Isabela Quintela Matos

Solvation Free Energy Calculations of Molecules Mimicking Asphaltenes Using The SAFT-γ Mie Force Field

Rio de Janeiro

2018

Isabela Quintela Matos

Solvation Free Energy Calculations of Molecules Mimicking Asphaltenes Using The SAFT-γ Mie Force Field

Masther's thesis presented to Engenharia de Processos Químicos e Bioquímicos graduate program, Escola de Química, Universidade Federal do Rio de Janeiro, as required for obtaining a Master's degree in Chemical Engineering.

Universidade Federal do Rio de Janeiro

Escola de Química

Engenharia de Processos Químicos e Bioquímicos Graduate Program

Supervisor: Charlles Rubber de Almeida Abreu
Co-supervisor: Papa Matar Ndiaye

Rio de Janeiro

2018

Isabela Quintela Matos

Solvation Free Energy Calculations of Molecules Mimicking Asphaltenes Using The SAFT-γ Mie Force Field

Masther's thesis presented to Engenharia de Processos Químicos e Bioquímicos graduate program, Escola de Química, Universidade Federal do Rio de Janeiro, as required for obtaining a Master's degree in Chemical Engineering.

Approved dissertation. Rio de Janeiro, February 2, 2018:

_____

Charlles Rubber de Almeida Abreu
Supervisor

_____

Professor
Guest 1

_____

Professor
Guest 2

_____

Professor
Guest 3

Rio de Janeiro

2018

Abstract

This dissertation studied the solvation free energy differences of molecules mimicking asphaltenes in different solvents with the SAFT-γ Mie force field. Solvation free energy differences were obtained by carrying out molecular dynamics simulations at the expanded ensemble. The output from these simulations was then used to estimate the differences with the MBAR method. The results with solvents other than water had low absolute deviations in relation to the experimental data. Meanwhile, the hydration free energy calculations required a binary interaction parameter estimated with output data from molecular dynamics in order to obtain accurate free energy differences. These results indicated some problems on the SAFT-γ Mie model for water, but, generally, proved that a coarse grained model can represent the free energy differences of the studied sets of solute-solvent.

Keywords: solvation free energy. asphaltenes. SAFT-γ Mie force field.

List of Figures

List of Tables

List of symbols

T	Temperature
P	Pressure
V	Volume
t	Time
p	Momentum
r	Coordinates
U, u	Potential Energy
m	mass
v	velocity
P	Pressure
$\mathcal{H}$	Hamiltonian
q	Generalized Coordinates
K	Kinetic Energy
N	Number of atoms/molecules
h	Planck Constant
S	Entropy
k_b	Boltzmann Constant
β	$1/k_b T$
μ	Chemical Potential
A	Helmholtz Free Energy
G	Gibbs Free Energy
f	Free Energy
F	Forces

ϵ	Depth of the potential well
σ	Distance correspondent to a zero intermolecular potential
λ_r	Repulsive exponent
λ_a	Attractive exponent
x_i	Molar fraction
w_i	Weight fraction
ρ	Density
λ	Coupling Parameter
η	Arbitrary Weight
k_{ij}	Binary Interaction Parameter

Contents

1 Introduction

Solvation free energy calculations with molecular dynamics (MD) have a variety of applications ranging from drug design in the pharmaceutical industry to development of separation technologies in the chemical industry. Solvation free energy is, more specifically, the Gibbs free energy difference between the solute alone in the gas phase and the solute interacting with the solvent. Through the study of the solvation phenomenon, it is possible to obtain information about the behavior of the solvent in different chemical environments and the influence of the solute's molecular geometry. It is also possible to calculate other important properties with the solvation free energy, namely the activity coefficient at infinite dilution, Henry constant, partition coefficients and solubility.

The solvation phenomenon is intrinsically complex. There are many competing forces interfering in the behavior of the solute-solvent interaction and free energy simulations are susceptible to sampling problems in high energy regions. With the intention of improving free energy calculations, simulation methodologies such as the expanded ensemble (LYUBARTSEV *et al.*, 1992), thermodynamic integration (KIRKWOOD, 1935), free energy perturbation (ZWANZIG, 1954; BENNETT, 1976; SHIRTS; CHODERA, 2008) and umbrella sampling (TORRIE; VALLEAU, 1977) have been developed in order to obtain accurate estimations for the energy differences.

Another influencing factor in the output of these calculations are the force fields chosen to describe the solvent and solute molecules. Force fields have different levels of description (quantum mechanics, atomistic, coarse grained). In the coarse grained description, molecules are grouped in pseudo atoms or beads. Coarse grained models generally reproduce free energy differences since the effects of reducing degrees of freedom in the entropy are counterbalanced by the reduction of enthalpic terms (KMIECIK *et al.*, 2016). Additionally, the success of a coarse grained force field is important to increase the scale of solvation free energy calculations and reveal deficiencies in the description of small molecules by these models (MOBLEY *et al.*, 2007; SHIRTS *et al.*, 2003). That's the reason we, at this study, try to provide information about free energy calculation on the expanded ensemble with the SAFT-γ Mie coarse grained force field. This force field uses the Mie Potential (MIE, 1903) and has a more straightforward method of obtaining its parameters than other models. It was initially parameterized with pure component equilibrium and interfacial tension data. This strategy has provided satisfactory results for prediction of phase equilibrium of aromatic compounds, alkanes, light gases and water (HERDES; TOTTON; MÜLLER, 2015; MÜLLER; MEJÍA, 2017; LOBANOVA *et al.*, 2015) , thermodynamic properties of carbon dioxide and

methane (AIMOLI; MAGINN; ABREU, 2014a), multiphase equilibrium of mixtures of water, carbon dioxide and n-alkanes (LOBANOVA *et al.*, 2016) and water/oil interfacial tension (HERDES *et al.*, 2017).

The solvents and solutes chosen in this study range from standard sets used as benchmark in solvation free energy calculations to ring substances used as a model to asphaltenes. Asphaltenes are complicated to characterize by determining their composition on a molecular basis, but it is generally accepted that they can be characterized as a fraction of crude oil soluble in toluene and insoluble in n-alkenes (pentane, hexane, heptane) (SJöBLOM *et al.*, 2003). They've motivated many studies interested in developing models for their structure due to all the problems they can cause during their transportation and refining (SJöBLOM; SIMON; XU, 2015). This work's strategy to test the efficiency of the SAFT-γ Mie force field in describing the solvation phenomenon was to use molecules that appear in asphaltenes models and that have similarities in terms of solubilities with asphaltenes (phenanthrene, anthracene, pyrene) for the solutes. Meanwhile, for the solvents, we chose compounds that are used to characterize asphaltenes (toluene, hexane). The anti solvent/solvent effect of carbon dioxide was also tested due to its influence in asphaltene precipitation during the oil processing (SOROUSH *et al.*, 2014). Asphaltenes are described by their solubility, hence simulation scale can be increased and studies with more complete models can be carried out if this coarse grained model can correctly describe solvation free energy differences of simple molecules mimicking asphaltenes.

2 Literature Review

2.1 Molecular Simulations of Molecules Mimicking Asphaltenes

Asphaltenes consist of polyfunctional molecules, and they are defined by their solubility: insoluble in n-alkanes (pentane, hexane and heptane) and solube in tolune. Due to uncertainties related to its structures, a lot work has been done to develop model compounds that have well defined structure and can represent an average asphaltene. The two category of models presented in the literature are the archipelago and continental models. In the archipelago, asphaltenes consist of polyaromatic parts linked together by aliphatic or naphthemic moieties and, in the continental, they consist of a single polyaromatic ring with linked aliphatic or naphthenic chains (MULLINS, 2010; MURGICH, 2003). Studies with molecular simulation uses this models or compounds, such as like toluene and pyrene, that have similar solubility properties. The model's structure as chemical bonding can cause high energies during the simulation and, consequently, low relative occurrence probabilities (LI; GREENFIELD, 2011) .

Ervik *et al.* (2016) obtained correct interfacial orientation of asphaltenes using coarse grained molecular dynamics simulations of the interface, with an accurate model for the asphaltene molecules. Using a coarse grained force field also, Jover *et al.* (2015) carried molecular simulations with a continental asphaltene model. The results reproduced experimental data if the strong aggregation of asphaltene molecules in n-heptane and high solubility in toluene. Meanwhile, Gao *et al.* (2014) performed a molecular dynamics study with the force field GROMOS 45a3 force field to identify the structural features of different asphaltene molecules (C5 Pe and anionic C5 Pe). The archipelago model was employed in the work of Mikami *et al.* (2013) to investigate interfacial behavior of asphaltene molecules at the oil - water interface using molecular dynamics simulations with the OPLS-AA force field. They found that asphaltenes are preferably distributed in the oil phase in the case of pure toluene and at the oil - water interface in

the case of pure heptane. A molecular oscillation behavior of asphaltene molecules at the oil - water interface was also discovered in the shape of a nanoscale aggregate.

A pyrelene based model was used in the work of Teklebrhan *et al.* (2012), Teklebrhan *et al.* (2014) to study molecular association and interaction as well as the adsorption properties of the pyrelene model at the water/toluene or heptane interface. Molecular dynamics simulations were also used to the study the nanoaggregation of four types of model asphaltene molecules in binary mixtures of toluene and water (KUZNICKI; MASLIYAH; BHATTACHARJEE, 0). This study observed that in thin films of toluene trapped between two aqueous phases, both interface-bound and core-bound asphaltenes have similar diffusion behavior. Headen *et al.* (2017) reported molecular dynamics simulations of four model asphaltenes. They aver that there is no formation of nanoaggregates, and the distribution of clusters is continuous for mixtures of asphaltene in heptane. Generally the works on molecular simulation with asphaltenes models showed a difference in packing tendencies according to the model (GREENFIELD, 2011). For a thorough review of studies of different models of asphaltenes, the reader is referred elsewhere (GREENFIELD, 2011).

2.2   Coarse Grained Force Fields

Molecular simulations can be carried out at different levels of description. The detailed atomistic level or *ab initio* level is described by the laws of quantum mechanics. The system consists of a set of subatomic particles in which Schrodinger's equation is solved for all of them. The next level is the atomistic description. It considers that the system is made up of atoms following the laws of statistical mechanics. Force fields at this level are based on pair potentials with Coulombic charged sites. Contributions due to intramolecular interactions such as bond-stretching, angle-bending and torsion are also usually accounted by these kinds of force fields. When the simulation scale needs to be increased and the atomistic simulations become too computationally expensive, the coarse-grained (CG) description is more suited. It considers that the system is made up of pseudo atoms or beads that contain multiple atoms or even an entire molecule.

There is an obvious loss of information in grouping atoms, hence it is necessary to assure that the process of eliminating unnecessary or unimportant information ('coarse graining') doesn't affect the system's physical behavior. Ideally, coarse grained models need to have accuracy, transferability, robustness, and computational efficiency. In order to achieve these goals, coarse grained force fields are usually developed by mapping the atomistic model in order to define the pseudo atoms, which are generally formed by similar functional groups. The level of coarse graining also needs to be defined, up to 6 heavy atoms (non-hydrogen atoms) per bead in order to not lose important details and maintain isotropic representations of the beads (SHINODA; DEVANE; KLEIN, 2007; MARRINK *et al.*, 2007; HADLEY; MCCABE, 2012). CG force fields can be parametrized following two different approaches, bottoms up and top down, to link the simulations on the coarse grained scale to a more detailed scale according to the schematic representation in Figure 2.2.1. The bottoms up approach uses information of a more detailed scale such as the *ab initio* description or the atomistic description to obtain information necessary to the parametrization. This method highly depends on the detailed model quality to succeed. Meanwhile, the top down methodology obtains parameters from larger scales, namely experimental thermodynamic properties or native-structure based properties.

H
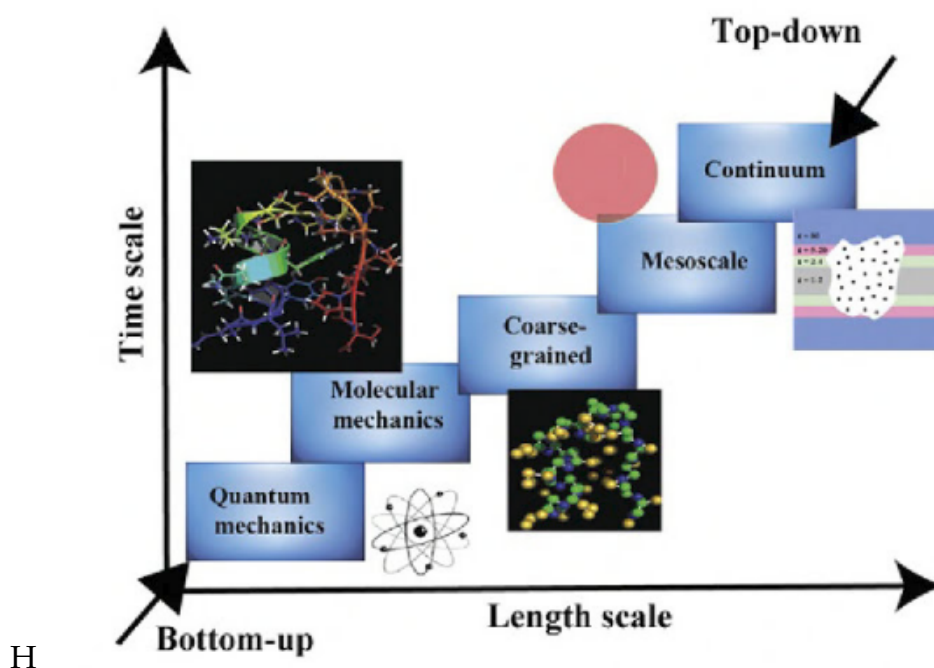
Figure 2.2.1 – Schematic representation for the two different approaches of coarse graining.Taken from Karabencheva-Christova (2014)

One of the first applications of coarse grained models is the study of protein folding (LEVITT; WARSHE, 1975; LEVITT, 1976). These earlier protein CG models were based on molecule structure, and they contributed for the knowledge of physicochemical forces associated with protein folding and protein interactions (KOGA; TAKADA, 2001). More recent, models focused on retaining the protein's chemical specificity. The Bereau and Deresmo model (BEREAU; DESERNO, 2009) has up to four-bead representation and was used in studies of protein folding and aggregation. However, this model still needs tuning to improve protein stability (BEREAU; BACHMANN; DESERNO, 2010). The OPEP (Optimized Potential for Efficient Protein Structure Prediction) model (STERPONE *et al.*, 2014; STERPONE; DERREUMAUX; MELCHIONNA, 2015) has up to six-bead representation. It was used to investigate a variety of phenomenon, ranging from protein folding to *ab initio* peptide structure prediction (BARDUCCI; BONOMI; DERREUMAUX, 2011; CHEBARO *et al.*, 2009b; CHEBARO *et al.*, 2009a). Another CG protein models used in the literature are the Scorpion (solvated coarse-grained protein interaction) (BASDEVANT; BORGIS; HA-DUONG, 2013), the UNRES (united residue) (ADAM *et al.*, 2014) and the MARTINI model (LARS *et al.*, 2013). The later one is the

most popular model for CG modeling of membrane proteins (MARRINK; TIELEMAN, 2013). The MARTINI model is also extensively used as CG model for water. This model represents four water molecules as one bead using a shifted Lennard Jones potential for non bonded interactions. Though its extensive use, the MARTINI water model doesn't properly represent properties as interfacial tension and compressibility (HE *et al.*, 2010) and can freeze at room temperature (WINGER *et al.*, 2009; MARRINK *et al.*, 2007), what requires the use of anti-freeze agents during the simulations. This behavior can be explained by the high level of coarse graining (4:1), the lack of explicit charges and the use of a 12-6 potential. Chiu, Scott and Jakobsson (2010) used the Morse Potential, which is softer than the LJ potential, to improve the MARTINI model. Meanwhile, Shinoda, Devane and Klein (2007) used different forms of the Mie potential. They concluded that a 12-4 Mie Potential was ideal for water cross interactions and a 9-6 Mie Potential was suited for solute-solute interactions.

Outside of the Martini framework, He *et al.* (2010) studied different levels of coarse-graining for water ranging for one to 4 molecules per bead using different Mie and Morse potentials. Works also assessed the use of Soft-core potentials to study aqueous solutions of surfactants (SHINODA; DEVANE; KLEIN, 2007), ionic liquids (BHARGAVA; KLEIN, 2009), lipids (SHINODA; DEVANE; KLEIN, 2010) and membranes (PANTANO; KLEIN, 2009). Another CG force field for water based on the Mie Potential is the SAFT-γ Mie (LOBANOVA *et al.*, 2015). In this strategy, there are two different models: the CGW1-vle and the CGW1-ift. Both of them represent the water molecule as one bead and the Mie Potential has a repulsive and attractive parameter equal to eight and six, respectively. The CGW1-vle model was parameterized using saturated-liquid density and vapor pressure data, and should be used for simulations of aqueous systems' fluid-phase equilibrium at high temperatures and pressures. This model still suffers from premature freezing with a triple point at 343 K. The other model, CGW1-ift, was parameterized using saturated-liquid density and vapor-liquid interfacial tension, hence it is best suited for interfacial properties calculations. Both models have temperature-dependent size and energy parameters and performed well for these

properties over the entire liquid temperature range. The SAFT-γ Mie force field have also been applied to other compounds with satisfactory results. Müller and Mejía (2017) parameterized the force field for aromatic compounds and tested it with simulations of fluid phase equilibrium. Herdes, Totton and Müller (2015) carried out simulations of alkanes and light gases. Binary and ternary mixtures of water, carbon-dioxide and water (LOBANOVA *et al.*, 2016), thermodynamic and transport properties of carbon dioxide and methane (AIMOLI; MAGINN; ABREU, 2014a; AIMOLI; MAGINN; ABREU, 2014b) and water/oil interfacial tension (HERDES *et al.*, 2017) were also studied with this force field.

2.3   Solvation Free Energies

Solvation free energy calculations with molecular dynamics have a variety of applications due the amount of information it can provide about the behavior of the solvent in different chemical environments and the influence of the solute's molecular geometry. Free energy calculations have an inherent complexity that attracted studies in order to improve free energy simulations and post processing methods (SHIRTS; CHODERA, 2008; PALIWAL; SHIRTS, 2011; SHIRTS; PANDE, 2005; YTREBERG; SWENDSEN; ZUCKERMAN, 2006) in the last decades.

Recent work (MOBLEY; GUTHRIE, 2014; MATOS *et al.*, 2017) made available a big database of hydration free energy of small molecules using the GAFF force field. Beckstein, Fourrier and Iorga (2014) also calculated the hydration free energies for 52 compounds with OPLS-AA force field. They obtained an overall root mean square deviation of the prediction from the experimental data of 1.75 kcal/mol and concluded that the precision of the results are mainly limited by the reproducibility of the Lennard-Jones contribution towards the solvation free energy. A comparison of polar and nonpolar contributions to these hydration free energy indicated the significance of each the terms (IZAIRI; KAMBERAJ, 2017). Garrido *et al.* (2009), Garrido *et al.* (2011) calculated the free energy of solvation of large alkanes in 1-octanol and water with three different force fields (TraPPE, Gromos, OPLS-AA/TraPPE) and the solvation free

energy of propane and benzene in non aqueous solvents like n-hexadecane, n-hexane, ethyl benzene and acetone with the force fields TraPPE-UA and TraPPE-AA. Roy, Blinov and Kovalenko (2017) addressed the choice of the Lennard Jones parameters for predicting solvation free energy in 1-octanol. Gonçalves and Stassen (2005) calculated the free energy of solvation in the solvents tetrachloride, chloroform and benzene with GROMOS force field. Using the GAFF and the polarizable AMOEBA force fields, Mohamed, Bradshaw and Essex (2016) evaluated the solvation free energy of small molecules in toluene, chloroform and acetonitrile, and obtained a mean unsigned error of 1.22 kcal/mol for AMOEBA and 0.66 kcal/mol for GAFF. In order to define the role of solvent water in the docking structure determination, Matubayasi (2017) developped a method to compute the solvation free energy of proteins while using OPLS-AA force field for the solutes and TIP3P for water.

Though these variety of data using the intramolecular Lennard-Jones potential, we are not aware of works using the Mie Potential(MIE, 1903) in free energy calculations. We, at this study, try to provide information about theses calculations with the SAFT-γ Mie coarse grained force field. The output of these calculations are highly dependent on the force field and deficiencies in the description of small molecules by the models can be revealed with these calculations (MOBLEY et al., 2007; SHIRTS et al., 2003). Other important reason for testing a coarse grained force field is that they can generally reproduce free energy difference since the effects of the degrees of freedom reduction in the entropy are counterbalanced by reduced enthalpic terms (KMIECIK et al., 2016). Hence, knowing if more coarse grained approaches have a similar performance to the all atoms force fields can help increase the scale of solvation free energy calculations.

2.4   Solvation Free Energies Calculation Methods

The data from molecular dynamics simulations provide potential energies that need to be post processed and analyzed in order to calculate the solvation free energies. Since these calculations can have slow convergences, a lot of papers in the las decades focused in developing methods to improve sampling and analysis techniques. Almost all

methods rely o three different approaches: free energy of perturbation (FEP), histogram and thermodynamic integration methods.

2.4.1 Thermodynamic integration

The thermodynamic integration method (KIRKWOOD, 1935) uses equilibrium averages to evaluate the potential energy derivative with respect to the coupling parameter. Then, the free energies are calculated by doing the derivative ($\frac{\partial G}{\partial \lambda}$):

$$\frac{\partial \beta G}{\partial \lambda} = -\frac{1}{Z(\lambda)}\frac{\partial Z}{\partial \lambda} \tag{2.1}$$

The Eq. (3.67) can be rewritten according to the Halmitonian of the system ($\mathcal{H}$):

$$\Delta G = -\kappa_b T ln\left(\frac{Z_1}{Z_0}\right) = -\kappa_b T ln \int \frac{exp(-\beta\mathcal{H}_1(r,p))drdp}{exp(-\beta\mathcal{H}_0(r,p))drdp} \tag{2.2}$$

Substituting Eq. (2.2) in (2.2) and adding the coupling parameter to the Halmitonian ($\mathcal{H}(r,p,\lambda)$) in order to describe the transition between the end states:

$$\frac{\partial \beta G}{\partial \lambda} = \int \frac{\frac{\partial \mathcal{H}(r,p,\lambda)}{\partial \lambda}exp(-\beta\mathcal{H}(r,p,\lambda))drdp}{exp(-\beta\mathcal{H}(r,p,\lambda))drdp} = \left\langle \frac{\partial \mathcal{H}(r,p,\lambda)}{\partial \lambda} \right\rangle \tag{2.3}$$

The integral in Eq. (2.3) is obtained by interpolating the output data between the states from simulations in different ways. Some examples of methods for the interpolations are the trapezoidal rule or natural cubic spline (PALIWAL; SHIRTS, 2011). There are also more complex schemes that are usually system specific as the works of Jorge *et al.* (2010) and Shyu and Ytreberg (2010). MD simulations at each λ_k and the average over of the derivative at each state is compute in order to obtain the final solvation free energy:

$$\Delta G \approx \sum_k \left\langle \frac{\partial \mathcal{H}(r,p,\lambda)}{\partial \lambda} \right\rangle \tag{2.4}$$

2.4.2 Histograms

Histograms are used to compute probability distributions. Usually every histogram bin is treated as number of visits to a specific state.The standard practice when using histograms is to use the weighted histogram analysis method (WHAM) developed by Ferrenberg and Swendsen (1989) and generalized by Kumar *et al.* (1992) (CHIPOT; POHORILLE, 2007) to put together different histograms by minimizing the statistical error in the computed density of states and entropy function. This method describes the total probability distribution as a weighted sum of probability distributions without bias obtained from biased simulations. This method was developed to avoid problems related to data loss, high uncertainties and the calculation of the constant of the free energy added to the system by the biased potential (ROUX, 1995). The probability distribution dependent on the potential energy and temperature ($\tilde{\varrho}_r^*(U,T)$) for the WHAM is:

$$\tilde{\varrho}_r^*(U,T) = \frac{\sum_i f_i(U) exp(-\beta U)}{\sum_i f_{tot,i} exp(\beta_i \tilde{A}_i - \beta_i U)} \tag{2.5}$$

$$exp(-\beta_i \tilde{A}_i) = \sum_U \tilde{\varrho}_r^*(U,T) \tag{2.6}$$

$$\tilde{\varrho}_r^*(U,T) = \frac{\tilde{\varrho}_r^*(U,T)}{\sum_U \tilde{\varrho}_r^*(U,T)} \tag{2.7}$$

where $\tilde{A}_i$ gives the free energy for run i, $f_i(U)$ is the number of counts of energy U for run i and $f_{tot,i}$ is the total number of counts in run i. The Eq. (2.5) and (2.6) are solved self consistently with the initial value for $\tilde{A}_i$ equal to zero. The final probability distribution is then given by Eq. (2.7), and is used to calculate various moments of the potential energy (CHIPOT; POHORILLE, 2007).

2.4.3   Free Energy of Perturbation (FEP)

The free energy of perturbation method (ZWANZIG, 1954) is the oldest and one of the most general purpose strategy to calculate free energy differences. In this method, the thermodynamics of two different systems(A and B) are related with the intention of evaluating differences in intermolecular potentials. This energy change from state A to state B is calculated by:

$$\Delta G_{AB} = -\frac{1}{\beta} ln \langle e^{-\beta(U_B - U_A)} \rangle_A \qquad (2.8)$$

According to the equation above, the free energy difference is calculated by doing an average over the potential energies of state A and B obtained during the simulation of state A. This method requires a great overlap between states(the state B needs to represent a small perturbation in state A) in order to obtain a rapid convergence of the free energy difference. To assure overlap, it is possible to carry out simulations in N intermediate states between A and B, so Eq. (2.8) becomes:

$$\Delta G_{AB} = -\frac{1}{\beta} ln \left(\frac{1}{N} \sum_{i=0}^{N+1} e^{-\beta(U_{i+1} - U_i)} \right)_i \qquad (2.9)$$

The way of calculating ΔG of Eq. (2.9) is called Exponential Averaging (EXP) (ZWANZIG, 1955; PALIWAL; SHIRTS, 2011). The direction of the transformation is also important in this method. If the direction is of decreasing entropy, the step is of insertion (ΔG_{AB}) and the method is called insertion exponential averaging (IEXP). The direction of increasing entropy is called a deletion step (ΔG_{BA}) and the method is labeled as deletion exponential averaging (DEXP). These directions can yield different values of free energy differences due to under sampling in the tail regions of the ΔG_{AB} distribution (KLIMOVICH; SHIRTS; MOBLEY, 2015; POHORILLE; JARZYNSKI; CHIPOT, 2010). These problems make the EXP methods not suited to calculate free energy differences when the system hasn't a sufficient overlap. For these cases, the Bennet Acceptance Ratio or the Multi-State Bennet Acceptance Ratio is more indicated.

2.4.3.1 Bennet Acceptance Ratio (BAR)

The BAR method (BENNETT, 1976) was developed with the intent of eliminating the bias in the free energy estimation. It uses the uncorrelated samples of the potential energy in both directions ($A \to B$ and $B \to A$) to obtain the free energy differences using the information in a statically optimal way. The free energy difference between two intermediate states (i and j) is calculated by the self-consistent solution of the following equations:

$$\Delta G_{ij} = \frac{1}{\beta} ln \left(\frac{\sum_{k=1}^{N_j} \frac{1}{1 + \exp[-\beta(\Delta U_k^j - C)]}}{\sum_{l=1}^{N_i} \frac{1}{1 + \exp[-\beta(\Delta U_l^i - C)]}} \right) + C - \frac{1}{\beta} ln \left(\frac{N_j}{N_i} \right) \qquad (2.10)$$

$$C = \Delta G_{ij} + \frac{1}{\beta} ln \left(\frac{N_j}{N_i} \right) \qquad (2.11)$$

The total free energy difference between end states is then given by the sum over differences of consecutive intermediate states. This method also provides a function to obtain the minimum variance for free energy differences. The variance equation for any value of C is given by:

$$s_{ij}^2 = \frac{1}{\beta^2 N_i} \left[\frac{\langle f^2(x) \rangle_i}{\langle f(x) \rangle_i^2} - 1 \right] + \frac{1}{\beta^2 N_j} \left[\frac{\langle f^2(x) \rangle_j}{\langle f(x) \rangle_j^2} - 1 \right] \qquad (2.12)$$

where $f(x) = 1/(1+x)$ is the Fermi equation and $x = \exp[\beta(\Delta U - C)]$. The variance of the free energy difference between end states can be calculated by assuming independent errors and summing over the variance of consecutive intermediate states. However, this assumption is not correct and there is no general formula to obtain a statistically unbiased estimate of an entire transformation with the BAR method (PALIWAL; SHIRTS, 2011).

There are two other methods related to the BAR method that don't solve Eqs. (2.10) and (2.11) self consistently. By doing that, free energy differences will not have

minimum variance, but significant space and disk memory can be saved since the averages of Eqs. (2.10) - (2.12) are accumulated.The two methods are the Unoptimized Bennett Acceptance Ratio (UBAR) and the Range-Based Bennett Acceptance Ratio (RBAR). The first one avoids the self consistently resolution of the BAR equations by defining $C = \beta^{-1} ln(N_j/N_i)$. The UBAR method also requires that the intermediate free energy differences are approximately equal to zero to obtain optimal estimations. Meanwhile, the RBAR method selects a range of initial guesses of the constant C in order to calculate a range of ΔG_{ij}. The value of free energy difference correspondent to the minimum variance is then used as input in Eq. (2.11) to calculate the value of C. Hence, this method requires a good estimation of the initial range for the values of C. In terms of accuracy, the UBAR method can be as accurate as the BAR method, but it may end up being as computational costly (PALIWAL; SHIRTS, 2011).

2.4.3.2 Multistate Bennet Acceptance Ratio (MBAR)

The MBAR method (SHIRTS; CHODERA, 2008) is a further development of the BAR method. It proposes an estimator that computes free energies and their uncertainties of all K states by minimizing the KxK matrix of variances simultaneously. The estimator solves the following equation for each G_i self consistently:

$$G_i = \frac{1}{\beta} ln \sum_{k=1}^{K} \sum_{n=1}^{N_k} \frac{\exp[-\beta U_i(x_{kn})]}{\sum_{l=1}^{K} N_l \exp[\beta(G_l - U_l(x_{kn}))]} \tag{2.13}$$

The equation above requires the evaluation of the potential energy of uncorrelated configuration n for all K states ($U_i(x_{kn})$) and for all uncorrelated configuration snapshots (N_k) from state k. Free energy changes between states are given then by $\Delta G_{ij} = G_j - G_i$. The statistical variance of $S_{ij}^2 \Delta G_{ij}$ is given by the matrix covariances:

$$s_{ij}^2 \Delta G_{ij} \equiv cov(-ln\hat{Z}_j/\hat{Z}_i, -ln\hat{Z}_j/\hat{Z}_i) \tag{2.14}$$

where $\hat{Z}_j$ and $\hat{Z}_i$) are the partition functions of states i and j.

3 Fundamentals of the Computational Methods

3.1 Molecular Dynamics

3.1.1 Background and Formalution

Molecular Dynamics (MD) is a kind of molecular simulation that uses molecular configurations (Cartesian coordinates and momentum) to extract structural , thermodynamical and dynamical information of a system. This information is extracted from the time evolution of a syequationsstem, which are obtained through the Newton motion numerical integration (TUCKERMAN, 2010):

$$\frac{d\vec{p}_i}{dt} = -\frac{\partial U(\vec{r}_N)}{\partial \vec{r}_i} \tag{3.1}$$

where p_i is the momentum and r^N are the coordinates of all the atoms $(x_1, y_1, z_1, ..., x_N, y_N, z_N)$. Alternatively, we can write the equation relative to the velocity (v_i):

$$m_i \frac{\vec{v}_i}{dt} = -\frac{\partial U(\vec{r}_N)}{\partial \vec{r}_i} \tag{3.2}$$

In order to develop equations for any coordinate system ($q^N = (r_1, \theta_1, \phi_1)$), the Halmitonian formulation, a more generalized formulation of classical mechanics, is used to develop the equations of motion:

$$\mathcal{H}(q^N, p^N) = K(p^N) + U(q^N) \tag{3.3}$$

In the equation above, K is the kinetic energy and U is the potential energy. The

equations of motion can now be represented by:

$$\frac{d\vec{p}_i}{dt} = -\frac{\partial \mathcal{H}}{\partial \vec{q}_i} \tag{3.4}$$

$$\frac{d\vec{q}_i}{dt} = -\frac{\partial \mathcal{H}}{\partial \vec{q}_i} \tag{3.5}$$

The coordinates and momentum axes fo each atoms in a 6N dimensional space is defined as the phase space. The trajectory through the phase space is then the time evolution of a system in a molecular dynamics simulation. This evolution of the simulation may be then used to calculate the thermodynamic properties if the system is ergodic. That is, a trajectory in this system's will explore with the same probability all regions of the phase space of microstates with the same energy.

3.1.2 Statistical Ensembles

A statistical ensemble corresponds to the set of configurations of a system with control variables defined. For a isolated system at equilibrium, the ensemble will correspond to the microcanonical ensemble. The microcanonical ensemble have constant values of number of particles (N), volume (V) and total energy (E). Following the ergodic hypothesis, the system at these conditions will then spend the same amount of time in each of the microstates (points in phase space) with the fixed E. The number of accessible microstates is defined by the partition function or the density of states, and, for the microcanonical ensemble, is given by:

$$\Omega(N, V, E) = \frac{1}{h^{3N}N!} \int dp^N dr^N \delta[\mathcal{H})(p^N, r^N) - E] \tag{3.6}$$

where h is the Planck constant and the delta functions is a Dirac delta function. As mentioned above, the system has equal probabilities an these can be written as:

$$\varrho(p^N, r^N) = \frac{[\mathcal{H})(p^N, r^N) - E]}{\Omega(N, V, E)} \tag{3.7}$$

The relation of the microcanonical partition function to the macroscopic entropy is:

$$S = k_b ln\Omega(N, V, E) \tag{3.8}$$

where k_b is the Boltzmann constant. Other properties can be derived form the equation using the fundamental equation or its energy version:

$$dS = \frac{1}{T}dE + \frac{P}{T}dV + \frac{\mu}{T}dN \tag{3.9}$$

$$dE = TdS + PdV + \mu dN \tag{3.10}$$

As said above, the microcanonical ensemble have N,V,E as its control variable. Other ensembles can also be defined according to the macroscopic properties held constant. In the canonical ensemble, N, V and the temperature (T) are held constant and N, pressure (P) and T are held constant in the isothermal-isobaric ensemble. Other ensembles include the isenthalpic-isobaric (constant number of particles, pressure and enthalpy) and the grand canonical (constant chemical potential, volume and temperature). A variety of physical properties are measure at the conditions of the isenthalpic-isobaric ensemble such as enthalpies, entropies, redox potential, equilibrium constants and free energies, what makes this ensemble one of the most important (TUCKERMAN, 2010). This is also the ensemble in which solvation free energy simulations are carried at this work, hence we going to briefly talk about it. This ensemble is obtained from a Legendre transformation on the canonical ensemble. The Helmholtz free energy $A(N, V, T)$ is becomes the Gibbs free energy $G(N, P, T)$ by transforming the volume into the external pressure:

$$G(N, P, T) = A(N, V(P), T) + PV(P) \tag{3.11}$$

The Gibbs free energy is related to its the partition function $\Delta(N, P, T)$ (the

number of accessible microstates) by:

$$G(N, P, T) = -\kappa_b T ln \Delta(N, P, T) \qquad (3.12)$$

$$\Delta(N, P, T) = \frac{1}{V_0} \int_0^\infty dV \int d^N p d^N r \exp\left[-\beta\left(\sum_{i=1}^N \frac{p_i^2}{2m_i} + U(r^N) + PV(r^N)\right)\right] \quad (3.13)$$

where $Q(N, V, T)$ is the partition function of the canonical ensemble. From this relations and a differential change in G, we can obtain the chemical potential (μ), volume and entropy (S) relations for this ensemble:

$$\mu = (\frac{\partial G}{\partial N})_{P,T} = -\kappa_b T(\frac{\partial ln\Delta(N, P, T)}{\partial N})_{N,T} \qquad (3.14)$$

$$\langle V \rangle = (\frac{\partial G}{\partial P})_{N,T} = \kappa_b T(\frac{\partial ln\Delta(N, P, T)}{\partial N})_{N,P} \qquad (3.15)$$

$$S = (\frac{\partial G}{\partial T})_{N,P} = \kappa_b(lnQ(N, V, T) + T(\frac{\partial lnQ(N, V, T)}{\partial T})_{V,N}) \qquad (3.16)$$

The isothermal-isobaric ensembles have external conditional being applied to it (temperature and pressure, respectively). For temperature control, the method employed is to mimic the effect of a thermal reservoir through a thermostat. The most widely used thermostat are the Berendsen and Nosé-Hoover (BERENDSEN *et al.*, 1984; HOOVER, 1985). Meanwhile, the pressure is controlled with a barostat, that allows volume fluctuation. Among the more used barostats are the Berendsen, Hoover, Martina-Tobias-Klein and Parrinnello-Rahman (BERENDSEN *et al.*, 1984; HOOVER, 1985; MARTYNA; TOBIAS; KLEIN, 1994; PARRINELLO; RAHMAN, 1981). It is possible to impose movements restrictions on the molecule or even consider the molecule as a rigid body in order to increase the integration step or to follow the modeling of the molecules. Methods for this include the SHAKE, RATTLE and Kamberaj algorithms (RYCKAERT;

CICCOTTI; BERENDSEN, 1977; ANDERSEN, 1983; KAMBERAJ; LOW; NEAL, 2005).

3.1.3 Integration of the motion's equations

With the formalism defined for the equations of motions an for the statistical ensemble, we can then derive discrete-time numerical approximations for them. The basic idea is to solve the trajectory of atoms as function of time ($r^N(t)$) by updating the positions in discrete time intervals or time steps. One of the possible algorithms is The Verlet (VERLET, 1967). Its proposed equation for the coordinates update is:

$$r(t + \Delta t) = 2r(t) - r(t + \Delta t) + \frac{F(t)}{m}\Delta t^2 + \vartheta(\Delta t^4) \tag{3.17}$$

In this equation, the positions are updated by a time step of Δt by using the positions at the previous two time steps and the forces at Δt. The forces (F) acting on every particle are calculated by doing directional derivatives of the intermolecular potential $U(r)$, generally described by a force field, in relation to the atoms' positions:

$$\vec{F}(r)_i = -\nabla U(r) \tag{3.18}$$

The velocities of this method can be approximated by:

$$v(t + \Delta t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} + \vartheta(\Delta t^3) \tag{3.19}$$

The Verlet method doesn't use the velocity directly to update the positions. Hence, the Velocity Verlet algorithm was proposed to serve as a reformulation of the Verlet algorithm that uses the velocity:

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{F(t)}{2m}\Delta t^2 \tag{3.20}$$

The velocity in the equation above is updated following:

$$v(t + \Delta t) = v(t) + \frac{F(t + \Delta t) + F(t)}{2m} \Delta t \qquad (3.21)$$

Instead of using a time step of Δt, the velocities can be updated at $\Delta t/2$. This is the strategy proposed by the leap frog algorithm:

$$v(t + \Delta t/2) = v(t - \Delta t/2) + \frac{F(t + \Delta t) + F(t)}{m} \Delta t \qquad (3.22)$$

$$r(t + \Delta t) = r(t) + v(t + \Delta t/2)\Delta t \qquad (3.23)$$

3.1.4 Initial Configuration

In order to use the equations above, every MD simulation need a non overlapping initial positions and velocities for every atom in the system. The initial velocities follow a temperature dependent Maxwell-Boltzmann distribution:

$$\varrho(v_{x,i}) = \frac{m_i}{2\pi k_b T}^{\frac{1}{2}} exp(-\frac{m_i v_{x,i}^2}{2\pi k_b T}) \qquad (3.24)$$

Random velocities are then for each of the 3N components of the velocity. Meanwhile, the initial positions can be obtained by several approaches. The initial configuration can be takes from x ray or NMR spectroscopy, the atoms cab be placed randomly in the simulation volume or the atoms in molecules can be placed in idealized or approximate geometries. The generally used method to acquire the configurations places the molecules on a cubic lattice. Independent of the method used, certain restrictions should be applied in the initial configurations in order to carry out molecular simulations. As an example, the cubic lattice has a finite size, but a finite size box would result in simulations dominated by surface effects. To avoid that, we can create a box periodically repeated in all directions by applying the so called Periodic Boundary Conditions (PBC). The periodic box has a primitive cell, which contains the system N particles, replicated
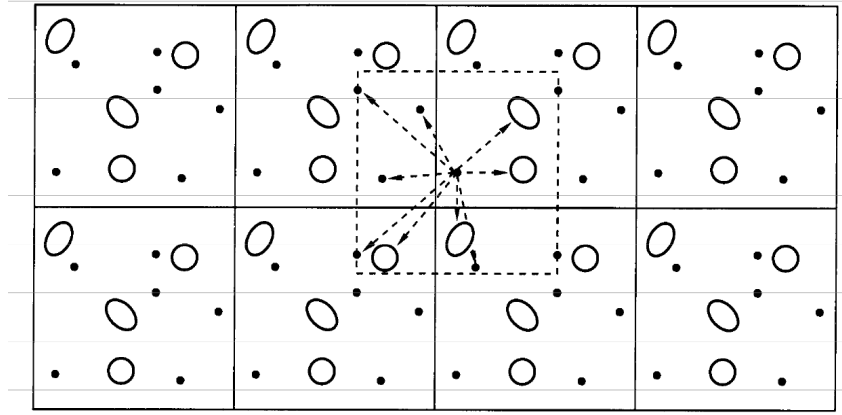
Figure 3.1.1 – Periodic boundary condition representation. Taken from Frenkel and Smit (2001)

in periodic lattice of infinite cells as represented in Figure 3.1.1(FRENKEL; SMIT, 2001).

3.1.5 Force Fields

Force fields are models for the description of structural characteristics such as van der Waals interactions, bond lengths, bond angles and torsion. The description is done by approximating the potential energy function ($U(r^N)$). $U(r^N)$ has contributions due to intermolecular and intramolecular interactions. The intramolecular interactions include bond stretching, bond angle bending and bond torsion. The contribution for the bond stretching is usually given by the following Taylor expansion around the energy minimum:

$$u_{bs}(d) = k_{bs}(d - d_0)^2 \tag{3.25}$$

Here, d is the bond length, d_0 is the equilibrium bond length and k_{bs} is bond stretching constant. The bond angle bending contribution corresponds to deviations from the preferred geometry, and is given by:

$$u_{ab}(\theta) = k_{ab}(\theta - \theta_0)^2 \tag{3.26}$$

where k_{ab} and θ_0 are constants defined by the force field and θ is the bond angle between

three atoms. The bond torsion interactions corresponds to the energies of rotations along bonds and it happens among four atoms:

$$u_{bt}(\omega) = \sum_{n=0}^{N} c_n cos(\omega)^n \tag{3.27}$$

where N is the number of terms, c_n is the summation index defined by the force field and ω is the torsional angle defined by the force field.

The intermolecular interactions include electrostatics interactions, van der Waal interactions and volume repulsions. The electrostatics interactions of two atoms i and j with partial charges follow the Coulomb's Law:

$$u_q(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{3.28}$$

In the equation above, q_i and q_j are the partial charges and ϵ_0 is the free space permittivity constant. In many force fields, the van der Waal interactions and volume repulsions between particles i and j are modeled in the same equation by the Lennard Jones Potential:

$$u_{LJ}(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right] \tag{3.29}$$

where r is the distance the molecules, ϵ is the depth of the potential well, σ is the distance correspondent to a zero intermolecular potential. The graphical representation of the potential is:
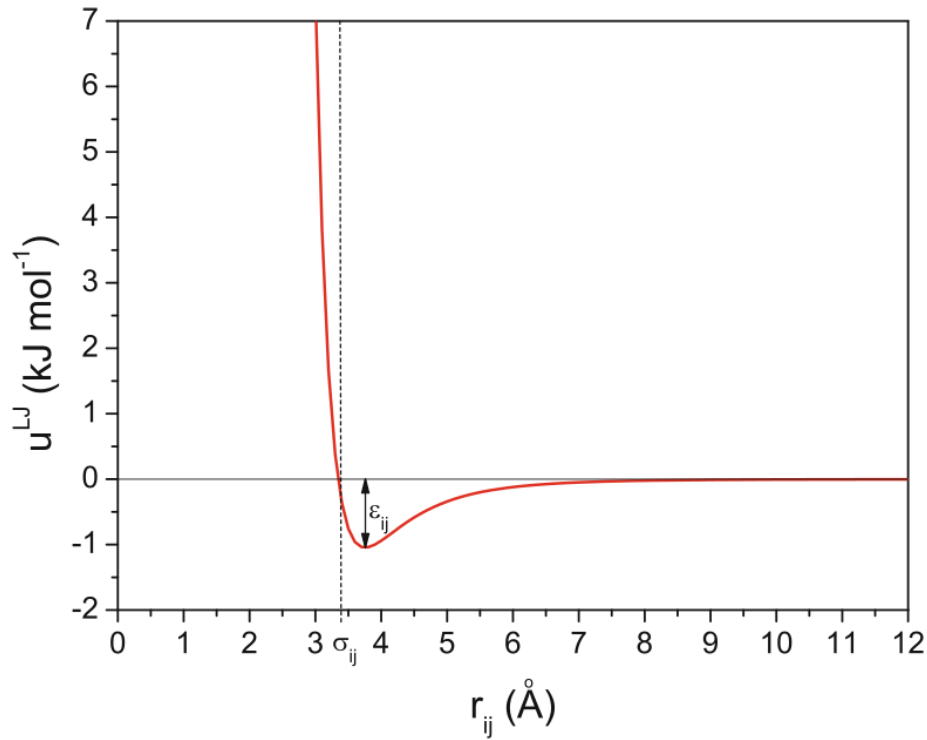
Figure 3.1.2 – Lennard Jones potential representation for two atoms of argon (i,j). Taken from Raabe (2017)

The interactions in Figure 3.1.2 tend to zero after a certain value of r, so this value can be set as the cutoff radius in which the potential energy is considered to be zero after it. Also, only interactions with te nearest periodic image of the cell is considered in order for short range interactions (Minimum image convention condition). With this conditions, the calculations of forces and velocities are viable.

The final potential energy function defined by the force field can then be expressed by summing all the interactions above:

$$U(r^N) = u_{bs}(d) + u_{ab}(\theta) + u_{bt}(\omega) + u_q(r_{ij}) + u_{LJ}(r_{ij}) \tag{3.30}$$

3.2   SAFT-γ Mie Force Field

3.2.1   SAFT-VR Mie EoS

The SAFT-VR Mie equation of state (LAFITTE *et al.*, 2013) is the basis for the SAFT-γ Mie coarse grained force field (AVENDAÑO *et al.*, 2011). This EoS was initially developed to describe chain molecule formed from fused Mie segments using the Mie attractive and repulsive potential. The Mie potential is a type of generalized Lennard-Jones potential that can be used to explicitly describe repulsive interactions of different hardness/softness and attractive interactions of different ranges, and is given by:

$$U_{Mie}(r) = \epsilon \frac{\lambda_r}{\lambda_r - \lambda_a} \left(\frac{\lambda_r}{\lambda_a}\right)^{\left(\frac{\lambda_a}{\lambda_r - \lambda_a}\right)} \left[\left(\frac{\sigma}{r}\right)^{\lambda_r} - \left(\frac{\sigma}{r}\right)^{\lambda_a}\right] \tag{3.31}$$

where λ_r is the repulsive exponent and λ_a is the attractive exponent. This equation uses the Barker and Henderson (1976) high perturbation expansion of the Helmholtz free energy up to third order and an improved expression for the radial distribution function (RDF) of Mie monomers at contact to obtain an equation able to give an accurate theoretical description of the vapor-liquid equilibrium and second derivative properties (LAFITTE *et al.*, 2013). For a non-associating fluid, the Helmholtz free energy is:

$$\frac{A}{N\kappa_b T} = a = a^{IDEAL} + a^{MONO} + a^{CHAIN} \tag{3.32}$$

3.2.1.1   Ideal Contribution

The ideal contribution for a mixture is given by:

$$a^{IDEAL} = \sum_{i=1}^{N_c} x_i \ln\left(\rho_i \Lambda_i^3\right) - 1 \tag{3.33}$$

where $x_i = N_i/N$ is the molar fraction of component i, $\rho_i = N_i/V$ is the number density, N_i is the number of molecules of each component and Λ_i^3 is the de Broglie wavelength.

3.2.1.2 Monomer Contribution

The monomer contribution describes interactions between Mie segments and can be expressed, for a mixture, as:

$$a^{MONO} = \left(\sum_{i=1}^{N_c} x_i m_{s,i} \right) a^M \tag{3.34}$$

In the equation above, $m_{s,i}$ is the number of spherical segments making up the molecule i and a^M is the monomer dimensionless Helmholtz free energy and it is expressed as a third order perturbation expansion in the inverse temperature (BARKER; HENDERSON, 1976):

$$a^M = a^{HS} + \beta a_1 + \beta a_2{}^2 + \beta a_3{}^3 \tag{3.35}$$

where $\beta = \kappa_b T$ and a^{HS} is the hard-sphere dimensionless Helmholtz free energy for a mixture :

$$a^{HS} = \frac{6}{\pi \rho_s} \left[\left(\frac{\zeta_2^3}{\zeta_3^2} - \zeta_0 \right) \ln(1 - \zeta_3) + \frac{3\zeta_1\zeta_2}{1 - \zeta_3} + \frac{\zeta_2^3}{\zeta_3(1 - \zeta_3)^2} \right] \tag{3.36}$$

The variable $\rho_s = \rho \sum_i^{N_c} x_i m s, i$ is the total number density of spherical segments and ζ_l are the moments of the number density:

$$\zeta_l = \frac{\pi \rho_s}{6} \left(\sum_{i=1}^{N_c} x_{s,i} d_{ii}^l \right), l = 0, 1, 2, 3 \tag{3.37}$$

where $x_{s,i}$ is the mole fraction of segments and is related through the mole fraction of component i (x_i) by:

$$x_{s,i} = \frac{m_{s,i} x_i}{\sum_{k=1}^{N_c} m_{s,k} x_k} \tag{3.38}$$

The effective hard-sphere diameter d_{ii} for the segments is:

$$d_{ii} = \int_0^{\sigma_{ii}} (1 - \exp(-\beta U_{ii}^{Mie}(r))) dr \tag{3.39}$$

The integral in Eq. (3.39) is normally obtained by means of Gauss-Legendre with a 5-point quadrature (PAPAIOANNOU et al., 2014). The detailing of terms of Eq. (3.34)

can be found in Lafitte *et al.* (2013).

3.2.1.3 Chain Contribution

The chain formation of m_s tangentially bonded Mie segments contribution is based on the first-order perturbation theory (TPT1) (PAPAIOANNOU *et al.*, 2014) and can be expressed as:

$$a^{CHAIN} = -\sum_{i=1}^{N_c} x_i(m_{s,i} - 1)\ln(g_{ii}^{Mie}(\sigma_{ii}))$$ (3.40)

The $g_{ij}^{Mie}(\sigma_{ij})$ term correspond to the radial distribution function (RDF) of the hypothetical Mie system evaluated at the effective diameter and can be obtained with the perturbation expansion:

$$g_{ij}^{Mie}(\sigma_{ij}) = g_{d,ij}^{HS}(\sigma_{ij})\exp[\beta\epsilon g_{1,ij}(\sigma_{ij})/g_{d,ij}^{HS}(\sigma_{ij}) + (\beta\epsilon)^2 g_{2,ij}(\sigma_{ij})/g_{d,ij}^{HS}(\sigma_{ij})]$$ (3.41)

The other terms in the equations above are explicitly exposed in the original article (LAFITTE *et al.*, 2013).

3.2.1.4 Ring Contribution

There are two forms for the Helmholtz free energy for rings formed from m_s tangentially bonded segments in the literature. The first one (LAFITTE *et al.*, 2012) considered that the difference between a chain and a ring molecule is that the latter has one more bond that is connecting the first segment to the last. With this assumption, Eq. (3.40) can be adapted to rings by:

$$a^{RING} = -\sum_{i=1}^{N_c} x_i m_{s,i}\ln(g_{ii}^{Mie}(\sigma_{ii}))$$ (3.42)

According to Lafitte *et al.* (2012), Eq. (3.42) needs an additional parameterization with molecular simulation data so the EoS can be used in molecular simulations, but this procedure is not necessary for chain molecules. Recently, Müller and Mejía (2017)

tried to correct this inconsistency by means of developing the ring free energy based on the work of Müller and Gubbins (1993), who obtained rigorous expressions for hard fluids with molecular geometries of rings of $m_s = 3$. The final expression developed for the ring dimensionless Helmholtz free energy is:

$$a^{RING} = -\sum_{i=1}^{N_c} x_i(m_{s,i} - 1 + \chi_i\eta_i)\ln(g_{ii}^{Mie}(\sigma_{ii})) \qquad (3.43)$$

$\eta_i = m_{s,i}\rho_i\sigma_{ii}^3/6$ is the packing fraction and χ_i is a parameter which depends on $m_{s,i}$ and on the geometry of the ring of each component i. For a value of $\chi = 0$, Eq. (3.43) is equal to Eq. (3.40). Meanwhile, the equation corresponds to a hard sphere system of triangles when $\chi = 1.3827$. Müller and Mejía (2017) also calculated values of ζ for $m_s = 3, m_s = 4, m_s = 5, m_s = 7$ with pseudo-experimental data from molecular dynamics (MD) for a defined pure fluid. The values of χ for each geometry estimated can be seen in Figure 3.2.1.

(a) $m_s = 3$; $\chi = 1.4938$

(b) $m_s = 4$; $\chi = 2.9833$

(c) $m_s = 5$; $\chi = 3.2222$

(d) $m_s = 5$; $\chi = 4.7188$

(e) $m_s = 7$; $\chi = 9.0958$

(f) $m_s = 7$; $\chi = 8.8640$
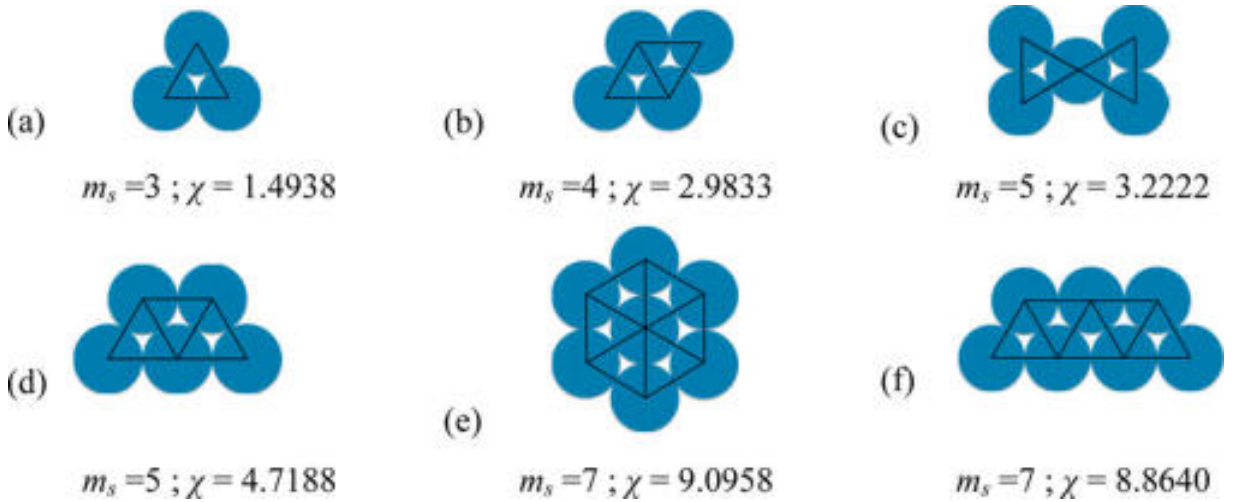
Figure 3.2.1 – Values for parameter χ according to the ring geometry (MÜLLER; MEJÍA, 2017)

3.2.1.5 Combining rules for the intermolecular potential parameters

Lafitte *et al.* (2013) also suggested mixing rules for the potential parameters based on Lorentz-Berthelot combining rules (ROWLINSON; SWINTON, 1982):

$$\sigma_{ij} = \frac{\sigma ii + \sigma jj}{2} \tag{3.44}$$

$$\lambda_{k,ij} - 3 = \sqrt{(\lambda_{k,ii} - 3)(\lambda_{k,jj} - 3)}, k = r, a \tag{3.45}$$

$$\epsilon_{ij} = (1 - k_{ij}) \frac{\sqrt{\sigma_{ii}^3 \sigma_{jj}^3}}{\sigma_{ij}^3} \sqrt{\epsilon_{ii} \epsilon_{jj}} \tag{3.46}$$

The k_{ij} is a binary interaction parameter to correct the deviations of the Lorentz-Berthelot rule for chemically distinct compounds. This parameter can be fitted to experimental data or pseudo experimental data.

3.2.2 Parameter Estimation for the SAFT-γ Mie Force Field

The SAFT-γ Mie Force Field uses a coarse graining top down methodology in its parameterization. This methodology aims to obtain the intermolecular parameters from macroscopic experimental data such as fluid-phase equilibrium or interfacial tension data. The idea is that the force field parameters estimated with the SAFT-VR Mie EoS can be used on molecular simulations since both the equation of state and the force field use the same explicit intermolecular potential model (Mie potential). This correspondence between models has already been seen for a variety of fluids in which this force field was parameterized. This success in the representation of the properties of real fluids can be imputed to the degrees of freedom of Mie Potential (HERDES; TOTTON; MÜLLER, 2015). Furthermore, this flexibility provides the exploration of a very large parameter space without using an iterative simulation scheme (AVENDAÑO *et al.*, 2011).

Each substance has initially five parameters to be estimated (m_s,σ,ϵ,λ_r and λ_a) according to Eq. (3.31). The number of segments are usually fixed in an integer value

since each segment represents one pseudo atom. The attractive parameter is generally fixed due to its high correlation with the repulsive parameter. Usually, the parameter is fixed in the London value of 6, which is a good representation of the dispersion scale of most simple fluids that don't have strong polar interactions (RAMRATTAN *et al.*, 2015; HERDES; TOTTON; MÜLLER, 2015). There are two strategies to obtain the parameters: one is by fitting the Saft-Vr Mie EoS to experimental data as vapor pressure and liquid density and the other one is using correspondent state parametrization. The first, generally, minimizes the following unweighted least-squares objective function:

$$
\min_{\sigma,\epsilon,\lambda_r} F_{obj}(\sigma,\epsilon,\lambda_r) = \sum_{i=1}^{N_p} \left(\frac{P_v^{SAFT}(T_i,\sigma,\epsilon,\lambda_r) - P_v^{exp}(T_i)}{P_v^{exp}(T_i)} \right)^2 + \\
\sum_{i=1}^{N_p} \left(\frac{\rho_l^{SAFT}(T_i,\sigma,\epsilon,\lambda_r) - \rho_l^{exp}(T_i)}{\rho_l^{exp}(T_i)} \right)^2
\tag{3.47}
$$

where N_p is the number of experimental points, P_v is the vapor pressure and ρ_l is the saturated liquid density. Another properties that can be used in the estimation are superficial tension and speed of sound. The multiple parameters of the model make it necessary the use of a wide range of experimental data since multiple solutions may be found for the fit. Therefore, one needs to be careful in deciding the level of coarse graining (i.e. the parameter m_s) and the subsequent parameter space that will not result in some physical inconsistencies such as a premature freezing fluid.

Lafitte *et al.* (2012) suggested that two corrections factors (c_σ and c_ϵ) should be estimated with simulation data when using Eq. (3.42) for the ring contribution. They are related to the EoS parameters by scaled parameters:

$$
\sigma^{scaled} = c_\sigma \sigma^{SAFT}
\tag{3.48}
$$

$$
\epsilon^{scaled} = c_\epsilon \epsilon^{SAFT}
\tag{3.49}
$$

According to Lafitte *et al.* (2012), these corrections are necessary because the

approximations employed in the EoS theory generate discrepancies between molecular simulations and the EoS for ring molecules modeled with Eq. (3.42). The objective function for this estimation is given by:

$$\min_{c_\sigma, c_\epsilon} F_{obj}(c_\sigma, c_\epsilon) = \sum_{i=1}^{N_p} \left(\frac{P_v^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT}) - P_v^{SAFT}(T_i, \sigma^{scaled}, \epsilon^{scaled})}{P_v^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT})} \right)^2 +$$
$$\sum_{i=1}^{N_p} \left(\frac{\rho_{liq}^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT}) - \rho_{liq}^{SAFT}(T_i, \sigma^{scaled}, \epsilon^{scaled})}{\rho_{liq}^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT})} \right)^2 \quad (3.50)$$

The repulsive parameter is maintained in the value found on the minimization of Eq. (3.47). The refined values for σ and ϵ are:

$$\sigma^{sim} = \sigma^{SAFT}/c_\sigma \quad (3.51)$$

$$\epsilon^{sim} = \epsilon^{SAFT}/c_\epsilon \quad (3.52)$$

It is interesting to point out that this new parameterization is not necessary when using Eq. (3.43) as the ring contribution. The other method to obtain the force field parameters is the correspondent state parametrization (MEJÍA; HERDES; MÜLLER, 2014). This method considers that the unweighted volume average of the attractive contribution to the Mie intermolecular potential,a_1, is a mean field approximation:

$$a_1 = 2\pi \rho \sigma^3 \epsilon \alpha \quad (3.53)$$

The van der Waals constant, α, considering $\lambda_a = 6$ is related by the Mie exponents by:

$$\alpha = \frac{1}{\epsilon \sigma^3} \int_\sigma^\infty \phi(r) r^2 dr = \frac{\lambda_r}{3(\lambda_r - 3)} \left(\frac{\lambda_r}{6} \right)^{6/(\lambda_r - 6)} \quad (3.54)$$

The parameterization in this method starts by using the experimental acentric factor, ω, for each molecule with a fixed value of m_s to obtain the value of the repulsive exponent with the following Padé series:

$$\lambda_r = \frac{\sum_{i=0} a_i \omega^i}{1 + \sum_{i=1} b_i \omega^i} \qquad (3.55)$$

a_i and b_i are dependent parameters of the number of segments and a table with their values is presented in the original paper (MEJÍA; HERDES; MÜLLER, 2014). The van der Waals constant can be found substituting λ_r into Eq. (3.54). The reduced critical potential T_c^* is related to α by a Padé series:

$$T_c^* = \frac{\sum_{i=0} c_i \alpha^i}{1 + \sum_{i=1} d_i \alpha^i} \qquad (3.56)$$

The values of c_i and d_i are also available in the original paper. The reduced temperature of the equation above is used in conjunction with the experimental critical temperature, T_c, to find the energy parameter with the relation below:

$$T_c^* = \frac{\kappa_b T_c}{\epsilon} \qquad (3.57)$$

The diameter parameter, however, is not obtained with the critical properties, but with the reduced liquid density,$\rho_{T_r=0.7}$, at the reduced temperature ,T_r, of 0.7. This density is also obtained with a Padé series using parameters by Mejía, Herdes and Müller (2014):

$$\rho_{T_r=0.7}^* = \frac{\sum_{i=0} j_i \alpha^i}{1 + \sum_{i=1} k_i \alpha^i} \qquad (3.58)$$

The relation among the equation above, σ and the experimental density is given by:

$$\rho_{T_r=0.7}^* = \rho_{T_r=0.7} \sigma^3 N_{av} \qquad (3.59)$$

where N_{av} is The Avogadro number. This correspondent state method has the advantage of only requiring critical data, which is available for a great range of fluids, and liquid density data. The parameters found with this strategy are available at an online database (ERVIK; MEJÍA; MÜLLER, 2016).

The binary interaction parameter k_{ij} of Eq. (3.46) is necessary to adjust the mixture behavior of chemically distinct components. Normally, it is estimated minimizing the difference between experimental binary vapor liquid equilibrium or interfacial tension data and the SAFT-VR Mie EoS output data (MÜLLER; MEJÍA, 2017; LOBANOVA *et al.*, 2016). The objective function is similar to:

$$
\min_{k_{ij}} F_{obj}(k_{ij}) = \sum_{k=1}^{N_p} \left(\frac{P_v^{SAFT}(T_k, x, k_{ij}) - P_v^{exp}(T_k, x)}{P_v^{exp}(T_k, x)} \right)^2 +
$$
$$
\sum_{k=1}^{N_p} \left(\frac{\rho_l^{SAFT}(T_k, x, k_{ij}) - \rho_l^{exp}(T_i)}{\rho_l^{exp}(T_i)} \right)^2
$$

(3.60)

However, Ervik *et al.* (2016) used molecular simulation results to fit the parameter to the superficial tension data. The strategy followed by them was to do simulations in three values of k_{ij} first and, after, they refined the parameter until a value in good agreement with the experimental data was found.

3.3   Solvation Free Energy Simulations Based on Molecular Dynamics

Free energies can be expressed as averages over ensembles of atomic configurations generated using Monte Carlo or molecular dynamics techniques. In the canonical ensemble, the free energy is given by:

$$
F(N, V, T) = -\kappa_b T ln Q(N, V, T)
$$

(3.61)

where $Q(N, V, T))$ is the partition function of the canonical ensemble:

$$Q(N, V, T) = \int d^n p d^n r \exp\left[-\beta\left(\sum_{i=1}^{N} \frac{p_i^2}{2m_i} + U(r_1, .., r_n)\right)\right] \qquad (3.62)$$

where $\beta = 1/k_b T$. Meanwhile, the average over the isothermal-isobaric ensemble gives the Gibbs free energy:

$$G(N, P, T) = -\kappa_b T ln \Delta(N, P, T) \qquad (3.63)$$

where $\Delta(N, P, T)$ is the partition function of the isothermal-isobaric ensemble:

$$\Delta(N, P, T) = \frac{1}{V_0}\int_0^\infty dV \int d^n p d^n r \exp\left[-\beta\left(\sum_{i=1}^{N} \frac{p_i^2}{2m_i} + U(r_1, .., r_n) + PV(r_1, .., r_n)\right)\right] \qquad (3.64)$$

Evaluating the partition function is a difficult task, but we are interested in calculating the Gibbs free energy difference between two states of a system, that is given by:

$$\Delta G_{AB} = G_B - G_A = -\kappa_b T ln\left(\frac{\Delta_B}{\Delta_A}\right) \qquad (3.65)$$

Since the masses of particles in systems 0 and 1 are the same, the moment integrals in the ratio Δ_B/Δ_A can be simplified into to the ratio of the configuration integrals:

$$\frac{Z_B}{Z_A} = \frac{\int_0^\infty dV \int d^n r \exp\left[-\beta\left(U(r_1, .., r_n) + PV(r_1, .., r_n)\right)_B\right]}{\int_0^\infty dV \int d^n r \exp\left[-\beta\left(U(r_1, .., r_n) + PV(r_1, .., r_n)\right)_A\right]} \qquad (3.66)$$

What results in the following equation for the Gibbs free energy difference:

$$\Delta G_{AB} = G_B - G_A = -\kappa_b T ln\left(\frac{Z_B}{Z_A}\right) \qquad (3.67)$$

The Gibbs free energy difference between end states A and B are, more specifically, the difference between the solute alone in the gas phase and the solute interacting with the solvent. In order to these differences be accurate, the states' phase integral must have sufficient overlap (KLIMOVICH; SHIRTS; MOBLEY, 2015). This can be achieved by calculating the free energy difference between a series of intermediates states. The result of these differences are independent of the path chosen since free energy is a state function. That's why alchemical states (no physical sense) are used to link physical states of interest. The solvation free energy calculations are done through a thermodynamic cycle to gradually insert the solute molecule into the solvent as illustrated in Figure 3.3.1. According to this cycle, the free energy of solvation is expressed as:

$$\Delta G_{solv} = \Delta G_{1\to4} = \Delta G_{1\to2} + \Delta G_{2\to3} + \Delta G_{3\to4} \qquad (3.68)$$
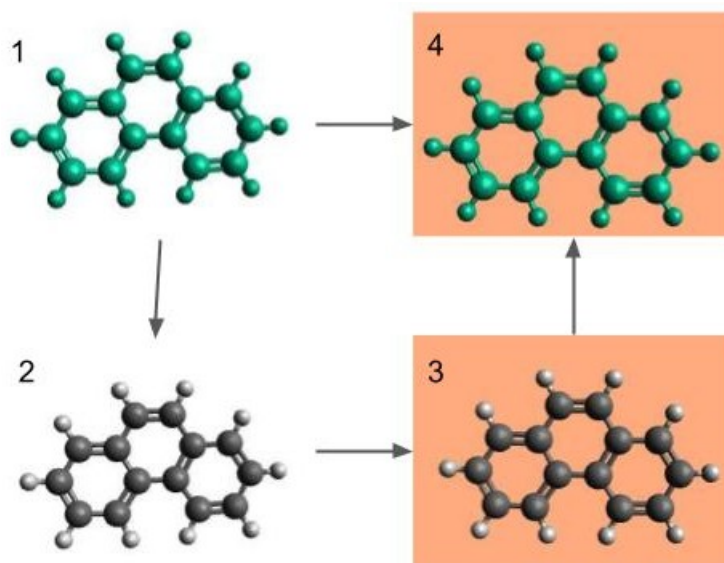
Figure 3.3.1 – Thermodynamic cycle for solvation free energy calculations with molecular dynamics (Adapted from Klimovich, Shirts and Mobley (2015))

The solvation free energy between states 1 and 2 in Figure 3.3.1 is associated with turning off the molecule's non bonded interactions in the gas phase. The following transformation, $\Delta G_{2\to3}$, is the free energy of moving the non-interacting molecule in the gas phase to the solvent, and is equal to zero since the transformation of a non-interacting molecule doesn't depend on the environment. Lastly, $\Delta G_{3\to4}$ is the free

energy required to the non-interaction molecule in the aqueous phase regain its non-bonded interactions. The solvation free energy calculation can be classified according to the types of non bonded interactions that are turned off in the $1 \to 2$ and $3 \to 4$ parts of the cycle. If both non-bonded interactions with the environment and internal interactions are turned off, this is an annihilation free energy calculation. Meanwhile, if only non-bonded interactions with the environment are turned off,this is a decoupling free energy calculation. In the later case, $\Delta G_{1 \to 2} = 0$ and the $\Delta G_{solv} = \Delta G_{3 \to 4}$. The methods used to carry out these transformations scale the solute charges to zero and then turn off the interactions corresponding to the Lennard Jones potential. In order to carry out the later process, a modified potential with a coupling parameter(λ) is used. Each λ represent an alchemical state and, when $\lambda = 0$, there is no interaction with the solvent and, when $\lambda = 1$, interactions are fully activated. The coupling of the λ parameter could be linear, but it could generate numerical problems related to the exponential part of the potential. That's why the non-linear soft-core scheme (BEUTLER *et al.*, 1994) is used to make the potential behave more smoothly in relation to the change of λ as can be seem in Figure 3.3.2. The generalized soft core potential is given by:

$$
U^{sc}(r) = \lambda \epsilon \frac{\lambda_r}{\lambda_r - \lambda_a} \left(\frac{\lambda_r}{\lambda_a} \right)^{\left(\frac{\lambda_a}{\lambda_r - \lambda_a} \right)}
$$
$$
\left\{ \frac{1}{[\alpha(1-\lambda) + (r/\sigma)^{\lambda_a}]^{\lambda_r/\lambda_a}} - \frac{1}{\alpha(1-\lambda) + (r/\sigma)^{\lambda_a}} \right\} \tag{3.69}
$$

where α is a constant which the value of 0.5 is normally assumed to it,r is the distance the molecules, ϵ is the depth of the potential well, σ is the distance correspondent to a zero intermolecular potential, λ_r is the repulsive exponent and λ_a is the attractive exponent. Using the Lennard Jones exponents ($\lambda_r = 12$ and $\lambda_a = 6$), Eq. (3.69) becomes:

$$
U^{sc}_{LJ}(r) = 4\lambda \epsilon \left\{ \frac{1}{[\alpha(1-\lambda) + (r/\sigma)^6]^2} - \frac{1}{\alpha(1-\lambda) + (r/\sigma)^6} \right\} \tag{3.70}
$$
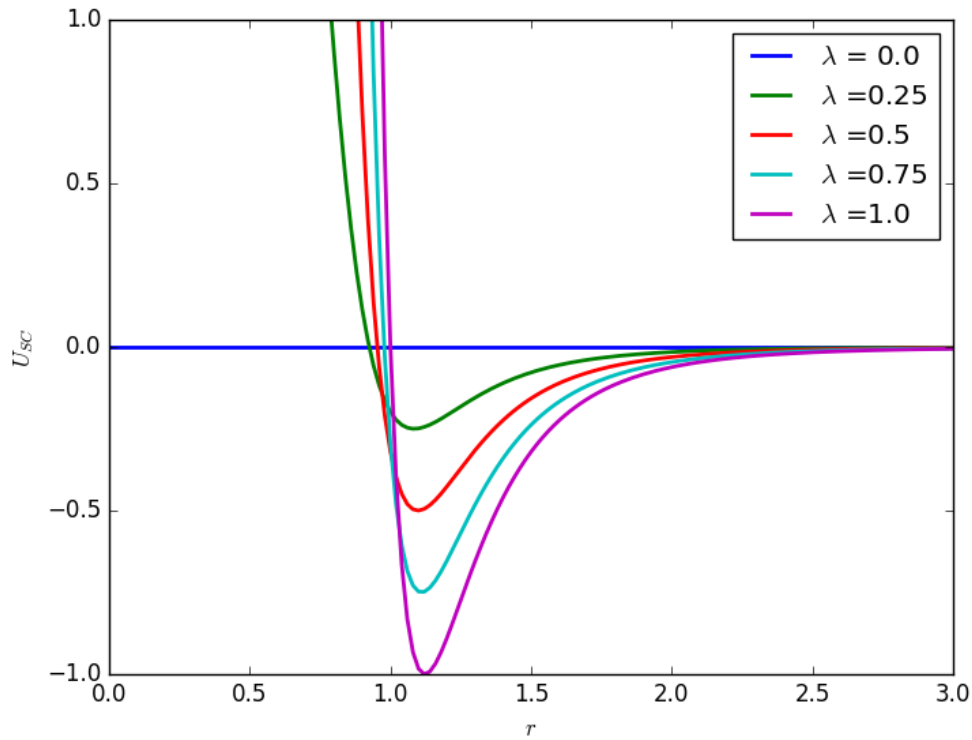
Figure 3.3.2 – Soft-core potential in reduced units.

The $\Delta G_{3 \to 4}$ can be then obtained by doing independent simulations in different values of λ or by doing expanded ensemble simulations (LYUBARTSEV *et al.*, 1992), which samples all states in a single simulation. This method allows a faster sampling across the alchemical states, provided that the kinetic barriers are not substantial. The free energies of solvation obtained can then be used to calculate other properties such as the partition coefficient, that is a measure of the partitioning of one solute in two solvents (a and b) at a temperature T:

$$\log P^{a/b} = \frac{\Delta G_{solv}^a - \Delta G_{solv}^b}{2.303RT} \tag{3.71}$$

3.4   Multistate Bennet Acceptance Ratio (MBAR)

The MBAR method is a free energy of perturbation based method, the MBAR is equal to the BAR method when the calculations are made between only two states. It proposes an estimator that computes free energies and their uncertainties of all K

states by minimizing the KxK matrix of variances simultaneously for a simulation with N_j uncorrelated samples in equilibrium. For each of the $x_{i,n_{n=1}}^{N_i}$ configurations of i, the following probability distributions is sampled:

$$p_i(x) = \frac{q_i(x)}{c_i} \qquad (3.72)$$

$$c_i = \int dx q_i(x) \qquad (3.73)$$

where $q_i(x) = exp(-u_i(x))$ and u_i is the potential energy of each state, defined by $u_i = \beta[U_i(x) + P_iV(x) + \mu_i^T n(x)]$. Meanwhile, c_i is normalization constant. The free energies are estimated with the ratio of this constant in each state:

$$\Delta f_{ij} = f_i - f_j = -ln\frac{c_j}{c_i} = -ln\frac{\int dx q_j(x)}{\int dx q_i(x)} \qquad (3.74)$$

Shirts and Chodera (2008) then proposed the following arbitrary function:

$$c_i \langle \alpha_{ij} q_j \rangle_i = c_j \langle \alpha_{ij} q_i \rangle_j \qquad (3.75)$$

Using the above equation for every state K, the relation is obtained:

$$\sum_{j=1}^{K} \frac{\hat{c}_i}{N_i} \sum_{n=1}^{N_i} \alpha_{ij} q_j(x_{i,n}) = \sum_{j=1}^{K} \frac{\hat{c}_j}{N_j} \sum_{n=1}^{N_j} \alpha_{ij} q_i(x_{j,n}) \qquad (3.76)$$

Shirts and Chodera (2008) suggested the following equation for the arbitrary term α_{ij} in order to decrease the variance:

$$\alpha_{ij}(x) = \frac{N_j \hat{c}_i^{-1}}{\sum_{k=1}^{K} N_k c_i^{-1} q_k(x)} \qquad (3.77)$$

Assuming the sampling is carried out following Boltzmann statistics, Eqs. (3.76) and (3.77) can be rearranged to obtain the free energy estimator, which is solved self

consistently:

$$f_i = \frac{1}{\beta} ln \sum_{k=1}^{K} \sum_{n=1}^{N_k} \frac{\exp[-\beta u_i(x_{kn})]}{\sum_{l=1}^{K} N_l \exp[\beta(f_l - u_l(x_{kn}))]} \tag{3.78}$$

The equation above requires the evaluation of the potential energy of uncorrelated configuration n for all K states ($u_i(x_{kn})$) and for all uncorrelated configuration snapshots (N_k) from state k. Free energy changes between states are then:

$$\Delta f_{ij} = f_j - f_i \tag{3.79}$$

The statistical variance resulting from free energy estimation $S_{ij}^2 \Delta f_{ij}$ is given by the matrix covariances:

$$s_{ij}^2 \Delta f_{ij} \equiv cov(-ln\hat{Z}_j/\hat{Z}_i, -ln\hat{Z}_j/\hat{Z}_i) \tag{3.80}$$

where $\hat{Z}_j$ and $\hat{Z}_i$ are the partition functions of states i and j. The MBAR method can be considered as limit case of the Weighted Histogram Analysis Method (WHAM) (KUMAR *et al.*, 1992) for computing free energies. Eq. (3.78) becomes equal to the WHAM equations if the histogram width tends to zero. Despite this, the MBAR is still more suited than the WHAM because it doesn't have the bias associated with the discretization, allows the calculation of an error estimate and allows free energy calculations of states that weren't sampled (SHIRTS; CHODERA, 2008).

3.5 Expanded Ensemble Method

Instead of doing various simulations in different values of λ, expanded ensemble simulations (LYUBARTSEV *et al.*, 1992) were developed to allow a non-Boltzmann sampling scheme of different states in only one simulation. In this scheme, the sampling is done by biasing the phase space exploration process with weights not related to the statistical ensemble. The statistical expanded ensemble, Z^{EE}, is obtained from the

probability distributions correspondent to each λ, hence Z^{EE} is defined as a sum of sub ensembles Z_i in different values of λ:

$$Z^{EE} = \sum_{i=1}^{N} Z_i(\lambda_i)exp(\eta_i) \tag{3.81}$$

where N is the number of alchemical states, η_i is the arbitrary weight of the sub ensemble at each state and Z_i is configuration partition function at state i. For the isothermal-isobaric ensemble, Z is:

$$Z = \int_0^\infty dV \int d^n r \exp\left[-\beta\left(U(\lambda, r_1, .., r_n) + PV(r_1, .., r_n)\right)\right] \tag{3.82}$$

In solvation free energy calculations with molecular dynamics, λ corresponds to the coupling parameter of the soft-core potential (Eq. 3.69). For molecular dynamic simulations, the sampling of the expanded ensemble is done by performing an arbitrary number of MD steps followed by a λ transition. Chodera and Shirts (2011) proved that type of sampling of the expanded ensemble is similar to the Gibbs sampling method (GEMAN; GEMAN, 1984; LIU, 2002). Following the Gibbs method, the sampling of the configuration space x for one state λ_k during the MD steps is done by using the conditional distribution:

$$\pi(x|\lambda_k) = \frac{\exp[-\beta u(x, \lambda_k)]}{\int dx \exp[-\beta u(x, \lambda_k)]} \tag{3.83}$$

Meanwhile, the state transition in the MD simulation uses the following conditional distribution:

$$\pi(\lambda_k|x) = \frac{\exp[-\beta u(x, \lambda_k) + \eta_k]}{\sum_{k=1}^{K} \exp[-\beta u(x, \lambda_k) + \eta_k]} \tag{3.84}$$

where $u(x, \lambda_k) = U(x, \lambda_k) + PV(x, \lambda_k)$ is the reduced potential function for the NPT ensemble. There are a variety of acceptance schemes to do the expanded sampling using Eq. (3.84), but Chodera and Shirts (2011) suggested that the independence sampling

(LIU, 2002) is the best strategy to increase the number of uncorrelated configurations. The implementation suggested by them updates the state index from i to j by first generating a uniform random number R on the interval $[0, 1)$ and then selecting the smallest new value of j that satisfies the relation below:

$$R < \sum_{i=1}^{j} \pi(\lambda_i | x) \qquad (3.85)$$

The sampling strategy above depends on the weights in order to assure an adequate sampling of the states. If there isn't a sufficient number of states sampled, the expanded ensemble becomes deficient in obtaining input data to estimate free energy differences with the methods exposed in Chapter 2. The weights can be calculated following the flat-histogram approach (BERG; NEUHAUS, 1992; LEE, 1993; DAYAL *et al.*, 2004). This strategy aims to obtain adequate sampling by assuring that all the states have an equal number of samples, i.e. the ratio of the probability of sampling state (π_i) to the probability of sampling state j (π_j) is equal to one. Given that π_i has the following equation:

$$\pi_i = \frac{Z_i(\lambda_i) exp(\eta_i)}{Z^{EE}} \qquad (3.86)$$

and using Eqs. 3.67 and 3.66, the following relation can be obtained for $\pi_i / \pi_j = 1$:

$$(\eta_i - \eta_j) = \beta(G_i - G_j) \qquad (3.87)$$

Eq. (3.87) is solved iteratively with trial simulations. For the first simulation, the values of η are chosen or set to zero and the histogram of the states visited is obtained. With this histogram, it is possible to estimate the free energy differences and, since the weights are related to the free energies by Eq. (3.87), the next values of η can be calculated. This iteration goes on until a uniform distribution is secured. The weights found are then used in a longer simulation to obtain the final solvation free energy differences.

The choice of the λ set correspondent to overlapping alchemical states are crucial to acquire accurate energy differences. In this work, the method chosen to obtain the optimal stage of the λ domain is the one developed by Escobedo and Martinez-Veracoechea (2007) with basis in the study of Trebst, Huse and Troyer (2004). This method targets "bottlenecks" in the simulation. It does that by optimizing λ through the minimization of the number of round trips per CPU time between the lowest λ ($= 0$) and highest λ ($= 1$). This is specifically done by maximizing the steady-state stream ϕ of the simulation, which "walks" among the values of λ. This stream is estimated from Fick's diffusion type of law:

$$\phi = D(\Lambda)\Pi(\Lambda)\frac{dx(\Lambda)}{d\Lambda} \tag{3.88}$$

In the equation above, Λ is the actual continue value of the coupling parameter. This continue function of λs is obtained by interpolating the λ set linearly. $D(\Lambda)$ is the diffusivity at state Λ and $x(\Lambda)$ is the fraction of times that the trial simulation at state Λ_i has most recently visited the state $\lambda = 1$ as opposed to state $\lambda = 0$. The derivative $dx(\Lambda)/d\Lambda$ is approximated with the central finite differences method. Finally, $\Pi(\Lambda)$ is the probability of visiting Λ:

$$\Pi(\Lambda) = \frac{C'\bar{\Pi}(\lambda)}{\Lambda_{i+1} - \Lambda_i} \tag{3.89}$$

The C' term in the equation above represents a constant and $\bar{\Pi}(\lambda)$ is the arithmetic average of visiting the Λ states:

$$\bar{\Pi}(\lambda) = \frac{\pi_{i+1} - \pi_i}{2} \tag{3.90}$$

The ϕ is maximum when the probability $\Pi'(\Lambda_i)$ is proportional to $1/\sqrt{D(\Lambda)}$. With that information, it is possible to estimate the diffusivity using one trial simulation

with the following equation:

$$D(\Lambda) = \frac{\Lambda_{i+1} - \Lambda_i}{\bar{\bar{\Pi}}(\lambda)dx(\Lambda)/d\Lambda} \tag{3.91}$$

Hence, we can calculate $\bar{\bar{\Pi}}$ and, consequently, the cumulative probability, which is used to obtain the new λ states:

$$\Phi = \int_{\lambda=0}^{\lambda=1} \Pi'(\Lambda_i)d\Lambda = \frac{i}{K} \tag{3.92}$$

where, K is the total number of λ states.

3.6 Gibbs Ensemble Monte Carlo (GEMC)

The Monte Carlo (MC) approach is another method for generating atomic trajectories in order to obtain macroscopic properties. The trajectories are obtained stochastically, unlike the molecular dynamics approach. The positions are evolved by random moves (MC steps) or perturbations obtained with the Metropolis method (Metropolis et al., 1953), hence the trajectories are not predictable from the set of initial positions. The Metropolis method is a Markov process, that is a stochastic process in which the configurations change randomly with time and have no memories of the previous configurations. The random move is constructed in such way that the probability of visiting a particular point r^N is proportional to the Boltzmann factor ($exp[-\beta U(r^N)]$) (FRENKEL; SMIT, 2001). The construction of a particle displacement MC step according to Metropolis et al. (1953) can be briefly summed up in:

1. Pick a random particle, and calculate its energy $U(r^N)$.

2. Perturb the particle by randomly displacing its configuration, $r' = r + \Delta$. Where Δ is a perturbation chosen from a defined interval of maximum displacement ($[-\delta_{max}, \delta_{max}]$). Calculate the energy with the new positions $U(r'^N)$.

3. Accept the move from r^N to r'^N with the probability:

$$acc_{A \to B} = min(1, exp[-\beta U(r'^N) + \beta U(r^N)]) \qquad (3.93)$$

The values of maximum displacement are defined in order to obtain acceptance rates of 25-50 % in the step 3 (FRENKEL, 2013). Monte Carlo simulations are interesting when we need to calculate properties in different thermodynamic ensembles, such as the Gibbs Ensemble (PANAGIOTOPOULOS, 1987). This ensemble is used to study phase coexistence with simultaneous Monte Carlo (MC) simulations of two boxes, representing a two phase system, with periodic conditions. The boxes exchange molecules, energy and volume between them. Equilibrium is obtained through MC steps that consist of translation and rotation moves, volume exchange moves and randomly exchanges of molecules between the boxes. For the phase equilibrium of multi-component systems,the GEMC simulations should be carried out at the NPT (constant number of particles, pressure and temperature) ensemble to obey the requirement of an additional degree of freedom for mixtures. Meanwhile, the simulation of single component systems is carried out at constant number of particles, temperature and volume (NVT) since the two phase region would be a line for this system at constant pressure and temperature. The constant volume GEMC ensemble is rigorously equivalent to the canonical ensemble in the thermodynamic limit as demonstrated by Frenkel and Smit (2001). The partition function of the GEMC-NVT ensemble is obtained considering that the particles in both boxes are subjected to the same intermolecular interactions. Also, the boxes' volumes and number of particles (N_1,N_2,V_1 and V_2) can vary while the total volume (V) and total number of particles (N) remain constant ($N = N_1 + N_2$,$V = V_1 + V_2$):

$$Q(NVT) \equiv \sum_{N_1}^{N} \frac{1}{V \Lambda^{3N} N_1! (N - N_1)!} \int_0^V dV_1 V_1^{N_1} V_2^{N_2} \int dx_1^{N_1} \exp[-\beta U(x_1^{N_1})]$$
$$\int dx_2^{N_2} \exp[-\beta U(x_2^{N_2})] \qquad (3.94)$$

In order to define the acceptance rules for the MC moves and compute any property of interest, it is necessary to know the probability of finding the configuration with N_1 particles in box 1 with volume V_1 and positions $x_1^{N_1}$ and $x_2^{N_2}$. This probability is given by:

$$\pi(x_1^{N_1}, x_2^{N_2}, N_1, N_2, V_1, V_2) \propto \frac{V_1^{N_1} V_2^{N_2}}{N_1! N_2!} \exp[-\beta U(x_1^{N_1}) - \beta U(x_2^{N_2})] \qquad (3.95)$$

The acceptance criterion for the translation and rotation moves of configuration A to B is similar to the conventional NVT MC ensembles and is equal to:

$$acc_{A \to B} = \min(1, \exp[-\beta U(x_A^{N_1}) - \beta U(x_B^{N_1})]) \qquad (3.96)$$

The exchange volume moves happen by exchanging an amount ΔV between the boxes to achieve pressure equilibrium. ΔV can be chosen from a uniform distribution based on the maximum variation of volume defined (δV_{max}) with probability $1/\delta V_{max}$. The acceptance rule for these moves is:

$$acc_{A \to B} = \min \left(1, \left(\frac{V_1^B}{V_1^A}\right)^{N_1=1} \left(\frac{V_2^B}{V_2^A}\right)^{N_2+1} \exp[-\beta U(x_A^N) - \beta U(x_B^N)]\right) \qquad (3.97)$$

Particle exchange moves are carried out to obtain the equality of chemical potential between the boxes. One particle from one box is removed and then added to a random location in the other box. The criteria to accept or reject this type of move is:

$$acc_{A \to B} = \min \left(1, \frac{N_1 V_2}{N_2 V_1} \exp[-\beta U(x_A^N) - \beta U(x_B^N)]\right) \qquad (3.98)$$

This method has been widely used to calculate phase equilibrium, but it under performs for the region near the critical point due to large density fluctuations. The GEMC also has a poor performance for dense systems since the particle exchange moves have a low acceptance ratio.

4 Methodology

4.1 Phenanthrene Parameterization

The two parameterization strategies for ring molecules described in section 3.2.2 were implemented for phenanthrene. For both of them, only vapor pressure data (MORTIMER; MURPHY, 1923) were used due to the unavailability of saturated liquid density. The attractive parameter, λ_a, was set to six to avoid correlation with the repulsive parameter. The parameterization with the ring equation of Müller and Mejía (2017) was done with the number of segments fixed in 5 since this level of coarse graining was also used for the similar molecule anthracene:
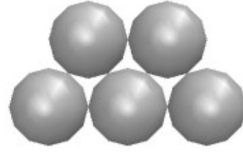
Figure 4.1.1 – Geometry for $m_s = 5$

The minimization was done using the Particle Swarm Optimization (PSO) method with the following objective function:

$$\min_{\sigma, \epsilon, \lambda_r} F_{obj}(\sigma, \epsilon, \lambda_r) = \sum_{i=1}^{N_p} \left(\frac{P_v^{SAFT}(T_i, \sigma, \epsilon, \lambda_r) - P_v^{exp}(T_i)}{P_v^{exp}(T_i)} \right)^2 \tag{4.1}$$

P_v^{exp} is the experimental vapor pressure and P_v^{SAFT} is the vapor pressure obtained with SAFT-VR Mie EoS. We used the routine proposed by Smith, Ness and Abbot (2007) routine to calculate the bubble point with the EoS. The resulting parameters σ, ϵ and λ_r of Eq. (4.1) are the final force field parameters used in molecular simulations.

The parameterization with the Lafitte *et al.* (2012) ring equation was done with $m_s = 3$ so every bead would represent one ring:

The first part of the estimation followed the same procedure described above
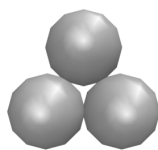
Figure 4.1.2 – Geometry for $m_s = 3$

for the Müller and Mejía (2017) equation. As explained in section 3.2.2, the Lafitte *et al.* (2012) equation requires the estimation of correction factors c_σ and c_ϵ (Eqs. (3.48) and (3.49)). The PSO method was used with Eq. (3.50). In this equation, the vapor pressures and saturated liquid densities from molecular simulations were obtained using the Gibbs Ensemble Monte Carlo method on the NVT ensemble (section 3.6).

The boxes for the GEMC-NVT simulations were generated inserting 400 molecules on the liquid box and 100 on the vapor one using the Packmol package (MARTÍNEZ *et al.*, 2009). Initial densities of each box were equal to the saturated densities found with the SAFT-VR Mie Eos in order to avoid the migration of all molecules to a single phase throughout the simulation. Simulations had equilibration and production times of 10^4 and 510^4 MC cycles respectively. Each MC cycle corresponded to 10^3 rotation trials, 10^3 translation trials, 10^2 molecule insertion trials, 10^2 molecule deletion trials and 10 volume exchange trials. The cut off distance was equal to $20\hat{A}$ without long range interactions. The saturated vapor density (ρ_{vap}), the saturated liquid density (ρ_{liq}) and the vapor pressure (P_v) were sampled at each 100 MC cycles and this data were divided in five blocks to calculate the averages and standard deviations. With the correction factors found after the estimations, we calculated the σ and ϵ parameters with Eqs. (3.51) and (3.52).

4.2   Solvation Free Energy Calculations

Molecular dynamic simulations to estimate the free energy differences with the SAFT-γ Mie force field were performed using the Lammps package (PLIMPTON, 1995). The motion equations were integrated with the velocity-Verlet algorithm (VERLET, 1967) with a time step of 1 fs. As required by the coarse grained model, molecules were

treated as rigid bodies. The thermostat and the barostat were the Nose/Hoover with chains with a damping factors of 100 and 1000 fs respectively. Electrostatics interactions are not explicitly accounted by the SAFT-γ Mie force field, hence there were no shifting of forces or long range corrections. The potential cutoff was equal to 20 $\dot{A}$ (MÜLLER; MEJÍA, 2017) with a neighbor skin of 2 $\dot{A}$. The initial configurations of the solvated systems were generated with the Packmol package (MARTÍNEZ *et al.*, 2009). For the binary mixtures, one molecule of solute and a varying number of solvent molecules-700 molecules for toluene and octanol, 1024 for hexane, 3000 for water - were randomly added to a cubic box. The simulations to study solvation free energy of phenanthrene in a mixture of toluene and carbon dioxide were done with different weight fractions of carbon dioxide. The system consisted of one molecule of phenanthrene for all the fractions and 123 molecules of CO_2 and 618 molecules of toluene for $w_{CO_2} = 0.087$; 166 molecules of CO_2; 589 molecules of toluene for $w_{CO_2} = 0.119$; 232 molecules of CO_2 and 545 molecules of toluene for $w_{CO_2} = 0.169$ and 380 molecules of CO_2 and 446 molecules of toluene for $w_{CO_2} = 0.289$.

All simulations were carried out maintaining the temperature and pressure constant at 298 K and 1 bar, except the ones containing carbon dioxide. These had the temperature of 298 K and the pressure of the liquid phase equilibrium correspondent to the CO_2 fraction (CHANG, 1992). For all the simulations, the initial box was equilibrated at the NPT ensemble for 2 ns and then the resulting configurations were used as the initial configuration of the expanded ensemble simulations. These were carried out with the Lammps user package for expanded ensemble simulations with the Mie Potential developed by our group, available at https://github.com/atoms-ufrj/USER-ALCHEMICAL. The sampling of a new state was tried at every 10 MD steps. In order to define the optimal values of λ and η related to each state, short trials simulations, having around 9 ns of production time, were carried out. In the first simulation, the group of λ for all the pairs solvent-solute was: (0.0,0.15,0.2,0.25,0.3,0.4,0.45,0.5,0.55,0.7,0.9,1.0) and the ηs were set to zero or were given the values of the ones found for similar pairs solute-solvent. The subsequent group of η were estimated with the flat histogram

approach (Eq. (3.87)). We then did another trial simulation with the new weights. The results of this simulation were used to optimize the group of λs by minimizing the number of round trips, as described in section 3.5. The ηs correspondent to the newest group of λs were interpolated linearly from the free energy differences. With the final values of η and λ defined for each mixture, larger simulations with a production time of 20 ns were carried out.

Since the force field considers that the beads don't have charges, there is no coulombic interaction and Eq. (3.68) becomes equal to $\Delta G_{3\rightarrow4}$. The post processing method used to calculate free energy differences was the Multisate Bennet Acceptance Ratio (MBAR) described in section 3.4. The software alchemical-analysis (KLIMOVICH; SHIRTS; MOBLEY, 2015) was used to obtain the ΔG_{solv} with MBAR and to assess the quality of the results. We only estimated the binary interaction parameter of Eq. (3.46) for pairs with water as a solvent. The estimation was done by performing trial expanded ensemble simulations in three values of k_{ij}. With the ΔG_{solv} obtained with these simulations, we did a linear fit to obtain the refined value of the parameter. We usede this strategy because the estimation with SAFT VR Mie EoS gave poor results for the free energy.

5 Results and Discussion

5.1 Solvation free energies

The first part of this work consisted of obtaining phenanthrene parameters for the SAFT-γ Mie Force Field, as described in section 4.1, since the parameters were not available for the ring geometry on the force field database. The parameters obtained and the mean percentage error (MPE) to the experimental data were:

Table 1 – Estimated SAFT-γ Mie Force Field parameters for phenanthrene

m_s	ϵ/k_B (K)	$\sigma(\dot{A})$	λ_r	MPE(%)
3 (LAFITTE *et al.*, 2012)	485.55	4.197	14.34	1.64 | 9.74
5 (MÜLLER; MEJÍA, 2017)	262.74	4.077	9.55	0.88

In the table above, the first value of MPE for the Lafitte *et al.* (2012) strategy corresponds to the estimation with experimental data and the second corresponds to the corrections factors estimation with MD data. This strategy has an inconsistency of requiring two estimations because parameters solely estimated with the EoS aren't accurate for molecular simulation. Hence, the solvation free energy of phenanthrene was only studied with the set of parameters estimated with Müller and Mejía (2017) strategy. In fact, the (LAFITTE *et al.*, 2012) strategy was only followed because it was the only one available when we first started this work. The sets of parameters for the other compounds were retrieved from the literature (LOBANOVA *et al.*, 2016; HERDES; TOTTON; MÜLLER, 2015; ERVIK; MEJÍA; MÜLLER, 2016; MÜLLER; MEJÍA, 2017):

The solvation free energies of aromatic solutes in nonpolar (hexane), aromatic (toluene) and hydrogen bonding (1-octanol) solvents were examined with binary interaction parameters equal to zero. Since the force field doesn't account for charges, the solvation free energy is equal to the Mie contribution (Eq. (3.69)). A total of 15-18, depending on the pairs solute-solvent, λs and their respective ηs were estimated as described in the methodology. The final λ set was found using the cumulative probability

Table 2 – SAFT-γ Mie Force Field for each substance used in this work

	m_s	ϵ/k_B (K)	$\sigma(\mathring{A})$	λ_r
Water	1	305.21	2.902	8.0
Propane	1	426.08	4.871	34.29
Carbon dioxide	2	194.94	2.848	14.65
Hexane	2	376.35	4.508	19.57
Octanol	3	495.71	4.341	28.79
Toluene	3	268.24	3.685	11.80
Benzene	3	230.30	3.441	10.45
Pyrene	4	459.04	4.134	15.79
Anthracene	5	259.68	3.631	9.55

distribution (Eq. (3.92)) for all pairs. The distribution for the pair hexane+benzene can be seem in Figure 5.1.1 and the optimized values of λ and η in Table 3. The λs and ηs for the other pairs are available at Appendix A. The coupling parameters are concentrated on the region with a steeper slope for all the pairs as it was expected.
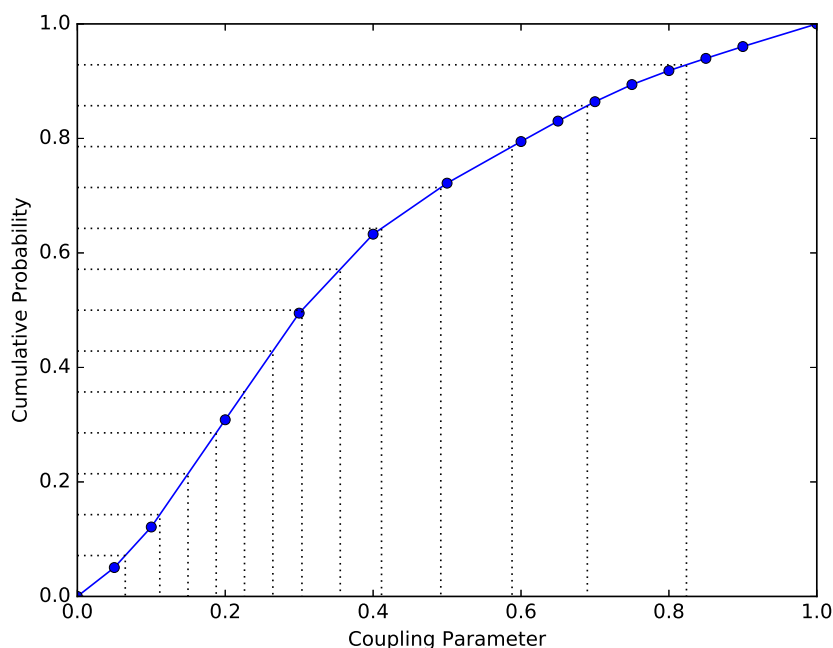
Figure 5.1.1 – Cumulative probability

It is also important to analyze the reliability of solvation free energy estimations through the overlapping of the intermediate states found with the optimization above. Insufficient overlap among states when using a FEP based such MBAR may result in the underestimation of variance, and, consequently, in substantially incorrect free energies. The overlapping matrix for the solvation free energy of benzene in hexane is

Table 3 – Optimized values of λ and η for the pair benzene + hexane

λ	η
0	0
0.065	0.708
0.112	1.385
0.15	1.892
0.188	2.399
0.226	2.519
0.264	2.457
0.304	2.367
0.356	1.921
0.411	1.411
0.492	0.524
0.588	-0.663
0.69	-2.016
0.824	-3.922
1	-6.583

presented in Figure 5.1.2 and the matrices for the other pairs are available at Appendix B. The elements of these matrices are the probabilities of observing a sample from state i in state j. As an example, the probability of observing state 6 in a simulation of state 8 is 0.16 in Figure 5.1.2. It can be observed in the figures that the matrices for all the pairs in this work have more than three diagonals. In addition to that, the probabilities in the three main diagonals are higher than 0.03, which indicates that the free energy estimations are reliable (KLIMOVICH; SHIRTS; MOBLEY, 2015).

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.42	.29	.16	.08	.03	.01	.01								
1	.33	.27	.18	.10	.06	.03	.02	.01							
2	.23	.23	.18	.12	.09	.06	.04	.02	.01	.01					
3	.14	.16	.16	.13	.11	.10	.08	.06	.03	.02	.01				
4	.06	.09	.11	.11	.12	.12	.13	.11	.07	.04	.02	.01	.01		
5	.02	.03	.05	.07	.09	.12	.16	.16	.12	.08	.05	.02	.01	.01	
6		.01	.02	.03	.06	.09	.15	.18	.16	.12	.08	.04	.02	.01	
7			.01	.02	.03	.07	.13	.17	.18	.16	.11	.07	.04	.02	.01
8				.01	.02	.04	.09	.14	.17	.18	.15	.10	.06	.03	.01
9					.01	.02	.06	.11	.15	.18	.17	.13	.09	.05	.02
10						.01	.04	.07	.11	.16	.18	.17	.14	.08	.03
11						.01	.02	.04	.08	.12	.17	.19	.19	.14	.06
12							.01	.02	.05	.08	.13	.18	.22	.19	.11
13							.01	.01	.03	.05	.09	.15	.22	.25	.19
14								.01	.01	.03	.06	.11	.19	.29	.30

Figure 5.1.2 – Overlapping matrix for hexane+benzene.

Table 4 – Calculated and experimental values for the solvation free energy differences (kcal/mol) of solutes in non aqueous solvents

Solvent	Solute	ΔG_{solv}^{exp}	ΔG_{solv}^{Mie}	Absolute Deviation
hexane	benzene	-3.96	-3.76 $\pm$ 0.01	0.20
hexane	pyrene	-11.53	-10.82 $\pm$ 0.02	0.71
hexane	phenanthrene	-10.01	-9.16 $\pm$ 0.01	0.85
1-octanol	propane	-1.32	-1.36 $\pm$ 0.02	0.04
1-octanol	anthracene	-11.72	-8.16 $\pm$ 0.03	3.61
1-octanol	phenanthrene	-10.22	-8.34 $\pm$ 0.03	1.47
toluene	pyrene	-12.86	-11.74 $\pm$ 0.01	1.11
toluene	anthracene	-11.31	-9.90 $\pm$ 0.01	1.41

The numerical values for solvation free energies estimated in Table 4 had smaller absolute deviations to experimental data, what shows that the SAFT-γ Mie force field performs better for a non polar solvent. Additionally, this force field presented better

results for the pair hexane+benzene than the Trappe force field (GARRIDO *et al.*, 2011).
We also observed the effect of molecule's size on the entropic region of the free energy
curve in Figure 5.1.3. It was expected that a force field based on an EoS that doesn't
explicitly account for hydrogen bond would not perform well for 1-octanol. Despite
this, the solvation free energies of propane and phenanthrene in 1-octanol stayed in
the desired deviation range of 1-2 kcal/mol (MOBLEY; GILSON, 2017). The deviation
was much smaller for propane. This can be attributed to propane's non polarity and
smoother free energy curve (Figure 5.1.4). The anthracene and phenanthrene molecules
have the same geometry in the model and similar physical properties, but the absolute
deviation of the solvation free energy of anthracene in 1-octanol is much higher than
the one of phenanthrene 1-octanol. This may indicate a problem in the parameterization
of anthracene.
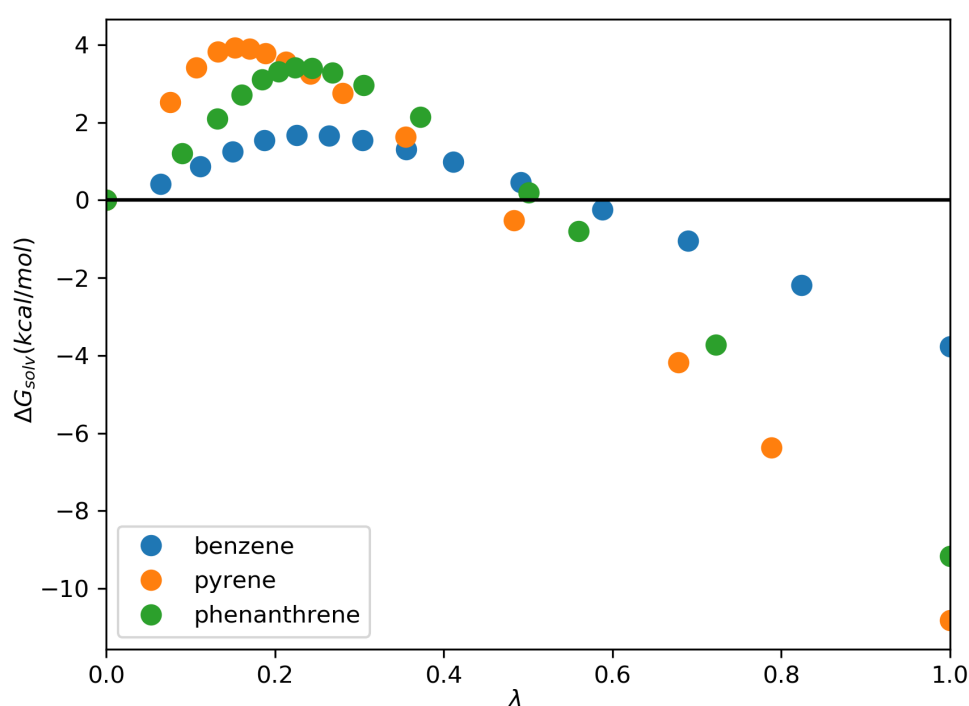
Figure 5.1.3 – Solvation free energy profiles of different solutes in hexane.
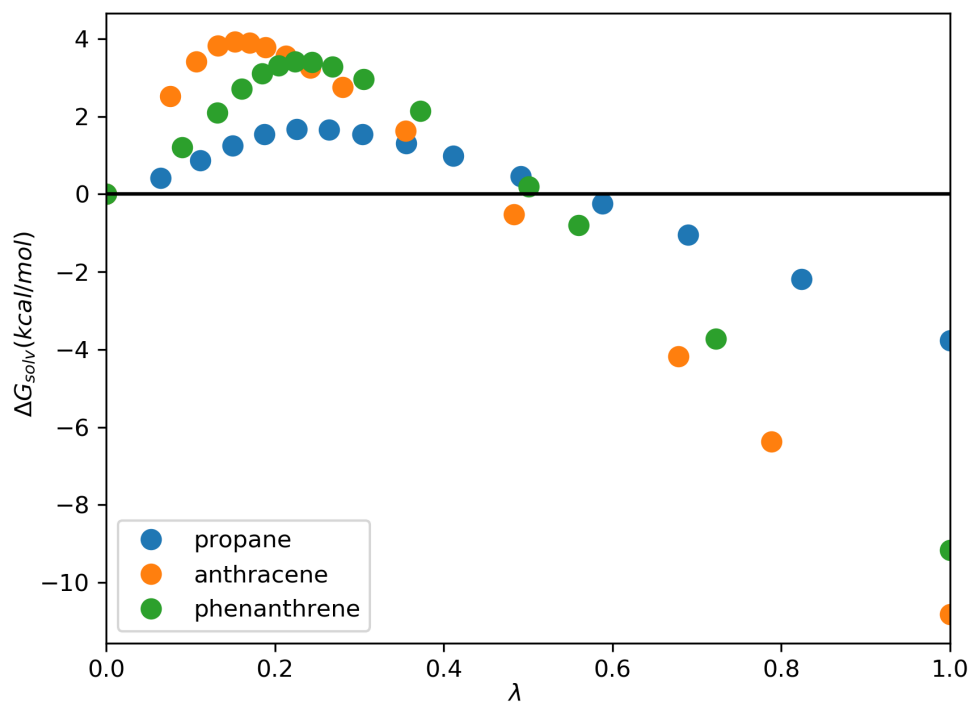
Figure 5.1.4 – Solvation free energy profiles of different solutes in 1-octanol.
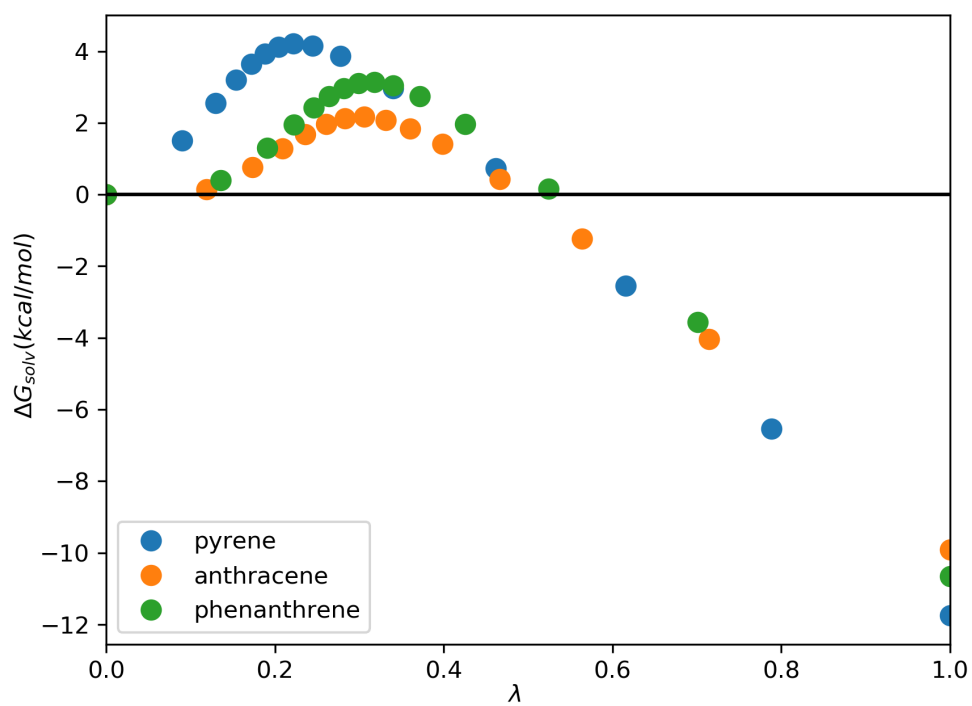
Figure 5.1.5 – Solvation free energy profiles of different solutes in toluene.

The results also indicated the prediction capability of the force field for pairs of aromatic solute and solvent. The influence of the molecule's geometry on free energy curves (Figure 5.1.5) was also observed for these pairs. ΔG_{solv} was calculated too for phenanthrene in toluene and in toluene+CO_2. To the best of our knowledge, there was no available experimental data for these solvation free energies, but the previous results for phenanthrene in other solvents and for the pair anthracene+toluene showed that the force field is adequate to describe the solvation phenomenon of phenanthrene in an aromatic solvent. The results for these sets are exposed below:

Table 5 – Calculated values for the solvation free energy differences (kcal/mol) of phenanthrene in toluene+CO_2.

w_{CO_2}	ΔG_{solv}^{Mie}
0.0	-10.65 $\pm$ 0.02
0.087	-10.73 $\pm$ 0.02
0.119	-10.78 $\pm$ 0.02
0.169	-10.71 $\pm$ 0.02
0.289	-10.69 $\pm$ 0.02

The increasing of CO_2 mass fraction in toluene caused a small effect on solvation free energies. First, the ΔG_{solv} decreased with the increase of w_{CO_2} , indicating a higher solubility. From the 0.169 fraction, the effect was reversed and carbon dioxide became an anti solvent. It was observed that asphaltene precipitation occurs when carbon dioxide mass fractions became higher than 0.10 in the system asphaltene+toluene+carbon dioxide (SOROUSH *et al.*, 2014), what is in accordance with the anti solvent effect of carbon dioxide observed on the calculated values. It is also important to point out that the small differences observed in the free energy profiles (Figure 5.1.6) may indicate the non-influence of CO_2 in solvation of phenanthrene in toluene when using the Saft-γ Mie force field. But, since this is a qualitative study due the lack of this system experimental data, more studies need to be done in order to make a secure assertion about it.
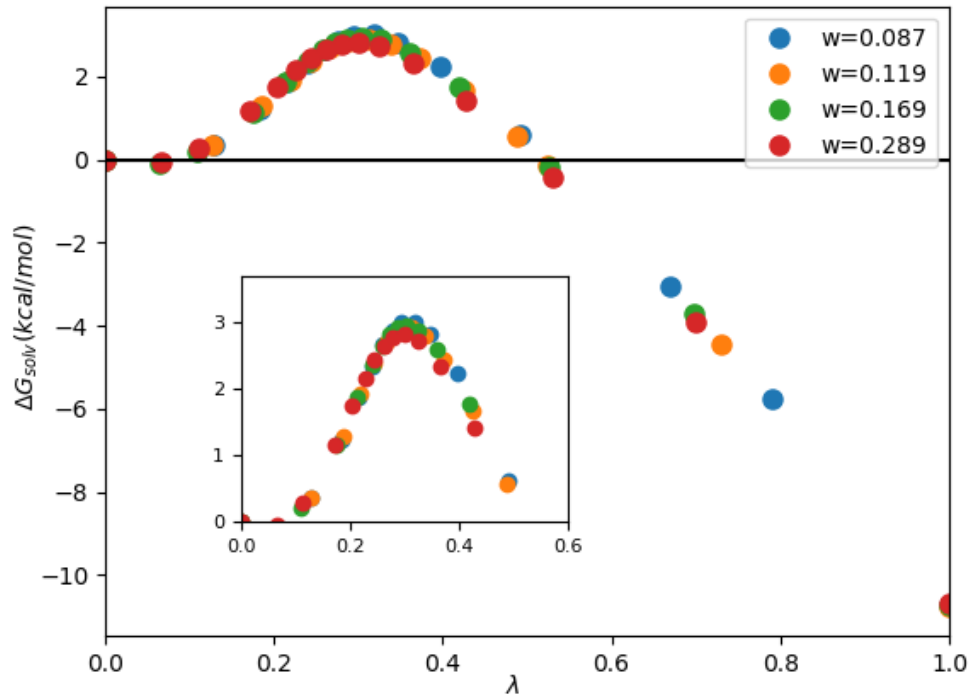
Figure 5.1.6 – Solvation free energy profiles of phenanthrene in toluene+CO_2.

5.2   Hydration free energies

The hydration free energies of widely studied solutes (propane, benzene) and aromatic solutes (toluene, phenanthrene) were calculated with a group of fifteen intermediate states. First, the binary interaction parameter was set to zero, but the results obtained deviated a lot from the experimental data as can be seen in the table below:

Table 6 – Calculated values for the hydration free energy differences (kcal/mol) of solutes in water for $k_{ij} = 0$.

Solute	ΔG_{solv}^{Mie}	ΔG_{solv}^{exp}
propane	2.00 ± 0.20	1.10 ± 0.01
benzene	-0.86 ± 0.20	-4.45 ± 0.03
toluene	-0.83 ± 0.20	-15.80 ± 0.06
phenanthrene	-3.88 ± 0.60	-10.90 ± 0.04

After these results, the need of the binary interaction parameter was clear. The k_{ij} estimation with the SAFT VR Mie EoS and experimental vapor pressure data also didn't provide good results, then the strategy of estimating the k_{ij} with the output from

solvation free energy calculations with molecular dynamics was used, as described in the last paragraph of section 4.2. Individual values for the interaction parameter for each solvent was initially found, but, since the parameters values for aromatic solutes were very similar (0.148, 0.162, 0.152), the average of these values was taken in order to obtain a general parameter for the water+aromatic pairs. The parameters estimated are:

Table 7 – Binary interaction parameters employed.

Pair	k_{ij}
water + propane	0.067
water + aromatic	0.154

The relatively large k_{ij} value for the aromatic solutes can be pinned on the lack of an explicit association term on the model and on the water model itself, since the model didn't need a k_{ij} for mixtures with 1-octanol. Nevertheless, the value obtained by fitting the parameter with molecular dynamics interfacial tension data was larger for the mixture water+toluene (0.241) (HERDES *et al.*, 2017). The SAFT-γ Mie model for water (LOBANOVA *et al.*, 2016) has actually two different of temperature dependent sets of parameters. The one used in this work was the set estimated with experimental interfacial tension data, but the binary interaction parameter estimated with MD phenomenon tension data could not be transferable to solvation free energy calculations. Modeling the water with a coarse grained methodology has a lot of difficulties because the water molecules move independently and only have non bonded interactions (HADLEY; MCCABE, 2010; HADLEY; MCCABE, 2012). The SAFT-γ Mie water model considers that one water molecule corresponds to one bead, this strategy only saves small simulation time, but it is able to predict properties at physiological temperatures unlike other more aggressive models, which consider that one bead represents various water molecules. In light of all this, the SAFT-γ Mie force field appears to be a good alternative when working close to room temperatures, but the necessity of additional parameters estimated with molecular simulation indicates problems on the model. The solvation free energy differences calculated with the final fitted binary interaction parameters are presented in Table 8.

Table 8 – Calculated and experimental values for the hydration free energy differences (kcal/mol) of solutes in water.

Solute	ΔG_{solv}^{exp}	ΔG_{solv}^{Mie}	Absolute Deviation	ΔG_{solv}^{GAFF}
propane	2.00 ± 0.20	2.01 ± 0.01	0.01	2.50 ± 0.02
benzene	-0.86 ± 0.20	-1.12 ± 0.01	0.26	-0.81 ± 0.02
toluene	-0.83 ± 0.20	-0.84 ± 0.01	0.01	-0.79 ± 0.03
phenanthrene	-3.88 ± 0.60	-3.47 ± 0.02	0.41	-5.26 ± 0.03

Hydration free energy differences with k_{ij} had low absolute deviations from the experimental data, as expected since the parameter was adjusted to fit the data. The root mean square error (RMSE) for the pairs tested with the SAFT-γ Mie model was equal to 0.24, meanwhile the RMSE for hydration free energy differences with the GAFF force field (MOBLEY; GUTHRIE, 2014) was 0.73. The difference in absolute deviations between the two force fields is significantly high for phenanthrene, hence the coarse grained force field with a binary parameter is preferred if the application requires a higher level of accuracy. Hydration free energy profiles in Figure 5.2.1 show the geometry dependence on the free energy profiles. We can observe also that the solvation free energy for the first non zero lambda is negative for benzene and toluene, when a positive value is expected.This can be caused by numerical errors or by the force field.
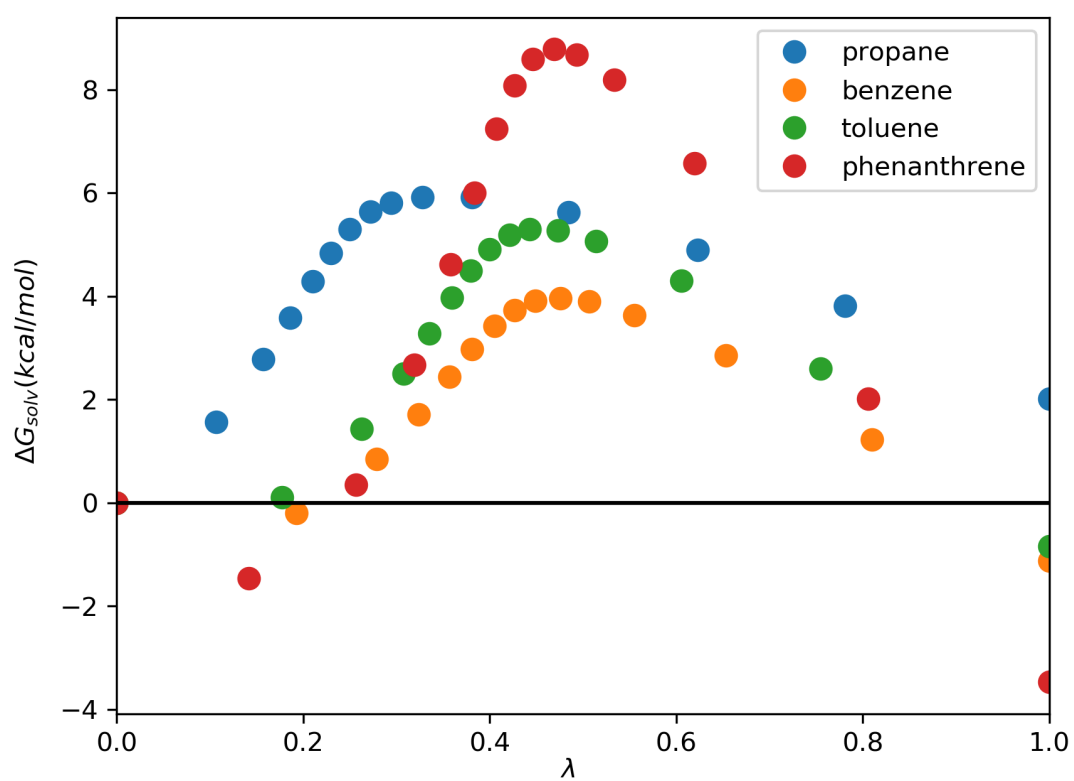
Figure 5.2.1 – Hydration free energy profiles for different solutes.

6 Conclusions

This dissertation consisted in the study of solvation free energy calculations of aromatic solutes that can mimic asphaltenes with the SAFT-γ Mie force field. By doing that, we provided information about these calculations since solvation free energy studies are mostly done using water as solvent and all atom force fields based on the Lennard Jones Potential. The parametrization of the SAFT-γ Mie force field is more straightforward when compared to other force fields since its parameters are obtained through the SAF-VR Mie EoS. Following this strategy, the phenanthrene parameters were obtained using two different ring equations and geometries. The ring equation proposed by Müller and Mejía (2017) provided the more adequate set of parameters. The potential energy data for every intermediate state were obtained with simulations at the expanded ensemble.

Solvation free energy differences were then estimated with these data using the MBAR method. The results for solvation free energy differences with non aqueous solvents had absolute deviations to the experimental data of less than 2.0 kcal/mol, except for the pair 1-octanol+anthracene. The geometry effect on the free energy curves was also observed- larger molecules had steeper curves and larger absolute deviations. The influence of carbon dioxide on the solvation free energy of phenanthrene in toluene was found to be minimum. The ΔG_{solv} decreased slightly until the mass fraction of CO_2 was equal to 0.119 and, after this point, solvation free energies increased.

Hydration free energy differences calculations with the SAFT-γ Mie model required the use of relatively larger values of k_{ij} in order to obtain satisfactory results. The parameter was estimated with the output from molecular dynamics data as the strategy of using the SAFT-VR Mie EoS also didn't provide good results. This necessity of one additional parameter happens probably due to the lack of an association term on the EoS that the model is based on. The results with k_{ij} estimated with MD output were great, the absolute deviations to the experimental data found were smaller than

the ones for the GAFF force field.

Generally, the SAFT-γ Mie force field proved to be a good model to represent the solvation phenomenon. It correctly described solvation free energy differences of solutes mimicking asphaltenes in hexane, toluene, 1-octanol and water. The requirement of binary interaction parameter estimated with MD output for hydration free energies increases the simulation time, which is already larger for this water model due to its coarse graining level. Nevertheless, the SAFT-γ Mie force field for water used doesn't predict freezing at room temperature as other force fields, which is essential for hydration free energy calculations. This dissertation had success in using a coarse grained force field to perform free energy calculations. Based on this work, we have some ideas for future development. We intend to test the binary interaction parameter transferability to calculations with other aromatic solutes in water. Additionally, we want to use the SAFT-γ Mie force field to model more complex asphaltenes models and, consequently, increase the scale of these simulations. The final step to expand this work would be to develop new methodologies to use solvation free energies to effectively calculate solubility

Bibliography

ADAM, L.; MACIEJ, B.; CEZARY, C.; EWA, G.; YI, H.; DAWID, J.; PAWEL, K.; MACIEJ, M.; MARIUSZ, M.; A, M. M.; ANDREI, N.; STANISLAW, O.; A, S. H.; K, S. A.; RAFAL, S.; TOMASZ, W.; YANPING, Y.; BARTLOMIEJ, Z. A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *Journal of molecular modeling*, v. 20, p. 2306, 2014.

AIMOLI, C. G.; MAGINN, E. J.; ABREU, C. Force field comparison and thermodynamic property calculation of supercritical co_2 and ch_4 using molecular dynamics simulations. *Fluid Phase Equilibria*, v. 368, p. 80–90, 2014.

AIMOLI, C. G.; MAGINN, E. J.; ABREU, C. R. Transport properties of carbon dioxide and methane from molecular dynamics simulations. *The Journal of Chemical Physics*, v. 141, p. 134101, 2014.

ANDERSEN, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *JOURNAL OF COMPUTATIONAL PHYSICS*, v. 52, p. 24–34, 1983.

AVENDAÑO, C.; LAFITTE, T.; GALINDO, A.; ADJIMAN, C. S.; JACKSON, G.; MULLER, E. A. 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*, v. 115, p. 11154–11169, 2011.

BARDUCCI, A.; BONOMI, M.; DERREUMAUX, P. Assessing the quality of the opep coarse-grained force field. *Journal of Chemical Theory and Computation*, v. 7, p. 1928– 1934, 2011.

BARKER, J. A.; HENDERSON, D. What is "liquid"? understanding the states of matter. *Review of Modern Physics*, v. 48, p. 587–671, 1976.

BASDEVANT, N.; BORGIS, D.; HA-DUONG, T. Modeling protein-protein recognition in solution using the coarse-grained force field scorpion. *Journal of Chemical Theory and Computation*, v. 9, p. 803–813, 2013.

BECKSTEIN, O.; FOURRIER, A.; IORGA, B. I. Prediction of hydration free energies for the sampl4 diverse set of compounds using molecular dynamics simulations with the opls-aa force field. *Journal of Computer-Aided Molecular Design*, v. 28, n. 3, p. 265–276, Mar 2014.

BENNETT, C. Efficient estimation of free energy differences from monte carlo data. *Journal of Computational Physics*, v. 22, p. 245–268, 1976.

BEREAU, T.; BACHMANN, M.; DESERNO, M. Interplay between secondary and tertiary structure formation in protein folding cooperativity. *Journal of the American Chemical Society*, v. 132, p. 13129– 13131, 2010.

BEREAU, T.; DESERNO, M. Generic coarse-grained model for protein folding and aggregation. *Journal of Chemical Physics*, v. 130, p. 235106, 2009.

BERENDSEN, H. J. C.; POSTMA, J. P. M.; GUNSTEREN, W. F. van; DINOLA, A.; HAAK, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, v. 81, n. 8, p. 3684–3690, 1984.

BERG, B. A.; NEUHAUS, T. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Physical Review Letters*, American Physical Society, v. 68, p. 9–12, Jan 1992.

BEUTLER, T.; MARK, A.; SCHAIK, R. van; GERBER, P.; GUNSTEREN, W. van. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical Physics Letters.*, v. 222, p. 529–539, 1994.

BHARGAVA, B.; KLEIN, M. L. Formation of micelles in aqueous solutions of a room temperature ionic liquid: a study using coarse grained molecular dynamicss. *Molecular Simulation*, v. 107, p. 393–401, 2009.

CHANG, C. J. The solubility of carbon dioxide in organic solvents at elevated pressures. *Fluid Phase Equilibria*, v. 15, p. 235–242, 1992.

CHEBARO, Y.; DONG, X.; LAGHAEI, R.; DERREUMAUX, P.; MOUSSEAU, N. Pep-fold: an online resource for de novo peptide structure prediction. *Nucleic Acids Research*, v. 37, p. 498– 503, 2009.

CHEBARO, Y.; DONG, X.; LAGHAEI, R.; DERREUMAUX, P.; MOUSSEAU, N. Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent. *Journal of Physical Chemistry B*, v. 113, p. 267– 274, 2009.

CHIPOT, C.; POHORILLE, A. *Free Energy Calculations Theory and Applications in Chemistry and Biology*. 1. ed. New York: Springer, 2007.

CHIU, S.; SCOTT, H.; JAKOBSSON, E. A coarse-grained model based on morse potential for water and n-alkanes. *Journal of Chemical Theory and Computation*, v. 6, p. 851–863, 2010.

CHODERA, J. D.; SHIRTS, M. R. Replica exchange and expanded ensemble simulations as gibbs sampling: simple improvements for enhanced mixing. *Journal of Chemical Physics*, v. 135, p. 194110, 2011.

DAYAL, P.; TREBST, S.; WESSEL, S.; WÜRTZ, D.; TROYER, M.; SABHAPANDIT, S.; COPPERSMITH, S. N. Performance limitations of flat-histogram methods. *Physical Review Letters*, American Physical Society, v. 92, p. 097201, Mar 2004.

ERVIK, A.; LYSGAARD, M. O.; HERDES, C.; JIMÉNEZ-SERRATOS, G.; MÏLER, E. A.; MUNKEJORD, S. T.; MÜLLER, B. A multiscale method for simulating fluid interfaces covered with large molecules such as asphaltenes. *Journal of Computational Physics*, v. 327, p. 576–611, 2016.

ERVIK, A.; MEJÍA, A.; MÜLLER, E. A. Bottled saft: A web app providing saft-γ mie force field parameters for thousands of molecular fluids. *Journal of Chemical Information and Modeling*, v. 56, p. 1609–1614, 2016.

ERVIK Åsmund; LYSGAARD, M. O.; HERDES, C.; JIMéNEZ-SERRATOS, G.; MüLLER, E. A.; MUNKEJORD, S. T.; MüLLER, B. A multiscale method for simulating fluid interfaces covered with large molecules such as asphaltenes. *Journal of Computational Physics*, v. 327, p. 576 – 611, 2016. ISSN 0021-9991.

ESCOBEDO, F. A.; MARTINEZ-VERACOECHEA, F. J. Optimized expanded ensembles for simulations involving molecular insertions and deletions. i. closed systems. *Journal of Chemical Physics*, v. 127, p. 174103, 2007.

FERRENBERG, A. M.; SWENDSEN, R. H. Optimized monte carlo data analysis. *Phys. Rev. Lett.*, American Physical Society, v. 63, p. 1195–1198, Sep 1989.

FRENKEL, D. Simulations: The dark side. *The European Physical Journal Plus*, v. 128, n. 1, p. 10, 2013.

FRENKEL, D.; SMIT, B. *Understanding Molecular Simulation*. 2nd. ed. Orlando, FL, USA: Academic Press, Inc., 2001. ISBN 0122673514.

GAO, F.; XU, Z.; LIU, G.; YUAN, S. Molecular dynamics simulation: The behavior of asphaltene in crude oil and at the oil/water interface. *Energy & Fuels*, v. 28, n. 12, p. 7368–7376, 2014.

GARRIDO, N. M.; JORGE, M.; QUEIMADA, A. J.; MACEDO, E. A.; ECONOMOU, I. G. Using molecular simulation to predict solute solvation and partition coefficients in solvents of different polarity. *Physical Chemistry Chemical Physics*, v. 20, p. 9155–9164, 2011.

GARRIDO, N. M.; QUEIMADA, A. J.; JORGE, M.; MACEDO, E. A.; ECONOMOU, I. G. 1-octanol/water partition coefficients of n-alkanes from molecular simulations of absolute solvation free energies. *Journal of Chemical Theory and Computation*, v. 5, p. 2436–2446, 2009.

GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, p. 721 – 741, 1984.

GONçALVES, P. F. B.; STASSEN, H. Free energy of solvation from molecular dynamics simulation applying voronoi-delaunay triangulation to the cavity creation. *The Journal of Chemical Physics*, v. 123, p. 214109, 2005.

GREENFIELD, M. L. Molecular modelling and simulation of asphaltenes and bituminous materials. *International Journal of Pavement Engineering*, v. 12, n. 4, p. 325–341, 2011.

HADLEY, K. R.; MCCABE, C. On the investigation of coarse-grained models for water: Balancing computational efficiency and the retention of structural properties. *The Journal of Physical Chemistry B*, v. 114, n. 13, p. 4590–4599, 2010.

HADLEY, K. R.; MCCABE, C. Coarse-grained molecular models of water: A review. *Molecular simulation*, v. 38 8-9, p. 671–681, 2012.

HE, X.; SHINODA, W.; DEVANE, R.; KLEIN, M. L. Exploring the utility of coarse-grained water models for computational studies of interfacial systems. *Molecular Simulation*, v. 108, p. 2007–2020, 2010.

HEADEN, T. F.; BOEK, E. S.; JACKSON, G.; TOTTON, T. S.; MüLLER, E. A. Simulation of asphaltene aggregation through molecular dynamics: Insights and limitations. *Energy & Fuels*, v. 31, n. 2, p. 1108–1125, 2017.

HERDES, C.; ERVIK, A.; MEJÍA, A.; MÜLLER, E. A. Prediction of the water/oil interfacial tension from molecular simulations using the coarse-grained saft-γ mie force field. *Fluid Phase Equilib.*, 2017.

HERDES, C.; TOTTON, T. S.; MÜLLER, E. A. Coarse grained force field for the molecular simulation of natural gases and condensates. *Fluid Phase Equilibria*, v. 406, p. 91–100, 2015.

HOOVER, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, American Physical Society, v. 31, p. 1695–1697, Mar 1985.

IZAIRI, R.; KAMBERAJ, H. Comparison study of polar and nonpolar contributions to solvation free energy. *Journal of Chemical Information and Modeling*, v. 57, p. 2539–2553, 2017.

JORGE, M.; GARRIDO, N.; QUEIMADA, A.; ECONOMOU, I.; MACEDO, E. Effect of the integration method on the accuracy and computational efficiency of free energy calculations using thermodynamic integration. *Journal of Chemical Theory and Computation*, v. 6, p. 1018–1027, 2010.

JOVER, J. F.; MüLLER, E. A.; HASLAM, A. J.; GALINDO, A.; JACKSON, G.; TOULHOAT, H.; NIETO-DRAGHI, C. Aspects of asphaltene aggregation obtained from coarse-grained molecular modeling. *Energy and Fuels*, v. 29, n. 2, p. 556–566, 2015.

KAMBERAJ, H.; LOW, R.; NEAL, M. Time reversible and symplectic integrators for molecular dynamics simulations of rigid molecules. *Journal of Chemical Physics*, v. 122, p. 224114, 2005.

KARABENCHEVA-CHRISTOVA, T. *Advances in Protein Chemistry and Structural Biology: Biomolecular Modelling and Simulations*. 1. ed. Cambridge: Academic Press, 2014.

KIRKWOOD, J. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics*, v. 3, p. 300–313, 1935.

KLIMOVICH, P. V.; SHIRTS, M. R.; MOBLEY, D. L. Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design*, v. 29, p. 397–411, 2015.

KMIECIK, S.; GRONT, D.; KOLINSKI, M.; WIETESKA, L.; DAWID, A. E.; KOLINSKI, A. Coarse-grained protein models and their applications. *Chemical Reviews*, v. 116, p. 7898–7936, 2016.

KOGA, N.; TAKADA, S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a go-like model. *Journal of Molecular Biology*, v. 331, p. 171–180, 2001.

KUMAR, S.; BOUZIDA, D.; SWENDSEN, R.; KOLLMAN, P.; ROSENBERG, J. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. the method. *Journal of Computational Chemistry*, v. 13, p. 1011–1021, 1992.

KUZNICKI, T.; MASLIYAH, J. H.; BHATTACHARJEE, S. Aggregation and partitioning of model asphaltenes at toluene - water interfaces: Molecular dynamics simulations. *Energy and Fuels*, v. 0, n. proofing, 0.

LAFITTE, T.; APOSTOLAKOU, A.; AVENDANO, C.; GALINDO, A.; ADJIMAN, C. S.; MULLER, E. A.; JACKSON, G. Accurate statistical associating fluid theory for chain molecules formed from mie segments. *The Journal of Chemical Physics*, v. 139, p. 154504, 2013.

LAFITTE, T.; AVENDAÑO, C.; PAPAIOANNOU, V.; GALINDO, A.; ADJIMAN, C. S.; JACKSON, G.; MÜLLER, E. A. Saft-γ force field for the simulation of molecular fluids: 3. coarse-grained models of benzene and hetero-group models of n-decylbenzene. *Molecular Physics*, v. 110, p. 1189–1203, 2012.

LARS, V.; PERIOLE, X.; TIELEMAN, D. P.; MARRINK, S. J. Improved parameters for the martini coarse-grained protein force field. *Journal of Chemical Theory and Computation*, v. 9, p. 687–697, 2013.

LEE, J. New monte carlo algorithm: Entropic sampling. *Physical Review Letters*, American Physical Society, v. 71, p. 211–214, Jul 1993.

LEVITT, M. A simplified representation of protein conformations for rapid simulation of protein folding. *Journal of Molecular Biology*, v. 104, p. 59–107, 1976.

LEVITT, M.; WARSHE, A. Computer-simulation of protein folding. *Nature*, v. 253, p. 694–698, 1975.

LI, D. D.; GREENFIELD, M. L. High internal energies of proposed asphaltene structures. *Energy & Fuels*, v. 25, n. 8, p. 3698–3705, 2011.

LIU, J. S. *Monte Carlo strategies in Scientific Computing*. 2. ed. New York: Springer, 2002.

LOBANOVA, O.; AVENDAÑO, C.; LAFITTE, T.; MÜLLER, E. A.; JACKSON, G. Saft-γ force field for the simulation of molecular fluids: 4. a single-site coarse-grained model of water applicable over a wide temperature range. *Mol. Phys.*, v. 113, p. 1228–1249, 2015.

LOBANOVA, O.; MEJÍA, A.; JACKSON, G.; MLLER, E. A. Saft-γ force field for the simulation of molecular fluids 6: Binaryand ternary mixtures comprising water, carbon dioxide, and n-alkanes. *The Journal of Chemical Thermodynamics*, v. 93, p. 320–336, 2016.

LYUBARTSEV, A. P.; MARTSINOVSKI, A. A.; SHEVKUNOV, S. V.; VORONTSOV-VELYAMINOV, P. N. New approach to monte carlo calculation of the free energy: Method of expanded ensembles. *Journal of Chemical Physics*, v. 96, p. 1776–1783, 1992.

MARRINK, S. J.; RISSELADA, H. J.; YEFIMOV, S.; TIELEMAN, D. P.; VRIES, A. H. de. The martini force field: Coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B*, v. 111, p. 7812–7824, 2007.

MARRINK, S. J.; TIELEMAN, D. P. Perspective on the martini model. *Chemical Society Reviews*, v. 42, p. 6801– 6822, 2013.

MARTÍNEZ, L.; ANDRADE, R.; BIRGIN, E. G.; MARTÍNEZ, J. M. Packmol: a package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, v. 30, p. 2157–2164, 2009.

MARTYNA, G. J.; TOBIAS, D. J.; KLEIN, M. L. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, v. 101, n. 5, p. 4177–4189, 1994.

MATOS, G. D. R.; KYU, D. Y.; LOEFFLER, H. H.; CHODERA, J. D.; SHIRTS, M. R.; ; MOBLEY, D. L. Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the freesolv database. *Journal Chemical Engineering Data*, v. 62, p. 1559–1569, 2017.

MATUBAYASI, N. Free-energy analysis of protein solvation with all-atom molecular dynamics simulation combined with a theory of solutions. *Current Opinion in Structural Biology*, v. 43, p. 45 – 54, 2017. ISSN 0959-440X. Theory and simulation ● Macromolecular assemblies. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0959440X16301683>.

MEJÍA, A.; HERDES, C.; MÜLLER, E. A. Force fields for coarse-grained molecular simulations from a corresponding states correlation. *Industrial and Chemical Engineering Research*, v. 53, p. 4131–4141, 2014.

Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, v. 21, p. 1087–1092, 1953.

MIE, G. Zur kinetischen theorie der einatomigen korper. *Ann. Phys.*, v. 316, p. 657–697, 1903.

MIKAMI, Y.; LIANG, Y.; MATSUOKA, T.; BOEK, E. S. Molecular dynamics simulations of asphaltenes at the oil–water interface: From nanoaggregation to thin-film formation. *Energy & Fuels*, v. 27, n. 4, p. 1838–1845, 2013.

MOBLEY, D. L.; DUMONT, E.; CHODERA, J. D.; DILL, K. A. Comparison of charge models for fixed-charge force fields: Small-molecule hydration free energies in explicit solvent. *The Journal of Physical Chemistry B*, v. 111, n. 9, p. 2242–2254, 2007.

MOBLEY, D. L.; GILSON, M. K. Predicting binding free energies: Frontiers and benchmarks. *Annual Review of Biophysics*, v. 46, n. 1, p. 531–558, 2017.

MOBLEY, D. L.; GUTHRIE, J. P. Freesolv: a database of experimental and calculated hydration free energies, with input files. *Journal of computer-aided molecular design*, v. 28, n. 7, p. 711—720, July 2014. ISSN 0920-654X.

MOHAMED, N. A.; BRADSHAW, R. T.; ESSEX, J. W. Evaluation of solvation free energies for small molecules with the amoeba polarizable force field. *Journal of Computational Chemistry*, v. 37, p. 2749–2758, 2016.

MORTIMER, S.; MURPHY, R. The vapor pressures of some substances found in coal tar. *Industrial and Engineering Chemistry Research*, v. 14, p. 1140–1142, 1923.

MÜLLER, E. A.; GUBBINS, K. E. Simulation of hard triatomic and tetratomic molecules. a test of associating fluid theory. *Molecular Physics*, v. 80, p. 957–976, 1993.

MÜLLER, E. A.; MEJÍA, A. Extension of the saft-vr mie eos to model homonuclear rings and its parametrization based on the principle of corresponding states. *Langmuir*, -, p. A–L, 2017.

MULLINS, O. C. The modified yen model. *Energy & Fuels*, v. 24, n. 4, p. 2179–2207, 2010.

MURGICH, J. Molecular simulation and the aggregation of the heavy fractions in crude oils. *Molecular Simulation*, v. 29, n. 6-7, p. 451–461, 2003.

PALIWAL, H.; SHIRTS, M. R. 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*, v. 7, p. 4115–4134, 2011.

PANAGIOTOPOULOS, A. Direct determination of phase coexistence properties of fluids using *Molecular Physics*, v. 61, p. 813–826, 1987.

PANTANO, D.; KLEIN, M. L. Characterization of membrane-protein interactions for the leucine transporter from aquifex aeolicus by molecular dynamics calculations. *Journal of Physical Chemistry B*, v. 113, p. 13715–13722, 2009.

PAPAIOANNOU, V.; LAFITTE, T.; AVENDAÑO, C.; ADJIMAN, C. S.; JACKSON, G.; MÜLLER, E. A.; GALINDO, A. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from mie segments. *The Journal of Chemical Physics*, v. 140, p. 054107, 2014.

PARRINELLO, M.; RAHMAN, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, v. 52, n. 12, p. 7182–7190, 1981.

PLIMPTON, S. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, v. 117, p. 1–19, 1995.

POHORILLE, A.; JARZYNSKI, C.; CHIPOT, C. Good practices in freeenergy calculations. *Journal of Physical Chemistry B*, v. 114, p. 10235–10253, 2010.

RAABE, G. *Molecular Simulation Studies on Thermophysical Properties - With Application to Working Fluids. Springer Series Molecular Modeling and Simulation*. [S.l.: s.n.], 2017. ISBN 978-981-10-3544-9 (Print) 978-981-10-3545-6 (Online).

RAMRATTAN, N.; AVENDAÑO, C.; MÜLLER, E.; GALINDO, A. A corresponding-states framework for the description of the mie family of intermolecular potentials. *Molecular Physics*, v. 113, p. 1–16, 2015.

ROUX, B. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, v. 91, n. 1, p. 275 – 282, 1995. ISSN 0010-4655.

ROWLINSON, J. S.; SWINTON, F. L. *Liquid and Liquid Mixtures*. 3. ed. London: Butterworth Scientific, 1982.

ROY, D.; BLINOV, N.; KOVALENKO, A. Predicting accurate solvation free energy in n-octanol using 3d-rism-kh molecular theory of solvation: Making right choices. *Journal of Physical Chemistry B*, v. 121, p. 9268–9273, 2017.

RYCKAERT, J.-P.; CICCOTTI, G.; BERENDSEN, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, v. 23, n. 3, p. 327 – 341, 1977. ISSN 0021-9991.

SHINODA, W.; DEVANE, R.; KLEIN, M. L. Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Molecular Simulation*, v. 33, p. 27–36, 2007.

SHINODA, W.; DEVANE, R.; KLEIN, M. L. Zwitterionic lipid assemblies: molecular dynamics studies of monolayers, bilayers, and vesicles using a new coarse grain force field. *Journal of Physical Chemistry B*, v. 114, p. 6836–6849, 2010.

SHIRTS, M. R.; CHODERA, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *Journal of Chemical Physics*, v. 129, p. 124105, 2008.

SHIRTS, M. R.; PANDE, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.*, v. 122, p. 144107, 2005.

SHIRTS, M. R.; PITERA, J. W.; SWOPE, W. C.; PANDE, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *Journal of Chemical Physics*, v. 119, p. 5740, 2003.

SHYU, C.; YTREBERG, F. M. Reducing the bias and uncertainty of free energy estimates by using regression to fit thermodynamic integration data. *Journal of Computational Chemistry*, v. 6, p. 1018–1027, 2010.

SJöBLOM, J.; ASKE, N.; AUFLEM, I. H.; BRANDAL Øystein; HAVRE, T. E.; SæTHER Øystein; WESTVIK, A.; JOHNSEN, E. E.; KALLEVIK, H. Our current understanding of water-in-crude oil emulsions.: Recent characterization techniques and high pressure performance. *Advances in Colloid and Interface Science*, v. 100-102, n. Supplement C, p. 399 – 473, 2003. ISSN 0001-8686.

SJöBLOM, J.; SIMON, S.; XU, Z. Model molecules mimicking asphaltenes. *Advances in Colloid and Interface Science*, v. 218, n. Supplement C, p. 1 – 16, 2015. ISSN 0001-8686.

SMITH, J.; NESS, H. V.; ABBOT, M. M. *Introdução á Termodinãmica da Engenharia Química*. 2. ed. Rio de Janeiro: LTC, 2007.

SOROUSH, S.; STRAVER, E. J.; RUDOLPH, E. S. J.; PETERS, C. J.; LOOS, T. W. de; ZITHA, P. L.; VAFAIE-SEFTI, M. Phase behavior of the ternary system carbon dioxide+toluene+asphaltene. *Fuel*, v. 137, n. Supplement C, p. 405 – 411, 2014. ISSN 0016-2361.

STERPONE, F.; DERREUMAUX, P.; MELCHIONNA, S. Protein simulations in fluids: Coupling the opep coarse-grained force field with hydrodynamics. *ga*, v. 11, p. 1843–1853, 2015.

STERPONE, F.; MELCHIONNA, S.; TUFFERY, P.; PASQUALI, S.; MOUSSEAU, N.; CRAGNOLINI, T.; CHEBARO, Y.; ST-PIERRE, J.-F.; KALIMERI, M.; BARDUCCI, A.; LAURIN, Y.; TEK, A.; BAADEN, M.; NGUYEN, P. H.; DERREUMAUX, P. The opep protein model: from single molecules, amyloid formation, crowding and hydrodynamics to dna/rna systems. *Chemical Society Reviews*, v. 43, p. 4871– 4893, 2014.

TEKLEBRHAN, R. B.; GE, L.; BHATTACHARJEE, S.; XU, Z.; SJöBLOM, J. Probing structure–nanoaggregation relations of polyaromatic surfactants: A molecular dynamics simulation and dynamic light scattering study. *The Journal of Physical Chemistry B*, v. 116, n. 20, p. 5907–5918, 2012. PMID: 22512276.

TEKLEBRHAN, R. B.; GE, L.; BHATTACHARJEE, S.; XU, Z.; SJöBLOM, J. Initial partition and aggregation of uncharged polyaromatic molecules at the oil–water interface: A molecular dynamics simulation study. *The Journal of Physical Chemistry B*, v. 118, n. 4, p. 1040–1051, 2014. PMID: 24397444.

TORRIE, G.; VALLEAU, J. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, v. 23, n. 2, p. 187 – 199, 1977. ISSN 0021-9991.

TREBST, S.; HUSE, D. A.; TROYER, M. Optimizing the ensemble for equilibration in broad-histogram monte carlo simulations. *Physical Review E*, American Physical Society, v. 70, p. 046701, Oct 2004.

TUCKERMAN, M. *Statistical Mechanics: Theory And Molecular Simulation*. 1nd. ed. New York, NY, USA: Oxford University Press, 2010.

VERLET, L. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical Review*, v. 159, p. 98–103, 1967.

WINGER, M.; TRZESNIAK, D.; BARON, R.; GUNSTEREN, W. van. On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models. *Physical Chemistry Chemical Physics*, v. 11, p. 1934, 2009.

YTREBERG, F. M.; SWENDSEN, R. H.; ZUCKERMAN, D. M. Comparison of free energy methods for molecular systems. *Journal of Chemical Physics*, v. 125, p. 184114, 2006.

ZWANZIG, R. W. High-temperature equation of state by a perturbation method. i. nonpolar gases. *The Journal of Chemical Physics*, v. 22, p. 1420, 1954.

ZWANZIG, R. W. High-temperature equation of state by a perturbation method. ii. polar gases. *The Journal of Chemical Physics*, v. 23, p. 1915, 1955.

APPENDIX A – Optimized values of λ and η

Table 9 – Optimized values of λ and η for the solutes in hexane

pyrene		phenanthrene	
λ	η	λ	η
0.000	0.000	0.000	0.000
0.076	4.234	0.090	1.981
0.107	5.620	0.132	3.461
0.132	6.499	0.161	4.494
0.152	6.690	0.185	5.185
0.170	6.643	0.205	5.552
0.189	6.461	0.224	5.725
0.213	6.091	0.244	5.722
0.242	5.566	0.268	5.523
0.280	4.729	0.305	4.975
0.355	2.853	0.372	3.576
0.483	-0.778	0.500	0.297
0.678	-6.947	0.560	-1.390
0.788	-10.631	0.722	-6.309
1.000	-18.141	1.000	-15.448

Table 10 – Optimized values of λ and η for the solutes in 1-octanol

propane		anthracene		phenanthrene	
λ	η	λ	η	λ	η
0.000	0.000	0.000	0.000	0.000	0.000
0.027	3.126	0.078	3.932	0.049	2.578
0.050	5.109	0.111	6.178	0.091	5.663
0.073	6.093	0.130	7.426	0.125	8.575
0.095	6.570	0.143	8.201	0.144	10.069
0.117	6.826	0.154	8.717	0.157	10.978
0.142	6.956	0.164	9.085	0.169	11.599
0.174	6.969	0.174	9.357	0.180	12.040
0.215	6.847	0.184	9.556	0.192	12.340
0.269	6.554	0.197	9.676	0.206	12.499
0.337	6.050	0.214	9.681	0.225	12.478
0.427	5.228	0.238	9.490	0.253	12.161
0.545	3.955	0.274	8.958	0.298	11.280
0.720	1.843	0.326	7.906	0.371	9.406
1.000	-1.903	0.399	6.088	0.484	5.891
		0.515	2.777	0.664	-0.516
		0.695	-2.960	0.802	-5.908
		1.000	-13.768	1.000	-14.073

Table 11 – Optimized values of λ and η for the solutes in toluene

pyrene		anthracene		phenanthrene	
λ	η	λ	η	λ	η
0.000	0.000	0.000	0.000	0.000	0.000
0.090	2.563	0.119	0.218	0.136	0.726
0.130	4.338	0.174	1.210	0.191	2.307
0.154	5.439	0.209	2.052	0.223	3.430
0.172	6.181	0.236	2.664	0.246	4.233
0.188	6.670	0.261	3.122	0.264	4.780
0.204	6.986	0.283	3.378	0.281	5.149
0.222	7.121	0.306	3.449	0.299	5.354
0.244	7.025	0.332	3.311	0.318	5.389
0.278	6.520	0.360	2.936	0.340	5.222
0.340	5.010	0.399	2.209	0.372	4.717
0.462	1.247	0.466	0.567	0.425	3.440
0.616	-4.283	0.564	-2.211	0.524	0.444
0.788	-11.032	0.715	-6.983	0.701	-5.814
1.000	-19.814	1.000	-16.923	1.000	-17.803

Table 12 – Optimized values of λ and η for the phenanthrene in different mass fractions of CO_2 in toluene

0.087		0.119		0.169		0.289	
λ	η	λ	η	λ	η	λ	η
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.128	0.604	0.128	0.732	0.064	0.883	0.066	0.806
0.184	2.067	0.186	2.223	0.108	0.764	0.111	0.760
0.217	3.164	0.219	3.319	0.175	1.969	0.172	1.983
0.240	3.940	0.244	4.098	0.214	3.156	0.204	2.967
0.260	4.472	0.267	4.704	0.240	3.974	0.227	3.627
0.277	4.823	0.289	5.031	0.258	4.457	0.245	4.082
0.295	5.035	0.313	5.084	0.273	4.750	0.262	4.395
0.318	5.059	0.339	4.950	0.287	4.921	0.279	4.583
0.347	4.762	0.373	4.371	0.305	4.962	0.299	4.621
0.397	3.753	0.425	3.055	0.326	4.885	0.325	4.423
0.491	1.031	0.488	1.196	0.361	4.401	0.365	3.739
0.670	-5.148	0.525	-0.027	0.419	2.990	0.428	2.198
0.791	-9.713	0.730	-7.185	0.527	-0.299	0.530	-0.842
1.000	-18.098	1.000	-17.769	0.697	-6.180	0.701	-6.763
				1.000	-17.998	1.000	-18.163

Table 13 – Optimized values of λ and η for the solutes in water

propane		benzene		toluene		phenanthrene	
λ	η	λ	η	λ	η	λ	η
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.107	2.673	0.193	-0.295	0.177	0.182	0.142	-2.462
0.157	4.703	0.279	1.468	0.262	2.432	0.256	0.597
0.186	6.047	0.324	2.931	0.307	4.244	0.319	4.504
0.210	7.148	0.357	4.168	0.336	5.552	0.358	7.762
0.230	8.017	0.381	5.091	0.360	6.696	0.384	10.104
0.250	8.883	0.405	5.891	0.380	7.558	0.407	12.185
0.272	9.291	0.427	6.443	0.400	8.233	0.427	13.607
0.294	9.700	0.449	6.770	0.422	8.678	0.446	14.490
0.328	9.900	0.476	6.900	0.443	8.859	0.469	14.834
0.381	9.930	0.506	6.805	0.473	8.810	0.494	14.667
0.484	9.463	0.555	6.392	0.514	8.452	0.533	13.832
0.623	8.195	0.653	5.109	0.606	7.148	0.620	11.069
0.781	6.378	0.810	2.421	0.755	4.273	0.806	3.279
1.000	3.333	1.000	-1.480	1.000	-1.547	1.000	-6.122

APPENDIX B – Overlapping Matrices

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.75	.19	.05	.01											
1	.28	.33	.19	.09	.05	.03	.02	.01							
2	.08	.21	.19	.15	.12	.10	.07	.04	.02	.01					
3	.01	.08	.11	.13	.14	.15	.15	.11	.07	.03	.01				
4		.03	.05	.08	.12	.16	.18	.16	.12	.07	.03	.01			
5		.01	.02	.05	.09	.14	.18	.18	.15	.10	.05	.01			
6			.01	.03	.07	.12	.17	.19	.18	.13	.07	.02			
7				.02	.04	.09	.14	.19	.20	.16	.11	.04	.01		
8			.01	.03	.06	.12	.17	.20	.19	.15	.06	.01			
9				.02	.04	.08	.14	.19	.20	.19	.09	.02	.01		
10					.01	.02	.05	.09	.14	.19	.25	.17	.06	.03	
11						.01	.02	.04	.07	.11	.21	.27	.17	.09	.01
12								.01	.02	.03	.08	.18	.32	.29	.07
13									.01	.01	.03	.10	.30	.38	.16
14											.01	.03	.16	.36	.44

Figure B.0.1 – Overlapping matrix for hexane+pyrene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.54	.29	.11	.04	.01										
1	.29	.31	.19	.11	.05	.03	.01	.01							
2	.14	.25	.20	.15	.10	.07	.04	.02	.01						
3	.06	.15	.17	.16	.14	.11	.09	.06	.04	.02					
4	.02	.07	.11	.13	.14	.13	.13	.11	.09	.05	.02				
5	.01	.03	.06	.09	.12	.13	.14	.14	.13	.10	.04	.01			
6		.01	.03	.05	.08	.11	.13	.15	.17	.16	.07	.02	.01		
7			.01	.03	.05	.08	.11	.15	.19	.21	.12	.03	.02		
8				.01	.03	.05	.08	.12	.19	.25	.17	.06	.03		
9					.01	.02	.05	.09	.16	.26	.24	.10	.06	.01	
10						.01	.02	.04	.09	.20	.27	.19	.14	.03	
11								.01	.03	.08	.17	.28	.28	.13	.02
12									.01	.04	.12	.27	.31	.20	.03
13										.01	.04	.17	.27	.37	.14
14											.01	.05	.09	.31	.55

Figure B.0.2 – Overlapping matrix for hexane+phenanthrene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.66	.26	.06	.01											
1	.40	.30	.14	.08	.04	.02	.01								
2	.10	.14	.17	.22	.15	.09	.06	.03	.02	.01					
3	.01	.03	.09	.20	.20	.16	.12	.08	.05	.03	.01	.01			
4		.01	.05	.14	.18	.17	.15	.12	.08	.05	.03	.01	.01		
5			.02	.09	.15	.16	.16	.14	.11	.08	.04	.02	.01		
6			.01	.06	.11	.14	.15	.16	.13	.11	.06	.04	.02	.01	
7			.01	.04	.07	.11	.13	.16	.15	.14	.09	.06	.03	.01	
8				.02	.05	.07	.10	.14	.16	.17	.13	.09	.05	.02	
9				.01	.02	.04	.07	.11	.14	.18	.16	.14	.09	.03	.01
10					.01	.02	.04	.07	.10	.16	.18	.18	.15	.06	.01
11					.01	.01	.02	.04	.06	.11	.15	.21	.24	.13	.02
12							.01	.02	.03	.06	.10	.19	.30	.24	.05
13								.01	.01	.02	.05	.11	.26	.38	.16
14										.01	.02	.04	.13	.34	.47

Figure B.0.3 – Overlapping matrix for 1-octanol+propane.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	.57	.28	.09	.03	.01	.01												
1	.28	.32	.17	.09	.06	.03	.02	.01	.01									
2	.13	.25	.19	.12	.09	.07	.05	.04	.03	.02	.01							
3	.06	.17	.17	.13	.11	.09	.07	.06	.05	.04	.03	.01						
4	.03	.11	.13	.12	.11	.10	.09	.08	.08	.06	.05	.03	.01					
5	.01	.07	.10	.10	.10	.10	.09	.09	.09	.09	.08	.05	.02	.01				
6	.01	.04	.07	.07	.08	.09	.09	.10	.11	.11	.10	.07	.04	.02				
7		.02	.04	.05	.07	.08	.08	.10	.11	.12	.13	.10	.06	.03	.01			
8		.01	.02	.03	.05	.06	.07	.09	.11	.13	.15	.13	.09	.04	.01			
9			.01	.02	.03	.04	.05	.07	.10	.13	.17	.16	.12	.06	.02	.01		
10				.01	.02	.02	.04	.05	.08	.11	.17	.19	.16	.10	.04	.01		
11					.01	.01	.02	.03	.05	.09	.14	.19	.20	.15	.07	.02		
12							.01	.01	.03	.05	.10	.17	.22	.21	.13	.05	.01	
13								.01	.01	.03	.06	.11	.19	.25	.22	.11	.02	
14										.01	.02	.05	.12	.21	.29	.22	.06	
15											.01	.02	.04	.11	.24	.36	.21	.01
16													.01	.03	.09	.28	.49	.10
17															.01	.06	.33	.59

Figure B.0.4 – Overlapping matrix for 1-octanol+anthracene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	.51	.34	.11	.03	.01													
1	.37	.36	.17	.06	.02	.01												
2	.22	.32	.23	.12	.06	.03	.02	.01										
3	.10	.20	.22	.18	.11	.07	.05	.03	.02	.01	.01							
4	.04	.12	.16	.18	.13	.11	.08	.07	.05	.03	.02	.01						
5	.02	.06	.10	.14	.13	.12	.10	.10	.09	.07	.05	.02	.01					
6	.01	.03	.06	.10	.11	.11	.11	.11	.11	.10	.08	.05	.02					
7		.01	.03	.06	.07	.09	.09	.11	.13	.13	.13	.09	.05	.01				
8			.01	.03	.04	.06	.07	.10	.13	.14	.17	.14	.09	.02				
9				.01	.02	.03	.05	.07	.11	.14	.19	.19	.15	.04				
10						.01	.01	.02	.04	.07	.11	.19	.23	.22	.07	.01		
11							.01	.02	.04	.07	.15	.23	.30	.14	.02			
12								.01	.02	.04	.09	.18	.35	.24	.06	.01		
13										.01	.03	.10	.28	.34	.16	.06	.01	
14											.01	.03	.12	.27	.28	.23	.06	
15													.02	.06	.14	.44	.31	.03
16														.01	.04	.33	.49	.13
17															.01	.13	.45	.41

Figure B.0.5 – Overlapping matrix for 1-octanol+phenanthrene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.53	.30	.11	.04	.01	.01									
1	.21	.34	.22	.12	.06	.03	.01	.01							
2	.09	.25	.24	.16	.11	.07	.04	.02	.01						
3	.04	.17	.20	.17	.14	.11	.08	.05	.03	.01					
4	.02	.10	.15	.15	.14	.13	.11	.09	.07	.04	.01				
5	.01	.05	.10	.12	.13	.13	.13	.12	.11	.08	.02				
6		.02	.05	.08	.10	.12	.13	.14	.15	.14	.05	.01			
7		.01	.02	.04	.06	.09	.11	.14	.19	.21	.11	.02			
8			.01	.02	.03	.05	.08	.12	.19	.27	.18	.05	.01		
9				.01	.02	.04	.07	.15	.29	.28	.10	.02			
10							.01	.03	.08	.22	.35	.23	.06	.01	
11								.01	.02	.08	.22	.38	.22	.07	.01
12										.02	.07	.25	.37	.25	.05
13											.01	.07	.26	.45	.20
14												.01	.08	.36	.55

Figure B.0.6 – Overlapping matrix for toluene+pyrene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.48	.29	.14	.05	.02	.01									
1	.24	.30	.23	.12	.06	.03	.01								
2	.12	.24	.24	.17	.11	.06	.03	.02	.01						
3	.06	.16	.21	.19	.14	.09	.07	.04	.02	.01					
4	.03	.10	.16	.17	.15	.12	.10	.07	.05	.03	.02				
5	.01	.05	.10	.13	.14	.13	.13	.11	.09	.06	.04	.01			
6		.02	.05	.08	.10	.11	.13	.13	.12	.11	.09	.03	.01		
7		.01	.02	.04	.06	.08	.11	.13	.15	.15	.15	.07	.02		
8			.01	.02	.03	.05	.08	.11	.14	.18	.21	.12	.05	.01	
9				.01	.01	.02	.05	.08	.12	.18	.26	.18	.08	.02	
10						.01	.02	.04	.08	.15	.27	.24	.14	.04	
11							.01	.02	.04	.09	.21	.28	.25	.10	.01
12								.01	.01	.04	.12	.23	.33	.23	.03
13										.01	.04	.11	.28	.41	.14
14												.02	.08	.29	.60

Figure B.0.7 – Overlapping matrix for toluene+anthracene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.52	.31	.11	.04	.01										
1	.20	.34	.23	.12	.06	.03	.01	.01							
2	.07	.24	.25	.17	.11	.07	.04	.02	.01						
3	.03	.15	.21	.18	.14	.11	.07	.05	.03	.01	.01				
4	.01	.09	.16	.16	.15	.13	.11	.08	.06	.03	.02				
5		.05	.11	.13	.14	.13	.13	.11	.09	.06	.04	.01			
6		.02	.06	.09	.11	.12	.13	.13	.12	.10	.08	.03	.01		
7		.01	.03	.05	.08	.10	.12	.13	.14	.14	.13	.07	.02		
8			.01	.03	.04	.07	.09	.11	.14	.16	.18	.13	.04		
9				.01	.02	.04	.06	.08	.12	.16	.22	.20	.07	.01	
10					.01	.01	.03	.05	.08	.13	.24	.29	.14	.02	
11						.01	.02	.04	.08	.19	.34	.26	.07		
12									.01	.03	.09	.25	.39	.22	.01
13										.02	.07	.25	.53	.13	
14													.03	.27	.69

Figure B.0.8 – Overlapping matrix for toluene+phenanthrene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.54	.30	.11	.04	.01										
1	.23	.34	.23	.11	.05	.03	.01	.01							
2	.09	.25	.25	.17	.10	.06	.04	.02	.01						
3	.04	.16	.22	.18	.13	.10	.08	.05	.03	.01					
4	.02	.10	.16	.17	.14	.12	.11	.08	.05	.03	.01				
5	.01	.05	.11	.13	.13	.13	.13	.12	.09	.06	.03				
6		.02	.06	.10	.11	.12	.14	.14	.13	.11	.06	.01			
7		.01	.03	.06	.07	.10	.13	.14	.15	.16	.11	.03			
8			.01	.02	.04	.06	.09	.13	.16	.21	.20	.07	.01		
9				.01	.01	.03	.05	.08	.13	.22	.29	.14	.02		
10						.01	.02	.03	.07	.17	.35	.27	.06	.02	
11							.01	.02	.07	.24	.39	.19	.07	.01	
12									.01	.05	.19	.37	.32	.07	
13										.01	.07	.29	.43	.20	
14											.01	.11	.34	.54	

Figure B.0.9 – Overlapping matrix for toluene+CO_2(0.087)+phenanthrene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.55	.31	.10	.03	.01										
1	.24	.36	.22	.10	.05	.02	.01								
2	.09	.27	.26	.17	.10	.06	.03	.01	.01						
3	.04	.17	.22	.18	.14	.10	.07	.04	.02	.01					
4	.01	.09	.16	.17	.15	.14	.12	.08	.05	.03	.01				
5		.04	.09	.12	.13	.15	.15	.12	.10	.07	.03	.01			
6		.01	.04	.07	.10	.12	.15	.15	.15	.12	.06	.02	.01		
7			.01	.03	.05	.08	.12	.14	.17	.19	.12	.05	.03		
8				.01	.02	.04	.08	.11	.17	.22	.19	.09	.05	.01	
9					.01	.02	.03	.06	.12	.22	.26	.16	.10	.01	
10							.01	.02	.06	.15	.27	.24	.19	.04	
11								.01	.03	.09	.22	.28	.26	.11	.01
12								.01	.02	.06	.18	.28	.28	.16	.01
13										.01	.05	.13	.18	.48	.14
14											.01	.02	.03	.32	.62

Figure B.0.10 – Overlapping matrix for toluene+CO_2(0.119)+phenanthrene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	.32	.36	.24	.06	.02											
1	.24	.34	.27	.09	.03	.01	.01									
2	.18	.30	.28	.13	.06	.02	.01	.01								
3	.08	.19	.24	.19	.12	.06	.04	.03	.02	.01						
4	.03	.10	.16	.19	.16	.10	.08	.06	.05	.03	.02	.01				
5	.01	.05	.09	.14	.15	.12	.11	.09	.08	.07	.05	.02	.01			
6	.01	.03	.05	.10	.12	.11	.11	.11	.11	.10	.08	.05	.02			
7		.01	.03	.07	.09	.09	.11	.11	.12	.12	.11	.08	.04	.01		
8		.01	.01	.04	.06	.07	.09	.11	.12	.13	.14	.12	.07	.01		
9				.02	.03	.05	.07	.09	.11	.14	.16	.17	.13	.03		
10				.01	.01	.02	.04	.06	.08	.12	.17	.21	.21	.06	.01	
11						.01	.02	.03	.05	.08	.14	.23	.31	.12	.02	
12							.01	.02	.03	.07	.18	.37	.24	.07		
13										.01	.02	.07	.26	.37	.25	.01
14												.01	.07	.24	.55	.12
15													.01	.04	.30	.66

Figure B.0.11 – Overlapping matrix for toluene+CO_2(0.169)+phenanthrene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	.33	.33	.22	.07	.02	.01										
1	.25	.31	.25	.11	.05	.02	.01									
2	.17	.27	.26	.15	.08	.04	.02	.01								
3	.08	.17	.22	.20	.13	.09	.05	.03	.02	.01						
4	.04	.10	.15	.19	.16	.12	.09	.07	.04	.03	.01	.01				
5	.02	.06	.10	.15	.15	.13	.11	.10	.07	.05	.03	.02				
6	.01	.03	.06	.11	.13	.13	.12	.12	.09	.08	.06	.03	.01			
7		.01	.03	.07	.10	.11	.12	.12	.11	.11	.10	.07	.03	.01		
8		.01	.01	.04	.07	.09	.10	.12	.12	.14	.13	.11	.05	.01		
9				.02	.04	.05	.07	.09	.11	.14	.17	.17	.10	.03		
10				.01	.01	.02	.04	.06	.08	.12	.17	.23	.18	.06	.01	
11						.01	.01	.02	.04	.08	.14	.25	.28	.14	.03	
12							.01	.01	.03	.07	.19	.33	.27	.08		
13										.01	.02	.09	.25	.38	.23	.02
14												.02	.08	.27	.50	.12
15													.01	.05	.30	.65

Figure B.0.12 – Overlapping matrix for toluene+CO_2(0.289)+phenanthrene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.45	.28	.16	.08	.02	.01									
1	.24	.26	.24	.16	.06	.02	.01	.01							
2	.13	.21	.26	.21	.09	.05	.02	.02	.01						
3	.07	.16	.24	.23	.12	.06	.04	.03	.03	.01					
4	.04	.11	.21	.22	.13	.08	.05	.05	.06	.04	.01				
5	.02	.07	.16	.19	.12	.08	.06	.08	.10	.08	.03	.01			
6	.01	.04	.09	.13	.09	.07	.07	.10	.16	.16	.07	.02			
7		.01	.04	.06	.05	.05	.06	.10	.20	.24	.14	.04	.01		
8			.01	.02	.03	.03	.04	.09	.20	.29	.19	.07	.02		
9				.01	.01	.01	.02	.06	.17	.29	.25	.13	.03	.01	
10							.01	.03	.11	.24	.28	.21	.08	.02	
11								.01	.04	.12	.21	.31	.22	.09	.01
12									.01	.03	.08	.22	.33	.26	.07
13										.01	.02	.08	.25	.40	.25
14												.02	.08	.31	.59

Figure B.0.13 – Overlapping matrix for water+propane.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.44	.33	.16	.05	.02	.01									
1	.19	.32	.27	.13	.05	.03	.01								
2	.08	.23	.30	.19	.10	.06	.03	.01	.01						
3	.04	.16	.27	.20	.13	.09	.05	.03	.02	.01					
4	.02	.10	.21	.19	.14	.11	.08	.06	.04	.03	.02				
5	.01	.06	.15	.16	.13	.12	.11	.08	.07	.06	.04	.01			
6		.03	.08	.11	.11	.11	.11	.10	.10	.10	.09	.04	.01		
7		.01	.04	.06	.07	.08	.10	.10	.12	.14	.16	.09	.02		
8			.01	.03	.04	.05	.07	.09	.11	.17	.23	.15	.05	.01	
9				.01	.01	.02	.04	.06	.09	.16	.27	.22	.09	.02	
10						.01	.02	.03	.06	.14	.27	.27	.14	.04	.01
11							.01	.01	.03	.09	.23	.29	.23	.08	.01
12									.01	.04	.12	.22	.31	.22	.06
13										.01	.03	.08	.22	.39	.28
14												.02	.07	.29	.62

Figure B.0.14 – Overlapping matrix for water+benzene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.46	.32	.15	.04	.01										
1	.19	.31	.28	.13	.06	.02	.01								
2	.07	.21	.32	.20	.11	.05	.02	.01							
3	.03	.13	.28	.23	.15	.08	.05	.03	.01	.01					
4	.01	.08	.23	.22	.17	.11	.07	.05	.03	.02	.01				
5	.01	.05	.16	.19	.16	.12	.09	.07	.05	.04	.03	.01			
6		.02	.10	.14	.14	.12	.11	.09	.08	.08	.07	.03	.01		
7		.01	.05	.08	.09	.10	.10	.10	.11	.12	.14	.08	.02		
8			.01	.03	.05	.06	.07	.09	.11	.15	.22	.16	.04	.01	
9				.01	.02	.03	.04	.06	.09	.14	.27	.24	.08	.01	
10						.01	.02	.03	.06	.12	.27	.31	.15	.03	
11							.01	.01	.03	.08	.23	.33	.23	.07	.01
12									.01	.03	.12	.25	.34	.22	.03
13										.01	.03	.08	.24	.45	.19
14												.01	.04	.25	.69

Figure B.0.15 – Overlapping matrix for water+toluene.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	.50	.37	.10	.02											
1	.29	.38	.23	.08	.02										
2	.08	.21	.33	.26	.10	.03	.01								
3	.02	.07	.24	.34	.22	.09	.02	.01							
4		.02	.13	.31	.29	.16	.07	.02	.01						
5		.01	.06	.22	.29	.20	.12	.06	.03	.01					
6			.02	.12	.21	.21	.17	.12	.08	.04	.02	.01			
7			.01	.04	.11	.14	.16	.16	.15	.11	.09	.04			
8				.01	.03	.06	.09	.13	.17	.17	.19	.14	.02		
9						.01	.03	.06	.11	.16	.26	.30	.06		
10							.01	.02	.05	.11	.25	.43	.13	.01	
11									.01	.05	.17	.48	.27	.02	
12										.01	.05	.29	.48	.15	.01
13												.03	.17	.58	.22
14													.01	.24	.75

Figure B.0.16 – Overlapping matrix for water+phenanthrene.

APPENDIX C − Work Published in Scientific Conference

de Termodinâmica Aplicada
e V Escola de Termodinâmica

Porto Alegre-RS
23 a 27 de Outubro de 2017

ar este cabeçalho (número de linhas e formatação). Será substituído no pós-processamento.

PARAMETRIZAÇÃO E AVALIAÇÃO DO CAMPO DE FORÇA SAFT-γ CG PARA SIMULAÇÃO MOLECULAR DE FENANTRENO

Isabela Q. Matos[1], Charlles R. A. Abreu[1*]

[1] Escola de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil

*Autor para correspondencia: abreu@eq.ufrj.br

O fenantreno é um hidrocarboneto policíclico aromático que pode ser usado para estudar moléculas mais complexas como os asfaltenos. Portanto, é interessante desenvolver campos de forças para simulação molecular eficientes computacionalmente que consigam descrever esses tipos de molécula. O campo de força SAFT-γ CG, escolhido para o estudo, é do tipo *coarse-grained* e foi desenvolvido com base na equação de estado SAFT-VR Mie. Essa equação modela uma substância subdividindo sua estrutura em segmentos que interagem através de potencial de Mie. Os parâmetros desse campo de força para o fenantreno foram estimados e avaliados em relação à determinação de propriedades de equilíbrio líquido-vapor (ELV). A estratégia de estimação possui duas etapas. A primeira consiste em minimizar o erro quadrático entre a pressão de vapor calculada com a equação de estado (EdE) SAFT-VR Mie e os valores experimentais. Os parâmetros dessa minimização foram, então, usados no cálculo de ELV com o método de *Gibbs Ensemble Monte Carlo* com volume total constante (GEMC-NVT). A segunda etapa consistiu em estimar novamente os parâmetros através do método de mínimos quadrados envolvendo a pressão de vapor e a densidade de líquido saturado calculadas por Monte Carlo (MC) e as calculadas com a EdE SAFT-VR Mie. Essa segunda parte é necessária devido a aproximações teóricas que geram diferenças entre os resultados da simulação molecular e da EdE. Esse campo de força apresentou uma boa descrição da pressão de vapor do fenantreno com método de GEMC quando comparado com os dados experimentais. As densidades de líquido e vapor saturados e as propriedades críticas apresentaram resultados similares aos obtidos com o campo de força atomístico TraPPE-EH, sendo que o SAFT-γ CG exige menor esforço computacional. Para continuação do trabalho, pretende-se usar esses parâmetros para calcular a energia de solvatação do fenantreno, em diluição infinita, em tolueno e em solução tolueno + CO_2 com dinâmica molecular.

Palavras-chave: campo de força, SAFT-γ CG, simulação molecular, fenantreno.

Introdução

Os campos de força do tipo *coarse-grained* parametrizados a partir de propriedades experimentais são uma alternativa aos campo de força desenvolvidos com cálculos *ab initio* quando a escala das simulações moleculares precisa ser aumentada. O método *coarse-grained* consiste, basicamente, em dividir a substância em pseudo-átomos representativos de grupos de átomos. Normalmente, a parametrização do campo de força é feita a partir de informações obtidas em uma escala mais detalhada. Quando as informações utilizadas provêm do comportamento em uma escala maior, a estratégia para o modelo *coarse-grained* é dita ser do tipo *top down*. Um dos campos de força que possuem essa estratégia é o SAFT-γ CG (Avendaño *et al*., 2011), o qual foi desenvolvido com base na equação de estado (EdE) SAFT-VR Mie (Lafitte et al., 2013). Esse modelo usa o potencial de Mie para descrever moléculas formadas por segmentos conectados. A vantagem de usar essa equação como base

para o campo de força é a sua capacidade de descrever bem as propriedades do fluido, incluindo derivadas de segunda ordem, de uma variedade de sistemas (Lafitte *et al*., 2013). Com base nessas ideias, o presente trabalho teve como objetivo parametrizar o campo de força SAFT-γ CG para o fenantreno, cuja estrutura é apresentada na Figura 1, utilizando dados de equilíbrio líquido-vapor e considerando que ele é formado por três segmentos esféricos. O fenantreno foi escolhido como forma de testar a capacidade de representação desse campo de força simplificado, já que sua parametrização para moléculas menores tem sido bem sucedida (Lafitte *et al*., 2012). Outra razão para o estudo do fenantreno é o fato de ele servir como modelo simplificado para moléculas mais complexas, tais como os asfaltenos, já que possui uma estrutura de anéis policondensados e é solúvel em tolueno.
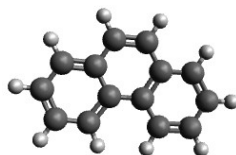

Figura1. Estrutura molecular do fenantreno.

Metodologia

Estimação com a Equação de Estado SAFT-VR Mie

A EdE SAFT-VR Mie descreve moléculas formadas por segmentos conectados e que interagem através do potencial de Mie, dado por:

$$U_{Mie}(r) = \varepsilon \frac{\lambda_r}{\lambda_r - \lambda_a} \left(\frac{\lambda_r}{\lambda_a} \right)^{\lambda_a / \lambda_r - \lambda_a} \left[\left(\frac{\sigma}{r} \right)^{\lambda_r} - \left(\frac{\sigma}{r} \right)^{\lambda_a} \right] \tag{1}$$

em que λ_r é o expoente repulsivo, λ_a é o expoente atrativo, σ corresponde à distância entre os centros dos segmentos e ε é o parâmetro de energia do segmento. A energia livre de Helmholtz específica para a EdE SAFT-VR Mie para um fluido não-associativo é definida como

$$a = a^{IDEAL} + a^{MONO} + a^{CHAIN} \quad , \tag{2}$$

em que a^{IDEAL} é a contribuição de gás ideal, a^{MONO} é contribuição dos monômeros (segmentos desconectados) e a^{CHAIN} é a contribuição relativa à formação das cadeias de segmentos. Para cadeias que possuem m_s segmentos do mesmo tipo tangencialmente ligados, a contribuição da cadeia é

$$a^{CHAIN} = -(m_s - 1) \ln g^{Mie}(\sigma), \tag{3}$$

onde $\ln g^{Mie}(\sigma)$ é o valor da função de distribuição radial de pares para o fluido monomérico de referência (fluido de Mie). Devido à estrutura aromática do fenantreno, usou-se a seguinte expressão para anéis formados por m_s segmentos no lugar da contribuição de cadeia:

$$a^{RING} = -m_s \ln g^{Mie}(\sigma).$$
(4)

Essa substituição é feita porque a diferença entre uma molécula formada por cadeias e uma formada por anéis, tendo ambas o mesmo número de segmentos, é que a última possui uma ligação a mais (Lafitte *et al.*, 2012). Para o equacionamento completo da EdE, o leitor é referido ao artigo de Lafitte *et al.* (2013). Seguindo essa formulação, os parâmetros ε, λ_r e σ foram estimados e os parâmetros m_s e λ_a foram fixados em 3 e 6, respectivamente. A razão para fixar o número de segmentos em três deve-se à própria estrutura do fenantreno, que consiste em três anéis aromáticos condensados. Já o parâmetro atrativo foi fixado no valor London para facilitar a estimação, já que é comprovada a alta correlação entre os parâmetros repulsivo e atrativo do potencial (Ramrattan *et al.*, 2015). A minimização foi feita através do método PSO (*Particle Swarm Optimization*) e com apenas dados de pressão de vapor, por uma questão de indisponibilidade de dados experimentais de densidade do fenantreno como líquido saturado. A função objetivo possuiu a seguinte forma:

$$F_{exp}\left(\sigma^{SAFT}, \varepsilon^{SAFT}, \lambda_r^{SAFT}\right) = \sum_{i=1}^{N_P} \left(\frac{P_v^{exp}(T_i) - P_v^{SAFT}(T_i)}{P_v^{exp}(T_i)}\right)^2$$
(5)

em que N_P corresponde ao número de pontos experimentais, P_v^{exp} aos pontos de pressão de vapor experimental (Mortimer e Murphy, 1923) e P_v^{SAFT} à pressão de vapor calculada com a EdE SAFT-VR Mie. Esse cálculo de equilíbrio foi feito usando como base a rotina do ponto de bolha proposta por Smith *et al.* (2007).

Cálculo com o Método "Gibbs Ensemble Monte Carlo" (GEMC)

Os parâmetros estimados com equação SAFT-VR Mie foram usados para realizar simulações no GEMC (Panagiotopoulos, 1987) com o simulador CASSANDRA (Shah e Maginn, 2011). O método de GEMC foi desenvolvido com o intuito de estudar a coexistência entre fases através da simulação simultânea de duas caixas com condições de contorno periódicas e que trocam moléculas, energia e volume entre si, mas de forma a manter o volume total constante. O equilíbrio entre elas é obtido através de passos de Monte Carlo (MC) que incluem o deslocamento aleatório das moléculas, mudança de volume e transferências aleatórias de moléculas entre as caixas. Para sistemas com apenas um componente, os cálculos são realizados mantendo-se o volume e o número de partículas total constantes (NVT), mas de maneira a permitir a variação de volume ($V = V^{liq} + V^{vap}$) e partículas ($N = N^{liq} + N^{vap}$) dentro de cada caixa. Essas simulações no GEMC-NVT foram feitas inserindo-se aleatoriamente 400 moléculas na caixa líquida e 100 moléculas na caixa vapor. As densidades iniciais das caixas foram escolhidas ajustando-as às densidades obtidas com a EdE SAFT-VR Mie, para evitar que todas as moléculas migrassem pra uma única caixa ao longo da simulação. A simulação consistiu em, no mínimo, 10000 ciclos de equilibração e 100000 ciclos de produção, sendo que cada ciclo de MC corresponde a 1000 tentativas de rotação, 1000 de translação, 100 de inserção, 100 de exclusão e 10 de variação de volume. A distância de corte usada foi igual a quatro vezes o valor do diâmetro do segmento estimado e as interações de van der Waals foram calculadas através do potencial Mie com correção de longa distância (*tail correction*). As propriedades densidade de vapor (ρ_{vap}), densidade de líquido (ρ_{liq}) e pressão (P_v^{sim}) foram amostradas a cada 100 ciclos de MC e essas amostragens foram divididas em cinco blocos para os cálculos da média e do desvio padrão. Os resultados

obtidos nessas simulações foram usados para estimar coeficientes de correção para os parâmetros conformacionais do campo de força (c_σ e c_ε), que são relacionados aos parâmetros provenientes da EdE SAFT-VR Mie através de parâmetros oriundos de ajuste,

$$\sigma^{ajuste} = c_\sigma \sigma^{SAFT} \tag{6}$$

$$\varepsilon^{ajuste} = c_\varepsilon \varepsilon^{SAFT} . \tag{7}$$

Os parâmetros de ajuste são necessários porque as aproximações teóricas feitas na EdE geram discrepâncias entre os resultados da simulação molecular e os da EdE (Lafitte *et al.*, 2012). Por isso, a estimação desses parâmetros foi feita de maneira a diminuir as diferenças entre a pressão de vapor e a densidade de líquido saturado obtidas com a equação de estado e as obtidas via simulação molecular. A função objetivo, minimizada através do método PSO, possuiu a seguinte forma:

$$
\begin{aligned}
F_{\text{sim}}\left(c_\sigma, c_\varepsilon\right) = &\sum_{i=1}^{N_P} \left(\frac{P_v^{\text{sim}}\left(T_i, \sigma^{SAFT}, \varepsilon^{SAFT}\right) - P_v^{\text{SAFT}}\left(T_i, \sigma^{ajuste}, \varepsilon^{ajuste}\right)}{P_v^{\text{sim}}\left(T_i, \sigma^{SAFT}, \varepsilon^{SAFT}\right)} \right)^2 \\
&+ \sum_{i=1}^{N_P} \left(\frac{\rho_{liq}^{\text{sim}}\left(T_i, \sigma^{SAFT}, \varepsilon^{SAFT}\right) - \rho_{liq}^{\text{SAFT}}\left(T_i, \sigma^{ajuste}, \varepsilon^{ajuste}\right)}{\rho_{liq}^{\text{sim}}\left(T_i, \sigma^{SAFT}, \varepsilon^{SAFT}\right)} \right)^2
\end{aligned}
\tag{8}
$$

Com os parâmetros atrativo e repulsivo fixados nos valores encontrados com a EdE SAFT-VR Mie, o espaço paramétrico pode ser redefinido de maneira a se encontrar um conjunto de parâmetros refinado para o campo de força SAFT-γ CG (Lafitte *et al.*, 2012), que são:

$$\sigma^{\text{sim}} = \sigma^{SAFT} / c_\sigma \tag{9}$$

$$\varepsilon^{\text{sim}} = \varepsilon^{SAFT} / c_\varepsilon \tag{10}$$

As simulações pelo método GEMC-NVT foram refeitas seguindo-se a mesma metodologia e os resultados foram comparados com os dados de referência. A determinação do ponto crítico não foi realizada por meio do método GEMC-NVT, pois ele apresenta grandes flutuações perto do ponto crítico, que fazem as caixas de simulação mudarem de identidade durante a simulação. A temperatura crítica (T_C) foi então ajustada através do método PSO com a seguinte equação:

$$\left(\rho_{liq} - \rho_{vap}\right) = A\left(T_C - T\right)^\eta \tag{11}$$

As densidades de líquido e vapor no equilíbrio na equação acima foram provenientes de simulações feitas em uma faixa de temperatura entre 476,75 e 825 K. O expoente crítico (η) foi fixado no valor correspondente ao seu valor universal (0,325) (Hansen e McDonald, 2006) e a constante A foi determinada pelo ajuste. A densidade crítica (ρ_c) foi obtida através da lei linear dos retângulos (Equação 12) na mesma faixa de temperatura usada para obter T_C.

$$\frac{\left(\rho_{liq} - \rho_{vap}\right)}{2} = \rho_c + D\left(T_C - T\right) \tag{12}$$

A curva de coexistência líquido-vapor do campo de força SAFT-γ CG foi comparada com os resultados disponíveis na literatura para o campo de força TraPPE-EH (Rai e Siepmann, 2013), e as demais propriedades foram avaliadas por comparação com dados experimentais (Mortimer e Murphy, 1923; Nelson e Senseman, 1922; Linstrom e Mallard, 2017). A equação de desvio relativo absoluto médio usada possui a seguinte forma:

$$\Delta X = \frac{1}{N_p} \sum_{i=1}^{N_P} \left| \frac{X_{i^{ref}} - X_{i^{calc}}}{X_{i^{ref}}} \right| 100\% , \tag{13}$$

X^{ref} é o valor da propriedade de referência e X^{calc} é o valor da propriedade calculado.

Resultados e Discussão

Os parâmetros do fenantreno com três segmentos para o campo de força SAFT-γ CG estimados com a metodologia descrita estão expostos na Tabela 1.

Tabela 1. Parâmetros do campo de força SAFT-γ CG para o fenantreno.

σ^{sim} /A	$\left(\varepsilon^{sim} / \kappa_b \right)$/K	λ_r^{sim}
4,008	529,646	14,339

A Figura 2 mostra a pressão de vapor do fenantreno em função de temperatura. O campo de força com parâmetros estimados conseguiu apresentar um comportamento similar ao observado nos dados experimentais, considerando a sua simplicidade e eficiência computacional. O valor da pressão de vapor para temperaturas mais próximas do ponto normal de ebulição foi subestimado, enquanto houve uma superestimação dos valores de pressão de vapor para temperaturas mais baixas. Esse mesmo comportamento e o desvio relativo da pressão de vapor, exposto na Tabela 2, foram similares aos resultados observados para os cálculos de propriedades de equilíbrio e transporte do dióxido de carbono e metano com o campo de força SAFT-γ CG em trabalhos anteriores (Aimoli *et al.*, 2014; Aimoli, 2015). Observa-se, também, que os resultados obtidos ficaram próximos dos resultados do campo de força TraPPE-EH disponíveis na literatura, o que indica que a simplificação do modelo acarreta somente em pequenos desvios no cálculo da pressão de vapor. O desvio relativo médio está disponível na Tabela 2.

Tabela 2. Desvio relativo médio entre os resultados obtidos com o campo de força SAFT-γ CG e os dados experimentais.
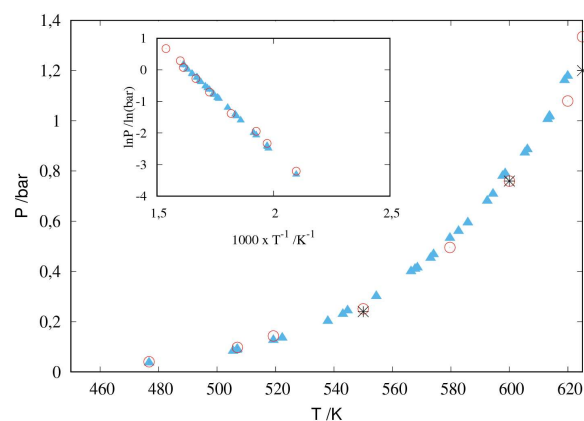
	Δ / %
P_v	9,01

Figura 2. Pressão de vapor em função da temperatura calculada com o campo de força SAFT-γ CG (círculos vermelhos), dados experimentais (triângulos azuis) e resultados do campo de força TraPPE-EH (asteriscos pretos). A figura inserida é o mesmo gráfico, porém no formato de Clausius-Clapeyron.

Na Figura 3, as curvas de densidade de líquido e vapor saturados obtidas para o campo de força SAFT-γ CG foram comparadas com os dados disponíveis do TraPPE-EH. O campo de força estimado apresentou comportamento razoavelmente similar, mas com discrepâncias maiores do que aquelas observadas para os resultados de pressão de vapor. Adicionalmente, a densidade do fenantreno foi calculada fora do ELV no ensemble NPT-MC a uma temperatura de 298 K e a uma pressão de 1 bar como forma de comparar os resultados com o valor experimental disponível que é igual 1180 kg/m^3 (Linstrom e Mallard, 2017). O campo de força TraPPE-EH forneceu um valor de densidade igual 1110 kg/m^3 e o campo de força SAFT-γ CG forneceu um valor igual a 1455 kg/m^3. Esses resultados mostram que o campo de força SAFT-γ CG tem mais dificuldade em predizer propriedades em estados diferentes daquele que foi estimado do que o TraPPE-EH por ser um modelo coarse-grained top down. Então, essa diferença na densidade pode ter sido causada pelo uso de apenas a pressão de vapor na estimação dos parâmetros da equação de estado tida como base.
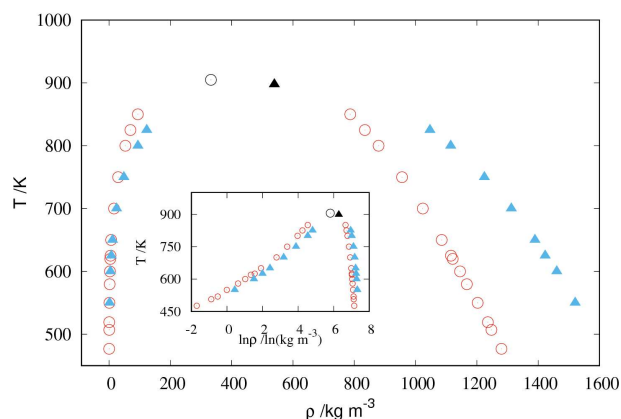
Figura 3. Curva de coexistência líquido vapor calculada com campo de força SAFT-γ CG (círculos vermelhos) e com o campo de força TraPPE-EH (triângulos azuis). O círculo e o triângulo pretos representam as propriedades críticas calculadas com o ajuste.

Os desvios relativos da densidade de líquido e de vapor saturados e das propriedades críticas encontradas com o ajuste das Equações 12-13 estão resumidos na Tabela 3. Os desvios deixam clara a existência de uma diferença significativa para densidade no equilíbrio ocasionados pela diferença de rigor teórico entre os modelos. Essas discrepâncias talvez possam ser reduzidas com uso dos dados da densidade do campo de força TraPPE-EH para estimação inicial, o que permitiria um melhor ajuste do expoente atrativo do potencial de Mie. Já com relação à previsão da temperatura crítica, os resultados entre os campos de força foram mais próximos.

Tabela 3. Propriedades críticas e desvios relativos entre o campo de força SAFT-γ CG e os dados de referência para o fenantreno.

| | Δ / % | | |
T_C	ρ_c	ρ_{liq}	ρ_{vap}
4,08	38,17	21,42	38,64

Conclusão

O campo de força SAFT-γ CG foi obtido para o fenantreno com uma metodologia em que se usa a equação de estado SAFT-VR Mie como base para parametrização desse campo de força. Os parâmetros encontrados foram avaliados através de cálculos das propriedades de equilíbrio com o método GEMC-NVT. Os resultados para pressão de vapor tiveram concordância com os dados experimentais e com o resultados da literatura obtidos para o campo de força atomístico TraPPE-EH. As densidades de líquido e vapor no equilíbrio apresentaram maiores desvios em relação ao TraPPE-EH do que a pressão de vapor. Isso mostra que a simplificação da molécula não representou uma grande perda na representação da pressão de vapor do fenantreno e que esse modelo de campo de força pode ser uma alternativa a modelos atomísticos. Com esse campo de força avaliado, pretende-se fazer cálculos de energia de solvatação do fenantreno em tolueno e em soluções tolueno+CO_2 utilizando-se dinâmica molecular.

Referências Bibliográficas

Aimoli, C.G; Maginn, E.J; Abreu, C.R.A. Transport properties of carbon dioxide and methane from molecular dynamics simulations. *J. Chem. Phys.,* 141 (2014), 134101.

Aimoli, C.G. Thermodynamic and Transport Properties of Methane and Carbon Dioxide: A Molecular Simulation Study; D.Sc Thesis, School of Chemical Engineering, University of Campina, 2015.

Avendaño, C.; Lafitte, T.; Galindo, A.; Adjiman, C.S.; Jackson, G.; Müller, E.A. SAFT-γ Force Field for the Simulation of Molecular Fluids. 1. A Single-Site Coarse Grained Model of Carbon Dioxide. *J. Phys. Chem. B.,* 115 (2011), 11154–11169.

Hansen, J.P.; McDonald, I.R. *Theory of Simple Liquids*. Academic Press, Amsterdam, 2006.

Lafitte, T.; Avendaño, C.; Papaioannou, V.; Galindo, A.; Adjiman, C.S.; Jackson, G.; Müller, E.A. SAFT-γ force field for the simulation of molecular fluids: 3. Coarse-grained models of benzene and hetero-group models of n-decylbenzene. *Mol. Phys.*, 110 (2012), 1189–1203.

Lafitte, T.; Apostolakou, A.; Avendaño, C.; Galindo, A.; Adjiman, C.S.; Müller, E.A.; Jackson, G. Accurate statistical associating fluid theory for chain molecules formed from Mie segments. *J. Chem. Phys.,* 139 (2013), 154504.

Linstrom, P.J.; Mallard, W.G. NIST Chemistry WebBook, NIST Standard Reference Database Number 69. Disponível em: <http://webbook.nist.gov>. Acesso em: 12 abr. 2017.

Panagiotopoulos, A.Z. Direct determination of phase coexistence properties of fluids using Monte Carlo simulation in a new ensemble. *Mol. Phys.*, 61 (1987), 813–826.

Mortimer, S.M.; Murphy, R.V. The Vapor Pressures of Some Substances Found in Coal Tar. *Ind. Eng. Chem. Res.,* 14 (1923), 1140-1142.

Nelson, O.A.; Senseman, C.E. Vapor Pressure Determinations on Naphthalene, Anthracene, Phenanthrene, and Anthraquinone between Their Melting and Boiling Points. *J. Ind. Eng. Chem.*, 14 (1922), 58-62.

Rai, N.; Siepmann, J.I. Transferable potentials for phase equilibria. 10. Explicit-hydrogen description of substituted benzenes and polycyclic aromatic compounds. *J. Phys. Chem. B*, 117 (2013), 273-288.

Ramrattan, N.; Avendaño, C.; Müller, E.; Galindo, A. A corresponding-states framework for the description of the Mie family of intermolecular potentials. *Mol. Phys.,* 113 (2015), 1–16.

Shah, J.R.; Maginn, E.J. A general and efficient Monte Carlo method for sampling intramolecular degrees of freedom of branched and cyclic molecules. *J. Chem. Phys.*, 135 (2011), 134121.

Smith, J.M.; Van Ness, H.C.; Abbot, M. M. *Introdução à Termodinâmica da Engenharia Química*. LTC, Rio de Janeiro, 2007.

APPENDIX D – Paper for Publication in Scientific Journal