the free energy of self-solvation of decylamine at 298.15 K and 1 bar, obtained using PolCA, as a function of simulation time. Error bars were estimated as described in section 2.8, and the horizontal dashed line corresponds to the solvation free energy value of the last point.

Ketones

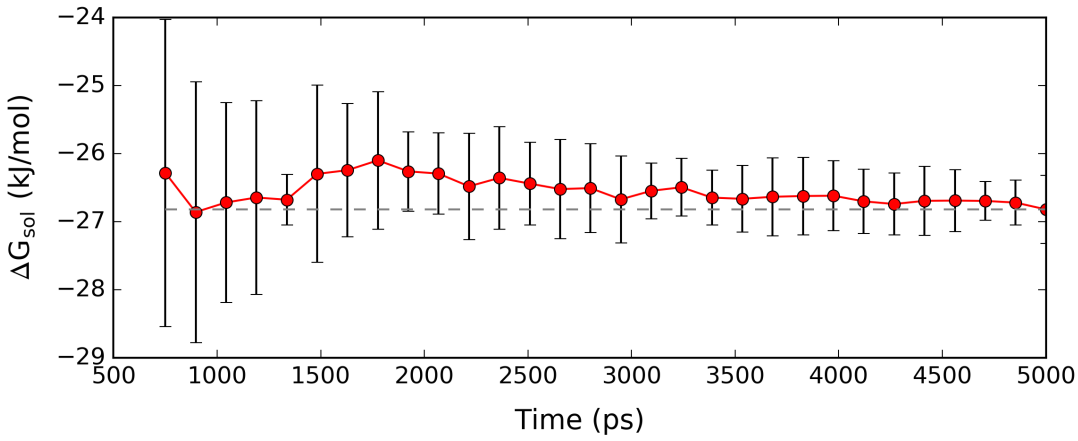

Figure A3: LJ component of the free energy of self-solvation of decanone at 298.15 K and 1 bar, obtained using PolCA, as a function of simulation time. Error bars were estimated as described in section 2.8, and the horizontal dashed line corresponds to the solvation free energy value of the last point.

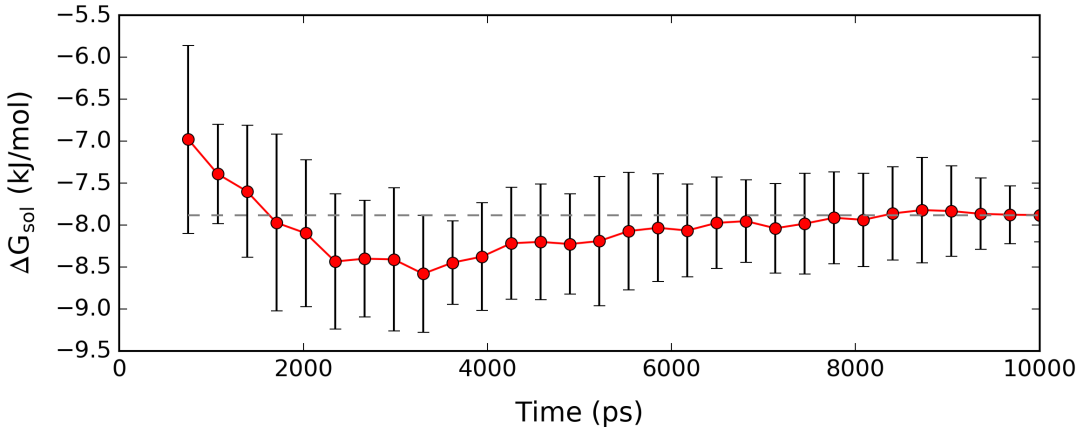

Figure A4: Electrostatic component of the free energy of self-solvation of decanone at 298.15 K and 1 bar, obtained using PolCA, as a function of simulation time. Error bars were estimated as described in section 2.8, and the horizontal dashed line corresponds to the solvation free energy value of the last point.

Mixtures

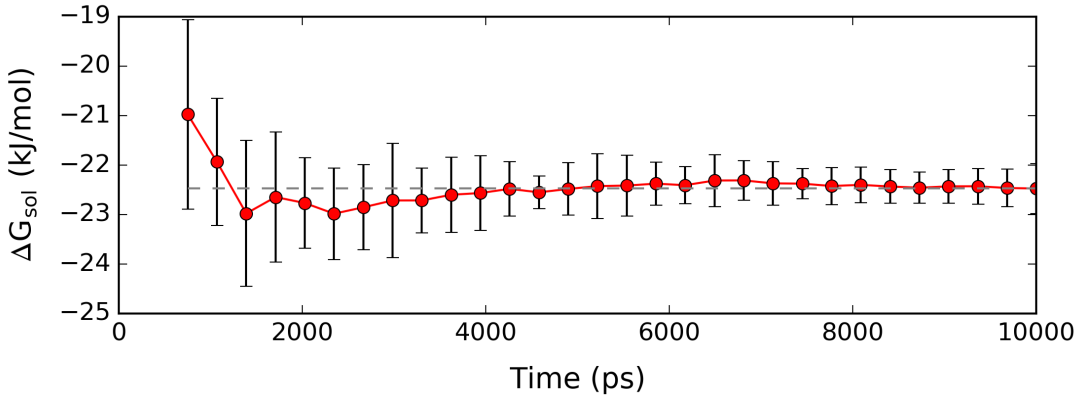

Figure A5: LJ component of the free energy of solvation of octylamine in octanol, obtained using PolCA, as a function of simulation time. Error bars were estimated as described in section 2.8, and the horizontal dashed line corresponds to the solvation free energy value of the last point.

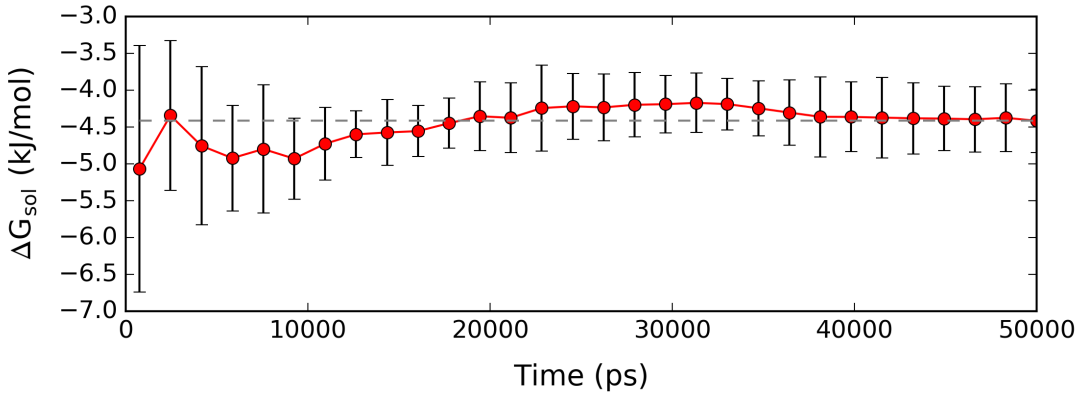

Figure A6: Electrostatic component of the free energy of solvation of octylamine in octanol, obtained using PolCA, as a function of simulation time. Error bars were estimated as described in section 2.8, and the horizontal dashed line corresponds to the solvation free energy value of the last point.

A2. Meta-models' performance

Alcohols

Model 1

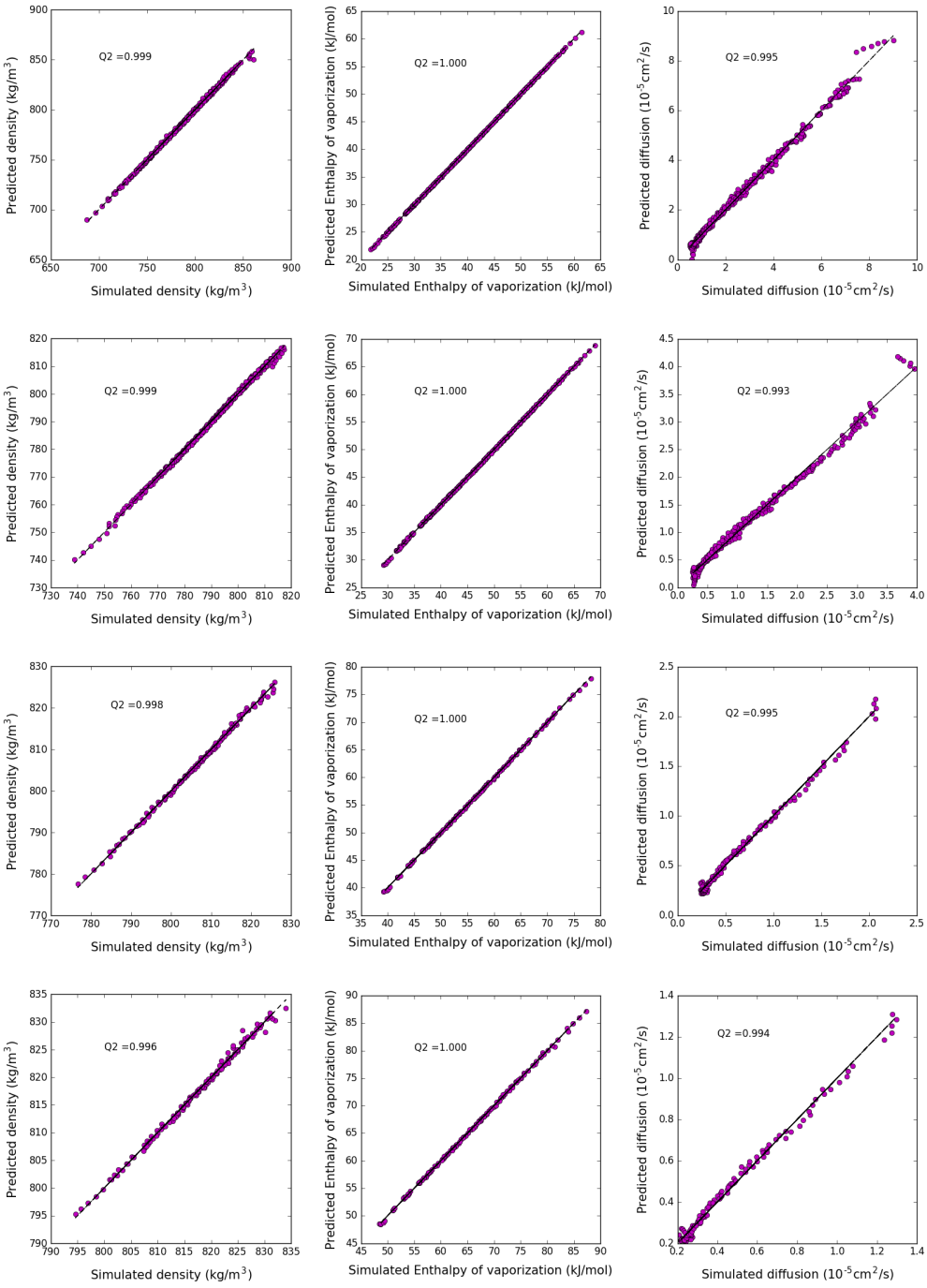

Figure A7: Plot of simulated values at 298.15 K and 1 bar versus the values predicted using Model 1's meta-models. The first column is the density, the second column the enthalpy of vaporisation, and the last column the diffusion constant. The rows are methanol, propanol, pentanol and heptanol; from top to bottom.

Model 2

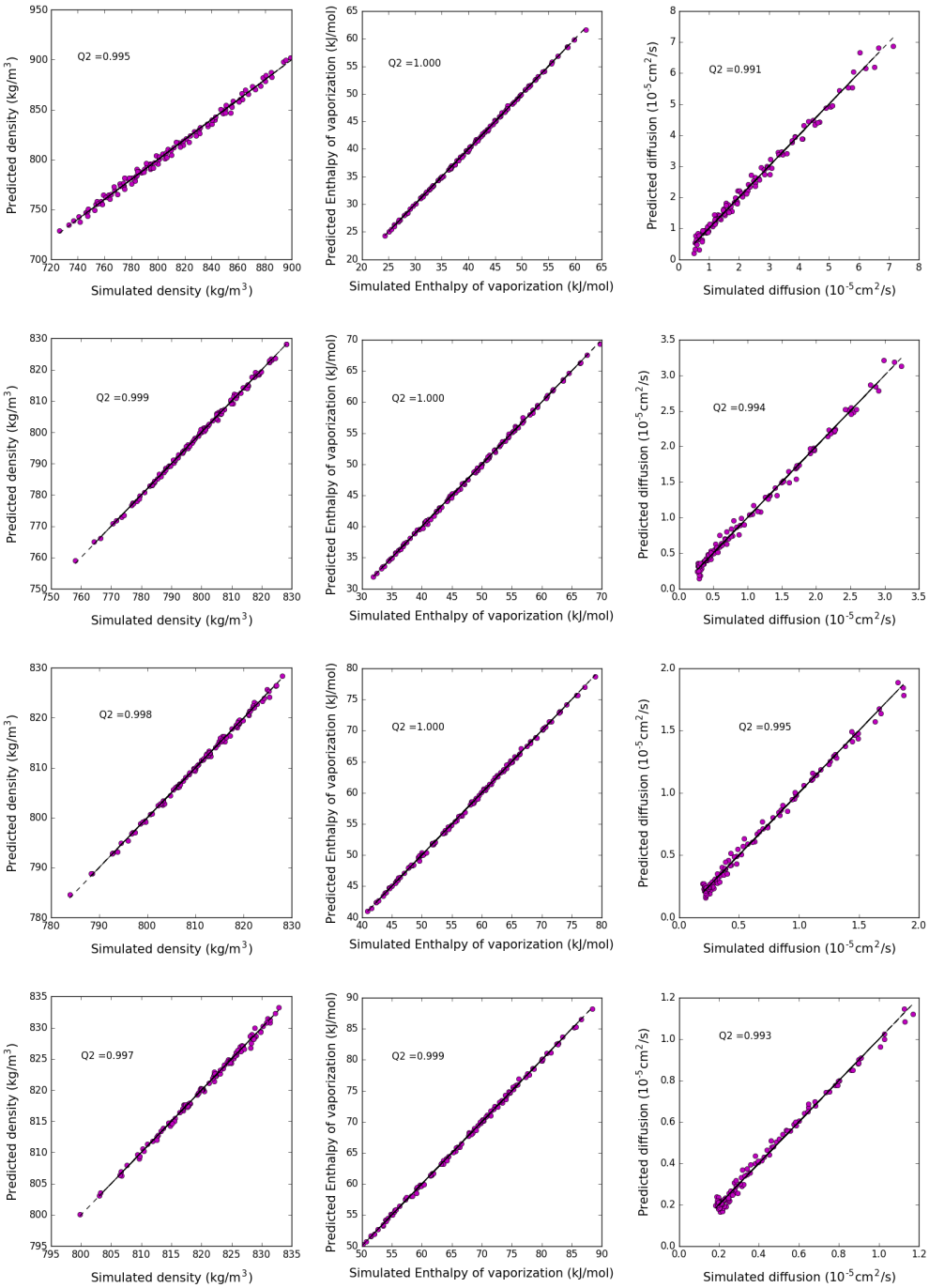

Figure A8: Plot of simulated values at 298.15 K and 1 bar versus the values predicted using Model 2's meta-models. The first column is the density, the second column the enthalpy of vaporisation and the last column the diffusion constant. The rows are methanol, propanol, pentanol and heptanol, from top to bottom.

Model 3

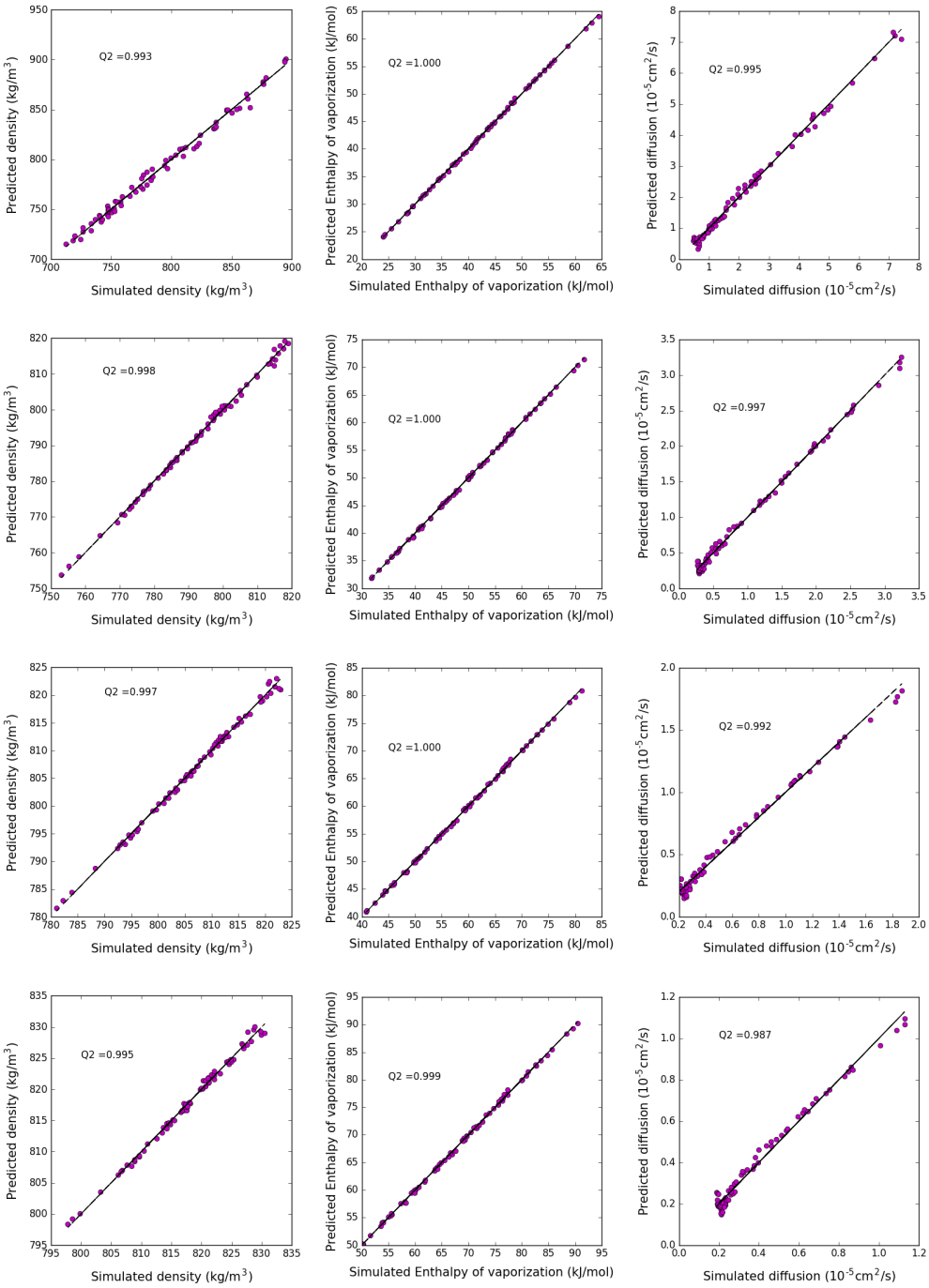

Figure A9: Plot of simulated values at 298.15 K and 1 bar versus the values predicted using Model 3's meta-models. The first column is the density, the second column the enthalpy of vaporisation, and the last column the diffusion constant. The rows are methanol, propanol, pentanol and heptanol; from top to bottom.

Amines

Model 1

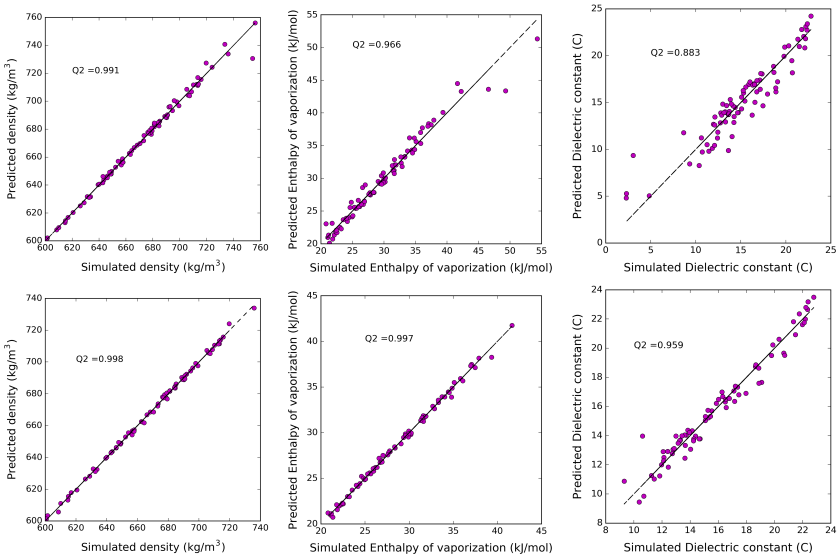

Figure A10: Performance of methylamine's meta-models from Model 1's learning set obtained using 76 simulations (bottom) and 81 simulations (top) per molecule at 298.15 K and 1 bar. The meta-models in the bottom do not include points with very low σ and ϵ and q_N = -0.992 (see text for details).

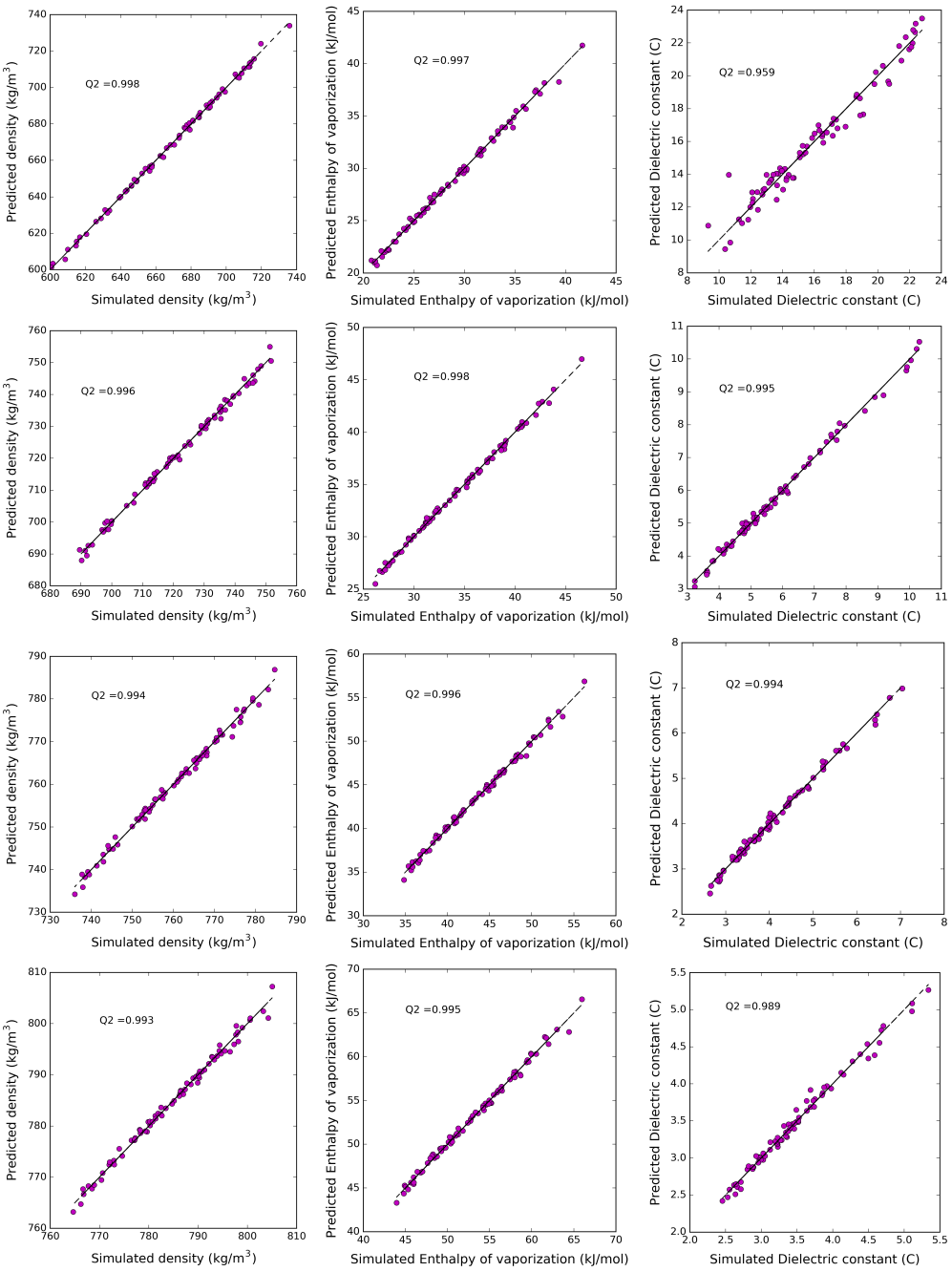

Figure A11: Plot of simulated values at 298.15 K and 1 bar versus predicted values obtained using the meta-models for Model 1. The first column is the density, the second column the enthalpy of vaporisation, and the last column the corrected dielectric constant. The rows are methylamine, propylamine, pentylamine and heptylamine; from top to bottom.

Model 2

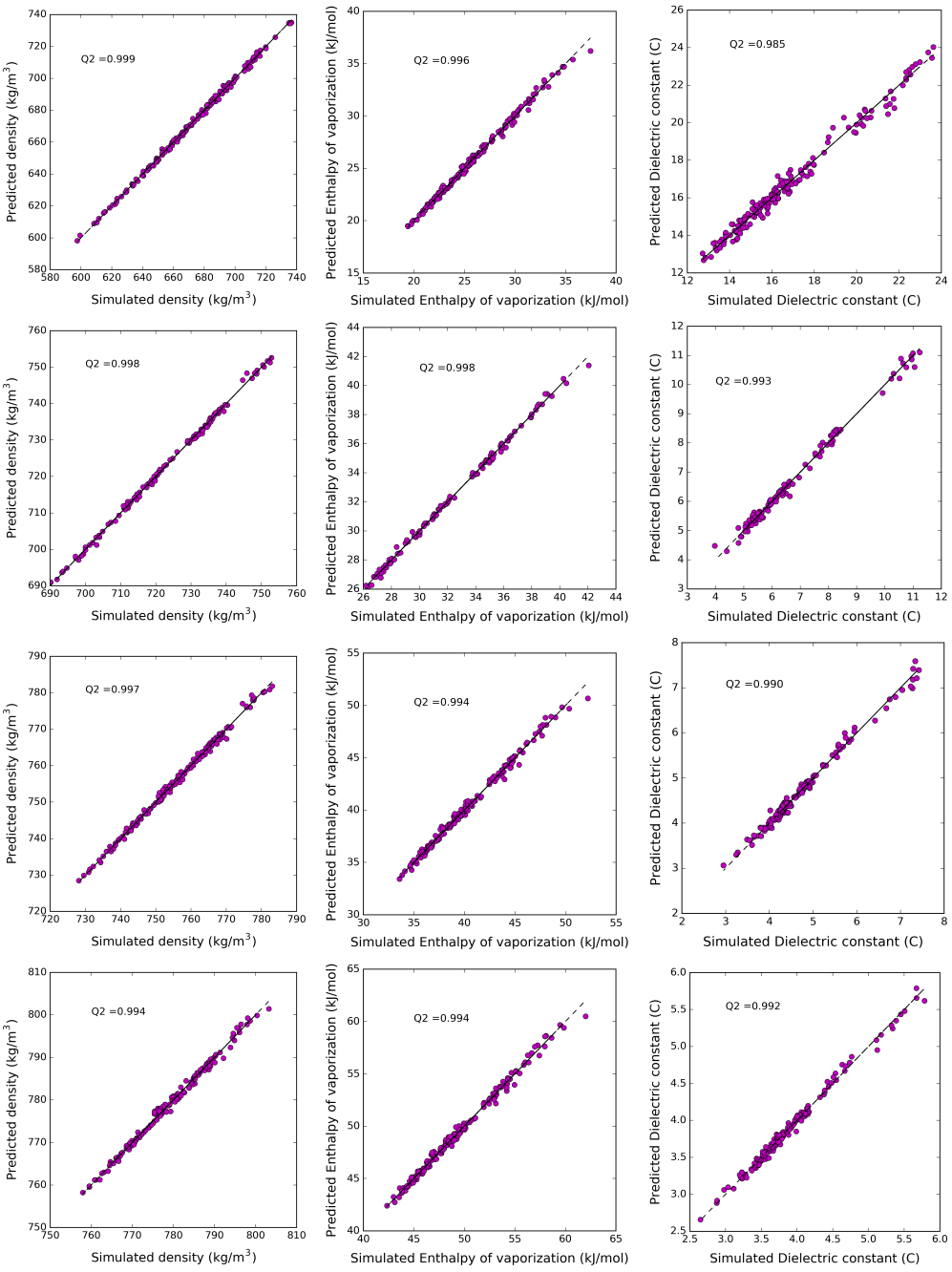

Figure A12: Plot of simulated values at 298.15 K and 1 bar versus predicted values obtained using the meta-models for Model 2. The first column is the density, the second column the enthalpy of vaporisation, and the last column the corrected dielectric constant. The rows are methylamine, propylamine, pentylamine and heptylamine; from top to bottom.

Model 3

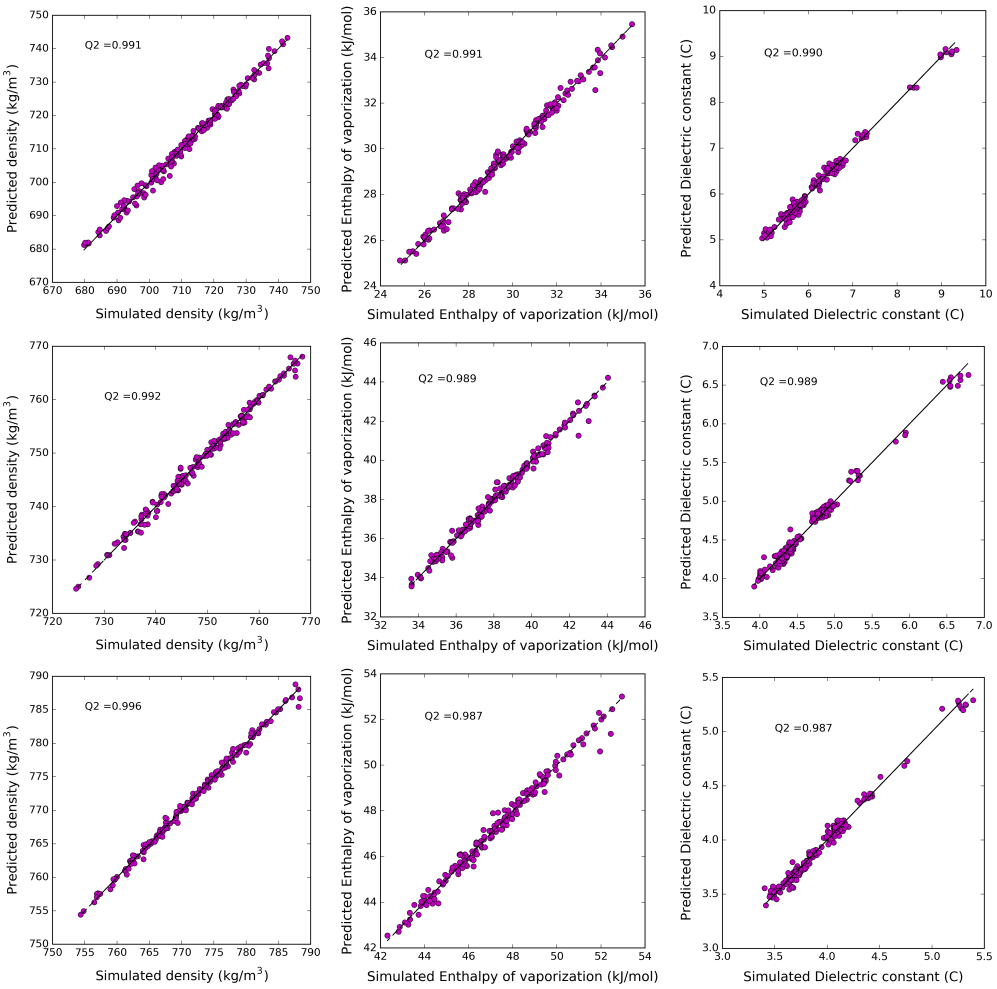

Figure A13: Plot of simulated values at 298.15 K and 1 bar versus predicted values obtained using the meta-models for Model 3. The first column is the density, the second column the enthalpy of vaporisation, and the last column the corrected dielectric constant. The rows are propylamine, pentylamine and heptylamine; from top to bottom.

Ketones

Model 1

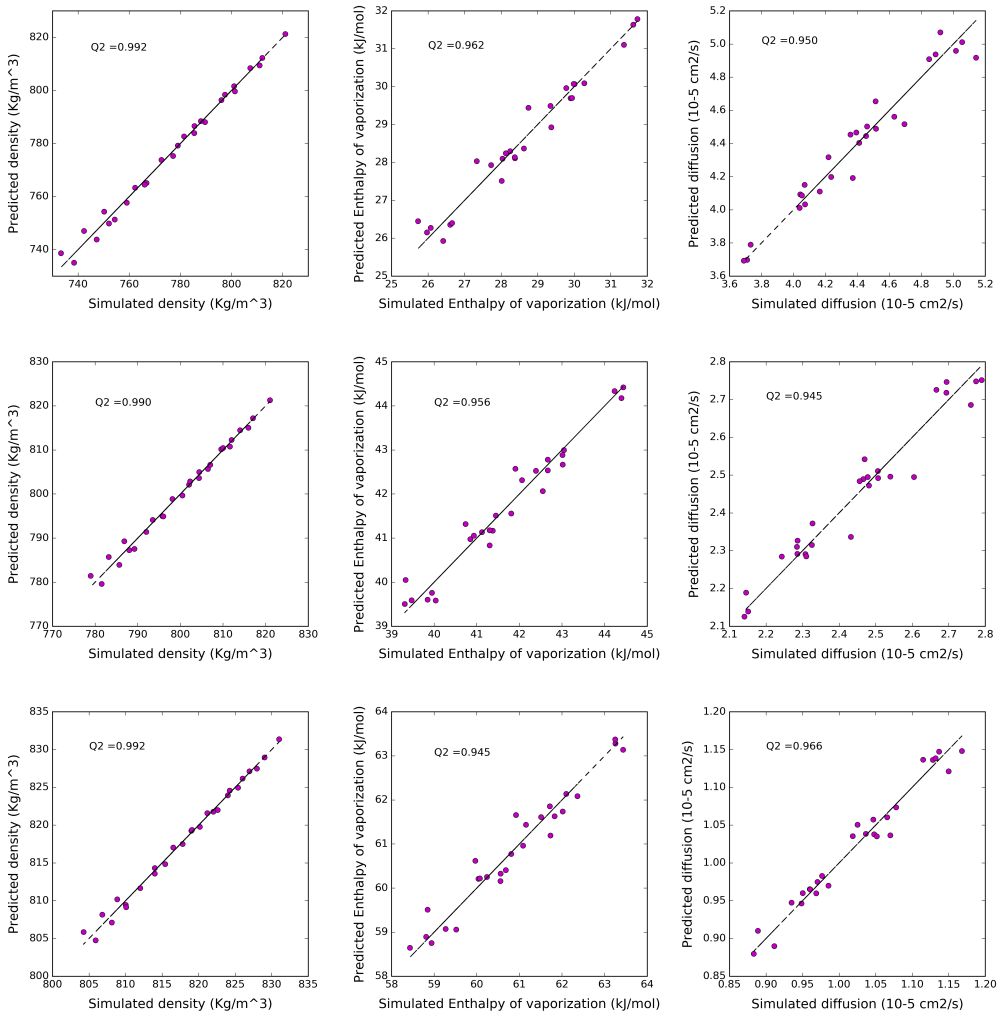

Figure A14: Plot of simulated values at 298.15 K and 1 bar versus the values predicted using the meta-models for Model 1. The first column is the density, the second column the enthalpy of vaporisation, and the last column the diffusion constant. The rows are propanone, hexanone and decanone; from top to bottom.

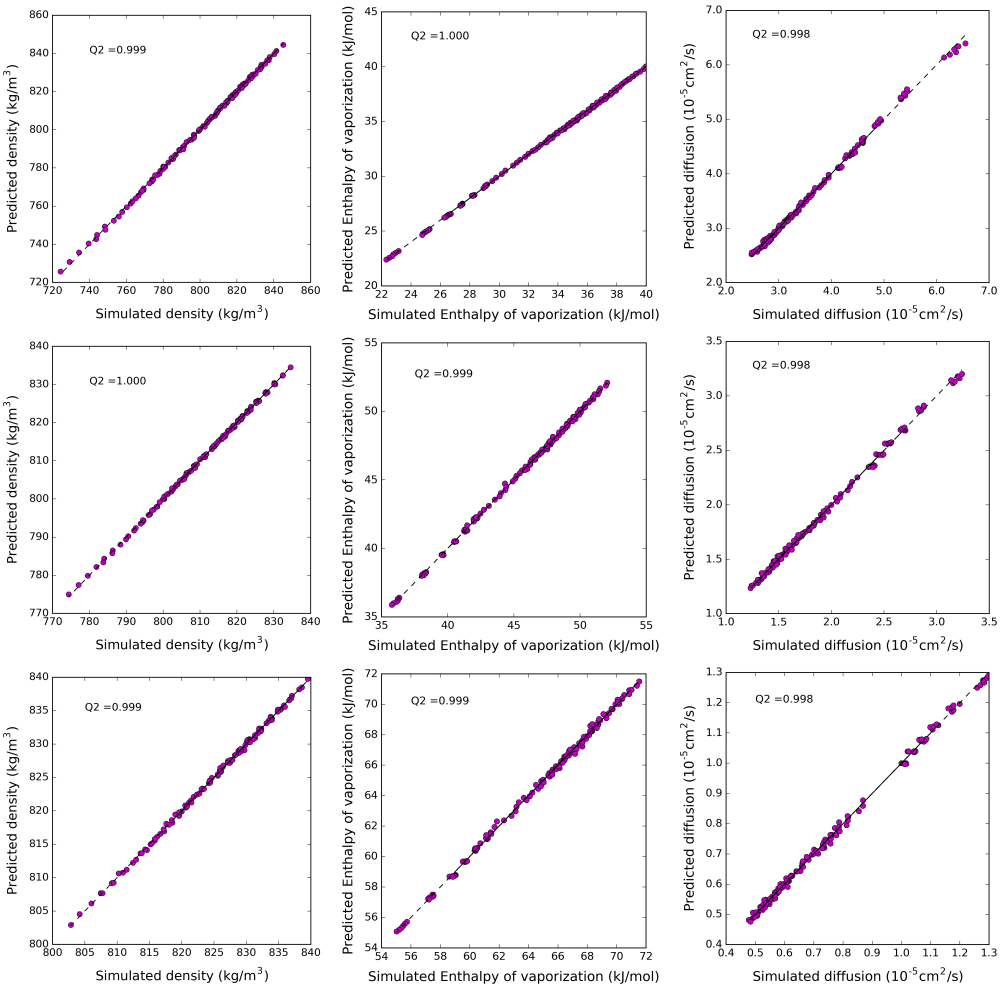

Figure A15: Plot of simulated values at 298.15 K and 1 bar versus the values predicted using the meta-models for Model 2 and Model 3. The first column is the density, the second column the enthalpy of vaporisation, and the last column the diffusion constant. The rows are propanone, hexanone and decanone; from top to bottom.

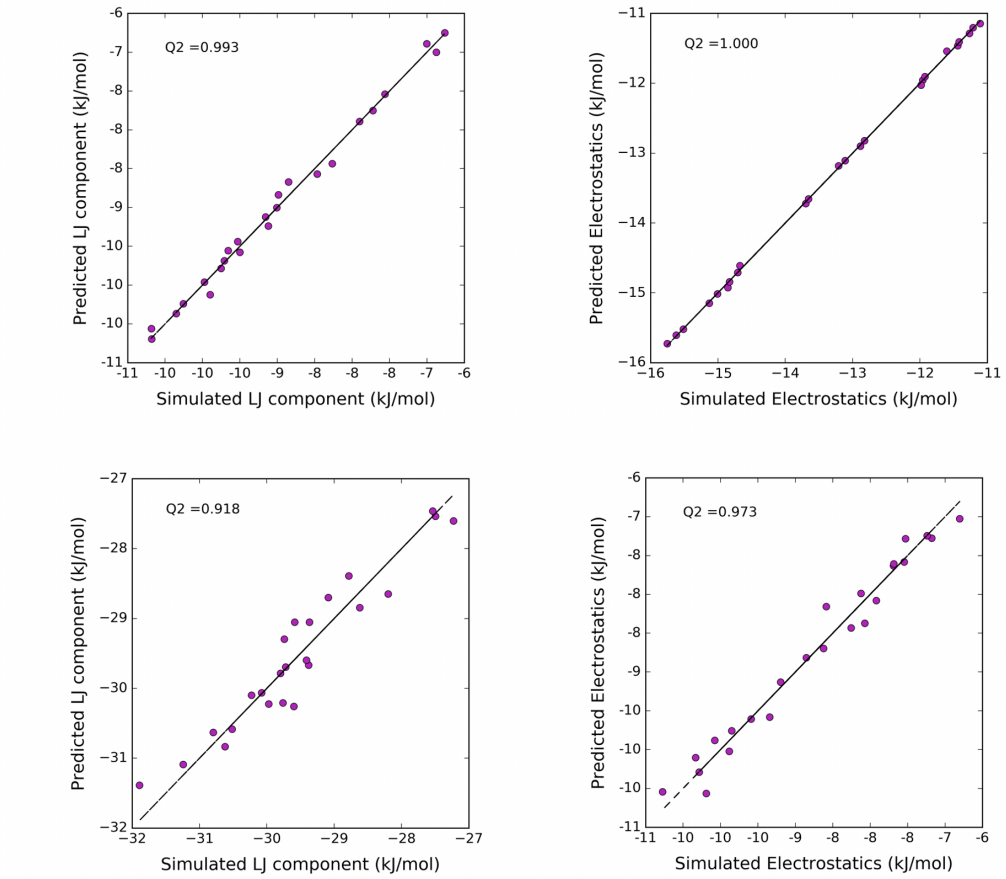

Figure A16: Meta-models performance for the free energy of self-solvation of propanone (top) and decanone (bottom) at 298.15 K and 1 bar. The graphs on the left show the LJ component's meta-models while the plots on the right are the meta-models for the electrostatic component.

A3. Experimental solvation free energies

Experimental free energies of self-solvation obtained from vapour pressure data are shown in Table A1. The majority of these values were taken from the Yaws Handbook [228].

Table A1: Free energy of self-solvation at 298 K and 1 atm for the values obtained using Equation 2.66. Vapour pressures were taken from reference [228], unless stated otherwise.

Molecule	M_w **[g/mol]**	**Density [kg/m3]**	P_{vap} **[atm]**	**ΔG [kJ/mol]**
Methanol	32.042	786.63	1.671e-01 [198]	-20.30
Ethanol	46.069	785.25	7.803e-02 [196]	-21.28
Propanol	60.096	799.54	2.763e-02 [199]	-23.24
Butanol	74.123	805.75	9.211e-03 [196]	-25.46
Pentanol	88.15	810.98	2.895e-03 [199]	-27.92
Hexanol	102.162	815.01	1.115e-03[195]	-29.93
Heptanol	116.204	818.65	2.846e-04[197]	-33.01
Octanol	130.23	821.69	7.895e-05 [195]	-35.91
Nonanol	144.26	824.15	2.987e-05[196]	-38.08
Decanol	158.285	826.15	1.120e-05[196]	-40.29
2-Butanol	74.123	803	2.408e-02[196]	-23.07
2-methylpropan-2-ol	74.123	781.14	5.355e-02 [196]	-21.02
Methylamine	31.05	655	3.472e+00	-12.399
Ethylamine	45.08	677	1.348e+00	-13.902
Propylamine	59.11	714	3.892e-01	-16.443
Butylamine	73.14	741	1.245e-01	-18.833
Pentylamine	87.16	751	4.112e-02	-21.178
Hexylamine	101.19	761	1.149e-02	-24.003
Heptylamine	115.22	772	2.832e-03 [229]	-27.188
Octylamine	129.24	779	1.036e-03 [229]	-29.418
Nonylamine	143.27	785	N/A	N/A
Decylamine	157.30	791	1.184e-04 [229]	-34.347
Isopropylamine	59.11	684	7.407e-01	-14.741
Dimethylamine	45.08	650	1.934e+00	-12.907
Diethylamine	73.14	702	2.661e-01	-16.815
Trimethylamine	59.11	629	2.302e+00	-11.721
Triethylamine	101.19	723	5.955e-02	-19.795
Tertbutylamine	73.14	688	4.800e-01	-15.303
1-1-dimethylpropylamine	87.17	727.6	1.090e-01	-18.681
2-amino-3-methyl-pentane	101.19	750	6.203e-03	-25.494
2-Butanamine	73.14	722	2.035e-01	-17.550
Propanone	58.08	785	3.025e-01	-17.35
Butanone	72.107	799	1.154e-01	-19.24
Pentanone	86.134	802	5.188e-02	-20.79
Hexanone	100.161	807	9.848e-03	-24.56
Heptanone	114.188	811	4.153e-03	-26.38
Octanone	128.214	815	1.318e-03	-28.95
Nonanone	142.241	816.7	4.220e-04	-31.53
Decanone	156.268	819.8	1.693e-04	-33.57
2-aminoethanol	61.084	1014	4.437e-04 [256]	-34.03

Solvation free energies for solutes in non-aqueous solvents were obtained using Equation A1. Here, $P_{naq/air}$ is the partition coefficient measured between the gas phase

and the dilute non-aqueous liquids. The subscript naq/air denotes that the solute is moving from the gas to the liquid phase.

$$\Delta G_{naq/air} = -2.303\, R\, T \log P_{naq/air} = (-5.709 \text{ kJ/mol}) \log P_{naq/air} \qquad \text{(A1)}$$

Table A2 contains the partition coefficients of amines in hexadecane and the resulting free energies of solvation.

Table A2: Free energy of solvation of amines in hexadecane at 298 K and 1 atm. The hexadecane/air partition coefficients (P_{hex}) were taken from the supporting information of reference [230].

	log P_{hex}	**ΔG (kJ/mol)**
Methylamine	1.3	-7.42
Ethylamine	1.68	-9.59
Propylamine	2.14	-12.22
Butylamine	2.62	-14.96
Pentylamine	3.14	-17.93
Hexylamine	3.66	-20.89
Heptylamine	4.17	-23.81
Octylamine	4.52	-25.80

The free energies of solvation of amines in octanol were obtained as the sum between their free energies of hydration and transfer free energies from water to octanol. The octanol/water (P_{oct}) and water/air partition coefficients (P_w) are shown in Table A3. This table also presents the calculated free energies of solvation of amines in octanol.

Table A3: Octanol/water and water/air partition coefficients used to obtain solvation free energies of amines in octanol. ΔG values are in kJ/mol.

	log P_w [230]	**ΔG_{hyd}**	**log P_{oct}** [251]	**$\Delta G_{wat\text{-}oct}$**	**ΔG_{oct}**
Methylamine	3.34	-19.07	-0.57	3.25	-15.81
Ethylamine	3.3	-18.84	-0.13	0.74	-18.10
Propylamine	3.22	-18.38	0.48	-2.74	-21.12
Butylamine	3.11	-17.75	0.88	-5.02	-22.78
Pentylamine	3	-17.13	1.45	-8.28	-25.41
Hexylamine	2.9	-16.56	2.06	-11.76	-28.32
Heptylamine	2.78	-15.87	2.57	-14.67	-30.54
Octylamine	2.68	-15.30	N/A	N/A	N/A

To corroborate the hydration free energies from Table A3, solvation free energies of amines in octanol were also calculated using hydration free energies from the FreeSolv

database [260] and are shown in Table A4. Two other values for butylamine's P_{oct} were found in reference [65], and thus, they are also included in this table.

Table A4: Solvation free energies of amines in octanol obtained using hydration free energies from the FreeSolv database [260]. ΔG values are in kJ/mol.

	$\mathbf{\Delta G_{hyd}}$	$\mathbf{log}\ P_{oct}$ [251]	$\mathbf{\Delta G_{wat\text{-}oct}}$	$\mathbf{\Delta G_{oct}}$
Methylamine	-19.04	-0.57	-3.25	-15.79
Ethylamine	-18.83	-0.13	-0.74	-18.09
Propylamine	-18.37	0.48	2.74	-21.11
Butylamine	-17.74	0.88	5.02	-22.76
Butylamine	-17.74	0.81 [65]	4.62	-22.36
Butylamine	-17.74	0.68 [65]	3.88	-21.62
Pentylamine	-17.11	1.45	8.27	-25.38
Hexylamine	-16.53	2.06	11.75	-28.28
Heptylamine	-15.86	2.57	14.66	-30.52
Octylamine	-15.27	N/A		

A4. Polarisation corrections

Table A5: Experimental data used to obtain the polarisation corrections of the alcohols studied in this thesis. $C_{Dist.}$, $C_{Elect.}$ and C_{pol} are in kJ/mol, and C_{pol} corresponds to a liquid to vapour transition. Polarisation corrections for solvation free energies are equal to - C_{pol}. Experimental dipoles in the gas phase were taken from reference [246]

Solute	μ_g **(D)**	$\alpha(\AA^3)$ **[261]**	**Solvent**	ϵ_{sol}	n	ϵ_{el}	μ_l **(D)**	$C_{Dist.}$	$C_{Elect.}$	C_{pol}
methanol	1.700	3.260	methanol	32.258	1.329	1.766	2.701	-9.251	8.842	-0.409
ethanol	1.690	5.130	ethanol	24.713	1.361	1.853	2.805	-7.297	7.074	-0.224
propanol	1.680	6.960	propanol	20.423	1.385	1.918	2.874	-6.169	6.073	-0.096
butanol	1.660	8.790	butanol	17.430	1.399	1.957	2.878	-5.079	5.101	0.022
pentanol	1.700	10.610	pentanol	14.960	1.410	1.988	2.969	-4.568	4.702	0.134
hexanol	1.650	12.460	hexanol	12.500	1.418	2.010	2.872	-3.611	3.860	0.249
heptanol	1.710	14.300	heptanol	11.500	1.425	2.030	2.985	-3.425	3.731	0.306
octanol	1.680	16.140	octanol	9.780	1.430	2.043	2.901	-2.783	3.175	0.393
nonanol	1.600	17.972	nonanol	8.600	1.433	2.054	2.734	-2.155	2.568	0.413
decanol	1.600	19.830	decanol	7.620	1.437	2.066	2.703	-1.848	2.307	0.459
2-propanol	1.580	6.980	2-propanol	19.430	1.375	1.891	2.652	-4.960	4.943	-0.017
2-butanol	1.600	8.770	2-butanol	16.600	1.398	1.954	2.760	-4.621	4.685	0.063
2-pentanol	1.660	10.570	2-pentanol	13.710	1.404	1.971	2.851	-4.042	4.251	0.209
tert-butanol	1.670	8.820	tert-butanol	12.490	1.382	1.911	2.753	-4.005	4.348	0.343
tert-amyl	1.700	10.640	tert-amyl	5.780	1.403	1.967	2.633	-2.465	3.583	1.117
methanol	1.700	3.260	hexadecane	2.090	1.433	2.053	2.149	-1.860	8.730	6.870
ethanol	1.690	5.130	hexadecane	2.090	1.433	2.053	2.136	-1.168	5.483	4.314
propanol	1.680	6.960	hexadecane	2.090	1.433	2.053	2.124	-0.851	3.993	3.142
butanol	1.660	8.790	hexadecane	2.090	1.433	2.053	2.098	-0.658	3.087	2.429
pentanol	1.700	10.610	hexadecane	2.090	1.433	2.053	2.149	-0.572	2.682	2.111
hexanol	1.650	12.460	hexadecane	2.090	1.433	2.053	2.086	-0.458	2.152	1.693
heptanol	1.710	14.300	hexadecane	2.090	1.433	2.053	2.161	-0.429	2.014	1.585
octanol	1.680	16.140	hexadecane	2.090	1.433	2.053	2.124	-0.367	1.722	1.355
nonanol	1.600	17.972	hexadecane	2.090	1.433	2.053	2.022	-0.299	1.403	1.104
decanol	1.600	19.830	hexadecane	2.090	1.433	2.053	2.022	-0.271	1.271	1.000
methane	0.001	2.600	octanol	9.780	1.430	2.043	0.002	0.000	0.000	0.000
ethane	0.001	4.470	octanol	9.780	1.430	2.043	0.002	0.000	0.000	0.000
propane	0.083	6.290	octanol	9.780	1.430	2.043	0.143	-0.017	0.020	0.002
butane	0.050	8.200	octanol	9.780	1.430	2.043	0.086	-0.005	0.006	0.001
pentane	0.100	10.111	octanol	9.780	1.430	2.043	0.173	-0.016	0.018	0.002
hexane	0.100	11.941	octanol	9.780	1.430	2.043	0.173	-0.013	0.015	0.002
heptane	0.100	13.810	octanol	9.780	1.430	2.043	0.173	-0.012	0.013	0.002
octane	0.100	15.600	octanol	9.780	1.430	2.043	0.173	-0.010	0.012	0.001
nonane	0.100	17.451	octanol	9.780	1.430	2.043	0.173	-0.009	0.010	0.001
decane	0.100	19.331	octanol	9.780	1.430	2.043	0.173	-0.008	0.009	0.001

A5. Simulated values

The tables in this section contain the simulated values with polarisation corrections when applicable. Experimental values are also presented here for completion and, unless stated otherwise, they are the average obtained from different sources (sources are given in the main body of this thesis). These values were used when calculating RMSDs.

Alcohols

Table A6: Density (kg/m^3).

	Model 1	Model 2	PolCA	TraPPE	Experimental
Methanol	788.25 +/- 0.08	784.13 +/- 0.07	785.52 +/- 0.06	785.24 +/- 0.06	786.63
Ethanol	782.63 +/- 0.08	782.64 +/- 0.05	781.75 +/- 0.09	783.51 +/- 0.06	785.25
Propanol	794.9 +/- 0.1	796.4 +/- 0.04	794.3 +/- 0.1	795.24 +/- 0.06	799.54
Butanol	804.2 +/- 0.2	804.84 +/- 0.08	803.1 +/- 0.1	804.18 +/- 0.04	805.75
Pentanol	810.3 +/- 0.1	811.14 +/- 0.07	809.6 +/- 0.1	811.23 +/- 0.07	810.98
Hexanol	816.0 +/- 0.2	816.75 +/- 0.09	815.5 +/- 0.2	817.72 +/- 0.09	815.01
Heptanol	820.8 +/- 0.2	821.08 +/- 0.06	820.22 +/- 0.1	822.8 +/- 0.1	818.65
Octanol	824.3 +/- 0.3	824.7 +/- 0.1	823.9 +/- 0.2	827.0 +/- 0.1	821.69
Nonanol	827.2 +/- 0.1	828.0 +/- 0.1	826.9 +/- 0.2	830.6 +/- 0.2	824.15
Decanol	829.7 +/- 0.2	830.5 +/- 0.2	829.7 +/- 0.2	833.8 +/- 0.1	826.15

Table A7: Densities (kg/m^3) at different temperatures obtained using Model 1.

Temperature (K)	Methanol	Ethanol	Butanol	Hexanol	Decanol
283	804.4 +/- 0.1	796.6 +/- 0.2	815.5 +/- 0.2	827.0 +/- 0.4	840.3 +/- 0.4
298	788.2 +/- 0.1	782.7 +/- 0.1	804.27 +/- 0.07	815.6 +/- 0.2	829.7 +/- 0.2
303	782.8 +/- 0.1	777.97 +/- 0.06	800.3 +/- 0.2	812.4 +/- 0.3	826.6 +/- 0.2
313	771.9 +/- 0.2	768.37 +/- 0.07	791.9 +/- 0.1	805.0 +/- 0.3	819.9 +/- 0.2
323	760.14 +/- 0.08	758.4 +/- 0.1	783.2 +/- 0.1	797.3 +/- 0.1	813.1 +/- 0.2
333	748.27 +/- 0.09	747.92 +/- 0.06	774.85 +/- 0.07	789.9 +/- 0.1	806.1 +/- 0.2

Table A8: Densities (kg/m^3) at different temperatures obtained using Model 2.

Temperature (K)	Methanol	Ethanol	Butanol	Hexanol	Decanol
283	798.1 +/- 0.2	794.8 +/- 0.08	815.65 +/- 0.06	827.0 +/- 0.1	840.5 +/- 0.3
298	784.0 +/- 0.1	782.51 +/- 0.08	804.99 +/- 0.09	816.7 +/- 0.1	830.6 +/- 0.2
303	779.39 +/- 0.08	778.6 +/- 0.1	801.2 +/- 0.2	813.4 +/- 0.2	827.5 +/- 0.1
313	770.09 +/- 0.08	770.26 +/- 0.08	793.93 +/- 0.06	806.58 +/- 0.06	821.1 +/- 0.1
323	760.49 +/- 0.08	762.03 +/- 0.07	786.6 +/- 0.1	799.8 +/- 0.2	814.2 +/- 0.1
333	751.19 +/- 0.09	753.25 +/- 0.07	779.1 +/- 0.1	792.86 +/- 0.08	807.9 +/- 0.1

Table A9: Densities (kg/m^3) at different temperatures obtained using the PolCA model.

Temperature (K)	Methanol	Ethanol	Butanol	Hexanol	Decanol
283	800.9 +/- 0.2	794.8 +/- 0.1	814.7 +/- 0.2	826.1 +/- 0.2	839.7 +/- 0.5
298	785.6 +/- 0.2	781.74 +/- 0.09	803.3 +/- 0.2	815.6 +/- 0.2	829.7 +/- 0.4
303	780.3 +/- 0.1	777.4+/- 0.1	799.51 +/- 0.08	812.1 +/- 0.1	826.4 +/- 0.3
313	769.8 +/- 0.1	768.3 +/- 0.1	791.81 +/- 0.09	805.1 +/- 0.2	819.6 +/- 0.2
323	759.2 +/- 0.2	759.02 +/- 0.09	783.4 +/- 0.2	797.6 +/- 0.1	813.0 +/- 0.2
333	748.2 +/- 0.2	749.21 +/- 0.08	775.66 +/- 0.07	790.2 +/- 0.2	806.2 +/- 0.2

Table A10: Enthalpy of vaporisation (kJ/mol) with polarisation corrections.

	Model 1	Model 2	PolCA	TraPPE(C)	Experimental
Methanol	38.05 +/- 0.01	38.43 +/- 0.01	38.83 +/- 0.01	37.3 +/- 0.02	37.6 +/- 0.5
Ethanol	42.24 +/- 0.02	43.08 +/- 0.02	43.08 +/- 0.02	41.18 +/- 0.03	42.3 +/- 0.4
Propanol	46.82 +/- 0.03	47.92 +/- 0.04	47.76 +/- 0.04	45.4 +/- 0.03	47.0 +/- 1.0
Butanol	51.58 +/- 0.08	52.71 +/- 0.04	52.51 +/- 0.07	49.95 +/- 0.05	52.0 +/- 3.0
Pentanol	56.31 +/- 0.08	57.48 +/- 0.05	57.2 +/- 0.05	54.34 +/- 0.04	57.0 +/- 2.0
Hexanol	61.11 +/- 0.1	62.38 +/- 0.05	62.08 +/- 0.07	58.91 +/- 0.06	61.0 +/- 2.0
Heptanol	65.9 +/- 0.1	67.18 +/- 0.08	66.9 +/- 0.1	63.48 +/- 0.08	67.0 +/- 2.0
Octanol	70.66 +/- 0.06	72.03 +/- 0.05	71.74 +/- 0.08	68.09 +/- 0.07	72.0 +/- 2.0
Nonanol	75.56 +/- 0.09	76.91 +/- 0.07	76.43 +/- 0.09	72.69 +/- 0.08	77.0 +/- 6.0
Decanol	80.4 +/- 0.1	81.9 +/- 0.1	81.32 +/- 0.08	77.35 +/- 0.09	82.0 +/- 6.0

Table A11: Diffusion constant (10^{-5} cm^2/s).

	Model 1	Model 2	PolCA	TraPPE	Experimental
Methanol	1.722 +/- 0.009	2.58 +/- 0.01	1.736 +/- 0.009	2.391 +/- 0.005	2.35
Ethanol	0.987 +/- 0.006	1.227 +/- 0.005	0.937 +/- 0.009	1.325 +/- 0.006	1.01
Propanol	0.678 +/- 0.008	0.726 +/- 0.005	0.616 +/- 0.006	0.851 +/- 0.007	0.55
Butanol	0.49 +/- 0.006	0.503 +/- 0.002	0.45 +/- 0.004	0.619 +/- 0.004	0.42
Pentanol	0.394 +/- 0.005	0.379 +/- 0.002	0.351 +/- 0.004	0.473 +/- 0.004	0.30
Hexanol	0.327 +/- 0.006	0.299 +/- 0.003	0.292 +/- 0.006	0.378 +/- 0.003	0.22
Heptanol	0.266 +/- 0.005	0.244 +/- 0.003	0.241 +/- 0.004	0.338 +/- 0.004	0.17
Octanol	0.245 +/- 0.005	0.223 +/- 0.003	0.22 +/- 0.004	0.262 +/- 0.003	0.14
Nonanol	0.208 +/- 0.002	0.185 +/- 0.003	0.191 +/- 0.004	0.236 +/- 0.003	N/A
Decanol	0.185 +/- 0.003	0.163 +/- 0.003	0.168 +/- 0.003	0.203 +/- 0.001	N/A

Table A12: Dielectric constant with polarisation corrections.

	Model 1	Model 2	PolCA	TraPPE(C)	Experimental
Methanol	34.0 +/- 1.0	35.9 +/- 0.7	34.4 +/- 0.5	34.6 +/- 0.5	31.9
Ethanol	26.8 +/- 0.7	29.7 +/- 0.8	26.7 +/- 0.7	27.7 +/- 0.6	24.85
Propanol	21.0 +/- 1.0	25.0 +/- 1.0	22.0 +/- 1.0	23.0 +/- 1.0	20.3
Butanol	17.0 +/- 1.0	20.0 +/- 1.0	19.0 +/- 1.0	20.0 +/- 1.0	17.56
Pentanol	13.3 +/- 1.0	18.0 +/- 1.0	15.0 +/- 2.0	16.8 +/- 0.8	14.6
Hexanol	9.4 +/- 0.9	16.0 +/- 2.0	11.0 +/- 1.0	14.0 +/- 1.0	12.5
Heptanol	9.3 +/- 0.8	13.0 +/- 1.0	10.0 +/- 1.0	12.3 +/- 0.9	11.5
Octanol	6.5 +/- 0.7	10.3 +/- 0.7	8.0 +/- 1.0	12.0 +/- 1.0	9.7
Nonanol	4.8 +/- 0.5	9.0 +/- 1.0	6.2 +/- 0.8	8.2 +/- 0.4	8.6
Decanol	3.8 +/- 0.5	8.0 +/- 1.0	5.1 +/- 0.2	7.3 +/- 0.6	7.62

Table A13: Dipole (D). Simulated errors were all lower than 0.001 D

	Model 1	Model 2	PolCA	TraPPE	Estimated liquid dipole
Methanol	1.9834	2.0503	2.0732	2.2589	2.70
Ethanol	1.9809	2.0482	2.0704	2.2554	2.81
Propanol	1.9809	2.0482	2.0705	2.2554	2.87
Butanol	1.9808	2.048	2.0704	2.2552	2.88
Pentanol	1.9804	2.0479	2.0702	2.255	2.97
Hexanol	1.9802	2.0479	2.0701	2.2548	2.87
Heptanol	1.9803	2.0479	2.07	2.2547	2.99
Octanol	1.98	2.0477	2.0699	2.2546	2.90
Nonanol	1.9799	2.0477	2.0697	2.2545	2.73
Decanol	1.9798	2.0477	2.0697	2.2545	2.70

Table A14: Free energy of self-solvation (kJ/mol) with polarisation corrections.

	Model 1	Model 2	PolCA	TraPPE(C)	Experimental
Methanol	-19.8 +/- 0.4	-21.3 +/- 0.3	-20.7 +/- 0.4	-20.4 +/- 0.2	-20.52
Ethanol	-20.6 +/- 0.6	-22.7 +/- 0.3	-22.1 +/- 0.3	-21.0 +/- 0.3	-21.10
Propanol	-22.9 +/- 0.7	-24.9 +/- 0.4	-23.2 +/- 0.4	-22.6 +/- 0.6	-22.41
Butanol	-24.8 +/- 0.4	-27.4 +/- 0.3	-26.2 +/- 0.4	-24.9 +/- 0.2	-25.36
Pentanol	-27.5 +/- 0.6	-30.0 +/- 0.4	-28.5 +/- 0.5	-27.1 +/- 0.3	-28.22
Hexanol	-30.3 +/- 0.7	-32.3 +/- 0.5	-31.4 +/- 0.6	-29.4 +/- 0.5	-29.72
Heptanol	-33.0 +/- 0.7	-34.9 +/- 0.6	-33.7 +/- 0.4	-32.0 +/- 0.6	-32.82
Octanol	-35.8 +/- 0.3	-38.0 +/- 0.7	-36.4 +/- 0.5	-33.9 +/- 0.7	-34.97
Nonanol	-38.3 +/- 0.9	-40.2 +/- 0.8	-39.2 +/- 0.6	-36.4 +/- 0.7	-38.08
Decanol	-40.3 +/- 0.7	-43.3 +/- 0.8	-41.8 +/- 0.4	-38.4 +/- 0.9	-40.10

Table A15: Free energy of solvation in hexadecane without and with polarisation corrections (kJ/mol).

	PolCA	TraPPE	TraPPE(C)	Minnesota	Katryzky
Methanol	-8.9 +/- 0.1	-2.24 +/- 0.09	-9.11 +/- 0.09	-5.53	-8.49
Ethanol	-9.2 +/- 0.1	-5.0 +/- 0.2	-9.4 +/- 0.2	-8.50	-8.49
Propanol	-11.1 +/- 0.2	-7.9 +/- 0.1	-11.0 +/- 0.1	-11.60	-11.57
Butanol	-13.6 +/- 0.2	-10.7 +/- 0.2	-13.2 +/- 0.2	-14.86	-14.82
Pentanol	-16.5 +/- 0.2	-13.3 +/- 0.1	-15.4 +/- 0.1	-17.75	-17.73
Hexanol	-18.9 +/- 0.3	-16.5 +/- 0.1	-18.2 +/- 0.1	-20.60	-20.58
Heptanol	-21.6 +/- 0.2	-18.9 +/- 0.3	-20.4 +/- 0.3	-23.53	-23.49
Octanol	-24.4 +/- 0.3	-21.6 +/- 0.3	-22.9 +/- 0.3	-26.37	-26.34
Nonanol	-27.2 +/- 0.5	-24.6 +/- 0.4	-25.7 +/- 0.4	N/A	-29.19
Decanol	-30.1 +/- 0.3	-26.9 +/- 0.3	-27.9 +/- 0.3	-32.15	-32.10

Amines

Gromos 2016H66

Table A16: Density (kg/m^3).

	Gromos-2016-RF	Gromos-2016-RF-Dispcorr	Gromos-2016-PME	Experimental
Methylamine	707.30 +/- 0.07	721.5 +/- 0.1	728.23 +/- 0.09	655
Ethylamine	688.67 +/- 0.08	704.72 +/- 0.08	716.70 +/- 0.09	677
Propylamine	717.59 +/- 0.07	733.13 +/- 0.09	739.46 +/- 0.09	714
Butylamine	740.01 +/- 0.07	754.8 +/- 0.1	758.85 +/- 0.06	741
Pentylamine	755.53 +/- 0.09	769.89 +/- 0.05	772.46 +/- 0.08	751
Hexylamine	767.06 +/- 0.07	781.0 +/- 0.1	783.4 +/- 0.1	761
Heptylamine	775.75 +/- 0.08	789.46 +/- 0.09	791.66 +/- 0.08	772
Octylamine	782.77 +/- 0.08	796.15 +/- 0.08	798.04 +/- 0.07	779
Nonylamine	788.4 +/- 0.1	801.6 +/- 0.1	803.3 +/- 0.1	785
Decylamine	792.81 +/- 0.09	806.0 +/- 0.1	807.5 +/- 0.1	791

Table A17: Enthalpy of vaporisation with polarisation corrections (kJ/mol).

	Gromos-2016-RF	Gromos-2016-RF-Dispcorr	Gromos-2016-PME	Experimental
Methylamine	25.5 +/- 0.2	27.25 +/- 0.03	26.41 +/- 0.04	23.85
Ethylamine	26.3 +/- 0.2	28.36 +/- 0.03	27.96 +/- 0.05	29.0 +/- 2.0
Propylamine	30.5 +/- 0.2	33.02 +/- 0.03	32.21 +/- 0.07	31.3 +/- 0.1
Butylamine	35.4 +/- 0.2	38.45 +/- 0.04	37.5 +/- 0.1	35.8 +/- 0.2
Pentylamine	39.9 +/- 0.2	43.44 +/- 0.04	42.5 +/- 0.2	40.3 +/- 0.7
Hexylamine	44.6 +/- 0.3	48.60 +/- 0.06	47.8 +/- 0.2	45.1 +/- 0.1
Heptylamine	49.2 +/- 0.2	53.50 +/- 0.06	52.8 +/- 0.2	50.0 +/- 0.1
Octylamine	53.9 +/- 0.3	58.71 +/- 0.04	58.2 +/- 0.2	54.6 +/- 0.2
Nonylamine	58.3 +/- 0.3	63.68 +/- 0.06	63.2 +/- 0.2	59.3
Decylamine	62.8 +/- 0.4	68.86 +/- 0.06	68.6 +/- 0.3	64

Table A18: Dielectric constant with polarisation corrections.

	Gromos-2016-RF	Gromos-2016-RF-Dispcorr	Gromos-2016-PME	Experimental
Methylamine	12.6 +/- 0.1	12.96 +/- 0.09	13.0 +/- 0.2	9.40
Ethylamine	6.07 +/- 0.04	6.27 +/- 0.04	6.17 +/- 0.04	6.61
Propylamine	4.82 +/- 0.03	4.9 +/- 0.02	4.89 +/- 0.03	4.99
Butylamine	4.44 +/- 0.03	4.47 +/- 0.04	4.46 +/- 0.02	4.62
Pentylamine	3.71 +/- 0.03	3.73 +/- 0.03	3.7 +/- 0.01	4.20
Hexylamine	3.47 +/- 0.02	3.48 +/- 0.02	3.47 +/- 0.03	4.03
Heptylamine	3.209 +/- 0.008	3.21 +/- 0.02	3.2 +/- 0.02	3.77
Octylamine	3.07 +/- 0.02	3.07 +/- 0.009	3.07 +/- 0.02	3.55
Nonylamine	2.95 +/- 0.009	2.96 +/- 0.01	2.95 +/- 0.01	3.38
Decylamine	2.868 +/- 0.009	2.87 +/- 0.01	2.87 +/- 0.01	3.28

Table A19: Dipole (D). Simulated errors were all lower than 0.001 D

	Gromos-2016-RF	Gromos-2016-RF-Dispcorr	Gromos-2016-PME	Est. Liq. Dipole
Methylamine	1.8945	1.8962	1.8979	2.01
Ethylamine	1.8446	1.8462	1.8491	1.83
Propylamine	1.8371	1.8387	1.8403	1.73
Butylamine	1.8358	1.8375	1.838	1.97
Pentylamine	1.8343	1.836	1.8363	1.75
Hexylamine	1.8327	1.8345	1.8348	1.87
Heptylamine	1.8313	1.8333	1.8336	1.72
Octylamine	1.8298	1.8319	1.8322	1.84
Nonylamine	1.828	1.8301	1.8309	1.69
Decylamine	1.8265	1.8289	1.8289	1.82

Table A20: Free energy of self-solvation with polarisation corrections (kJ/mol).

	Gromos-2016-RF	Gromos-2016-PME	Experimental
Methylamine	-13.23 +/- 0.1	-13.7 +/- 0.1	-12.40
Ethylamine	-13.5 +/- 0.2	-14.4 +/- 0.2	-13.90
Propylamine	-15.6 +/- 0.2	-16.9 +/- 0.3	-16.44
Butylamine	-18.4 +/- 0.4	-19.6 +/- 0.2	-18.83
Pentylamine	-20.7 +/- 0.2	-22.5 +/- 0.2	-21.18
Hexylamine	-22.8 +/- 0.4	-25.0 +/- 0.4	-24.00
Heptylamine	-24.9 +/- 0.2	-28.3 +/- 0.5	-27.19
Octylamine	-28.1 +/- 0.4	-31.1 +/- 0.3	-29.42
Nonylamine	-30.2 +/- 0.4	-33.9 +/- 0.6	N/A
Decylamine	-32.9 +/- 0.5	-37.1 +/- 0.4	-34.35

Table A21: Free energy of solvation in hexadecane with polarisation corrections (kJ/mol).

	Gromos-2016-RF	Gromos-2016-PME	Minnesota	Katryzky	Bordner
Methylamine	-4.09 +/- 0.08	-4.4 +/- 0.1	N/A	N/A	-7.42
Ethylamine	-5.8 +/- 0.3	-6.4 +/- 0.2	-9.58	-9.58	-9.59
Propylamine	-8.4 +/- 0.2	-9.5 +/- 0.2	-12.22	-12.20	-12.22
Butylamine	-11.9 +/- 0.3	-13.0 +/- 0.2	N/A	-14.94	-14.96
Pentylamine	-14.2 +/- 0.2	-16.2 +/- 0.2	-17.91	-17.90	-17.93
Hexylamine	-17.0 +/- 0.3	-19.3 +/- 0.3	N/A	N/A	-20.89
Heptylamine	-20.0 +/- 0.4	-22.8 +/- 0.6	N/A	N/A	-23.81
Octylamine	-22.7 +/- 0.5	-26.0 +/- 0.3	N/A	N/A	-25.80
Nonylamine	-25.8 +/- 0.3	-29.0 +/- 0.4	N/A	N/A	N/A
Decylamine	-28.7 +/- 0.4	-32.7 +/- 0.3	N/A	N/A	N/A

Table A22: Density (kg/m^3).

	Model 1	Model 2	Model 3	Experimental
Methylamine	655.5 +/- 0.2	654.1 +/- 0.1	662.7 +/- 0.1	655.00
Ethylamine	680.7 +/- 0.1	684.99 +/- 0.07	690.00 +/- 0.07	677.00
Propylamine	711.45 +/- 0.08	713.8 +/- 0.1	717.30 +/- 0.05	714.00
Butylamine	735.70 +/- 0.06	734.43 +/- 0.08	735.62 +/- 0.06	741.00
Pentylamine	753.19 +/- 0.07	750.57 +/- 0.07	750.80 +/- 0.05	751.00
Hexylamine	766.82 +/- 0.06	763.14 +/- 0.04	762.29 +/- 0.07	761.00
Heptylamine	777.13 +/- 0.09	773.10 +/- 0.05	771.73 +/- 0.09	772.00
Octylamine	785.3 +/- 0.1	781.10 +/- 0.09	779.54 +/- 0.04	779.00
Nonylamine	792.0 +/- 0.1	787.72 +/- 0.07	786.10 +/- 0.09	785.00
Decylamine	797.7 +/- 0.1	793.4 +/- 0.1	791.65 +/- 0.07	791.00

Table A23: Enthalpy of vaporisation with polarisation corrections (kJ/mol).

	Model 1	Model 2	Model 3	Experimental
Methylamine	25.63 +/- 0.02	24.49 +/- 0.02	25.62 +/- 0.01	23.85
Ethylamine	27.57 +/- 0.06	27.24 +/- 0.04	28.15 +/- 0.04	29.0 +/- 2.0
Propylamine	30.75 +/- 0.08	30.60 +/- 0.04	31.68 +/- 0.06	31.3 +/- 0.1
Butylamine	35.51 +/- 0.09	35.03 +/- 0.08	35.95 +/- 0.05	35.8 +/- 0.2
Pentylamine	39.97 +/- 0.07	39.26 +/- 0.03	40.12 +/- 0.05	40.3 +/- 0.7
Hexylamine	44.68 +/- 0.06	43.75 +/- 0.05	44.68 +/- 0.07	45.1 +/- 0.1
Heptylamine	49.2 +/- 0.09	48.24 +/- 0.05	49.1 +/- 0.1	50.0 +/- 0.1
Octylamine	53.98 +/- 0.06	52.92 +/- 0.08	53.72 +/- 0.07	54.6 +/- 0.2
Nonylamine	58.6 +/- 0.1	57.39 +/- 0.07	58.2 +/- 0.1	59.3
Decylamine	63.41 +/- 0.07	62.09 +/- 0.08	62.99 +/- 0.08	64

Table A24: Dielectric constant with polarisation corrections.

	Model 1	Model 2	Model 3	Experimental
Methylamine	15.3 +/- 0.1	16.7 +/- 0.1	15.8 +/- 0.1	9.40
Ethylamine	7.34 +/- 0.04	9.54 +/- 0.07	7.77 +/- 0.05	6.61
Propylamine	5.51 +/- 0.02	7.19 +/- 0.04	5.54 +/- 0.04	4.99
Butylamine	4.89 +/- 0.04	6.20 +/- 0.06	5.63 +/- 0.03	4.62
Pentylamine	4.04 +/- 0.02	5.22 +/- 0.03	4.33 +/- 0.02	4.20
Hexylamine	3.68 +/- 0.03	4.68 +/- 0.03	4.25 +/- 0.03	4.03
Heptylamine	3.41 +/- 0.01	4.28 +/- 0.03	3.68 +/- 0.02	3.77
Octylamine	3.23 +/- 0.01	3.99 +/- 0.02	3.71 +/- 0.02	3.55
Nonylamine	3.080 +/- 0.008	3.77 +/- 0.02	3.32 +/- 0.01	3.38
Decylamine	2.978 +/- 0.009	3.56 +/- 0.02	3.38 +/- 0.01	3.28

Table A25: Dipole (D). Simulated errors were all lower than 0.001 D

	Model 1	Model 2	Model 3
Methylamine	1.9487	1.9684	1.7841
Ethylamine	1.8942	1.9334	1.7409
Propylamine	1.88034	1.9233	1.7285
Butylamine	1.8798	1.9211	1.7239
Pentylamine	1.8784	1.9187	1.7198
Hexylamine	1.8776	1.9172	1.7168
Heptylamine	1.8768	1.9156	1.7138
Octylamine	1.8762	1.9142	1.7116
Nonylamine	1.875	1.9129	1.7099
Decylamine	1.8741	1.9118	1.7075

Table A26: Free energy of self-solvation with polarisation corrections (kJ/mol).

	Model 1	Model 2	Model 3	Experimental
Methylamine	-12.9 +/- 0.1	-13.7 +/- 0.2	-15.7 +/- 0.1	-12.40
Ethylamine	-13.6 +/- 0.1	-15.0 +/- 0.3	-17.19 +/- 0.07	-13.90
Propylamine	-15.1 +/- 0.3	-17.0 +/- 0.2	-19.5 +/- 0.3	-16.44
Butylamine	-18.0 +/- 0.2	-19.7 +/- 0.4	-21.9 +/- 0.1	-18.83
Pentylamine	-20.2 +/- 0.3	-21.8 +/- 0.3	-24.2 +/- 0.4	-21.18
Hexylamine	-22.9 +/- 0.5	-23.9 +/- 0.2	-26.9 +/- 0.2	-24.00
Heptylamine	-25.1 +/- 0.5	-26.6 +/- 0.4	-29.1 +/- 0.3	-27.19
Octylamine	-27.8 +/- 0.5	-28.9 +/- 0.3	-32.0 +/- 0.4	-29.42
Nonylamine	-30.6 +/- 0.5	-31.2 +/- 0.5	-34.6 +/- 0.2	N/A
Decylamine	-32.4 +/- 0.4	-34.3 +/- 0.5	-37.5 +/- 0.3	-34.35

Table A27: Free energy of solvation in hexadecane with polarisation corrections (kJ/mol).

	Model 1	Model 2	Model 3	Minnesota	Katryzky	Bordner
Methylamine	-3.1 +/- 0.1	-6.5 +/- 0.1	-11.51 +/- 0.09	N/A	N/A	-7.42
Ethylamine	-4.9 +/- 0.1	-8.0 +/- 0.1	-13.0 +/- 0.3	-9.58	-9.58	-9.59
Propylamine	-7.5 +/- 0.2	-10.8 +/- 0.2	-15.4 +/- 0.3	-12.22	-12.20	-12.22
Butylamine	-10.9 +/- 0.3	-13.5 +/- 0.2	-18.4 +/- 0.2	N/A	-14.94	-14.96
Pentylamine	-13.2 +/- 0.3	-16.2 +/- 0.2	-20.9 +/- 0.2	-17.91	-17.90	-17.93
Hexylamine	-16.2 +/- 0.3	-19.0 +/- 0.3	-24.0 +/- 0.2	N/A	N/A	-20.89
Heptylamine	-18.8 +/- 0.5	-22.0 +/- 0.3	-26.4 +/- 0.4	N/A	N/A	-23.81
Octylamine	-21.7 +/- 0.4	-24.7 +/- 0.3	-29.0 +/- 0.3	N/A	N/A	-25.80
Nonylamine	-24.6 +/- 0.5	-27.6 +/- 0.3	-32.3 +/- 0.6	N/A	N/A	N/A
Decylamine	-27.9 +/- 0.4	-30.4 +/- 0.3	-35.2 +/- 0.4	N/A	N/A	N/A

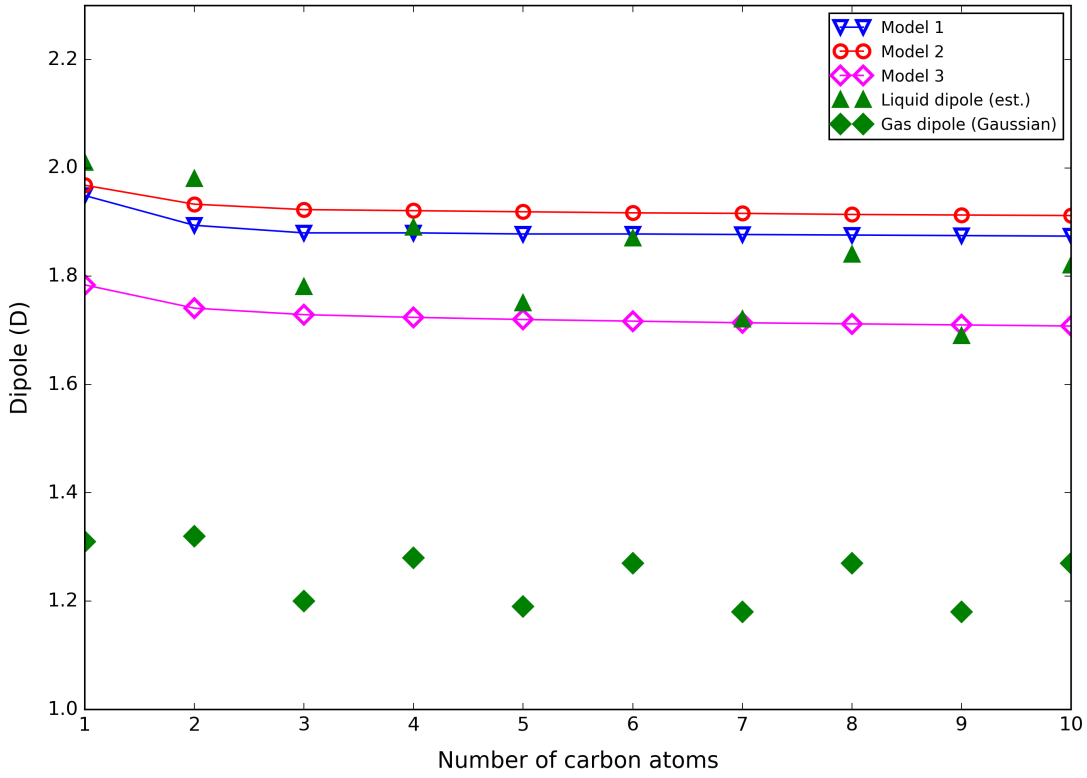

Figure A17: Liquid dipole of linear primary amines at 298.15 K and 1 bar obtained using Model 1 (blue triangles), Model 2 (red circles) and Model 3 (purple rhombus). The green triangles are the liquid dipoles obtained with the equation proposed by Leontyev and Stuchebrukhov (Equation 2.49) [92] and the green diamonds are the gas dipoles obtained using Gaussian 09.

Ketones

Table A28: Density (kg/m^3).

	Model 1	Model 2	Model 3	TraPPE	Experimental
Propanone	786.75 +/- 0.07	784.29 +/- 0.08	787.10 +/- 0.05	777.55 +/- 0.04	785.00
Butanone	794.04 +/- 0.04	792.79 +/- 0.04	791.34 +/- 0.06	788.98 +/- 0.04	799.00
Pentanone	798.37 +/- 0.07	797.11 +/- 0.04	795.76 +/- 0.03	795.83 +/- 0.06	802.00
Hexanone	805.61 +/- 0.05	804.50 +/- 0.05	803.41 +/- 0.04	804.72 +/- 0.04	807.00
Heptanone	810.41 +/- 0.08	809.28 +/- 0.06	808.64 +/- 0.08	810.69 +/- 0.06	811.00
Octanone	814.60 +/- 0.07	813.58 +/- 0.06	813.33 +/- 0.09	815.76 +/- 0.08	815.00
Nonanone	818.07 +/- 0.07	817.12 +/- 0.09	816.98 +/- 0.07	819.92 +/- 0.07	816.70
Decanone	821.17 +/- 0.06	820.24 +/- 0.05	820.26 +/- 0.09	823.5 +/- 0.1	819.80

Table A29: Diffusion constant (10^{-5} cm^2/s)

	Model 1	Model 2	Model 3	TraPPE	Experimental
Propanone	4.14 +/- 0.02	3.95 +/- 0.02	4.30 +/- 0.02	4.94 +/- 0.02	4.2
Butanone	3.69 +/- 0.02	3.54 +/- 0.02	3.69 +/- 0.01	4.36 +/- 0.02	3.58
Pentanone	3.00 +/- 0.01	2.89 +/- 0.01	2.77 +/- 0.01	3.57 +/- 0.02	2.97
Hexanone	2.37 +/- 0.01	2.275 +/- 0.008	2.09 +/- 0.01	2.82 +/- 0.02	2.06
Heptanone	1.88 +/- 0.01	1.81 +/- 0.01	1.61 +/- 0.01	2.239 +/- 0.009	1.61
Octanone	1.49 +/- 0.01	1.429 +/- 0.005	1.22 +/- 0.01	1.79 +/- 0.01	1.18
Nonanone	1.232 +/- 0.007	1.197 +/- 0.009	1.015 +/- 0.005	1.49 +/- 0.01	0.957
Decanone	0.994 +/- 0.008	0.963 +/- 0.007	0.804 +/- 0.007	1.2 +/- 0.01	0.747

Table A30: Enthalpy of vaporisation with polarisation corrections (kJ/mol).

	Model 1	Model 2	Model 3	TraPPE (C)	Experimental
Propanone	29.36 +/- 0.02	30.04 +/- 0.02	30.17 +/- 0.02	26.17 +/- 0.02	31.30
Butanone	33.09 +/- 0.02	33.77 +/- 0.02	33.65 +/- 0.04	29.88 +/- 0.03	34.70
Pentanone	37.77 +/- 0.05	38.39 +/- 0.04	38.28 +/- 0.03	34.35 +/- 0.04	38.40
Hexanone	42.29 +/- 0.03	42.98 +/- 0.08	42.76 +/- 0.07	38.69 +/- 0.06	43.10
Heptanone	47.22 +/- 0.07	47.76 +/- 0.04	47.60 +/- 0.05	43.36 +/- 0.06	48.50
Octanone	51.97 +/- 0.05	52.71 +/- 0.09	52.45 +/- 0.07	47.98 +/- 0.04	52.60
Nonanone	56.72 +/- 0.09	57.43 +/- 0.04	57.21 +/- 0.07	52.52 +/- 0.05	56.40
Decanone	61.43 +/- 0.08	62.02 +/- 0.09	61.82 +/- 0.06	56.90 +/- 0.07	60.90

Table A31: Dielectric constant with polarisation corrections.

	Model 1	Model 2	Model 3	TraPPE (C)	Experimental
Propanone	31.9 +/- 0.4	30.0 +/- 0.4	32.2 +/- 0.4	30.9 +/- 0.3	20.49
Butanone	27.8 +/- 0.2	25.6 +/- 0.3	27.5 +/- 0.3	27.1 +/- 0.4	18.25
Pentanone	22.1 +/- 0.3	20.7 +/- 0.2	22.4 +/- 0.3	21.6 +/- 0.2	15.20
Hexanone	19.0 +/- 0.2	17.8 +/- 0.2	19.4 +/- 0.2	18.6 +/- 0.3	14.14
Heptanone	15.6 +/- 0.2	15.0 +/- 0.2	16.3 +/- 0.3	15.4 +/- 0.1	11.66
Octanone	14.92 +/- 0.09	14.4 +/- 0.2	15.7 +/- 0.2	14.7 +/- 0.1	10.50
Nonanone	13.7 +/- 0.2	12.9 +/- 0.2	14.4 +/- 0.1	13.5 +/- 0.2	9.57
Decanone	11.1 +/- 0.2	10.55 +/- 0.07	11.6 +/- 0.2	10.9 +/- 0.1	8.88

Table A32: Dipole (D). Simulated errors were all lower than 0.001 D

	Model 1	Model 2	Model 3	TraPPE	Estimated liquid dipole
Propanone	2.7233	2.642	3.23416	2.5029	4.72
Butanone	2.7233	2.63829	3.22901	2.5029	4.70
Pentanone	2.7233	2.63822	3.22892	2.5029	4.58
Hexanone	2.7233	2.63818	3.22889	2.5029	4.57
Heptanone	2.7233	2.63812	3.22883	2.5029	4.42
Octanone	2.7233	2.63808	3.2288	2.5029	4.59
Nonanone	2.7233	2.63806	3.2288	2.5029	4.61
Decanone	2.7233	2.63803	3.22876	2.5029	4.29

Table A33: Free energy of self-solvation with polarisation corrections (kJ/mol).

	Model 1	Model 2	Model 3	TraPPE (C)	Experimental
Propanone	-18.1 +/- 0.1	-18.55 +/- 0.09	-18.2 +/- 0.1	-16.0 +/- 0.2	-17.35
Butanone	-20.30 +/- 0.05	-20.8 +/- 0.2	-20.0 +/- 0.1	-20.1 +/- 0.2	-19.24
Pentanone	-23.0 +/- 0.1	-23.3 +/- 0.3	-22.6 +/- 0.3	-20.7 +/- 0.3	-20.79
Hexanone	-25.3 +/- 0.1	-25.8 +/- 0.3	-25.1 +/- 0.3	-23.0 +/- 0.4	-24.55
Heptanone	-28.5 +/- 0.1	-29.0 +/- 0.2	-27.7 +/- 0.2	-26.3 +/- 0.3	-26.38
Octanone	-31.3 +/- 0.2	-31.6 +/- 0.3	-30.3 +/- 0.6	-28.7 +/- 0.3	-28.95
Nonanone	-34.3 +/- 0.3	-34.7 +/- 0.3	-33.2 +/- 0.4	-31.2 +/- 0.6	-31.53
Decanone	-37.3 +/- 0.3	-37.3 +/- 0.9	-35.7 +/- 0.6	-33.5 +/- 0.5	-33.57

Table A34: Free energy of solvation in hexadecane with and without polarisation corrections (kJ/mol).

	PolCA	TraPPE	TraPPE(C)	Minnesota	Katryzky
Propanone	-17.6 +/- 0.2	-10.5 +/- 0.2	-20.6 +/- 0.2	-9.66	-9.64
Butanone	-17.9 +/- 0.3	-13.2 +/- 0.2	-20.5 +/- 0.2	-13.05	-13.06
Pentanone	-19.3 +/- 0.2	-16.3 +/- 0.3	-21.9 +/- 0.3	-15.73	-15.74
Hexanone	-21.2 +/- 0.3	-18.7 +/- 0.2	-23.3 +/- 0.2	-18.62	-18.59
Heptanone	-23.3 +/- 0.4	-21.9 +/- 0.6	-25.7 +/- 0.6	-21.46	-21.44
Octanone	-26.6 +/- 0.2	-24.2 +/- 0.5	-27.8 +/- 0.5	-24.31	-24.29

Mixtures

Table A35: Free energy of solvation of alkanes in octanol (kJ/mol).

Model	PolCA	TraPPE (C)
Methane	1.9 +/- 0.2	2.1 +/- 0.2
Ethane	-2.6 +/- 0.3	-2.0 +/- 0.1
Propane	-5.6 +/- 0.2	-4.75 +/- 0.08
Butane	-8.4 +/- 0.3	-7.4 +/- 0.2
Pentane	-11.4 +/- 0.4	-9.7 +/- 0.3
Hexane	-14.3 +/- 0.4	-12.8 +/- 0.3
Heptane	-16.8 +/- 0.7	-15.0 +/- 0.2
Octane	-20.0 +/- 0.3	-17.7 +/- 0.4
Nonane	-22.4 +/- 0.3	-20.0 +/- 0.3

Table A36: Solvation free energy of amines in octanol without and with polarisation corrections (kJ/mol).

	PolCA	Gromos-PME	Gromos-PME(C)	Minnesota	Katryzky	Bordner
Methylamine	-8.6 +/- 0.3	-7.1 +/- 0.5	-8.1 +/- 0.5	-15.82	-10.83	-15.81
Ethylamine	-10.5 +/- 0.3	-8.6 +/- 0.3	-9.2 +/- 0.3	-17.11	N/A	-18.10
Propylamine	-12.8 +/- 0.3	-12.6 +/- 0.3	-13.0 +/- 0.3	-19.96	N/A	-21.12
Butylamine	-15.6 +/- 0.3	-15.6 +/- 0.5	-16.0 +/- 0.5	-22.30	N/A	-22.78
Pentylamine	-18.5 +/- 0.4	-19.2 +/- 0.6	-19.5 +/- 0.6	N/A	-20.58	-25.41
Hexylamine	-21.1 +/- 0.4	-22.2 +/- 0.7	-22.5 +/- 0.7	N/A	N/A	-28.32
Heptylamine	-23.4 +/- 0.6	-26.0 +/- 0.7	-26.2 +/- 0.7	N/A	N/A	-30.54
Octylamine	-27.1 +/- 0.7	-29.4 +/- 0.9	-29.6 +/- 0.9	N/A	N/A	N/A

Table A37: Solvation free energy of ketones in octanol without and with polarisation corrections (kJ/mol).

	PolCA	TraPPE	TraPPE(C)	Minnesota	Katryzky
Propanone	-21.2 +/- 0.6	-12.1 +/- 0.1	-15.1 +/- 0.1	-13.18	-13.23
Butanone	-23.2 +/- 0.9	-14.3 +/- 0.2	-16.4 +/- 0.2	-15.82	-16.03
Pentanone	-25.1 +/- 0.8	-16.7 +/- 0.3	-18.3 +/- 0.3	-18.20	-18.19
Hexanone	-27.8 +/- 0.8	-19.7 +/- 0.3	-21.1 +/- 0.3	-21.00	-20.98
Heptanone	-30.1 +/- 1.0	-22.4 +/- 0.4	-23.5 +/- 0.4	-23.64	-23.66
Octanone	-34.2 +/- 0.8	-24.9 +/- 0.5	-25.9 +/- 0.5	-26.69	N/A

Other simulated values

Table A39: Enthalpy of vaporisation and self-solvation free energy of 2-aminoethanol at 298.15 K simulated with PolCA.

	PolCA	**Experimental**
$\mathbf{\Delta H_{vap}}$	52.70 +/- 0.06	59.63
$\mathbf{\Delta G_{solv}}$	-27.7 +/- 0.4	-34.03

Table A38: Solvation free energy of alcohols in butanone with polarisation corrections (kJ/mol). Polarisation corrections are almost zero, and thus TraPPE(C) is the only one shown here.

	PolCA	TraPPE(C)	Katryzky
Methanol	-21.7 +/- 0.2	-12.2 +/- 0.1	-16.02
Ethanol	-23.8 +/- 0.3	-14.6 +/- 0.1	-18.07
Propanol	-26.4 +/- 0.3	-17.1 +/- 0.1	-20.70
Butanol	-28.9 +/- 0.2	-20.1 +/- 0.3	-23.49
Pentanol	-31.6 +/- 0.4	-22.8 +/- 0.2	N/A
Hexanol	-34.5 +/- 0.3	-25.6 +/- 0.2	-28.22
Heptanol	-36.8 +/- 0.4	-28.2 +/- 0.3	N/A
Octanol	-39.5 +/- 0.4	-30.5 +/- 0.4	-34.15