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/ Isabela Quintela Matos. – Rio de Janeiro, 2018-
109

Supervisor: Charlles Rubber de Almeida Abreu

Dissertation (Master) – Universidade Federal do Rio de Janeiro
Escola de Química
Engenharia de Processos Químicos e Bioquímicos Graduate Program, 2018.

1. Solvation free energy. 2. Asphaltenes. 3. SAFT-γ Mie force field.

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, March 7, 2018:

Charlles Rubber de Almeida Abreu, D.Sc. -EPQB/UFRJ
Supervisor

Frederico Wanderley Tavares, D.Sc. - EPQB/UFRJ
Guest 1

Bruno Araujo Cautiero Horta, D.Sc. - PGQU/UFRJ
Guest 2

Luciano Tavares da Costa, D.Sc. - PPGQ/UFF
Guest 3

Rio de Janeiro
2018

Acknowledgements

I would like to give some words of thanks to my parents for the support during the movie from Aracaju to Rio de Janeiro, and for supporting my decisions. I also would like to thank my advisers. Charlles Rubber de Almeida Abreu for the elucidative meetings and for taking the time to explain in such clearly way the concepts of molecular simulations and free energy calculations. To my co adviser Papa Matar Ndiaye, I would like to thank him for presenting the ATOMS group to me and for accepting my participation in his project about asphaltenes. I also would like to thank all the ATOMS group members, specially Ana Jorgelina Silveira for helping me with any theoretical and software doubts I had. Finally, I would like to thank COPPETEC foundation for the financial support.

Abstract

We, at this work, studied the solvation free energy differences of molecules mimicking asphaltenes in different solvents with the SAFT-γ Mie force field. We obtained solvation free energy differences 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 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 this coarse-grained model could 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
κ_b	Boltzmann Constant
β	$1/\kappa_b T$
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 the development of separation technologies in the chemical industry. Solvation free energy is, more specifically, the difference in free energy related to the process of transferring the solute from the ideal gas phase condition to the liquid solvent phase condition (SHIRTS *et al.*, 2003). 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 and partition coefficients. Additionally, solvation free energy calculations can be part of the process of calculating solubility with molecular dynamics.

The solvation free energy calculations described above are intrinsically complex due to the many competing forces interfering in the behavior of the solute-solvent interaction. In addition to that, free energy simulations are susceptible to sampling problems in low energy regions and results from the simulations need to be correctly-post processed in order to obtain free energy differences with low variances. Another influencing factor in the output of these calculations is the choice of force field used to model the solvent and solute molecules. Force field is the name given to the group of parameters and equations used to represent the potential energy function of a system in molecular simulations. They have different levels of description, such as quantum mechanics, atomistic and coarse-grained. The quantum mechanics approach describes the motion of electrons and requires the solution of the Schroedinger equation during the simulation. While, in the atomistic description, only the atomic motions are represented, and this is done by solving the Newton's equations of motion. Finally, in the coarse-grained description, atoms are grouped into pseudo atoms or beads, and the equations of motion are solved for them.

These 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). This fact makes these models a viable option to decrease the computational time of solvation free energy calculations. Additionally, deficiencies in the description of small molecules by coarse-grained models can be revealed by free energy calculations (MOBLEY *et al.*, 2007; SHIRTS *et al.*, 2003). Hence, we, in this study, assess the efficiencies and shortcomings of the SAFT-γ Mie coarse-grained force field (AVENDAÑO *et al.*, 2011) with free energy calculations of a variety of pairs solute-solvent. This coarse-grained force field was chosen because it uses, unlike the majority of the force fields, the Mie potential (MIE, 1903) and because its method of obtaining parameters is more straightforward than other models. It was initially parameterized with pure component equilibrium and interfacial tension data (AVENDAÑO *et al.*, 2011), and this strategy has provided satisfactory results. Examples include the prediction of phase equilibrium of aromatic compounds (MÜLLER; MEJÍA, 2017), alkanes, light gases (HERDES; TOTTON; MÜLLER, 2015), and water (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 in our free energy calculations were selected with the intention of testing the force field with standard sets used as a benchmark in solvation free energy calculations and with aromatic substances used as models to asphaltenes. Asphaltenes are complicated to characterize by determining their composition on a molecular basis, but the literature broadly accepts that they can be described as a fraction of crude oil soluble in toluene and insoluble in n-alkenes (pentane, hexane, heptane) (SJÖBLOM *et al.*, 2003). They have motivated many studies with interest in developing models for their structure and behavior due to all the problems they can cause during their transportation and refining such as precipitation during the oil processing (SJÖBLOM; SIMON; XU, 2015). This precipitation issue is a recurrent problem due to the growing market of the production of crude oil in deep waters, which have conditions favorable to precipitation, such as pressure depletion and acid stimulation (BUENROSTRO-GONZALEZ *et al.*, 2004). As an example, asphaltene precipitation due to pressure drop can clog oil production equipment and cause an almost exponential increase in the cost of production (JOSHI *et al.*, 2001). All these factors make the understanding of the behavior of asphaltenes in different chemical and physical environments relevant to the oil industry.

As said in the previous paragraph, asphaltene characterization still faces some issues. Hence, we choose to use polycyclic aromatic hydrocarbons (PAH'S), which have well-defined characteristics, to initially test the efficiency of the SAFT-γ Mie force field in describing the solvation phenomenon. PAH's are a group of organic compounds that have fused rings, carbon and hydrogen in their structure (RAVINDRA; SOKHI; GRIEKEN, 2008). The ones utilized in this work were phenanthrene, anthracene, and pyrene since they have similarities with asphaltenes regarding their solubility. Meanwhile, we selected compounds that are used to characterize asphaltenes (toluene, hexane) as solvents in our free energy calculations. We also tested the anti-solvent/solvent effect of carbon dioxide due to its influence in asphaltene precipitation during the oil processing (SOROUSH *et al.*, 2014). With these calculations of solvation free energies with the SAFT-γ Mie model, we intend to improve this force field and provide satisfactory solvation free energy estimates of PAH's with a coarse-grained model. The success of this description of smaller asphaltene-like compounds by this force field can then open up the possibility of obtaining satisfactory results for more complex asphaltene models with a less computational expensive force filed.

2 Literature Review

2.1 Molecular Simulations of Molecules Mimicking Asphaltenes

Asphaltenes, unlike the majority of the compounds, are not defined on a molecular basis. The most accepted definition is that they are a fraction of crude oil insoluble in n-alkanes (pentane, hexane, and heptane) and soluble in toluene (SJÖBLOM *et al.*, 2003). Due to uncertainties related to its structures, much work has been done to develop model compounds that have a well-defined structure and can represent asphaltenes. The two categories of models presented in the literature are the archipelago and continental models. In the archipelago model, asphaltenes consist of polyaromatic parts linked together by aliphatic or naphthenic moieties and, in the continental model, they consist of a single polyaromatic ring with linked aliphatic or naphthenic chains (MULLINS, 2010; MURGICH, 2003). The choice of the model's structure, such as chemical bonding, is highly essential since some structures can cause the occurrence of high energies regions during the simulation (LI; GREENFIELD, 2011) .

In order to evaluate the strengths and shortcomings of alphaltenes models, papers have been published about the calculations of various properties. There are some concordances in studies of theses models with molecular simulation, such as the influence of the model in the packing tendency of the molecule (GREENFIELD, 2011). Kuznicki, Masliyah and Bhattacharjee (2009) utilized molecular dynamics in the study of the nanoaggregation of four types of model asphaltene molecules in binary mixtures of toluene and water. The authors 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 alleged that there is no formation of nanoaggregates, and that the distribution of asphaltene clusters is continuous for mixtures of asphaltene in heptane.

Molecular dynamics simulations were also utilized in the study of Ervik *et al.* (2016) to obtain the correct interfacial orientation of asphaltenes using a coarse-grained model of the interface and a representative model for the asphaltene molecules. Also using a coarse-grained force field, Jover *et al.* (2015) carried molecular simulations with a continental asphaltene model. The results reproduced experimental data of the strong aggregation of asphaltene molecules in n-heptane and high solubility in toluene. Gao *et al.* (2014) performed a molecular dynamics study with the GROMOS 45a3 force field (SCHULER; DAURA; GUNSTEREN, 2001) to identify the structural features of different asphaltene molecules.

Mikami *et al.* (2013) employed the archipelago model in their work to investigate the interfacial behavior of asphaltene molecules at the oil-water interface using molecular dynamics simulations with the OPLS-AA force field (JORGENSEN; TIRADO-RIVES, 1988). 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. They also discovered an oscillatory behavior of asphaltene molecules at the oil-water interface when using the archipelago model. Teklebrhan *et al.* (2014) used a perylene based model to study molecular association and interaction as well as the adsorption properties of the perylene molecule at the water/toluene or water/heptane interface.

2.2 Coarse-Grained (CG) 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 classical mechanics. Force fields at this level are based on van der Waals interactions, which may include neutral or Coulombic charged sites. Contributions due to intramolecular interactions such as bond-stretching, angle-bending, and torsion are also usually accounted in these kinds of force fields. When the simulation scale needs to be increased, and the atomistic simulations become too computationally expensive such as in the study of biological systems, 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 evident loss of information in grouping atoms; hence it is necessary to assure that the process of eliminating unnecessary or unimportant information ('coarse-graining') does not affect the system's physical behavior. Ideally, coarse-grained models need to have representability, robustness, transferability, and computational efficiency. Representability means that you can use a model at a state point other than the one in which it was parameterized. The other characteristic, robustness, is related to the model's ability to enable reliable predictions for various structural, thermodynamic, or transport properties. Finally, a transferable model is one in which the representation of atomic or chemical moieties have the same behavior in different molecules- e. g., a pseudo-atom representing CH_2 should have the same properties both in an alkene molecule and in a polyethylene molecule (MÜLLER; JACKSON, 2014). To achieve the cited goals, coarse-grained force fields are usually developed by mapping the atomistic model to define the pseudo-atoms, which are generally formed by similar functional groups.

The level of coarse-graining also needs to be established. Hadley and McCabe (2012) affirm that six heavy atoms (non-hydrogen atoms) per bead are traditionally used in order to not lose valuable details and to maintain isotropic representations of the beads. After the mapping, CG force fields needs to be parametrized. There are two different approaches, bottom up and top down, to link the simulations on the coarse-grained scale to another scale (schematically represented in Figure 2.2.1). The bottom-up approach uses information of a more detailed scale such as the quantum mechanics description or the atomistic description to obtain information necessary to the parametrization. This method highly depends on the more detailed model quality to succeed. Meanwhile, the top down methodology obtains parameters from larger scales, namely experimental thermodynamic properties or structure based properties.

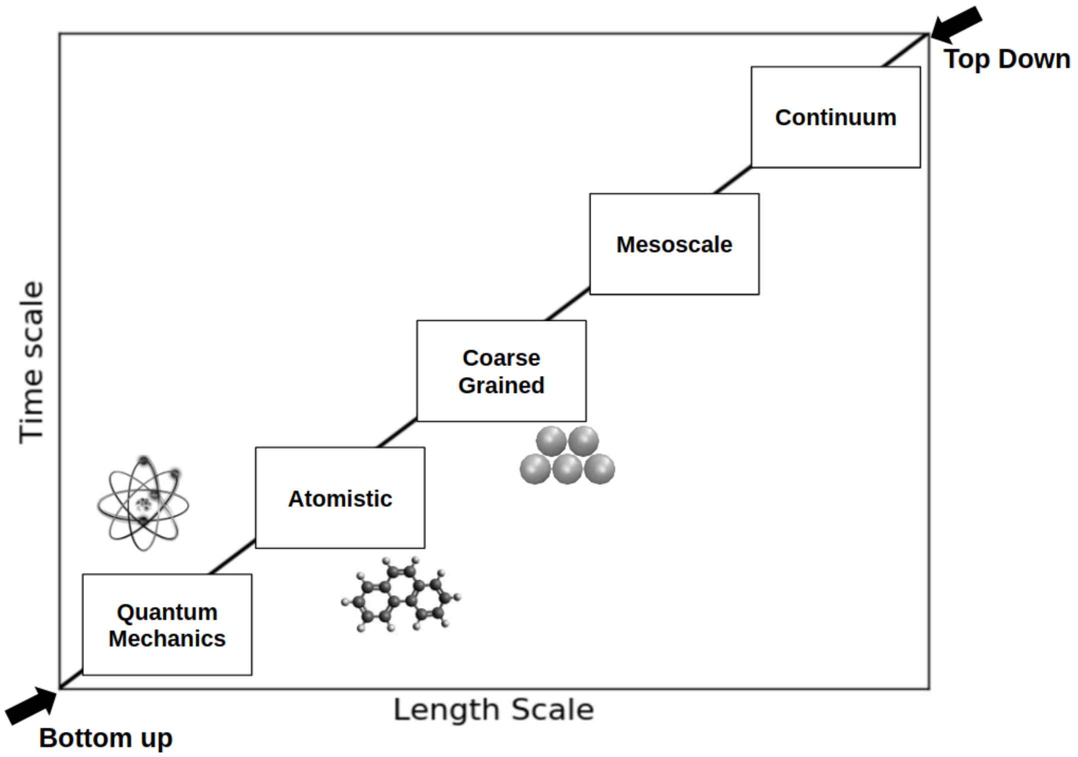


Figure 2.2.1 – Schematic representation of two approaches of coarse-graining.

One of the first applications of coarse-grained models is in the study of protein folding (LEVITT; WARSHE, 1975; LEVITT, 1976). These earlier protein CG models were based on known molecular structure, and they contributed to the knowledge of physicochemical forces associated with protein folding and protein interactions (KOGA; TAKADA, 2001). More recently, models focused on retaining the protein's chemical specificity. The Bereau and Deresmo model (BEREAU; DESERNO, 2009) represents a single amino acid with a maximum of four beads, and it 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) represents a single amino acid with a maximum of six beads. It was used to investigate a variety of phenomena, ranging from protein folding to the modeling of DNA-RNA complexes (BARDUCCI; BONOMI; DERREUMAUX, 2011; CHEBARO *et al.*, 2009; STERPONE *et al.*, 2014). Other 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 latter one is the most popular model for CG modeling of membrane proteins (MARRINK; TIELEMAN, 2013). The MARTINI force field is also extensively used as a CG model for water. This force field represents four water molecules as one bead using a shifted Lennard-Jones potential for non-bonded interactions. Despite its extensive use, the MARTINI water model does not properly reproduce properties such as interfacial tension and compressibility (HE *et al.*, 2010). Besides, it can freeze at room temperature (WINGER *et al.*, 2009; MARRINK *et al.*, 2007), which 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 to build a versatile and transferable coarse-grained model for surfactants/water systems using density, interfacial, and hydration free energies data. They selected a 12-4 Mie potential for water cross interactions and a 9-6 Mie potential for the surfactant (alkanes, oxyethylenes, ethylene glycols, ethers, and alcohols) interactions.

Outside of the Martini framework, Orsi and Essex (2011) proposed the ELBA coarse-grained model for molecular dynamics simulations of lipids membranes. In this model, electrostatics are modeled explicitly by charges, and water molecules are represented by a single Lennard-Jones bead embedded with a point dipole. Genheden (2016) expanded the Elba force field to model 1-hexanol, 1-nonanol, n-hexane and n-nonane by representing three carbons with a single bead. Using different Mie and Morse potentials, He *et al.* (2010) studied different levels of coarse-graining for water ranging from one to four molecules per bead. Other investigations 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: CGW1-vle and CGW1-ift. Both of them represent the water molecule as one bead, and the Mie Potential has a repulsive and attractive exponent 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 in 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 property calculations. Both models had temperature-dependent size and energy parameters and performed well for these properties over the entire liquid temperature range. The SAFT-γ Mie force field has 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. Lobanova *et al.* (2016) tested the force field with binary and ternary mixtures of water and carbon dioxide. There are also papers with the SAFT-γ Mie about 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).

2.3 Solvation Free Energies

Solvation free energy calculations with molecular dynamics can be used to evaluate the quality of a coarse-grained force field, such as the ones described in the section above, since these estimations can reveal deficiencies in a force field. Besides this application, solvation free energies are used to obtain information about the behavior of the solvent in different chemical environments and to assess the influence of the solute's molecular geometry on the solvation phenomenon. Due to their range of application and inherent complexity, free energy calculations were the subject of a variety of studies in the last decade interested in improving the free energy simulations and their post-processing methods (SHIRTS; CHODERA, 2008; PALIWAL; SHIRTS, 2011; SHIRTS; PANDE, 2005; YTREBERG; SWENDSEN; ZUCKERMAN, 2006)s.

Recent papers (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 fifty two compounds with the OPLS-AA force field. They obtained an overall root mean square deviation to the experimental data of 1.75 kcal/mol and concluded that the reproducibility of the Lennard-Jones parameters is the main limiter of the precision of their results. Izairi and Kamberaj (2017) also studied hydration free energies but with the intention of comparing the polar and nonpolar contributions. GARRIDO *et al.* (2009, 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, ethylbenzene, 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. They calculated the solvation free energy of a set of 205 small organic molecules in 1-octanol and found that the force field parametrization of n-octanol proposed by Kobryn and Kovalenko (2008) provided the best agreement. Goncalves and Stassen (2005) calculated the free energy of solvation using the polarizable continuum model coupled to molecular dynamics computer simulation with the GROMOS force field. These calculations were done with a representative set of solutes and with the solvents tetrachloride, chloroform, and benzene. 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. To define the role of water as solvent in the docking structure determination of proteins, Matubayasi (2017) developed a method to compute the solvation free energy of proteins while using OPLS-AA force field for the solutes and TIP3P for water. Genheden (2016) expanded the Elba force field to calculate solvation free energies of more than 150 solutes taken from the Minnesota solvation database in polar (water, hexanol, octanol and nonanol) and apolar (hexane,octane, and nonane) solvents. He obtained mean absolute deviations of 1 kcal/mol for water and 1.5 kcal/mol for hexane. In this model, three carbons are represented by a single bead and water is also represented by a single bead.

Though this variety of data using the intramolecular Lennard-Jones potential, we are not aware of work using the Mie Potential in free energy calculations. We, in the present study, try to provide information about these predictions with the SAFT-γ Mie coarse-grained force field. As said before, the output of these calculations are highly dependent on the force field, and some coarse-grained models obtained satisfactory results for these kinds of simulations. Hence, knowing if other coarse-grained approaches have similar performances to the all-atoms force fields can help increase the scale of solvation free energy calculations.

2.4 Solvation Free Energy Calculation Methods

Solvation free energy calculations account for the difference in free energy related to transferring the solute form ideal gas condition to the liquid solvent condition. To do that, we gradually insert the solute in the solvent. This process is mathematically carried out by using a coupling parameter (λ) on the total potential energy function, where λs represent the intermediate states in the transition from the ideal gal condition to the solvent condition. Hence, during the solvation free energy simulations, we obtain total potential energies correspondent to these coupling parameters [$(U(\lambda)$]. After the simulations, these potential energies need to be post-processed and analyzed to calculate the solvation free energies effectively. Since these calculations can have slow convergences, a lot of papers in the last decades focused on developing analysis methods to calculate free energies. Almost all methods rely on o the following approaches: free energy perturbation (FEP) based methods, thermodynamic integration, and histograms.

2.4.1 Thermodynamic integration

The thermodynamic integration method (KIRKWOOD, 1935) uses equilibrium averages to evaluate the energy derivative with respect to the coupling parameter (λ):

$$\frac{\partial(G\,1/\kappa_b T)}{\partial\lambda} = \left\langle \frac{\partial\mathcal{H}}{\partial\lambda} \right\rangle_{N,P,T}. \tag{2.1}$$

In Eq. (2.1), κ_b is the Boltzmann constant, G is the Gibbs free energy and $\mathcal{H}$ is the Halmitonian of the system. The derivative is obtained for every configuration data between the states from simulations through an analytical expression. Some examples of methods for obtaining theses expressions are the trapezoidal rule or natural cubic spline (PALIWAL; SHIRTS, 2011). There are also more complex schemes that are usually system specific, such as as those found in Jorge *et al.* (2010) and Shyu and Ytreberg (2010). MD simulations for each coupling parameter λ_k are carried out and the average over the derivative at each state is compute in order to obtain the final solvation free energy:

$$\Delta G \approx \int_0^1 \left\langle \frac{\partial\mathcal{H}}{\partial\lambda_k} \right\rangle d\lambda. \tag{2.2}$$

2.4.2 Histograms

Histograms are used to compute probability distributions. Usually, every histogram count is treated as the 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). It puts 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 unbiased sum of probability distributions from biased simulations. This method was developed to avoid problems related to data loss, high uncertainties, and the calculation of the constant added by the use of a biased potential (ROUX, 1995). The probability distribution dependent on the potential energy (U) and temperature (T) for the WHAM is

$$\begin{aligned} \tilde{\varrho}^*_r(U,T) =& \frac{\sum_i f_i(U)\exp(-\beta U)}{\sum_i f_{tot,i}\exp(\beta_i \tilde{A}_i - \beta_i U)}, \\ & \exp(-\beta_i \tilde{A}_i) = \sum_U \tilde{\varrho}^*_r(U,T), \; and \\ & \tilde{\varrho}_r(U,T) = \frac{\tilde{\varrho}^*_r(U,T)}{\sum_U \tilde{\varrho}^*_r(U,T)}, \end{aligned} \tag{2.3}$$

where $\beta = 1/\kappa_b T$, $\tilde{A}_i$ 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. Eq. (2.3) is solved self consistently with the initial value for $\tilde{A}_i$ equals to zero. The final unnormalized probability distribution is then given by $\tilde{\varrho}_r(U,T)$.

2.4.3 Free Energy Perturbation (FEP)

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

$$\Delta G_{AB} = -\kappa_b T ln \langle e^{-\beta(U_B - U_A)} \rangle_A. \tag{2.4}$$

According to the equation above, the free energy difference is calculated by doing an average over the configurations 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 ensure overlap, it is possible to carry out simulations in N intermediate states between A and B, so Eq. (2.4) becomes:

$$\Delta G_{AB} = -\kappa_b T \sum_{i=0}^{N} ln \langle e^{-\beta(U_{i+1} - U_i)} \rangle_i. \tag{2.5}$$

This way of calculation ΔG in Eq. 2.5 is also called Exponential Averaging (EXP) (ZWANZIG, 1955; PALIWAL; SHIRTS, 2011). The direction of the transformation is crucial 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). When the direction is of increasing entropy, the step is of deletion (ΔG_{BA}), and the method is labeled as deletion exponential averaging (DEXP). These directions can yield different values of free energy differences due to undersampling 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 does not have sufficient overlap. For these cases, the Bennett Acceptance Ratio or the Multi-State Bennett Acceptance Ratio is more adequate.

2.4.3.1 Bennett Acceptance Ratio (BAR)

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

$$\Delta G_{ij} = \frac{1}{\beta} ln \left(\frac{\sum_{k=1}^{N_j} \frac{1}{1+\exp[-\beta(\Delta U_k^j + C)]}}{\sum_{l=1}^{N_i} \frac{1}{1+\exp[-\beta(\Delta U_l^i - C)]}} \right) + C - \frac{1}{\beta} ln \left(\frac{N_j}{N_i} \right), \tag{2.6}$$

$$C = \Delta G_{ij} + \frac{1}{\beta} ln \left(\frac{N_j}{N_i} \right). \tag{2.7}$$

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

$$s_{ij}^2 = \frac{1}{\beta^2 N_i} \left[\frac{\langle f^2(x) \rangle_i}{\langle f(x) \rangle_i^2} - 1 \right] + \frac{1}{\beta^2 N_j} \left[\frac{\langle f^2(x) \rangle_j}{\langle f(x) \rangle_j^2} - 1 \right], \tag{2.8}$$

where $f(x) = 1/(1+x)$ is the Fermi function 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 variances 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 do not solve Eqs. (2.6) and (2.7) self consistently. By doing that, free energy differences will not have minimum variance, and the averages of Eqs. (2.6) - (2.8) are accumulated (PALIWAL; SHIRTS, 2011). The two methods are the Unoptimized Bennett Acceptance Ratio (UBAR) and the Range-Based Bennett Acceptance Ratio (RBAR). The first one avoids the self consistent solution of the BAR equations by defining $C = \beta^{-1}ln(N_j/N_i)$. The UBAR method will be close to optimal when each intermediate free energy is relatively near zero. In turn, the RBAR method selects a range of initial guesses of the constant C 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. (2.7) to calculate the value of C. Hence, this method requires a good estimation of the initial range of the values of C. The RBAR can be advantageous when compared to BAR since only the accumulated averages need to be retained for postprocessing (PALIWAL; SHIRTS, 2011).

2.4.3.2 Multistate Bennett Acceptance Ratio (MBAR)

The MBAR method (SHIRTS; CHODERA, 2008) is a further development of the BAR method, and is the one chosen to estimate the solvation free energies in this dissertation. This method consists of an estimator that computes free energies and their uncertainties of each K state by minimizing the $K \times K$ matrix of variances simultaneously. The estimator solves the following equation for each G_i self consistently:

$$G_i = \frac{1}{\beta} ln \sum_{k=1}^{K} \sum_{n=1}^{N_k} \frac{\exp[-\beta U_i(x_{kn})]}{\sum_{l=1}^{K} N_l \exp[\beta(G_l - U_l(x_{kn}))]}. \tag{2.9}$$

The equation above requires the evaluation of the potential energy $[U_i(x_{kn})]$ of the n_{th} uncorrelated configuration obtained at state K and 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 uncertainties can be computed by :

$$\delta_{ij}^2 s_{ij} = s_{ii}^2 + s_{jj}^2 - 2s_{ij}. \tag{2.10}$$

where s_{ij} is te covariance matrix. A further development of this method is available in Section 3.5.

3 Fundamentals of the Computational Methods

3.1 Molecular Dynamics

3.1.1 Background and Formalution

Molecular Dynamics (MD) uses molecular configurations (Cartesian coordinates and momentum) to extract structural , thermodynamic and dynamic information of a system. This information is extracted from the time evolution of the system, which is obtained through the numerical integration of the Newton's equations of motion (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, for instance $q^N = (r_1, \theta_1, \phi_1)$, the Halmitonian formulation, a more general 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 are then rewritten using the Hamiltonian:

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

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

In the dynamics described by the equations above, the Hamiltonian is preserved. The coordinate and momentum axes for each atom 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 used to calculate the thermodynamic properties if the system is ergodic. That is, a trajectory in this system will explore with the same probability all regions of the phase space of microstates with the same energy (SHELL, 2015).

3.1.2 Statistical Ensembles

In order to calculate thermodynamic properties, we need to define control variables of a system. For an isolated system at equilibrium, the control variables are the number of particles (N), volume (V), and total energy (E). The set of configurations under the control variables is then called statistical ensemble. In the example cited above, it is specifically called the microcanonical ensemble. Following the ergodic hypothesis, the system at these conditions will spend the same amount of time in each of the microstates (points in phase space) with the fixed Hamiltonian. The number of accessible microstates is defined by the partition function or the density of states, and is given by the following equation for the microcanonical ensemble:

$$\Omega(N, V, E) = \frac{\epsilon_0}{h^{3N} N!} \int dp^N dr^N \delta[\mathcal{H}(p^N, r^N) - E]. \tag{3.6}$$

Here, ϵ_0 is the energy unit, h is the Planck constant and δ is a Dirac delta function. As mentioned above, the system will spend the same amount of time at each of the microstates, i. e. each of these microstates have equal probabilities (ϱ) of being visited. Such probability is:

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

The macroscopic properties from molecular dynamics are then obtained from the relation of the microcanonical partition function to the entropy (S). Known as the Boltzmann equation. It is:

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

where κ_b is the Boltzmann constant. With this equations, we can derive other relations to macroscopic properties with the fundamental thermodynamic equations:

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

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

As said above, the microcanonical ensemble has N, V, and E as its control variables. 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 are the isoenthalpic-isobaric (constant number of particles, pressure, and enthalpy) and the grand canonical (constant chemical potential, volume, and temperature) ones. A variety of physical properties is measured at the conditions of the isothermal-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 are going to briefly talk about it. This ensemble is obtained from a Legendre transformation on the canonical ensemble. The Helmholtz free energy $A(N, V, T)$ becomes the Gibbs free energy $G(N, P, T)$ by transforming the volume into the external pressure:

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

where $V = V(P)$. The Gibbs free energy is related to its partition function $\Delta(N, P, T)$ by:

$$G(N, P, T) = -\kappa_b T ln\Delta(N, P, T), \tag{3.12}$$

where $\Delta(N, P, T)$ is given by:

$$\Delta(N, P, T) = \frac{1}{V_0} \int_0^\infty dV \int dp^N dr^N \exp\left[-\beta\left(\mathcal{H}(r^N, pN) + PV(r^N)\right)\right]. \tag{3.13}$$

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

$$Q(N, V, T) = \int d^N p d^N r \exp\left[-\beta\mathcal{H}(r^N, pN).\right] \tag{3.14}$$

From these relations and a differential change in G, we can obtain the chemical potential (μ), volume and entropy relations for isothermal-isobaric ensemble:

$$\mu = \left(\frac{\partial G}{\partial N}\right)_{P,T} = -\kappa_b T\left(\frac{\partial ln\Delta(N, P, T)}{\partial N}\right)_{N,T}, \tag{3.15}$$

$$\langle V\rangle = \left(\frac{\partial G}{\partial P}\right)_{N,T} = \kappa_b T\left(\frac{\partial ln\Delta(N, P, T)}{\partial N}\right)_{N,P}, \tag{3.16}$$

$$S = \left(\frac{\partial G}{\partial T}\right)_{N,P} = \kappa_b \left[lnQ(N,V,T) + T\left(\frac{\partial lnQ(N,V,T)}{\partial T}\right)_{V,N}\right]. \tag{3.17}$$

3.1.3 Thermostats and Barostats

The isothermal-isobaric and canonical ensembles have external conditions being applied to it (temperature and pressure). For temperature control, the method employed mimics the effect of a thermal reservoir through the use of a thermostat. The thermostats need to be capable of capturing the correct energy fluctuations in the system since the kinetic energy will fluctuate when using a heat bath to control the temperature in a canonical ensemble of a finite system (FRENKEL; SMIT, 2001).

Among the available thermostat are the Berendsen (BERENDSEN *et al.*, 1984), the Andersen (ANDERSEN, 1980) and the Nosé (NOSÉ, 1984) thermostats, but, here, we are going to discuss the most widely used thermostat: the Nosé-Hoover (HOOVER, 1985). This thermostat is based on the formulation of Nosé (NOSÉ, 1984), whom used a Lagrangian that contains additional and artificial coordinates and velocities (FRENKEL; SMIT, 2001). In this method, the Hamiltonian in a canonical ensemble is constructed as:

$$\mathcal{H} = K(p^N) + U(q^N) + \frac{\xi^2 \mathcal{Q}}{2} + 3N\kappa_b T lns, \tag{3.18}$$

where ξ is a friction coefficient related to the conjugate momentum of the thermal reservoir to which the system is coupled, s is the position of the thermal reservoir and $\mathcal{Q}$ is the effective mass associated with s. The velocity update is then done with the friction term added to the equations of motion (SHELL, 2015):

$$\frac{dr_i}{dt} = v_i, \tag{3.19}$$

$$\frac{dv_i}{dt} = -\frac{1}{m_i}\frac{\partial U(r^N)}{\partial r_i} - \xi v_i, \tag{3.20}$$

$$\frac{\xi}{dt} = \frac{\sum m_i v_i - 3N\kappa_b T}{\mathcal{Q}}, \tag{3.21}$$

$$\frac{\xi}{dt} = -\frac{1}{m_i}\frac{\partial U(r^N)}{\partial r_i} - \xi v_i. \tag{3.22}$$

To increase the robustness of the Nosé-Hoover thermostat, the Nosé-Hoover chains of thermostats method was developed. It proposes the use of multiple thermal reservoirs linked in order to enhance temperature equilibration (SHELL, 2015).

Meanwhile, the pressure is controlled with a barostat. It maintains the pressure constant during the simulation by adjusting the simulation volume. Among the available barostats methodologies are the Berendsen (BERENDSEN *et al.*, 1984), in which the pressure is coupled to a pressure bath and the volume is periodically rescaled, and the Anderson barostat (ANDERSEN, 1980), which serves as basis for other barostating methdos such as the ones developed by Hoover (HOOVER, 1985), Martina-Tobias-Klein (MARTYNA; TOBIAS; KLEIN, 1994) and Parrinnello-Rahman (PARRINELLO; RAHMAN, 1981). The Andersen's idea was to couple the system to a fictional pressure bath and incorporate the volume into the phase space as an additional degree of freedom (TUCKERMAN, 2010). As in the Nosé-Hoover thermostat, Andersen added terms into the Hamiltonian in order to control the variable of interest:

$$\mathcal{H} = \sum_i \frac{V^{-2/3}\pi_i^2}{2m_i} + U(V^{1/3}\mathbf{s}^N) + \frac{p_V^2}{2\mathcal{W}} + PV, \tag{3.23}$$

with s and π being scale transformations:

$$\mathbf{s}_i = V^{-1/3} r_i, \tag{3.24}$$

$$\pi_i = V^{1/3} p_i. \tag{3.25}$$

Here, p_V is the momentum conjugate to the volume and $\mathcal{W}$ is a mass parameter that determines the time scale of the volume motion. With the equations above, we can derive the equations of motion in Cartesian coordinates for an ensemble at constant pressure:

$$\frac{dr_i}{dt} = \frac{p_i}{m_i} + \frac{1}{3}\frac{dV}{dt}\frac{r_i}{V}, \tag{3.26}$$

$$\frac{dp_i}{dt} = -\frac{\partial U}{\partial r_i} - \frac{1}{3}\frac{dV}{dt}\frac{p_i}{V}, \tag{3.27}$$

$$\frac{dV_i}{dt} = \frac{p_V}{\mathcal{W}}, \tag{3.28}$$

$$\frac{dp_V}{dt} = \frac{1}{3V}\sum_i \left[\frac{p_i^2}{m_i} - \frac{\partial U}{\partial r_i} \cdot r_i\right] - P. \tag{3.29}$$

The equations of motion above are then numerically integrated using the methodologies described in the next section.

3.1.4 Integration of the equations of motion

With the formalism defined for the equations of motions and for the statistical ensemble, we can now derive discrete-time numerical approximations for them. The basic idea is to solve the trajectory of atoms as a function of time ($r^N(t)$) by updating the positions in discrete time intervals or time steps. To do that, the classical time evolution approach is used to preserve the Hamiltonian of the system in the numerical integration methods. In this approach, we start considering the time evolution of an arbitrary function $a(x_t)$ along a trajectory x_t. Doing the time derivative of $a(x_t)$:

$$\frac{da}{dt} = \sum_{\alpha=1}^{3N} \left[\frac{\partial a}{\partial q_\alpha} \frac{\partial \mathcal{H}}{\partial p_\alpha} - \frac{\partial a}{\partial p_\alpha} \frac{\partial \mathcal{H}}{\partial q_\alpha} \right]. \tag{3.30}$$

In the equation above, we can represent the time evolution of $a(x_t)$ with the Poisson bracket:

$$\frac{da}{dt} = \{a, \mathcal{H}\}. \tag{3.31}$$

The Poisson bracket is equal to applying the Liouville operator ($i\mathcal{L}$) on the phase space. Hence,

$$\frac{da}{dt} = i\mathcal{L}a. \tag{3.32}$$

Substituting the equation above in Eq. 3.30:

$$i\mathcal{L}a = \sum_{\alpha=1}^{3N} \left[\frac{\partial a}{\partial q_\alpha} \frac{\partial \mathcal{H}}{\partial p_\alpha} - \frac{\partial a}{\partial p_\alpha} \frac{\partial \mathcal{H}}{\partial q_\alpha} \right]. \tag{3.33}$$

The solution of Eq. 3.32 for $a(x_t)$ is given by

$$a(x_t) = \exp[i\mathcal{L}t]a(x_0). \tag{3.34}$$

Here, $\exp[i\mathcal{L}t]$ is known as the classical propagator. The effect of this operator in a function cannot be evaluated. However, we can develop approximate solutions for the Hamiltonian's equations with this operator. Rewriting Eq. 3.33 as

$$i\mathcal{L} = i\mathcal{L}_1 + i\mathcal{L}_2, \tag{3.35}$$

where

$$i\mathcal{L}_1 = \sum_{\alpha=1}^{N} \frac{\partial}{\partial q_\alpha} \frac{\partial \mathcal{H}}{\partial p_\alpha}$$
$$i\mathcal{L}_2 = -\sum_{\alpha=1}^{N} \frac{\partial}{\partial p_\alpha} \frac{\partial \mathcal{H}}{\partial q_\alpha}. \tag{3.36}$$

The operators $i\mathcal{L}_1$ and $i\mathcal{L}_2$ in the equations above are non-commuting operators, that is, the order in which the operator is applied is important (TUCKERMAN, 2010). This fact implies that we can not separate the classical propagator $\exp(i\mathcal{L}t)$ into the product $\exp(i\mathcal{L}_1 t)\exp(i\mathcal{L}_2 t])$. Though we can not do that, we can still express the propagator in terms of these two factors by using the symmetric Trotter theorem or Strang splitting formula (TROTTER, 1959; STRANG, 1968). Applying this theorem to the classical propagator, we then obtain

$$\exp(i\mathcal{L}t) = \exp(i\mathcal{L}_1 t + i\mathcal{L}_2 t) = \lim_{P\to\infty} \left[\exp(i\mathcal{L}_2 t/2P)\exp(i\mathcal{L}_1 t/P)\exp(i\mathcal{L}_2 t/2P)\right]^P, \tag{3.37}$$

where P is an integer. Defining a time step $\Delta t = t/P$ and using it in Eq. 3.37, we have

$$\exp(i\mathcal{L}t) = \lim_{P\to\infty,\Delta t\to 0} \left[\exp(i\mathcal{L}_2 \Delta t/2)\exp(i\mathcal{L}_1 \Delta t)\exp(i\mathcal{L}_2 \Delta t/2)\right]^P, \tag{3.38}$$

In order to obtain an approximation for $\exp(i\mathcal{L}t)$, we take the limits and consider that P is a finite number. The resulting approximation for the classical propagator is then

$$\exp(i\mathcal{L}t) \equiv \left[\exp(i\mathcal{L}_2 \Delta t/2)\exp(i\mathcal{L}_1 \Delta t)\exp(i\mathcal{L}_2 \Delta t/2)\right]^P + \vartheta(P\Delta t^3), \tag{3.39}$$

using $P = t/\Delta t$

$$\exp(i\mathcal{L}\Delta t) \equiv \exp(i\mathcal{L}_2 \Delta t/2)\exp(i\mathcal{L}_1 \Delta t)\exp(i\mathcal{L}_2 \Delta t/2) + \vartheta(\Delta t^3). \tag{3.40}$$

Now we can use Eq. 3.40 as a numerical propagation for a single time step (Δt). Using this propagation on a single particle moving with Hamiltonian, where $i\mathcal{L}_1 = K(p)$ and $i\mathcal{L}_2 = U(r)$, we obtain

$$\exp(i\mathcal{L}\Delta t) \equiv \exp\left(-\frac{\Delta t}{2}\frac{\partial U}{\partial r}\frac{\partial}{\partial p}\right)\exp\left(\Delta t \frac{p}{m}\frac{\partial}{\partial r}\right)\exp\left(-\frac{\Delta t}{2}\frac{\partial U}{\partial r}\frac{\partial}{\partial p}\right), \tag{3.41}$$

where the derivatives of the intermolecular potential $-\dfrac{\partial U(r)}{\partial r}$ is equal to the force (F) acting on the particle. We now are able to replace the exact solution of Eq. 3.34 with the approximation of Eq. 3.41. The approximation evolution, from a initial condition $(r(t), p(t))$, is then

$$\begin{aligned}\begin{bmatrix} r(t+\Delta t) \\ p(t+\Delta t) \end{bmatrix} &\equiv \exp\left(\frac{\Delta t}{2}F(r(t))\frac{\partial}{\partial p}\right) \\ &\times \exp\left(\Delta t\frac{p(t)}{m}\frac{\partial}{\partial r}\right) \\ \times \exp\left(\frac{\Delta t}{2}F(r(t))\frac{\partial}{\partial p}\right) &\begin{bmatrix} r(t) \\ p(t) \end{bmatrix}.\end{aligned} \tag{3.42}$$

The propagation is determined by acting each of the three operators starting from the right on r and p. The result of applying the operator in a function $g(r)$ is the Taylor expansion $g(r+c)$. Hence, we have the following equation after applying the first operator:

$$\exp\left(\frac{\Delta t}{2}F(r(t))\frac{\partial}{\partial p}\right)\begin{bmatrix} r(t) \\ p(t) \end{bmatrix} = \begin{bmatrix} r(t) \\ p(t) + \frac{\Delta t}{2}F(r(t)) \end{bmatrix}. \tag{3.43}$$

Acting the second operator in the equation above:

$$\exp\left(\Delta t\frac{p(t)}{m}\frac{\partial}{\partial r}\right)\begin{bmatrix} r(t) \\ p(t) + \frac{\Delta t}{2}F(r(t)) \end{bmatrix} = \begin{bmatrix} r(t) + \frac{\Delta t}{2m}p(t) \\ p(t) + \frac{\Delta t}{2}F(r(t) + \frac{\Delta t}{2m}p(t)) \end{bmatrix}. \tag{3.44}$$

and, finally:

$$\begin{aligned}\exp\left(\frac{\Delta t}{2}F(r(t))\frac{\partial}{\partial p}\right)\begin{bmatrix} r(t) + \frac{\Delta t}{2m}p(t) \\ p(t) + \frac{\Delta t}{2}F(r(t) + \frac{\Delta t}{2m}p(t)) \end{bmatrix} = \\ \begin{bmatrix} r(t) + \frac{\Delta t}{m}(p(t) + \frac{\Delta t}{2}F(r(t))) \\ p(t) + \frac{\Delta t}{2}F(r(t)) + \frac{\Delta t}{2}F\{r(t) + \frac{\Delta t}{m}[p + \frac{\Delta t}{2}F(r(t))]\} \end{bmatrix}.\end{aligned} \tag{3.45}$$

Using the equations above and substituting p/m for v, the final position $r(t+\Delta t)$ can be written as

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{F(t)}{2m}\Delta t^2. \tag{3.46}$$

Eq. 3.46 is the position update part of the velocity Verlet method. In this method, the positions are updated by a time step of Δt by using the positions at the previous time steps and forces. These recalculations of forces at each time step are the most computationally expensive part of the simulation since we have to take the derivative of the potential energy at each time step. The equation to update the velocity can also be derived from the equations above and is given by

$$v(t+\Delta t) = v(t) + \frac{F(t+\Delta t) + F(t)}{2m}\Delta t. \tag{3.47}$$

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:

$$v(t+\Delta t/2) = v(t-\Delta t/2) + \frac{F(t+\Delta t) + F(t)}{m}\Delta t, \tag{3.48}$$

$$r(t+\Delta t) = r(t) + v(t+\Delta t/2)\Delta t. \tag{3.49}$$

Although using different time steps or formulations, both methods described generate the same trajectory for a given initial configuration.

3.1.5 Initial Configuration and Periodic Boundary Condition

The equations above require an overlap-free initial configuration with positions and velocities for all atoms in the system. The initial velocities follow a temperature-dependent Maxwell-Boltzmann distribution, which is

$$\varrho(v_{x,i}) = \left(\frac{m_i}{2\pi\kappa_b T}\right)^{\frac{1}{2}} \exp\left(-\frac{m_i v_{x,i}^2}{2\pi\kappa_b T}\right). \tag{3.50}$$

Random velocities are then found with the equation above for each of the 3N components of the velocity. Meanwhile, the initial positions can be obtained by several approaches. The initial configuration can be taken from x-ray or nuclear magnetic resonance (NMR) spectroscopy, the atoms can be placed randomly in the simulation volume, or the atoms can be placed in idealized or approximate geometries. The generally used method to acquire the configurations places the molecules on a cubic lattice (SHELL, 2015). Among the available software to optimize this placement, there is the Packmol software (MARTÍNEZ *et al.*, 2009). It treats the initial configuration problem as a packing optimization problem. The molecules are packed in such way that atoms from different molecules keep a safe pairwise distance and, due to the optimization of function and gradient evaluations, this strategy enables the packing of millions of atoms in reasonable time (MARTÍNEZ *et al.*, 2009).

Independently of the technique or software used, certain restrictions should be

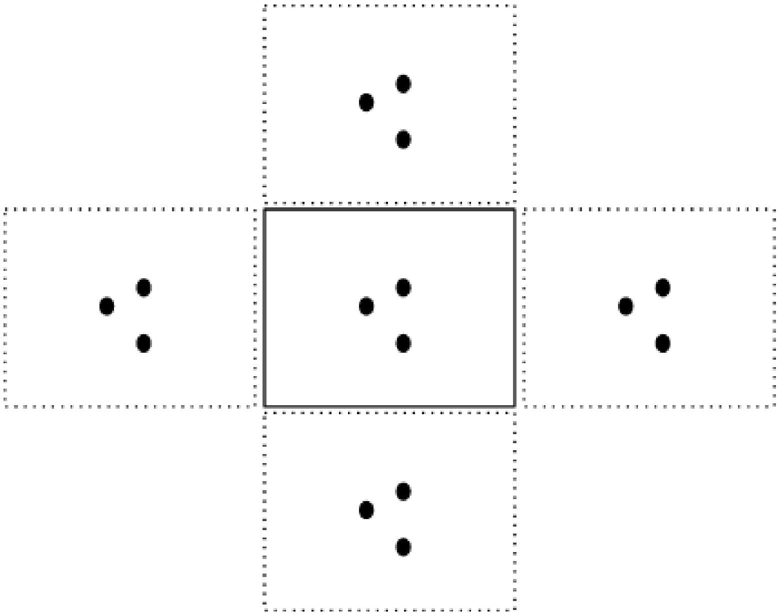


Figure 3.1.1 – Periodic boundary condition representation.

applied in the initial configurations to carry out molecular simulations. As an example, the cubic lattice has a finite size, but a finite 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 N particles, replicated in a periodic lattice of infinite cells as represented in Figure 3.1.1.

The application of the PBC results in particles interacting not only with each other but also with their images. This fact greatly increases the number of interacting pairs and, consequently, the computational time. To overcome that, we need to choose a limited range potential using the minimum image convention criterion. This criterion only allows a particle to interact with the nearest particle or image. This is technically done during the simulations by neglecting the interactions between two particles at or beyond a maximum length, which is given the name of cutoff radius (R_c). This cutoff should be equal to or less than half of the box length. This process of examining each pair separation can also be expensive depending on the number of distinct pairs. That is the reason molecular dynamics algorithms use pair listings. This method defines a 'skin' around the cutoff radius with a radius R_{List} (Figure 3.1.2). The pair list is initially built of all the neighbor particles within a distance R_{List} of each particle. Over the course of the simulation, only pairs in the pair list have their separation checked. This list can then be reconstructed in a specific time interval during the simulation but is essential to update this list before any unlisted pairs have come within R_{List}.

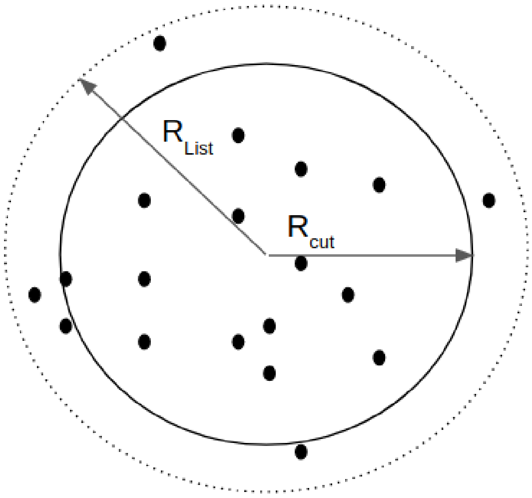


Figure 3.1.2 – Representation of the cutoff radius and the pair list radius.

3.1.6 Force Fields

Force fields are models used to describe structural characteristics such as van der Waals interactions, bond lengths, bond angles, and torsion. The description is done by approximating the potential energy function [$U(r^N)$], which has contributions due to intermolecular and intramolecular interactions. The intramolecular interactions include bond stretching, angle bending, and bond torsion (Figure 3.1.3).

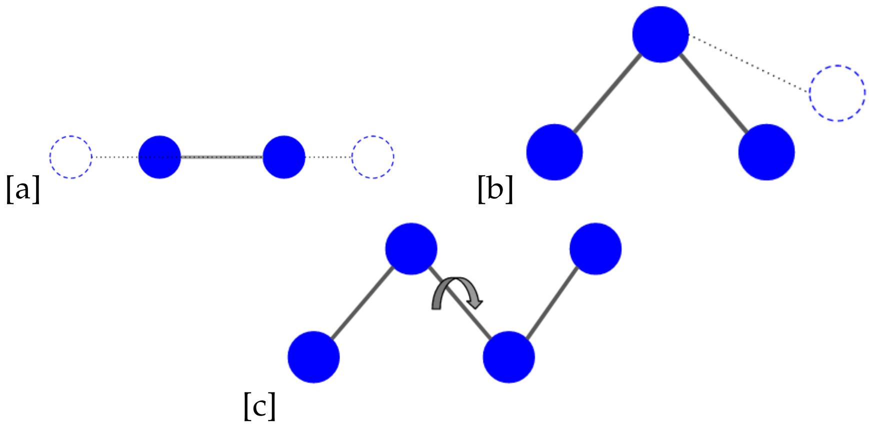


Figure 3.1.3 – Representation of bond stretching [a], angle bending [b], and bond torsion [c].

At this dissertation, we are going to present the equations which are most used to represent these interactions. The contribution to the bond stretching (bs) is usually given by the harmonic approximation around the energy minimum:

$$u_{bs}(d) = k_{bs}(d - d_0)^2. \tag{3.51}$$

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

$$u_{ab}(\theta) = k_{ab}(\theta - \theta_0)^2, \tag{3.52}$$

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 (bt) interactions correspond to the energies of rotations around bonds, and it happens among four atoms. A commonly used model is

$$u_{bt}(\omega) = \sum_{n=0}^{N} c_n cos^n(\omega), \tag{3.53}$$

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

The other type of interactions, the intermolecular interactions, include electrostatic and van der Waals interactions. The first one represents the interaction between two atoms i and j with partial charges ($q_i and q_j$) and they are usually represented by Coulomb's Law:

$$u_q(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \tag{3.54}$$

In the equation above, ϵ_0 is the free space permittivity constant and r_{ij} is the distance between atoms i and j. In many force fields, the van der Waals interaction between particles i and j is modeled by the Lennard-Jones Potential:

$$u_{LJ}(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right], \tag{3.55}$$

where ϵ is the depth of the potential well, σ is the distance correspondent to a zero intermolecular potential. The graphical representation of the Lennard-Jones potential is presented in Figure 3.1.4.

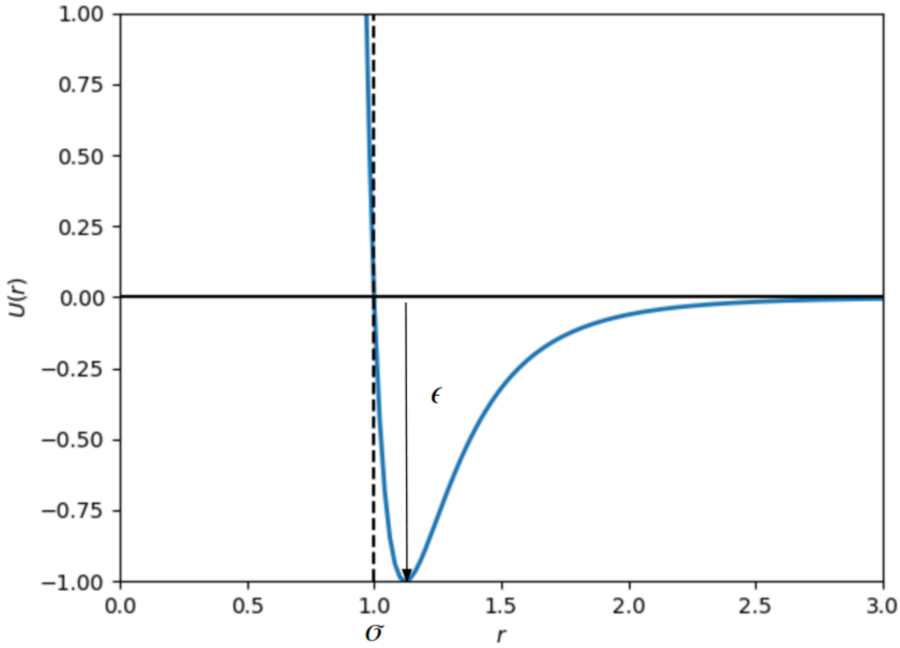


Figure 3.1.4 – Lennard-Jones potential representation for $\sigma = 1$ and $\epsilon = 1$.

The potential in Figure 3.1.4 tends to zero and becomes negligible after a specific value of r. Hence, we need to set a cutoff radius in which the potential energy is considered to be zero after it. The point in which the cutoff is defined is generally the one in which the radial distribution function ($g(r)$) is approximately constant. Also, only interactions with the nearest periodic image of the cell are considered for short-range interactions (minimum image convention condition) **??**. With this conditions, the calculations of forces and velocities are computationally feasible.

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

$$U(r^N) = u_{bs}(d) + u_{ab}(\theta) + u_{bt}(\omega) + u_q(r_{ij}) + u_{LJ}(r_{ij}). \tag{3.56}$$

3.2 SAFT-γ Mie Force Field

3.2.1 SAFT-VR Mie Equation of State (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 molecules formed from fused segments interacting via 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

$$U_{Mie}(r) = \epsilon \frac{\lambda_r}{\lambda_r - \lambda_a} \left(\frac{\lambda_r}{\lambda_a}\right)^{\left(\frac{\lambda_a}{\lambda_r - \lambda_a}\right)} \left[\left(\frac{\sigma}{r}\right)^{\lambda_r} - \left(\frac{\sigma}{r}\right)^{\lambda_a}\right], \tag{3.57}$$

where λ_r is the repulsive exponent and λ_a is the attractive exponent. The SAFT-VR Mie 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

$$\frac{A}{N\kappa_b T} = a = a^{IDEAL} + a^{MONO} + a^{CHAIN}, \tag{3.58}$$

or, depending on the molecule type, equal to

$$\frac{A}{N\kappa_b T} = a = a^{IDEAL} + a^{MONO} + a^{RING}. \tag{3.59}$$

Here, a^{IDEAL} is the ideal contribution for a mixture. It is given by

$$a^{IDEAL} = \sum_{i=1}^{N_c} x_i \ln\left(\rho_i {\Lambda_i}^3\right) - 1, \tag{3.60}$$

where $x_i = N_i/N$ is the molar fraction of component i, N_i is the number of molecules of each component, $\rho_i = N_i/V$ is the number density and Λ_i^3 is the de Broglie thermal wavelength. Also in Eq. 3.58, a^{MONO} is the monomer contribution, which describes interactions between Mie segments and can be expressed, for a mixture, as

$$a^{MONO} = \left(\sum_{i=1}^{N_c} x_i m_{s,i}\right) a^M. \tag{3.61}$$

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):

$$a^M = a^{HS} + \beta a_1 + \beta {a_2}^2 + \beta {a_3}^3. \tag{3.62}$$

Here, a^{HS} is the hard-sphere dimensionless Helmholtz free energy for a mixture and is given by:

$$a^{HS} = \frac{6}{\pi \rho_s}\left[\left(\frac{\zeta_2^3}{\zeta_3^2} - \zeta_0\right)\ln(1-\zeta_3) + \frac{3\zeta_1\zeta_2}{1-\zeta_3} + \frac{\zeta_2^3}{\zeta_3(1-\zeta_3)^2}\right]. \tag{3.63}$$

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:

$$\zeta_l = \frac{\pi \rho_s}{6} \left(\sum_{i=1}^{N_c} x_{s,i} d_{ii} \right), \quad l = 0, 1, 2, 3, \tag{3.64}$$

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

$$x_{s,i} = \frac{m_{s,i} x_i}{\sum_{k=1}^{N_c} m_{s,k} x_k}. \tag{3.65}$$

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

$$d_{ii} = \int_0^{\sigma_{ii}} \left\{ 1 - \exp\left[-\beta U_{ii}^{Mie}(r) \right] \right\} dr. \tag{3.66}$$

The integral in Eq. (3.66) is normally obtained by means of a 5-point Gauss-Legendre quadrature (PAPAIOANNOU *et al.*, 2014). For brevity, the detailing of the other terms of Eq. (3.62) are available at the Appendix A. The term a^{CHAIN} in Eq. 3.58 corresponds to the chain contribution. This chain formation of m_s tangentially bonded Mie segments is based on the first-order perturbation theory (TPT1) (PAPAIOANNOU *et al.*, 2014) and can be expressed as

$$a^{CHAIN} = - \sum_{i=1}^{N_c} x_i (m_{s,i} - 1) \ln \left[g_{ii}^{Mie}(\sigma_{ii}) \right]. \tag{3.67}$$

The term $g_{ij}^{Mie}(\sigma_{ij})$ correspond to the radial distribution function (RDF) of the hypothetical Mie system evaluated at the effective diameter. It is obtained with the following perturbation expansion

$$g_{ij}^{Mie}(\sigma_{ij}) = g_{d,ij}^{HS}(\sigma_{ij}) \exp \left[\beta\epsilon \frac{g_{1,ij}(\sigma_{ij})}{g_{d,ij}^{HS}(\sigma_{ij})} + (\beta\epsilon)^2 \frac{g_{2,ij}(\sigma_{ij})}{g_{d,ij}^{HS}(\sigma_{ij})} \right]. \tag{3.68}$$

In the equation above, $g_{d,ij}^{HS}$ is equal to

$$g_{d,ij}^{HS}(\sigma_{ij}) = \exp(k_0 + k_1 x_{0,ij} + k_2 x_{0,ij}^2 + k_3 x_{0,ij}^3), \tag{3.69}$$

where $x_{0,ij} = \sigma_{ij}/d_{ij}$ and k_1, k_2, and k_3 are density dependent coefficients. These coefficients and the other terms of Eq. 3.68 are available at Appendix A.

The ring contribution (a^{RING}) in Eq. 3.59 have two forms for rings formed from m_s tangentially bonded segments. The first one (LAFITTE *et al.*, 2012) considers 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. (3.67) can be adapted to rings by

$$a^{RING} = -\sum_{i=1}^{N_c} x_i m_{s,i} \ln[g_{ii}^{Mie}(\sigma_{ii})]. \tag{3.70}$$

According to Lafitte *et al.* (2012), Eq. (3.70) needs an additional parameterization with molecular simulation data so that the EoS can be used in molecular simulations, but this additional parameterization is not necessary when we are modeling chain molecules. Recently, Müller and Mejía (2017) tried to correct this inconsistency. They developed a ring free energy equation based on the work of Müller and Gubbins (1993), who obtained rigorous expressions for hard-sphere fluids with molecular geometries of rings with $m_s = 3$. The final expression developed for the dimensionless Helmholtz free energy due to ring formation is

$$a^{RING} = -\sum_{i=1}^{N_c} x_i \left(m_{s,i} - 1 + \chi_i \eta_i\right) \ln\left[g_{ii}^{Mie}(\sigma_{ii})\right], \tag{3.71}$$

where $\eta_i = m_{s,i}\rho_i\sigma_{ii}^3/6$ is the packing fraction of the atom i 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. (3.71) is equal to Eq. (3.67). In addition, the equation corresponds to a system of hard sphere triangles when $\chi = 1.3827$. Müller and Mejía (2017) also calculated values of ζ for $m_s = 3, m_s = 4, m_s = 5, and m_s = 7$ with pseudo-experimental data from molecular dynamics (MD) for a defined pure fluid with $\epsilon/\kappa_b = 250K$, $\sigma = 3.0\dot{A}$, $\lambda_r = 11$, and $\lambda_r = 6$. Values of χ for some of the geometry estimated in the article can be seen in Figure 3.2.1.

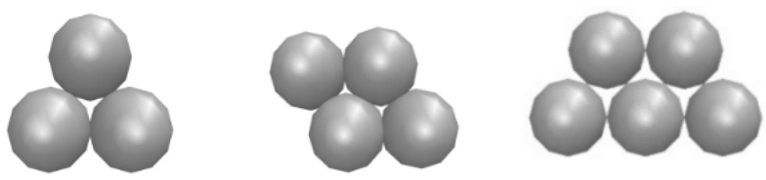


Figure 3.2.1 – Values for parameter χ according to the ring geometry. Adapted from Müller and Mejía (2017).

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

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2}, \tag{3.72}$$

$$d_{ij} = \frac{d_{ii} + d_{jj}}{2}, \tag{3.73}$$

$$\lambda_{k,ij} - 3 = \sqrt{(\lambda_{k,ii} - 3)(\lambda_{k,jj} - 3)}, \quad k = r, a, \tag{3.74}$$

$$\epsilon_{ij} = (1 - k_{ij}) \frac{\sqrt{\sigma_{ii}^3 \sigma_{jj}^3}}{\sigma_{ij}^3} \sqrt{\epsilon_{ii}\epsilon_{jj}}. \tag{3.75}$$

The k_{ij} is a binary interaction parameter to correct the deviations of the mixing rule for chemically distinct compounds that can be fitted to experimental or molecular simulation data. This necessity of an additional parameter brings the question of the quality of these mixing rules and the necessity of the new mixing rules to describe the mixing potential well parameter. Since these were the available mixing rules and the ones used by other papers that used this force field, we ended up using Eqs. 3.72 to 3.75 in our study. For our mixtures, the binary interaction parameter was only necessary for aqueous mixtures, and the k_{ij} was obtained with molecular simulation data.

3.2.2 Parameter Estimation for the SAFT-γ Mie Force Field

The SAFT-γ Mie Force Field uses a top-down coarse-graining 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 in 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 been used to parametrize a variety of fluids (ERVIK; MEJÍA; MÜLLER, 2016). This force field has the advantage of incorporating the degrees of freedom provided by the use of the Mie Potential (HERDES; TOTTON; MÜLLER, 2015). This flexibility offers the exploration of a vast parameter space without using an iterative simulation scheme (AVENDAÑO *et al.*, 2011). Despite these advantages, the force field can be restricted by the shortcomings of the equation of state. As an example, the lack of an association term in the equation can cause an inadequate representation of the properties of hydrogen bonding compounds.

Each substance has initially five parameters to be estimated ($m_s,\ \sigma,\ \epsilon,\ \lambda_r,\ and\, \lambda_a$) according to Eq. (3.57). 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 chosen value for this parameter is 6, corresponding to the London model, which is a good representation of the dispersion scale of most simple fluids that do not 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 such as vapor pressure, liquid density and the other one is by using correspondent state parametrization. The first was the one followed in this dissertation to obtain the parameters for. Generally, this approach minimizes the following unweighted least-squares objective function:

$$\min_{\sigma,\epsilon,\lambda_r} F_{obj} = \sum_{i=1}^{N_p} \left[\frac{P_v^{SAFT}(T_i, \sigma, \epsilon, \lambda_r) - P_v^{exp}(T_i)}{P_v^{exp}(T_i)} \right]^2 + \sum_{i=1}^{N_p} \left[\frac{\rho_l^{SAFT}(T_i, \sigma, \epsilon, \lambda_r) - \rho_l^{exp}(T_i)}{\rho_l^{exp}(T_i)} \right]^2, \tag{3.76}$$

where N_p is the number of experimental points, P_v is the vapor pressure and ρ_l is the saturated liquid density. Other properties that can be used in the estimation are interfacial tension and speed of sound, for instance. 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 choice of parameter m_s) and the subsequent parameter space so as to avoid some physical inconsistencies such as a premature freezing (LOBANOVA *et al.*, 2015).

Lafitte *et al.* (2012) suggested that two correction factors (c_σ and c_ϵ) should be estimated with simulation data when using Eq. (3.70) for the ring contribution. They are related to the EoS parameters by scaled parameters:

$$\sigma^{scaled} = c_\sigma \sigma^{SAFT}. \tag{3.77}$$

$$\epsilon^{scaled} = c_\epsilon \epsilon^{SAFT}. \tag{3.78}$$

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. (3.70). However, this new parameterization is not necessary when using Eq. (3.71) as the ring contribution or when we are modeling chain molecules with Eq. 3.67. This fact makes the strategy of Lafitte *et al.* (2012) inconsistent since parameterization with molecular simulation should not be necessary according to the overall idea of this force field. Furthermore, the use of molecular simulation data highly increases the time spent on the parameterization process. The objective function for the estimation of the correction parameter is given by

$$\min_{c_\sigma, c_\epsilon} F_{obj} = \sum_{i=1}^{N_p} \left[\frac{P_v^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT}) - P_v^{SAFT}(T_i, \sigma^{scaled}, \epsilon^{scaled})}{P_v^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT})} \right]^2 + \sum_{i=1}^{N_p} \left[\frac{\rho_{liq}^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT}) - \rho_{liq}^{SAFT}(T_i, \sigma^{scaled}, \epsilon^{scaled})}{\rho_{liq}^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT})} \right]^2 . \tag{3.79}$$

The repulsive parameter is maintained in the value found on the minimization of Eq. (3.76). The refined values for σ and ϵ are

$$\sigma^{sim} = \sigma^{SAFT} / c_\sigma, \tag{3.80}$$

$$\epsilon^{sim} = \epsilon^{SAFT} / c_\epsilon, \tag{3.81}$$

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 the following mean-field approximation

$$a_1 = 2\pi\rho\sigma^3\epsilon\alpha. \tag{3.82}$$

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

$$\alpha = \frac{1}{\epsilon\sigma^3}\int_\sigma^\infty \phi(r) r^2 dr = \frac{\lambda_r}{3(\lambda_r - 3)}\left(\frac{\lambda_r}{6}\right)^{6/(\lambda_r - 6)} . \tag{3.83}$$

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:

$$\lambda_r = \frac{\sum_{i=0} a_i \omega^i}{1 + \sum_{i=1} b_i \omega^i}, \tag{3.84}$$

where a_i and b_i are parameters that are dependent on 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 by substituting λ_r into Eq. (3.83). The reduced critical temperature T_c^* is related to α by a Padé series:

$$T_c^* = \frac{\sum_{i=0} c_i \alpha^i}{1 + \sum_{i=1} d_i \alpha^i}. \tag{3.85}$$

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:

$$T_c^* = \frac{\kappa_b T_c}{\epsilon}. \tag{3.86}$$

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

$$\rho^*_{T_r=0.7} = \frac{\sum_{i=0} j_i \alpha^i}{1 + \sum_{i=1} k_i \alpha^i}. \tag{3.87}$$

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

$$\rho^*_{T_r=0.7} = \rho_{T_r=0.7} \sigma^3 N_{av}, \tag{3.88}$$

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. (3.75) is necessary to adjust the mixture behavior of chemically distinct components. Normally, it is estimated by 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:

$$\begin{aligned} \min_{k_{ij}} F_{obj} = \sum_{k=1}^{N_p} \left(\frac{P_v^{SAFT}(T_k, x, k_{ij}) - P_v^{exp}(T_k, x)}{P_v^{exp}(T_k, x)} \right)^2 + \\ \sum_{k=1}^{N_p} \left(\frac{\rho_l^{SAFT}(T_k, x, k_{ij}) - \rho_l^{exp}(T_i)}{\rho_l^{exp}(T_i)} \right)^2. \end{aligned} \tag{3.89}$$

However, Ervik *et al.* (2016) used molecular simulation results to fit the parameter to the interfacial tension data. The strategy they followed was to carry out simulations in three values of k_{ij} first and, after, refine the parameter until a value in good agreement with the experimental data is found. We decided to follow this strategy in our estimations of k_{ij} since the estimation with the EoS did not provide satisfactory results for the hydration free energy calculations.

3.3 Solvation Free Energy Simulations Based on Molecular Dynamics

Using the SAFT-γ Mie Force Field described in the section above, we carried solvation free energy molecular dynamic simulations. The free energies we are trying to calculate 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

$$F(N, V, T) = -\kappa_b T \ln Q(N, V, T). \tag{3.90}$$

Recall that $Q(N, V, T)$ is the partition function of the canonical ensemble, expressed as

$$Q(N, V, T) = \frac{\epsilon_0}{h^{3N} N!} \int d^n p d^n r \exp\left[-\beta \left(\sum_{i=1}^{N} \frac{p_i^2}{2m_i} + U(r_1, .., r_n) \right)\right]. \tag{3.91}$$

The Gibbs free energy, the object of study in this dissertation, is given by

$$G(N, P, T) = -\kappa_b T ln \Delta(N, P, T), \tag{3.92}$$

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

$$\Delta(N, P, T) = \frac{1}{V_0} \int_0^\infty dV \int d^n p d^n r \exp\left[-\beta \left(\sum_{i=1}^{N} \frac{p_i^2}{2m_i} + U(r_1, .., r_n) + PV(r_1, .., r_n) \right)\right]. \tag{3.93}$$

Evaluating the partition function is an often unfeasible task, but we are interested in calculating only the Gibbs free energy difference between two states of a system, which is

$$\Delta G_{AB} = G_B - G_A = -\kappa_b T ln \left(\frac{\Delta_B}{\Delta_A} \right). \tag{3.94}$$

Since the masses of particles in at states A and B of a system are the same and the Hamiltonian is separable in $K(p)$ and $U(r)$, the moment integrals in the ratio Δ_B/Δ_A can be simplified into to the ratio of the configuration integrals:

$$\frac{Z_B}{Z_A} = \frac{\int_0^\infty dV \int d^n r \exp\left\{-\beta\left[U(r_1,..,r_n) + PV(r_1,..,r_n)\right]_B\right\}}{\int_0^\infty dV \int d^n r \exp\left\{-\beta\left[U(r_1,..,r_n) + PV(r_1,..,r_n)\right]_A\right\}}. \tag{3.95}$$

This identity results in the following equation for the Gibbs free energy difference, which does not require the calculation of the partition function at each state:

$$\Delta G_{AB} = G_B - G_A = -\kappa_b T ln\left(\frac{Z_B}{Z_A}\right). \tag{3.96}$$

In the case of a study concerning the solvation of a single molecule, 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 have accurate results for free energy differences, the phase space 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 is independent of the path chosen since free energy is a state function. That is why alchemical states (without physical sense) can be used to link physical states of interest. The solvation free energy calculations are done through a thermodynamic path in which the solute molecule is gradually inserted into the solvent as illustrated in Figure 3.3.1. According to this path, the free energy of solvation is expressed as

$$\Delta G_{solv} = \Delta G_{1\to 4} = \Delta G_{1\to 2} + \Delta G_{2\to 3} + \Delta G_{3\to 4} - \kappa_b T \ln\frac{V^*}{V^1}. \tag{3.97}$$

The last term in Eq. 3.97 accounts for the difference between the mean volume of the box with the solute inserted (V_1) and the mean volume of the box with only solvent molecules in it (V^*). Shirts *et al.* (2003) have shown that this term is negligible with respect to the statistical uncertainty of calculating ΔG. However, the post-processing software used in this dissertation included this term in the estimation of the solvation free energy differences. Hence, this difference of volume was accounted for in our calculations. When we have other solute molecules in our box, another term in Eq. 3.97 can appear in order to distinguish the inserted solute molecule from the other molecules of solute in the box (SHIRTS *et al.*, 2003). The term $\Delta G_{1\to 2}$, also represented in Figure 3.3.1, is the solvation free energy associated with turning off the molecule's non bonded interactions in the gas phase. The next transformation, $\Delta G_{2\to 3}$ is the free energy of moving the non-interacting molecule from the gas phase to the solvent, and

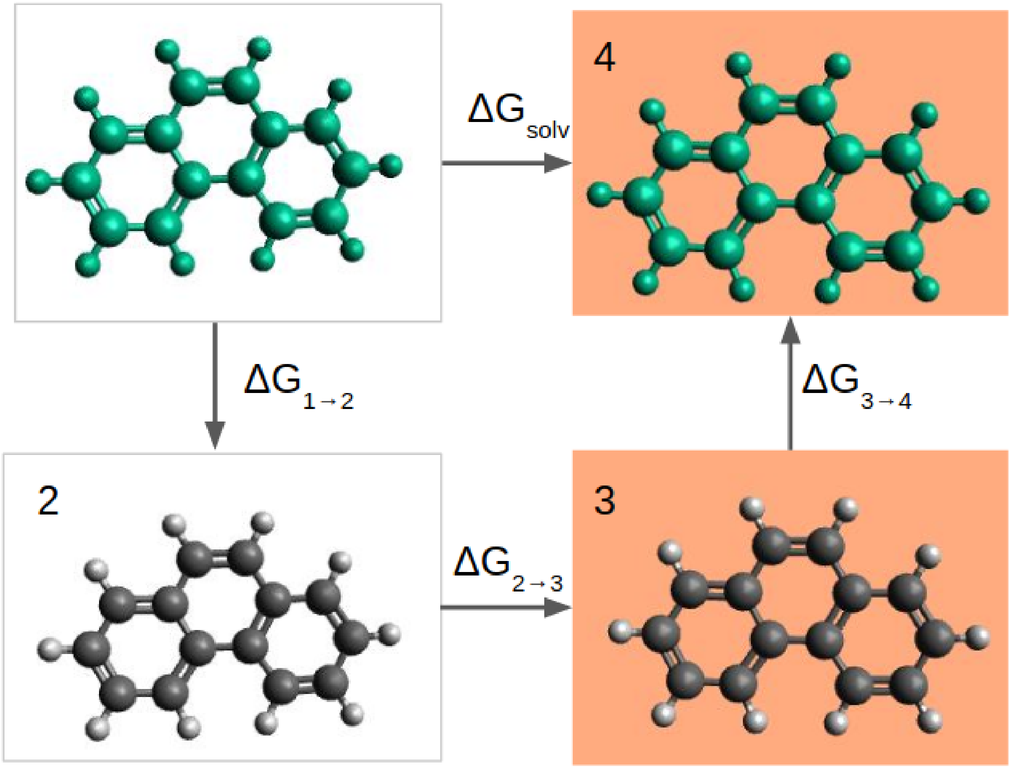


Figure 3.3.1 – Thermodynamic cycle for computing solvation free energy of a single solute molecule with molecular dynamics. Adapted from Klimovich, Shirts and Mobley (2015).

is equal to zero since the transformation of a non-interacting molecule does not depend on the environment. Lastly, $\Delta G_{3\to 4}$ is the free energy required for the non-interaction molecule in the aqueous phase to regain its non-bonded interactions with the solvent. The solvation free energy calculation can be classified according to the types of non-bonded interactions that are turned off/on in the $1 \to 2$ and $3 \to 4$ parts of the path. If both non-bonded interactions with the environment and internal interactions are turned off, this is an annihilation free energy calculation. On the other hand, if only non-bonded interactions with the environment are turned off, this is a decoupling free energy calculation (KLIMOVICH; SHIRTS; MOBLEY, 2015). In the latter case, $\Delta G_{1\to 2} = 0$ and the $\Delta G_{solv} = \Delta G_{3\to 4}$.

The methods used to carry out the transformations of Figure 3.3.1 during the simulation scale the solute charges to zero and then turn off the interactions corresponding to the Lennard-Jones or Mie potential. In order to carry out the latter process, a modified potential with a coupling parameter (λ) is used. Each λ represents an alchemical state. 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 (SHIRTS *et al.*, 2003). We can see in Figure 3.3.2 that a linear coupling of the parameter would cause an abrupt change in the potential energy when the λ reaches zero. That is the reason the non-linear softcore scheme (BEUTLER *et al.*, 1994) is used to couple the λ. This scheme makes the potential behave more smoothly in relation to the change of λ, as can be seen in Figure 3.3.3. The softcore potential is

$$U_{LJ}^{sc}(r) = 4\lambda\epsilon \left\{ \frac{1}{[\alpha(1-\lambda) + (r/\sigma)^6]^2} - \frac{1}{\alpha(1-\lambda) + (r/\sigma)^6} \right\}, \tag{3.98}$$

where α is a constant whose value is normally assumed to be 0.5. Based on Eq. 3.98, we propose a generalized softcore Mie potential for any value of λ_r and λ_a. It is given by:

$$U^{sc}(r) = \lambda\epsilon \frac{\lambda_r}{\lambda_r - \lambda_a} \left(\frac{\lambda_r}{\lambda_a}\right)^{\left(\frac{\lambda_a}{\lambda_r - \lambda_a}\right)} \left\{ \frac{1}{[\alpha(1-\lambda) + (r/\sigma)^{\lambda_a}]^{\lambda_r/\lambda_a}} - \frac{1}{\alpha(1-\lambda) + (r/\sigma)^{\lambda_a}} \right\}. \tag{3.99}$$

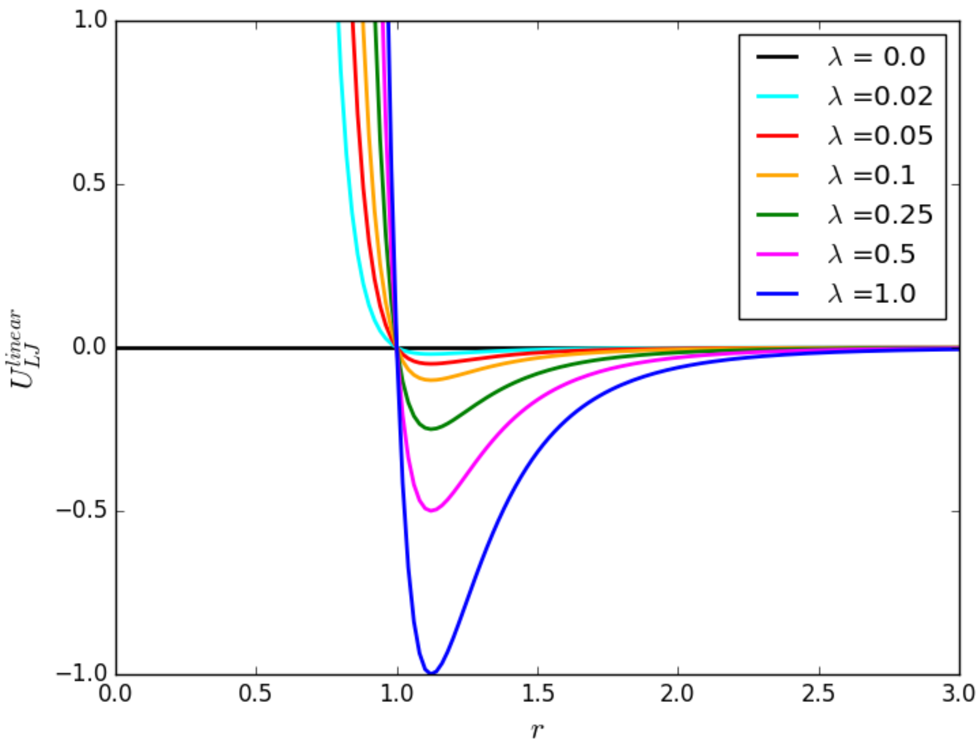


Figure 3.3.2 – Linear coupling of the potential energy, $U_{LJ}^{linear} = \lambda U_{LJ}$, for different values of λ in reduced units.

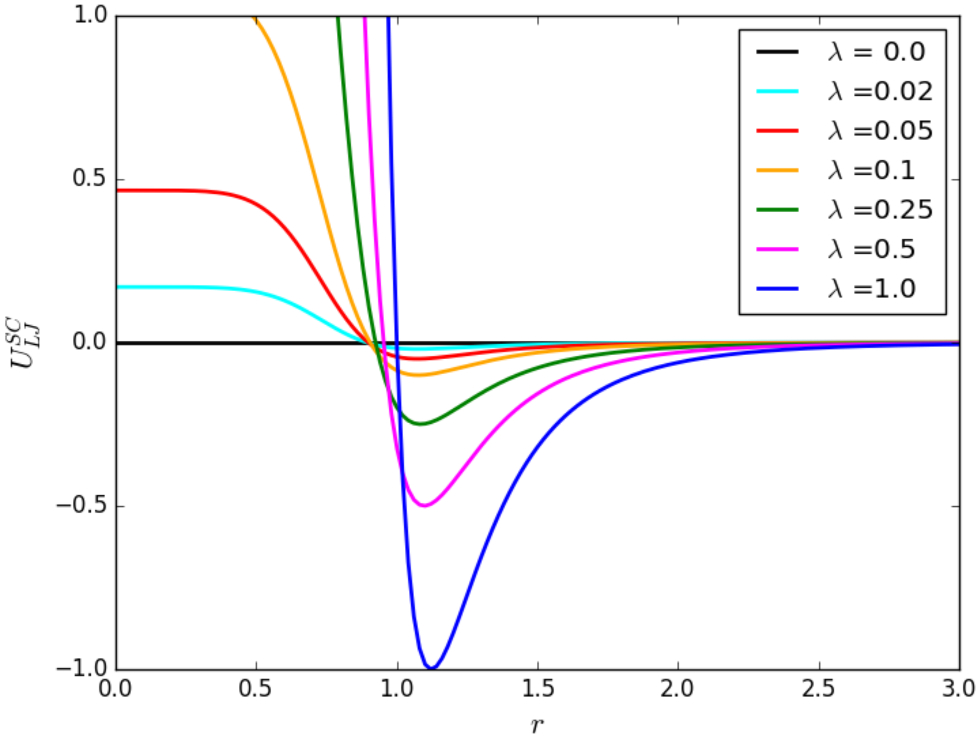


Figure 3.3.3 – Softcore potential, Eq. 3.98, for different values of λ in reduced units.

Now that we defined our coupled potential, we can then obtain the potential energies related to each alchemical state by doing independent simulations in different values of λ or by doing expanded ensemble simulations (LYUBARTSEV *et al.*, 1992). The latter was the method used in this dissertation and it is described in Section 3.4. Having the potential energies, the next step is to use post-processing methods, such as the MBAR used in this study, to effectively calculate $\Delta G_{3\to 4}$. The solvation free energy differences can then be used to calculate other properties such as the partition coefficient. This property is a measure of the partitioning of one solute in two solvents (a and b) with different physicochemical characteristics at a temperature T. It is defined by the following equation when the activity coefficients are assumed to be one:

$$P = \frac{[solute]_a}{[solute]_b}, \tag{3.100}$$

where $[solute]_a$ and $[solute]_b$ are the concentration of the solute in the solvent a and b, respectively. Since P is an equilibrium constant, it can be related to free energy change associated with transferring the solute from the phase a to the phase b. Hence, we can define the relation between the partition coefficient and the difference in free energy with the equation bellow (ESSEX; REYNOLDS; RICHARDS, 1992):

$$2.303RT\log P^{a/b} = \Delta G^{a}_{solv} - \Delta G^{b}_{solv}, \tag{3.101}$$

where the factor 2.303 is used to convert P into $\log P$.

3.4 Expanded Ensemble Method

We decided to use the Expanded Ensemble method (LYUBARTSEV *et al.*, 1992) in our solvation free energy simulations since it allows a non-Boltzmann sampling scheme of different states in a single simulation. Lyubartsev *et al.* (1992) initially proposed in their paper a sampling scheme of different temperatures, but this idea can be generalized to a sampling scheme of different states or $\lambda's$ (ESCOBEDO; MARTINEZ-VERACOECHEA, 2007). In this scheme, the sampling is done by biasing the phase space exploration process with weights not related to the statistical ensemble. The partition function of 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 subensembles Z_i in different values of λ, that is,

$$Z^{EE} = \sum_{i=1}^{N} Z_i(\lambda_i)exp(\eta_i), \tag{3.102}$$

where N is the number of alchemical states, η_i is the arbitrary weight of the subensemble at each state, and Z_i is the configurational partition function of state i. For the isothermal-isobaric ensemble, Z_i is given by

$$Z_i = \frac{1}{V_0}\int_0^{\infty} dV \int d^n r^n \exp\left\{-\beta_i\left[U(\lambda, r_1, .., r_n) + P_i V(r_1, .., r_n)\right]\right\}. \tag{3.103}$$

Since we are carrying out 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 this 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:

$$\pi(x|\lambda_k) = \frac{\exp[-\beta u(x, \lambda_k)]}{\int dx \exp[-\beta u(x, \lambda_k)]}. \tag{3.104}$$

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

$$\pi(\lambda_k|x) = \frac{\exp[-\beta u(x,\lambda_k) + \eta_k]}{\sum_{k=1}^{K} \exp[-\beta u(x,\lambda_k) + \eta_k]}, \tag{3.105}$$

where $u(x, \lambda_k)$ is the reduced potential function for the NPT ensemble. There is a variety of acceptance schemes to do the expanded sampling using Eq. (3.105), 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 they suggested consist of updating 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

$$R < \sum_{i=1}^{j} \pi(\lambda_i|x). \tag{3.106}$$

The sampling strategy above depends on a proper selection of weights in order to assure an adequate sampling of the states. If there is not a sufficient number of visits to each state, the expanded ensemble becomes deficient in obtaining input data to estimate free energy differences with the methods exposed in Section 2.4. Here, we followed the flat-histogram approach (BERG; NEUHAUS, 1992; LEE, 1993; DAYAL *et al.*, 2004) to calculate the weights. This strategy aims to obtain adequate sampling by ensuring that all the states have an equal number of visits, i.e. the ratio of the probability of sampling state i (π_i) to the probability of sampling state j (π_j) is equal to one. Given that π_i is equal to:

$$\pi_i = \frac{Z_i(\lambda_i)exp(\eta_i)}{Z^{EE}}, \tag{3.107}$$

and using Eqs. 3.96 and 3.95, the following relation can be obtained for $\pi_i/\pi_j = 1$:

$$(\eta_i - \eta_j) = \beta(G_i - G_j). \tag{3.108}$$

Eq. (3.108) proposes that the choice of the new weights is dependent on the free energies that we are attempting to obtain. This equation is then 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. (3.108), the next values of η can be calculated. This iteration goes on until a uniform distribution is attained. 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 free energy differences. In this work, the method chosen to obtain the optimal staging of the λ domain is the one developed by Escobedo and Martinez-Veracoechea (2007) with a basis in the study of Katzgraber *et al.* (2006). 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 (0) and highest (1) values of λ. This is specifically done by maximizing the steady-state stream ϕ of the simulation, which "walks" among the values of λ. This flow is estimated from a Fick's diffusion type of law:

$$\phi = D(\Lambda)\Pi(\Lambda)\frac{dx(\Lambda)}{d\Lambda}. \tag{3.109}$$

In the equation above, Λ is the actual continuous value of the coupling parameter. This continuous function of $\lambda's$ 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:

$$\Pi(\Lambda) = \frac{C'\bar{\Pi}(\lambda)}{\Lambda_{i+1} - \Lambda_i}. \tag{3.110}$$

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

$$\bar{\Pi}_i(\lambda) = \frac{\pi_{i+1} - \pi_i}{2}. \tag{3.111}$$

The ϕ is maximum when the optimal probability $\Pi'(\Lambda_i)$ of visiting state Λ_i is proportional to $1/\sqrt{D(\Lambda)}$ (TREBST; HUSE; TROYER, 2004). With that information, it is possible to estimate the diffusivity using one trial simulation with the following equation:

$$D(\Lambda) = \frac{\Lambda_{i+1} - \Lambda_i}{\bar{\Pi}(\lambda)dx(\Lambda)/d\Lambda}, \quad \Lambda_i < \Lambda < \Lambda_{i+1}. \tag{3.112}$$

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

$$\Phi = \int_{\lambda=0}^{\lambda=1} \Pi'(\Lambda_i)d\Lambda = \frac{i}{K}, \tag{3.113}$$

where K is the total number of λ states. We obtained these cumulative probabilities for every λ set we estimated in order to carry out our solvation free energy simulations. A graphical demonstration of the relation between the optimized coupling parameters and the cumulative probability of Eq. 3.113 is presented in Figure 3.4.1.

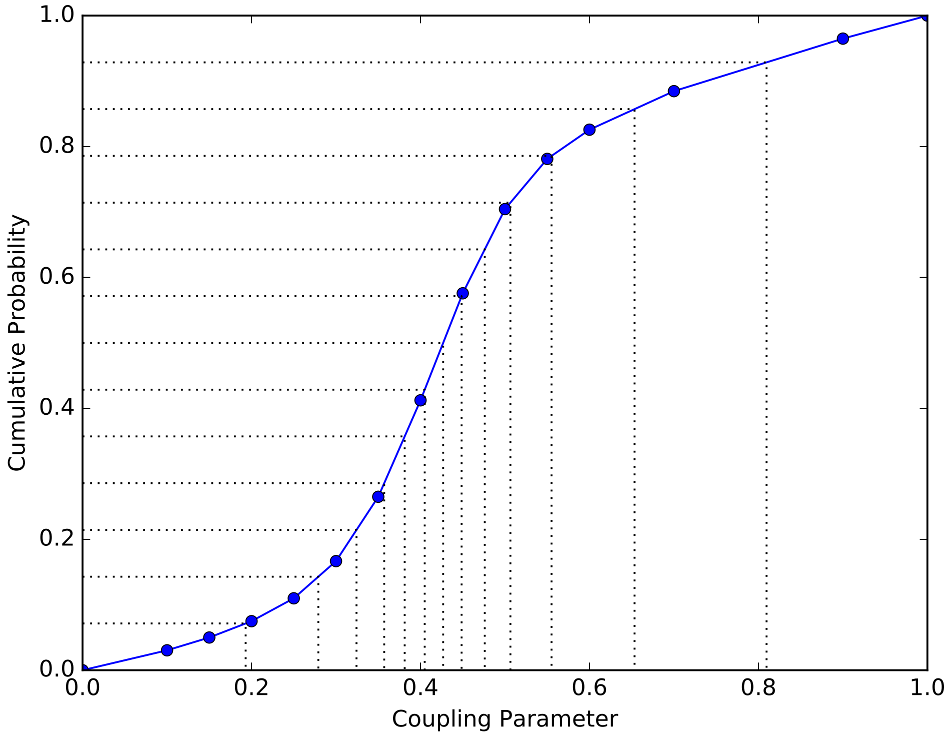


Figure 3.4.1 – Relation between the optimized coupling parameters and the cumulative probability used to obtain them.

3.5 Multistate Bennett Acceptance Ratio (MBAR)

We presented in the sections above the methods used to obtain the total potential energies of each alchemical state with molecular dynamics, but, in this section, we are going to discuss the methodology utilized to estimate solvation free energy differences with these data. The MBAR method (SHIRTS; CHODERA, 2008) is based on free energy of perturbation. It is a maximum likelihood method which proposes an estimator that computes free energies and their uncertainties of all K states by minimizing the $K \times K$ matrix of variances 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 are sampled:

$$p_i(x) = \frac{q_i(x)}{c_i}, \tag{3.114}$$

$$c_i = \int dx q_i(x), \tag{3.115}$$

where $q_i(x) = \exp[-u_i(x)]$ and u_i is the reduced potential energy of each state, defined for a an alchemical transformation by $u_i(x) = \beta_i[U_i(x) + P_iV(x) + \mu_i^T n(x)]$. In addition, c_i is a normalization constant. The free energies are estimated from the ratio of this constant in each state, since

$$\Delta f_{ij} = f_i - f_j = -\ln \frac{c_j}{c_i} = -\ln \frac{\int dx q_j(x)}{\int dx q_i(x)}. \tag{3.116}$$

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

$$c_i \langle \alpha_{ij} q_j \rangle_i = c_j \langle \alpha_{ij} q_i \rangle_j. \tag{3.117}$$

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

$$\sum_{j=1}^{K} \frac{\hat{c}_i}{N_i} \sum_{n=1}^{N_i} \alpha_{ij} q_j(x_{i,n}) = \sum_{j=1}^{K} \frac{\hat{c}_j}{N_j} \sum_{n=1}^{N_j} \alpha_{ij} q_i(x_{j,n}). \tag{3.118}$$

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

$$\alpha_{ij}(x) = \frac{N_j \hat{c_i}^{-1}}{\sum_{k=1}^{K} N_k c_i^{-1} q_k(x)}. \tag{3.119}$$

Assuming the sampling is carried out following Boltzmann statistics, Eqs. (3.118) and (3.119) can be rearranged to obtain the free energy estimator, which is solved self consistently:

$$f_i = \frac{1}{\beta} ln \sum_{k=1}^{K} \sum_{n=1}^{N_k} \frac{\exp[-\beta u_i(x_{kn})]}{\sum_{l=1}^{K} N_l \exp\{\beta[f_l - u_l(x_{kn})]\}}. \tag{3.120}$$

The equation above requires the evaluation of the potential energy of every uncorrelated configuration n for all K states $[u_i(x_{kn})]$ and for all uncorrelated configuration snapshots (N_k) from state k. With the free energies, we compute the free energy differences between states with Eq. 3.116. The statistical variance resulting from free energy estimation is given by the covariance matrix (s):

$$\delta_{ij}^2 s_{ij} = s_{ii}^2 + s_{jj}^2 - 2s_{ij}. \tag{3.121}$$

The MBAR method explained here can be considered as a limiting case of the Weighted Histogram Analysis Method (WHAM) (KUMAR *et al.*, 1992) for computing free energies. WHAM equations become equal to Eq. (3.120) if the histogram width tends to zero. Despite this, the MBAR is still more suited than the WHAM because it does not have the bias associated with the discretization and allows the calculation of an error estimate (SHIRTS; CHODERA, 2008).

3.6 Gibbs Ensemble Monte Carlo (GEMC)

In the initial steps of this research, we estimated the SAFT-γ Mie force field parameters of phenanthrene with the methodology proposed by Lafitte *et al.* (2012). This approach required liquid-vapor equilibrium data obtained with molecular simulation, as it is described in Section 3.2.2. Hence, we carried out Monte Carlo simulations at the Gibbs Ensemble (PANAGIOTOPOULOS, 1987) since this ensemble is commonly used to study phase coexistence with molecular simulation. In addition to that, this method does not use an explicit interface, which can hinder the determination of bulk phase behavior of small systems with long-range interactions (RAI; MAGINN, 2012).

Before talking in more detail about this ensemble, we are going to discourse on Monte Carlo simulations briefly. The Monte Carlo (MC) approach is another method for generating atomic trajectories in order to obtain macroscopic properties. Rather than using the numerical integration of Newton's equations of motion, the trajectories are obtained stochastically in the Monte Carlo approach. The positions are evolved by random moves or perturbations (MC steps) acquired 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 only depends on the states and their directly preceding states, but not on the previous configurations (RAABE, 2017). The random move is constructed in such a 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 as:

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

2. Perturb the particle by randomly displacing it, $r' = r + \Delta r$. Where Δr is a perturbation randomly 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:

$$acc_{A\rightarrow B} = min\{1, exp[-\beta U(r'^N) + \beta U(r^N)]\}. \tag{3.122}$$

The values of maximum displacement are defined iteratively in order to obtain acceptance rates of 25-50% in step 3 (FRENKEL, 2013). Monte Carlo simulations are interesting when we need to calculate properties in different thermodynamic ensembles, such as the Gibbs Ensemble used in this dissertation. The phase coexistence at this ensemble is obtained with simultaneous Monte Carlo (MC) simulations of two boxes with periodic boundary conditions, representing a two-phase system. 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 random 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. In turn, the simulation of single component systems is carried out at a 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 (FRENKEL; SMIT, 2001). The partition function of the GEMC-NVT ensemble is obtained by considering that the particles in both boxes are subjected to the same intermolecular interactions. Also, volumes and number of particles (N_1, N_2, V_1 and V_2) can vary while the total volume (V) and the total number of particles (N) remain constant ($N = N_1 + N_2, V = V_1 + V_2$). Thus, the partition function is

$$
\begin{aligned}
Q(NVT) \equiv \sum_{N_1}^{N} \frac{1}{V\Lambda^{3N} N_1!(N-N_1)!} \int_0^V V_1^{N_1} V_2^{N_2} dV_1 \\
\int \exp[-\beta U(x_1^{N_1})] dx_1^{N_1} \int \exp[-\beta U(x_2^{N_2})] dx_2^{N_2}.
\end{aligned}
\tag{3.123}
$$

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:

$$
\pi(x_1^{N_1}, x_2^{N_2}, N_1, N_2, V_1, V_2) \propto \frac{V_1^{N_1} V_2^{N_2}}{N_1! N_2!} \exp[-\beta U(x_1^{N_1}) - \beta U(x_2^{N_2})]. \tag{3.124}
$$

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

$$
acc_{A \to B} = \min\{1, \exp[-\beta U(x_A^{N_1}) - \beta U(x_B^{N_1})]\}. \tag{3.125}
$$

The volume exchange moves take place 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 (δV_{max}) defined with probability $1/\delta V_{max}$ (FRENKEL; SMIT, 2001). The acceptance rule for these moves is:

$$acc_{A\to B} = \min\left\{1, \left(\frac{V_1^B}{V_1^A}\right)^{N_1=1}\left(\frac{V_2^B}{V_2^A}\right)^{N_2+1}\exp[-\beta U(x_A^N) - \beta U(x_B^N)]\right\}. \quad (3.126)$$

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:

$$acc_{A\to B} = \min\left\{1, \frac{N_1 V_2}{N_2 V_1}\exp[-\beta U(x_A^N) - \beta U(x_B^N)]\right\}. \quad (3.127)$$

This method has been widely used to calculate phase equilibrium, but its performance is poor for the region near the critical point due to large density fluctuations. The GEMC method also has poor performance for dense systems since the particle exchange moves have a low acceptance rate (WESTMORELAND *et al.*, 2002).

4 Methodology

In this study, we had to first obtain the parameters of phenanthrene for the SAFT-γ Mie force field since they were not available on this force field database (ERVIK; MEJÍA; MÜLLER, 2016). Hence, this chapter is divided in two sections. The first one describes how we parametrized the phenanthrene molecule and the second one describes how we carried out the solvation free energy simulations.

4.1 Phenanthrene Parameterization

We implemented the two parameterization strategies for molecules with aromatic rings described in Section 3.2.2 for phenanthrene. For both of them, only vapor pressure data (MORTIMER; MURPHY, 1923) were used due to the unavailability of saturated liquid density. We did not estimate the attractive exponent, λ_a. Instead, the value of six was given to it due to its high correlation with the repulsive exponent. The parameterization with the ring equation of Müller and Mejía (2017) was carried out with the number of segments equal to five and with a geometry such as that in Figure 4.1.1, since this level of coarse-graining was also used for a similar molecule (anthracene) in the original paper.

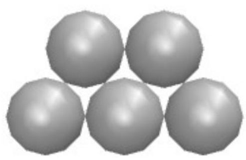


Figure 4.1.1 – Representation of phenanthrene with the geometry proposed by Müller and Mejía (2017).

The minimization was done using the Particle Swarm Optimization (PSO) method (SCHWAAB *et al.*, 2008) with the following objective function:

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

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

The parameterization with the ring equation the Lafitte *et al.* (2012) was carried out with $m_s = 3$, so that every bead would represent one aromatic ring, such as depicted in Fig:

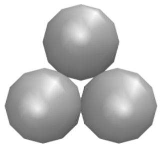


Figure 4.1.2 – Representation of phenanthrene with the geometry proposed by Lafitte *et al.* (2012).

The first part of the estimation followed the same procedure described above for the Müller and Mejía (2017) equation. However, 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. (3.77) and (3.78)). We then estimated these parameters by using the PSO method with Eq. (3.79). In this equation, vapor pressures and saturated liquid densities from molecular simulations are required. We then decided to use the Gibbs Ensemble Monte Carlo method on the NVT ensemble, explained in Section 3.6, to obtain these equilibrium properties at eight different temperatures since this method does not use an explicit interface.

The boxes for the GEMC-NVT simulations were generated by inserting 400 molecules of phenanthrene into one liquid box and 100 molecules of phenanthrene into the other one using the Playmol package (ABREU, 2017), which is integrated with the Packmol package (MARTÍNEZ *et al.*, 2009). Initial densities of each box were made equal to the saturated densities found with the SAFT-VR Mie Eos, aiming at avoiding the migration of all molecules to a single phase during the simulation. The GEMC-NVT simulations were carried using the Cassandra software (SHAH; MAGINN, 2011), which was developed to perform Monte Carlo simulations. The equilibration and production times lasted around 10^4 and 5×10^4 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 Å and we did not use 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. Later on, these data were divided into five blocks for calculation of their averages and standard deviations. With the correction factors found after the estimation with the simulation data, we calculated with Eqs. (3.80) and (3.81) the σ and ϵ parameters. Lafitte *et al.* (2012) proposes that are the final parameters to be used in molecular simulation. Hence, a iterative simulation is not required and the set of optimal parameters can be obtained with one group of molecular simulations.

4.2 Solvation Free Energy Simulations

Using the parameters for phenanthrene estimated with the Müller and Mejía (2017) approach and th SAFT-γ Mie force field parameters available for other compounds, we carried out molecular dynamic simulations to estimate solvation free energy differences. The chosen software package to perform the simulations was the LAMMPS (PLIMPTON, 1995). In this package, the equations of motion were integrated with the velocity-Verlet algorithm (VERLET, 1967) with a time step of 2 fs. As required by the coarse-grained model, molecules with more than one bead were treated as rigid bodies. The thermostat and the barostat were the Nosé Hoover chains as described originally in Hoover (1985) and Martyna, Klein and Tuckerman (1992) with damping factors of 100 and 1000 time steps, respectively. For the rigid bodies in our simulations, we used the rigid-body algorithm of Kamberaj, Low and Neal (2005). Electrostatics interactions are not explicitly accounted for the SAFT-γ Mie force field, hence we did not compute Coulombic interactions. The potential cutoff was equal to 20 Å (MÜLLER; MEJÍA, 2017) with a neighbor list skin of 2 Å. The initial configurations of the solvated systems were also generated using the Playmol package, which is integrated with the Packmol package. For the binary mixtures, one molecule of solute and a varying number of solvent molecules- 700 molecules of toluene, 700 molecules of octanol, 1024 molecules of hexane, 3000 molecules of water - were randomly added to a cubic box. Besides the systems with pure substances acting as solvents, we performed simulations to study solvation free energy of phenanthrene in a mixture of toluene and carbon dioxide with different weight fractions (w_{CO_2}). The system consisted of one molecule of phenanthrene for all the cases and 123 molecules of CO_2 and 618 molecules of toluene ($w_{CO_2} = 0.087$); 166 molecules of CO_2 and 589 molecules of toluene ($w_{CO_2} = 0.119$); 232 molecules of CO_2 and 545 molecules of toluene ($w_{CO_2} = 0.169$); 380 molecules of CO_2 and 446 molecules of toluene ($w_{CO_2} = 0.289$). The substances used in this study were selected with the intention of testing the force field with standard sets used as a benchmark in solvation free energy calculations, with aromatic substances used as models to asphaltenes and with water, which probably is the most used solvent in computational studies.

All simulations were performed with the constant temperature and pressure values of 298 K and 1 bar, except the ones containing carbon dioxide. These had the temperature of 298 K and the pressure of the experimental liquid-phase equilibrium correspondent to each composition of the system CO_2+toluene (CHANG, 1992). For all simulations, the initial box was equilibrated at the NPT ensemble for 2 ns, and 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 research group, available at https://github.com/atoms-ufrj/USER-ALCHEMICAL.

During these expanded ensemble simulations, the sampling of a new alchemical state was tried at every 10 MD steps. To define the optimal values of λ and η corresponding to each state, trial simulations, having around 9 ns of production time, were carried out. In the first simulation, we chose the group of λ values arbitrarily, and we either set all $\eta's$ to zero or assigned values previously found for similar solute-solvent pairs. The subsequent group of $\eta's$ were estimated with the flat histogram approach (Eq. (3.108)). We then did another trial simulation with the new weights. The results of this simulation were used to optimize the group of $\lambda's$ by minimizing the number of round trips, as described in Section 3.4. The $\eta's$ corresponding to the newest group of $\lambda's$ 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 employed force field considers that the beads do not have charges, there are no Coulombic interactions and the ΔG in Eq. (3.97) becomes equal to $\Delta G_{3\rightarrow 4}$. The post-processing method used to effectively calculate free energy differences with the potential energies obtained from the expanded ensemble simulations was the Multistate Bennett Acceptance Ratio (MBAR) method, described in Section 3.5. The software alchemical-analysis (KLIMOVICH; SHIRTS; MOBLEY, 2015) was utilized to obtain the ΔG_{solv} with MBAR and to assess the quality of the results. After the first estimations, we realized that the binary interaction parameter of Eq. (3.75) was necessary for systems containing water. Hence, we estimated k_{ij} for these pairs and, for all the other pairs, we set k_{ij} to zero. The estimation was done by performing trial expanded ensemble simulations in three values of k_{ij}, as suggested by Ervik *et al.* (2016). With the ΔG_{solv} obtained with these simulations, we did a linear fit to obtain the refined value of the parameter. We used this strategy because the estimation with SAFT VR Mie EoS gave poor results for the solvation free energies.

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. This part was necessary since these parameters were not available for the ring geometry on the force field database (ERVIK; MEJÍA; MÜLLER, 2016). The parameters obtained and the mean percentage error (MPE) of the vapor pressure found with the SAFT-VR Mie EoS to the experimental data (MORTIMER; MURPHY, 1923) were those observed in Table 5.1.1.

Table 5.1.1 – Estimated SAFT-γ Mie Force Field parameters for phenanthrene.

m_s	ϵ/κ_b (K)	σ (Å)	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

The MPE value of 1.69 for the Lafitte *et al.* (2012) strategy in the Table 5.1.1 is the error between the vapor pressure found with the equation of state and the experimental data. Meanwhile, the other MPE value for the Lafitte *et al.* (2012) strategy (9.74) is the error between the vapor pressure obtained with the equation of state and the vapor pressure obtained in the GEMC simulations. The Lafitte *et al.* (2012) strategy should not need an estimation with molecular simulation data since this additional procedure is not necessary when estimating parameters for the chain equation (AVENDAÑO *et al.*, 2011) or the ring equation of Müller and Mejía (2017). In addition to that, this use of molecular simulation data to acquire the parameters negates the overall idea proposed by (AVENDAÑO *et al.*, 2011). They developed this force field with the intention of obtaining the parameters in a more straightforward way than other force fields since the SAFT-γ Mie model would not have the computational time associated with doing molecular simulations in its parameterization. Due to these specific characteristics of the model of Lafitte *et al.* (2012), we only studied the solvation free energy of phenanthrene with the set of parameters estimated with the strategy of Müller and Mejía (2017). In fact, we only followed the strategy of Lafitte *et al.* (2012) because it was the only one available when we first started this research. 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), and all the utilized parameters are available in Table 5.1.2.

Table 5.1.2 – SAFT-γ Mie Force Field for each substance used in this work.

	m_s	ϵ/κ_b (K)	σ (Å)	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

Our primary intention with this study is to assess the capability of the SAFT-γ Mie force field to represent solvation free energies. Hence, we chose benchmark solutes used in the literature (benzene, propane) and polyaromatic solutes (benzene, pyrene, phenanthrene, anthracene), and, for the solvents, we picked non-polar (hexane), aromatic (toluene), and hydrogen bonding (1-octanol) substances. It would be interesting to do a study with a bigger database of pairs solvent-solute. However, the time required for performing each of the solvation free energy simulations, some difficulties related to the available computational structure and the fact that a better model of aromatic compounds with this force field was only published in the middle of our study prevented us of doing a more extensive study. The solvation free energy simulations for the pairs chosen were carried with binary interaction parameters equal to zero, since these parameters were not necessary according to our preliminary studies. Since the force field does not account for charges, we only calculated the Mie contribution (Eq. (3.99)) to the solvation free energy. A total of 15 to 18 $\lambda's$, depending on the solute-solvent pairs, and their respective $\eta's$ were estimated as described in Chapter 4. The final λ set for all the pairs was found using the cumulative probability distribution (Eq. (3.113)). The distribution for the hexane(solvent)+benzene(solute) pair can be seen in Figure 5.1.1. The optimized values of λ and η for this pair and all the other pairs are available in Tables 5.1.3 to 5.1.6. Observing the coupling parameters found for all the pairs, we can see that they are concentrated on the region with a steeper slope as it is expected in this method.

It is also essential to analyze the reliability of solvation free energy estimations through the overlapping of the intermediate states. Insufficient overlap among states when using FEP based methods such as MBAR may result in the underestimation of variance and, consequently, in substantially incorrect free energies (KLIMOVICH; SHIRTS; MOBLEY, 2015). The overlap matrix for the solvation free energy of benzene in hexane is presented in Figure 5.1.2 and the matrices for the other pairs are available at Appendix B. Each element ij of these matrices is the average probability of observing

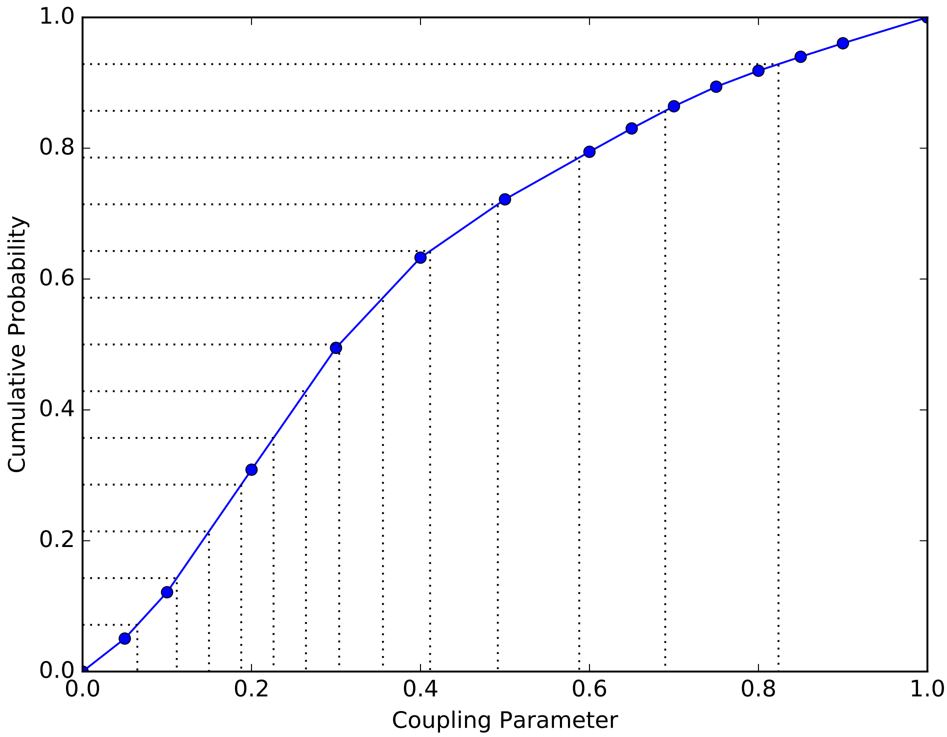


Figure 5.1.1 – Cumulative probability used to obtain the optimized values of $\lambda's$ for the pair hexane+benzene.

a configuration sampled from state i in state j. As an example, the average probability of finding a configuration sampled from state 3 in state 4 is 0.11 in Figure 5.1.2. According to Klimovich, Shirts and Mobley (2015), a tridiagonal overlap matrix is an indication of reliable free energy estimates, as long as the resulting error is sufficiently low. They define a tridiagonal matrix as one matrix with elements appreciable different from zero (the values should be as low as 0.03) in the main diagonal and the first diagonals above and below the main one. This requirement was met for all the pairs in our study. Some of the overlap matrices, including the one in Figure 5.1.2, had more than three diagonals, and, consequently, an apparent unnecessary number of intermediate states. However, this number of intermediate states were indispensable in our study because the error estimate of the solvation free energies significantly increased when we removed some of the intermediate states. Hence, we maintained these intermediate states in order to obtain low error values. After this analysis, we present in Table 5.1.7 the results for solvation free energy calculations and the absolute deviations to experimental data (KATRITZKY *et al.*, 2003).

Table 5.1.3 – Optimized values of λ and η for the hexane+solute pairs.

benzene		pyrene		phenanthrene	
λ	η	λ	η	λ	η
0.000	0.000	0.000	0.000	0.000	0.000
0.065	0.708	0.076	4.234	0.090	1.981
0.112	1.385	0.107	5.620	0.132	3.461
0.15	1.892	0.132	6.499	0.161	4.494
0.188	2.399	0.152	6.690	0.185	5.185
0.226	2.519	0.170	6.643	0.205	5.552
0.264	2.457	0.189	6.461	0.224	5.725
0.304	2.367	0.213	6.091	0.244	5.722
0.356	1.921	0.242	5.566	0.268	5.523
0.411	1.411	0.280	4.729	0.305	4.975
0.492	0.524	0.355	2.853	0.372	3.576
0.588	-0.663	0.483	-0.778	0.500	0.297
0.69	-2.016	0.678	-6.947	0.560	-1.390
0.824	-3.922	0.788	-10.631	0.722	-6.309
1.000	-6.583	1.000	-18.141	1.000	-15.448

Table 5.1.4 – Optimized values of λ and η for the 1-octanol+solute pairs.

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 5.1.5 – Optimized values of λ and η for the toluene+solute pairs.

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 5.1.6 – Optimized values of λ and η for the phenanthrene+CO_2+solute pairs with different values of w_{CO_2}.

$w_{CO_2} = 0.087$		$w_{CO_2} = 0.119$		$w_{CO_2} = 0.169$		$w_{CO_2} = 0.289$	
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

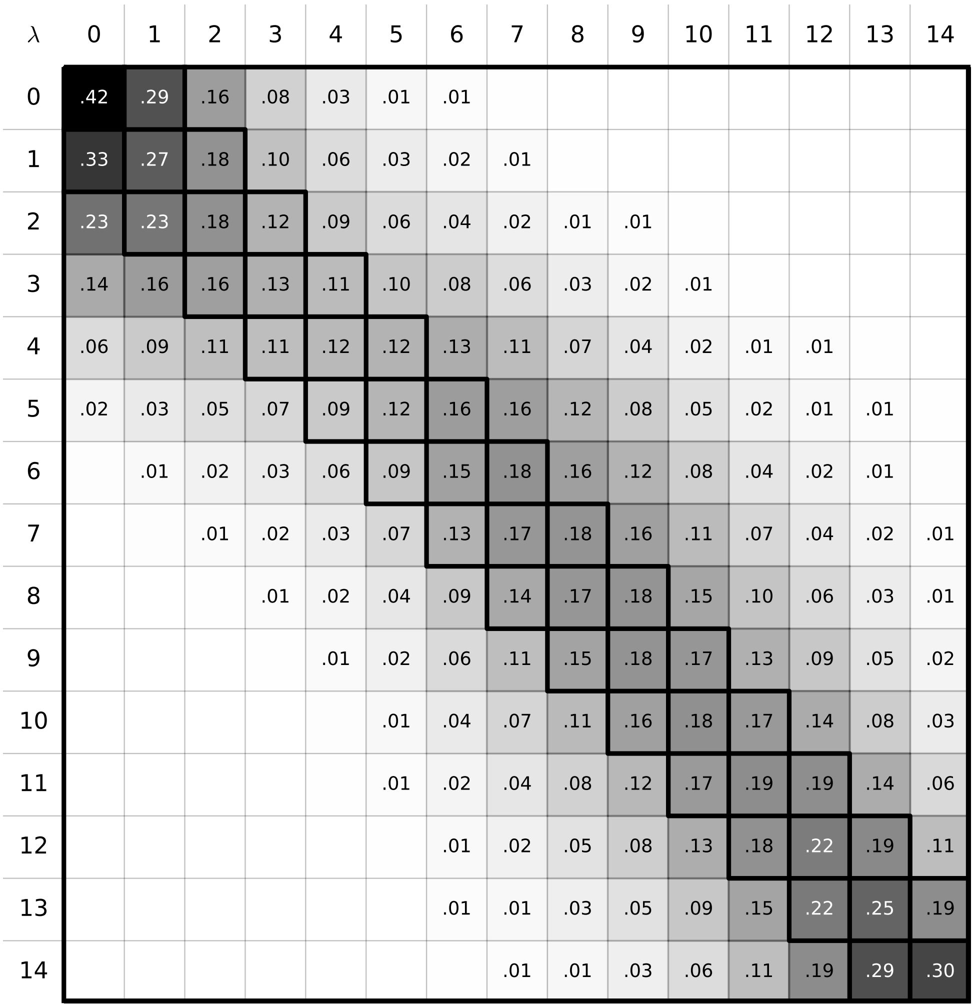


Figure 5.1.2 – Overlap matrix for hexane+benzene.

Table 5.1.7 – Calculated and experimental values for the solvation free energy differences (kcal/mol) of solutes in non-aqueous solvents.

Solute	Solvent	ΔG_{solv}^{exp}	ΔG_{solv}^{Mie}	Absolute Deviation
benzene	hexane	-3.96	-3.76 $\pm$ 0.01	0.20
pyrene	hexane	-11.53	-10.82 $\pm$ 0.02	0.71
phenanthrene	hexane	-10.01	-9.16 $\pm$ 0.01	0.85
propane	1-octanol	-1.32	-1.36 $\pm$ 0.02	0.04
anthracene	1-octanol	-11.72	-8.12 $\pm$ 0.03	3.61
phenanthrene	1-octanol	-10.22	-8.34 $\pm$ 0.03	1.47
pyrene	toluene	-12.86	-11.74 $\pm$ 0.01	1.11
anthracene	toluene	-11.31	-9.90 $\pm$ 0.01	1.41

The numerical values for solvation free energy differences in hexane had overall smaller absolute deviations to experimental data than the values in the other solvents. Additionally, this force field presented better results for the pair hexane+benzene than the TraPPE force field (- 4.35 $\pm$ 0.05 kcal/mol) (GARRIDO *et al.*, 2011) and the ELBA coarse-grained force field (-2.92 $\pm$ 0.01 kcal/mol) (GENHEDEN, 2016). TraPPE is a force field parametrized with fluid-phase equilibria data that uses the Lennard-Jones potential to describe the non-bonded interactions. In the cited paper, they used the united-atom description of the TraPPE force field for the alkyl group, the all-atom description for the polar groups and the explicit-hydrogen approach for the aromatic groups. In the explicit-hydrogen approach, the interaction sites for all hydrogen atoms, some lone pair electrons, and bond centers are accounted for (RAI; SIEPMANN, 2007). In turn, the ELBA force field is a coarse-grained model that comprises six independent parameters. This force field models three carbons as one Lennard-Jones site and one water molecule as a single Lennard Jones site with a point dipole. The free energy profiles for all the pairs studied here are presented in Figures 5.1.3 to 5.1.5. Specifically observing the solvation free energy profiles in hexane (Figure 5.1.3), we can see the effect of the molecule's size on the entropic region of the free energy curve, that is, the region corresponding to the first values of λ where space in the solvent is being 'opened' for the insertion of the solute.

We expected that a force field based on an EoS that does not explicitly account for hydrogen bond would not perform well for 1-octanol in mixtures since the parameterization of this molecule did not explicitly account for the association interactions. All the beads representing 1-octanol have the same intermolecular parameter, there is no distinction between the polar and apolar groups. Despite this, the solvation free energies of propane and phenanthrene in 1-octanol lied in the desired deviation range of 1-2 kcal/mol (MOBLEY; GILSON, 2017). For propane, the observed deviation in solvation free energies was much smaller when compared to the other solutes, which can be attributed to the non-polarity of propane and smoother free energy curve (Figure 5.1.4). Such solvation free energy of propane in 1-octanol also had a smaller deviation than the prediction of the ELBA force field (-0.92 $\pm$ 0.01) (GENHEDEN, 2016). The absolute deviation of the solvation free energy computed for anthracene in 1-octanol is much higher than the one calculated for phenanthrene in 1-octanol. The anthracene and phenanthrene molecules have the same geometry (Figure 4.1.1) in the SAFT-γ Mie model, although anthracene is a linear molecule and phenanthrene is not, and also similar physical properties. Hence, this high deviation of the solvation free energy of anthracene in 1-octanol may indicate a problem in the geometry chosen for anthracene in the SAFT-γ Mie force field and the importance of the geometry in modeling the molecules with this force field.

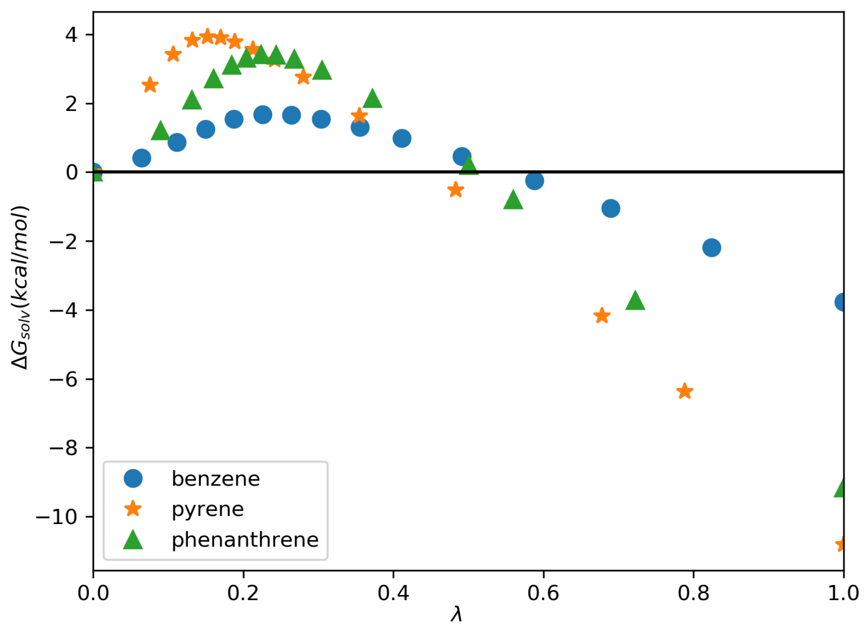


Figure 5.1.3 – Representation of solvation free energy profiles obtained with MD simulations of different solutes in hexane.

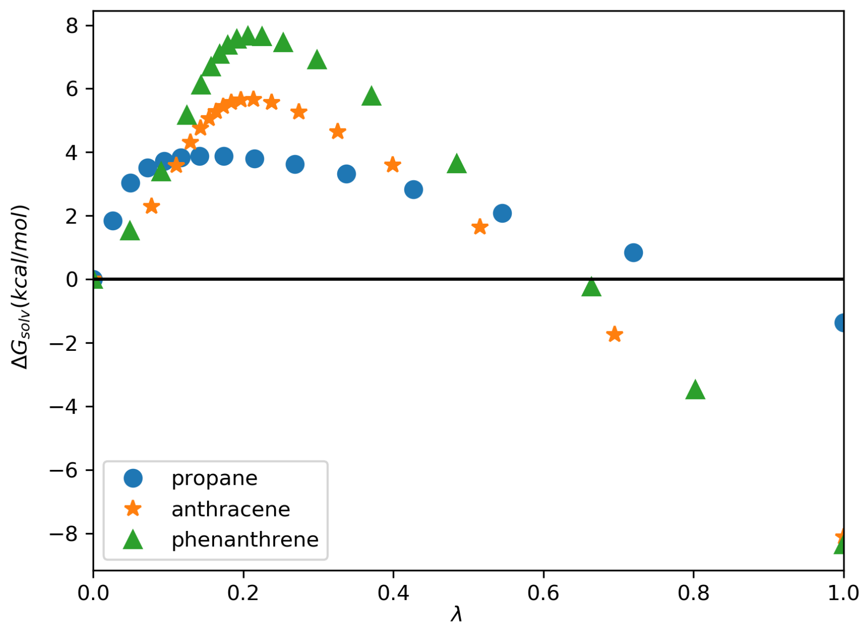


Figure 5.1.4 – Representation of solvation free energy profiles obtained with MD simulations of different solutes in 1-octanol.

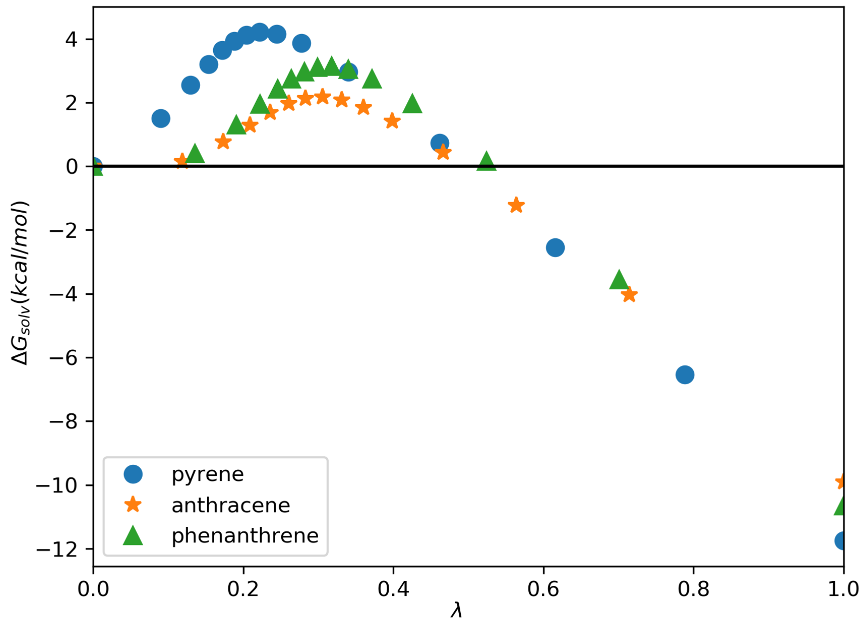


Figure 5.1.5 – Representation of solvation free energy profiles obtained with MD simulations of different solutes in toluene.

The results also indicate the capability of the force field for predicting the solvation free energies of polyaromatic solutes in aromatic solvents. The influence of the molecular geometry on the solvation free energy curves was the same as the one observed for other solvents (Figure 5.1.5). ΔG_{solv} was also calculated for phenanthrene in pure toluene and in toluene+CO_2 mixtures. To the best of our knowledge, there were no available experimental data for these solvation free energies, but the previous results for phenanthrene in other solvents showed that the force field is adequate to describe the solvation phenomenon of phenanthrene in a pure aromatic solvent. The results for these sets are exposed in Table 5.1.8.

Table 5.1.8 – 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 increase of the mass fraction of CO_2 in toluene caused a small effect on the solvation free energies in the range of fractions studied in this dissertation. First, the ΔG_{solv} decreased with the increase of w_{CO_2} up to 0.119. After this, the effect was reversed, and carbon dioxide became an anti-solvent. Soroush *et al.* (2014) reported that asphaltene precipitation occurs when carbon dioxide mass fractions became higher than 0.10 in the system asphaltene+toluene+carbon dioxide, which is in agreement with the anti-solvent effect of carbon dioxide observed in the values calculated here. In the Figure 5.1.6, we present the free energy profiles of the solvation free energies int the toluene + CO_2 mixtures. The small differences observed in this figure and in Table 5.1.8 may indicate that the effect of CO_2 is insignificant in the solvation of phenanthrene in toluene when using the SAFT-γ Mie force field. Nevertheless, more studies need to be done to make a safe assertion about it. It is also worth remarking that this is a qualitative study due to the lack of experimental data. Overall, the methodology proposed by the SAFT-γ Mie force field was satisfactory in predicting the solvation free energies of the pairs solvent-solute studied here. For the pair 1-octanol+anthracene, the performance was not as good as it was for the other pairs. This highlights the importance of choosing a correct geometry for this coarse-grained force field.

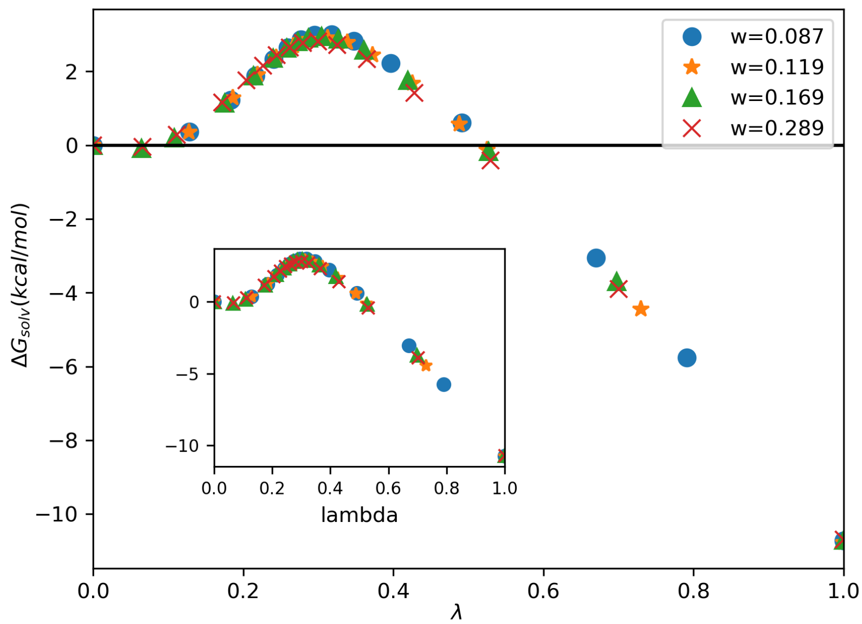


Figure 5.1.6 – Representation of solvation free energy profiles obtained with MD simulations of phenanthrene in toluene+CO_2.

5.2 Hydration free energies

Water is a solvent extensively used in experimental and computational studies. Because of this importance and the fact that water has unique properties, such as density maximum at 277 K and increased diffusivity upon compression, developing an accurate computational model for water is an ongoing quest (HADLEY; MCCABE, 2012). Hence, we also calculated the solvation free energies in water (hydration free energies) with the SAFT-γ Mie force field. With these calculations, we intend to verify if this coarse-grained model would represent the water molecule correctly and would be a good alternative to decrease the computational cost of solvation studies with asphaltene models. The simulations with water as a solvent were carried out using widely studied solutes (propane, benzene) and polyaromatic solutes (toluene, phenanthrene) with a set of fifteen intermediate states. We obtained these sets of λ and η with the same methodology used to acquire the sets for the solvation free energies with non-aqueous solvents, and they are exposed in Table 5.2.1. At first in our simulations, the binary interaction parameters of all aqueous mixtures were set to zero, but preliminary results for hydration free energies, displayed in Table 5.2.2, exhibited a high deviation from experimental data (ABRAHAM *et al.*, 1990; RIZZO *et al.*, 2006).

Table 5.2.1 – Optimized values of λ and η for the water+solute pairs.

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

Table 5.2.2 – Calculated values using $k_{ij} = 0$ and experimental values for the hydration free energy differences (kcal/mol) of solutes in water.

Solute	ΔG_{solv}^{exp}	ΔG_{solv}^{Mie}	Absolute Deviation
propane	2.00 $\pm$ 0.20	1.10 $\pm$ 0.01	0.90
benzene	-0.86 $\pm$ 0.20	-4.45 $\pm$ 0.03	3.59
toluene	-0.83 $\pm$ 0.20	-10.98 $\pm$ 0.30	10.15
phenanthrene	-3.88 $\pm$ 0.60	-10.90 $\pm$ 0.04	7.02

With these results, the need for binary interaction parameters became clear. First, we estimated k_{ij} with the SAFT-VR Mie EoS and experimental vapor pressure data, but this strategy also provided results that had high absolute deviations to the experimental data. Hence, we used the approach of estimating the k_{ij} with the output from solvation free energy calculations with molecular dynamics, as described in the last paragraph of Section 4.2. We initially found individual values for the interaction parameter of each pair, but, since the parameters for aromatic solutes were very similar (0.148, 0.162, 0.152), we averaged these values. By doing that, we obtained a general parameter for the water+aromatic pairs, which is exposed in Table 5.2.3. Also in this table, we display the binary interaction parameter for the pair water+propane.

Table 5.2.3 – Binary interaction parameters employed.

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

The relatively large k_{ij} value of the interaction between aromatic solutes and water can be related to the lack of an explicit association term in the equation of state used to obtain the parameters for water. Actually, the SAFT-VR Mie has an association term (LAFITTE *et al.*, 2013), but it was not incorporated in the force field. The SAFT-γ Mie model for water (LOBANOVA *et al.*, 2016) has two different temperature-dependent sets of parameters. The parameters utilized in this work were those estimated with experimental interfacial tension data. Hence, we tested the only binary interaction parameter for water+toluene estimated with MD interfacial data available in the literature (HERDES *et al.*, 2017). Nevertheless, the result also had high absolute deviation, and this parameter could not be transferred to the calculation of the solvation free energy of toluene in water.

These issues faced by SAFT-γ Mie model can also be related to the problems of modeling water with a coarse-grained force field. One of the main difficulties is the choice of which water molecules are going to be represented by which specific beads since water molecules move independently and are only bound by non-bonded interactions (HADLEY; MCCABE, 2010; HADLEY; MCCABE, 2012). The SAFT-γ Mie water considers that one water molecule corresponds to one bead. This strategy only saves a small amount of simulation time, but it can predict properties at physiological temperatures unlike other more aggressive models such as the MARTINI, which consider that one bead represents various water molecules. In light of all these facts, 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 severe flaws in the methodology. This estimation of the binary parameter increased significantly the simulation time required to calculate the hydration free energies, since we had to carry out three additional simulations for every pair water-solute and then more three additional simulations for the three water+polyaromatic solutes in order to test the averaged binary interaction parameter. If these simulations are necessary for every time a new mixture with water is going to be studied with the SAFT-γ Mie force field, the use of this model can become impractical. With this idea in mind, a useful investigation to be made is to check how much other pairs of water+aromatic solute can be modeled using the binary interaction parameter estimated here. Using these binary interaction parameters calculated with data from molecular dynamics, we then obtained the final hydration free energy differences presented in Table 5.2.4.

Table 5.2.4 – Calculated and experimental hydration free energy differences (kcal/mol) of solutes in water.

Solute	ΔG_{solv}^{GAFF}	ΔG_{solv}^{ELBA}	ΔG_{solv}^{exp}	ΔG_{solv}^{Mie}	Absolute Deviation
propane	2.50 ±0.02	2.76 ± 0.02	2.00 ± 0.20	2.01 ± 0.01	0.01
benzene	-0.81±0.02	-0.69 ± 0.01	-0.86 ± 0.20	-1.12 ± 0.01	0.26
toluene	-0.79±0.03	-0.76 ± 0.01	-0.83 ± 0.20	-0.84 ± 0.01	0.01
phenanthrene	-5.26±0.03	N/A	-3.88 ± 0.60	-3.47 ± 0.02	0.41

Hydration free energy differences calculated using the SAFT-γ Mie force field with $k_{ij} \neq 0$ had low absolute deviations to the experimental data, as expected since the parameters were adjusted to fit these experimental data. In the table above, we also show the results obtained by Genheden (2016) with the ELBA force field and by Mobley and Guthrie (2014) with the GAFF force field for the solutes and with the TIP3P model for water. The GAFF (General Amber Force Field) force field is an all-atom model that consists of bonded and non-bonded parameters and is suitable for the study of a significant number of molecules. In turn, the TIP3P model considers that water is a rigid monomer represented by three interacting sites with non-bonded interactions and Coulombic interactions (JORGENSEN *et al.*, 1983). Both the GAFF and the TIP3P models use the Lennard-Jones potential to calculate the non-bonded interactions.

Comparing the three mentioned force fields, the root mean square error (RMSE) of all the pairs tested with the SAFT-γ Mie model was 0.24, the RMSE for hydration free energy differences obtained with the GAFF force field was 0.73, and that for the ELBA coarse-grained force field was 0.44. The difference in absolute deviations between the GAFF and SAFT-γ Mie force fields is significantly high for phenanthrene, hence the coarse-grained force field with a binary parameter is preferred if the application requires a high level of accuracy. The results also indicated that the SAFT-γ Mie model with the binary interaction parameter performed better than the ELBA force field in modeling the solvation phenomenon of the pairs studied in this work, but performed worse with the binary parameter set to zero. This occurred despite the fact that both the SAFT-γ Mie and ELBA models have the same level of coarse-graining for the solvent (one bead represents one water molecule). Hence, the choice between the two coarse-grained models is dependent on the availability and transferability of binary interaction parameters of the Mie Model. We also present, for the SAFT-γ Mie force field, the hydration free energy profiles in Figure 5.2.1. Bigger molecules had steeper free energy profiles, as it was for the solvation free energy study in other solvents. We also observe that the hydration free energy for the first non-zero λ is negative for benzene and toluene when a positive value is expected since free energy is required to 'open space' in the solvent for the solute's insertion. This anomaly can be caused by numerical errors during the estimation of the solvation free energy or by the fact that the attractive term in the Mie potential compensates the need to open space.

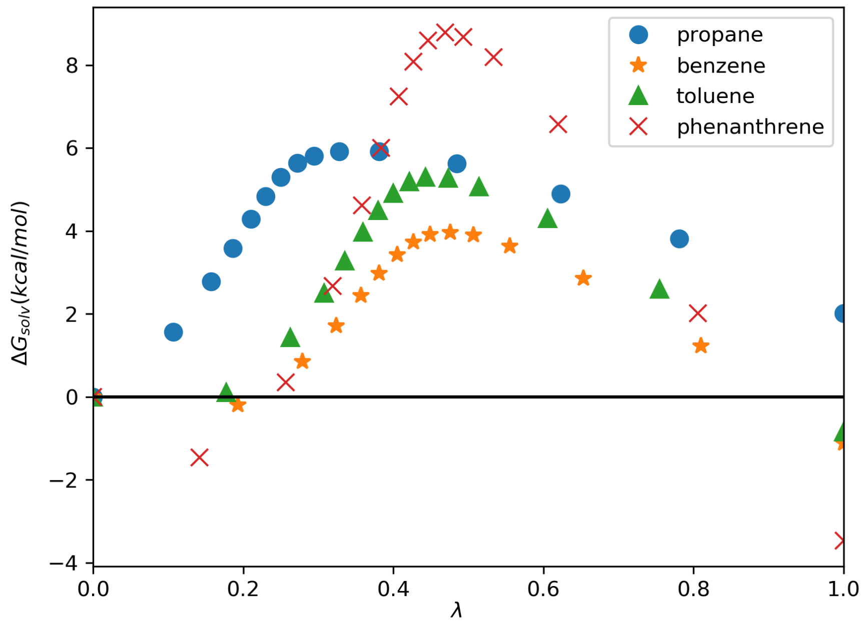


Figure 5.2.1 – Representation of hydration free energy profiles obtained with MD simulations for different solutes.

The results found here for both the solvation free energies and hydration free energies fulfilled the intentions of this dissertation. We assessed the prediction capability of the SAFT-γ Mie force field and provided satisfactory solvation free energy estimates of PAH's using a coarse-grained force field. In addition to that, we found flaws in the methodology used by the SAFT-γ Mie force field to model the water molecule. Hence, these shortcomings of this model can now be addressed, and the force field can even be improved by using other mixing rules to avoid the use of a binary parameter or, even, using hydration free energy estimates in the parameterization of water. These results also encourage us to calculate solvation free energies of more complex molecules mimicking asphaltenes in non-aqueous solvents in future studies.

5.3 Partition Coefficients

Using the solvation free energies estimated in the sections above, we also calculated partition coefficients by means of Eq. (3.101), for the pairs water/1-octanol and water/hexane with the intention of testing the modeling capabilities of SAFT-γ Mie model again. The partition functions studied here have many experimental data available in the literature due to their environmental importance (SANGSTER, 1997). Besides this, the calculations of these specific partition coefficients are relevant because 1-octanol is used to quantify hydrophobicity and can serve as a model for biological lipids and different soils (RUELLE, 2000), and hexane is a model for an apolar, hydrophobic phase. Calculated values and experimental data are shown in Table 5.3.1. The experimental data of the partition coefficients were taken from Poole, Durham and Kibbey (2000), Sangster (1997) for the coefficient of water/1-octanol and from Schulte *et al.* (1998) for the coefficient of water/hexane.

Table 5.3.1 – Partition Coefficient Calculated from MD simulations and from experimental data.

	Molecular Dynamics	Experimental	Absolute Deviation
	$\log P^{water/1-octanol}$		
propane	2.47	2.40	0.07
phenanthrene	3.57	4.46	0.89
	$\log P^{water/hexane}$		
benzene	1.93	2.06	0.13
phenanthrene	4.17	4.49	0.32

Overall absolute deviations were small for pairs with smaller solvation free energy deviations such as propane and benzene. The water/1-octanol partition coefficient of phenanthrene had higher deviation due to the higher deviation of the free energy of solvation of this compound in 1-octanol. Comparing with other force fields, Garrido *et al.* (2009) reported average absolute deviations for the water/1-octanol partition coefficient of 0.4 with the GROMOS 53a6 force field (OOSTENBRINK *et al.*, 2004), 0.3 for TraPPE, and 0.9 for OPLS-AA/TraPPE force fields. However, they attribute the low deviations of TraPPE to the cancellation of errors between the two solvation free energies. Additionally, Genheden (2016) found average absolute deviations of 0.86 for the water/hexane partition coefficients and of 0.75 for the water/1-octanol partition coefficients with the ELBA coarse-grained force field. At this dissertation, we performed a small study of partition coefficients with the SAFT-γ Mie force field. Hence, a larger set would be necessary to do a complete evaluation of the performance of this force field in the prediction of partition coefficients.

6 Conclusions

This dissertation consisted of the study of solvation free energy calculations of aromatic solutes that can mimic asphaltenes in different solvents with the SAFT-γ Mie coarse-grained force field. Solvation free energy studies are mostly done using water as a solvent and with all-atom force fields based on the Lennard-Jones Potential. Therefore, with this study, we were able to provide data that were lacking in the literature about the performance of a coarse-grained force field based on the Mie Potential for solvation free energy calculations. Additionally, the solvation free energy estimations carried out here can help improve the SAFT-γ Mie force field since these calculations are helpful in identifying errors in the modeling process. The SAFT-γ Mie uses the SAFT-VR Mie EoS in its parameterization, which results in a more straightforward method of obtaining parameters. Following this strategy, the phenanthrene parameters, which were not available in this force field database, were obtained using vapor-liquid equilibrium data and two different ring equations and geometries. However, only the parameters estimated with the ring equation proposed by Müller and Mejía (2017) were utilized in the solvation free energy simulations since this equation did not require molecular simulation data in its parameterization.

To obtain accurate solvation free energies, we carefully selected and optimized the coupling parameter and their respective simulation weights used in our Expanded Ensemble simulations. The resulting potential energies from these simulations were then served as input to estimate solvation free energy differences with 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. We also observed the geometry effect on the free energy curves - larger molecules had steeper curves and more substantial absolute deviations. The influence of carbon dioxide on the solvation free energy of phenanthrene in toluene was found to be minimum according to the SAFT-γ Mie force field.

Hydration free energy differences calculations with the SAFT-γ Mie model required the use of relatively large values of k_{ij} to obtain satisfactory results. We chose to estimate the parameter with the output from molecular dynamics data since the strategy of using the SAFT-VR Mie EoS provided high absolute deviations. This necessity of one additional parameter probably happens due to the lack of a term to account for the hydrogen bond on the EoS that this force field is based and due to the problems associated with the coarse-graining of water molecules. 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 and ELBA force field. We also used the solvation free energies to calculate partition coefficients in water/1-octanol and water/hexane. The absolute deviations to the experimental data obtained were similar to the ones found for all-atom force fields (GROMOS, TraPPE and OPLS-AA/TraPPE) and other coarse-grained force filed (ELBA).

Overall, the SAFT-γ Mie force field proved to be an excellent model to represent the solvation phenomenon of non-aqueous solvents. It correctly described solvation free energy differences of solutes mimicking asphaltenes in hexane, toluene, 1-octanol. However, the calculation of hydration free energies required the use of a binary interaction parameter estimated with MD output, what significantly increased the simulation time. This fact evidenced flaws in the methodology used by the SAFT-γ force field and made us question the feasibility of this model when studying hydration free energies. Nevertheless, the SAFT-γ Mie force field for water used here does not predict freezing at room temperature as other force fields, which is essential for our hydration free energy calculations. Hence, it would be relevant to test if the binary interaction parameter estimated here can be used in hydration free energy calculations of a variety of aromatic solutes. Also based on this dissertation, we have some ideas for future development. We intend to use the SAFT-γ Mie force field to model more complex asphaltenes models and, consequently, increase the scale of the simulations we performed. Additionally, we want to develop new methodologies to calculate solubility using solvation free energies.

Bibliography

ABRAHAM, M. H.; WHITING, G. S.; FUCHS, R.; CHAMBERS, E. J. Thermodynamics of solute transfer from water to hexadecane. *Journal of the Chemical Society, Perkin Transactions 2*, p. 291–300, 1990.

ABREU, C. R. A. *Playmol*. 2017. <http://atoms.peq.coppe.ufrj.br/playmol/index.html>. Accessed: 2017-03-10.

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. R. A. 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. A. Transport properties of carbon dioxide and methane from molecular dynamics simulations. *The Journal of Chemical Physics*, v. 141, p. 134101, 2014.

ANDERSEN, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *Journal of Chemical Physics*, v. 72, p. 2384–2393, 1980.

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. *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, p. 265–276, 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, p. 3684–3690, 1984.

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

BUENROSTRO-GONZALEZ, E.; LIRA-GALEANA, C.; GIL-VILLEGAS, A.; WU, J. Asphaltene precipitation in crude oils: Theory and experiments. *AIChE Journal*, v. 50, p. 2552–2570, 2004.

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. 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. W.; SCOTT, H. L.; 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*, v. 92, p. 097201, 2004.

ERVIK, A.; LYSGAARD, M. O.; HERDES, C.; JIMÉNEZ-SERRATOS, G.; ML̈LER, 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.

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.

ESSEX, J. W.; REYNOLDS, C. A.; RICHARDS, W. G. Theoretical determination of partition coefficients. *Journal of the American Chemical Society*, v. 114, p. 3634–3639, 1992.

FERRENBERG, A. M.; SWENDSEN, R. H. Optimized monte carlo data analysis. *Physical Review Letters*, v. 63, p. 1195–1198, 1989.

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

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

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, 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.

GENHEDEN, S. Predicting partition coefficients with a simple all-atom/coarse-grained hybrid model. *Journal of Chemical Theory and Computation*, v. 12, p. 297–304, 2016.

GONCALVES, 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, 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, p. 4590–4599, 2010.

HADLEY, K. R.; MCCABE, C. Coarse-grained molecular models of water: a review. *Molecular Simulation*, v. 38, 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, 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. *Physical Review A*, v. 31, p. 1695–1697, 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.

JORGENSEN, W. L.; CHANDRASEKHAR, J.; MADURA, J. D.; IMPEY, R. W.; KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, v. 79, p. 926–935, 1983.

JORGENSEN, W. L.; TIRADO-RIVES, J. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, v. 110, p. 1657–1666, 1988.

JOSHI, N. B.; MULLINS, O. C.; JAMALUDDIN, A.; CREEK, J.; MCFADDEN, J. Asphaltene precipitation from live crude oil. *Energy & Fuels*, v. 15, p. 979–986, 2001.

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 & Fuels*, v. 29, 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.

KATRITZKY, A. R.; OLIFERENKO, A. A.; OLIFERENKO, P. V.; PETRUKHIN, R.; TATHAM, D. B.; MARAN, U.; LOMAKA, A.; ACREE, W. E. A general treatment of solubility. 1. the qspr correlation of solvation free energies of single solutes in series of solvents. *Journal of Chemical Information and Modeling*, v. 43, p. 1794–1805, 2003.

KATZGRABER, H. G.; TREBST, S.; HUSE, D. A.; TROYER, M. Feedback-optimized parallel tempering monte carlo. *Journal of Statistical Mechanics: Theory and Experiment*, v. 2006, p. P03018, 2006.

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.

KOBRYN, A. E.; KOVALENKO, A. Molecular theory of hydrodynamic boundary conditions in nanofluidics. *The Journal of Chemical Physics*, v. 129, p. 134701, 2008.

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 & Fuels*, v. 23, p. 5027–5035, 2009.

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*, v. 71, p. 211–214, 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, 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. *Molecular Physics*, v. 113, p. 1228–1249, 2015.

LOBANOVA, O.; MEJÍA, A.; JACKSON, G.; MÜLER, 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.; KLEIN, M. L.; TUCKERMAN, M. Nosé - hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics*, v. 97, p. 2635, 1992.

MARTYNA, G. J.; TOBIAS, D. J.; KLEIN, M. L. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, v. 101, 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.

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. *Annals of Physics*, 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, 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, p. 2242–2254, 2007.

MOBLEY, D. L.; GILSON, M. K. Predicting binding free energies: Frontiers and benchmarks. *Annual Review of Biophysics*, v. 46, 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, p. 711—720, 2014.

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.; JACKSON, G. Force field parameters from the saft - γ equation of state for use in coarse-grained molecular simulations. *Annual Review of Chemical and Biomolecular Engineering*, v. 5, p. 405–427, 2014.

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, p. 2179–2207, 2010.

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

NOSÉ, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, v. 81, p. 511–519, 1984.

OOSTENBRINK, C.; VILLA, A.; MARK, A. E.; GUNSTEREN, W. F. V. A biomolecular force field based on the free enthalpy of hydration and solvation: The gromos force-field parameter sets 53a5 and 53a6. *Journal of Computational Chemistry*, v. 25, p. 1656–1676, 2004.

ORSI, M.; ESSEX, J. W. The elba force field for coarse-grain modeling of lipid membranes. *PLOS ONE*, v. 6, p. 1–22, 2011.

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, 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.

POOLE, S. K.; DURHAM, D.; KIBBEY, C. Rapid method for estimating the octanol–water partition coefficient (log pow) by microemulsion electrokinetic chromatography. *Journal of Chromatography B: Biomedical Sciences and Applications*, v. 745, p. 117 – 126, 2000.

RAABE, G. *Molecular Simulation Studies on Thermophysical Properties - With Application to Working Fluids. Springer Series Molecular Modeling and Simulation*. 1. ed. Singapore: Springer, 2017.

RAI, N.; MAGINN, E. J. Critical behaviour and vapour-liquid coexistence of 1-alkyl-3-methylimidazolium bis(trifluoromethylsulfonyl)amide ionic liquids via monte carlo simulations. *Faraday Discuss.*, v. 154, p. 53–69, 2012.

RAI, N.; SIEPMANN, J. I. Transferable potentials for phase equilibria. 9. explicit hydrogen description of benzene and five-membered and six-membered heterocyclic aromatic compounds. *The Journal of Physical Chemistry B*, v. 111, p. 10790–10799, 2007.

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.

RAVINDRA, K.; SOKHI, R.; GRIEKEN, R. V. Atmospheric polycyclic aromatic hydrocarbons: Source attribution, emission factors and regulation. *Atmospheric Environment*, v. 42, p. 2895 – 2921, 2008.

RIZZO, R. C.; AYNECHI, T.; CASE, D. A.; KUNTZ, I. D. Estimation of absolute free energies of hydration using continuum methods accuracy of partial charge models and optimization of nonpolar contributions. *Journal of Chemical Theory and Computation*, v. 2, p. 128–139, 2006.

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

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.

RUELLE, P. The n-octanol and n-hexane/water partition coefficient of environmentally relevant chemicals predicted from the mobile order and disorder (mod) thermodynamics. *Chemosphere*, v. 40, p. 457 – 512, 2000.

SANGSTER, J. *Octanol-Water Partitioning Coefficients: Fundamentals and Physical Chemistry*. 1. ed. [S.l.]: John Wiley and Sons, 1997.

SCHULER, L. D.; DAURA, X.; GUNSTEREN, W. F. van. An improved gromos96 force field for aliphatic hydrocarbons in the condensed phase. *Journal of Computational Chemistry*, v. 22, p. 1205–1218, 2001.

SCHULTE, J.; DüRR, J.; RITTER, S.; HAUTHAL, W. H.; QUITZSCH, K.; MAURER, G. Partition coefficients for environmentally important, multifunctional organic compounds in hexane + water. *Journal of Chemical & Engineering Data*, v. 43, p. 69–73, 1998.

SCHWAAB, M.; BISCAIA, E. C.; , J. ; MONTEIRO, J.; PINTO, J. C. Nonlinear parameter estimation through particle swarm optimization. *Chemical Engineering Science*, v. 63, p. 1542–1552, 2008.

SHAH, J. K.; MAGINN, E. J. A general and efficient monte carlo method for sampling intramolecular degrees of freedom of branched and cyclic molecules. *The Journal of Chemical Physics*, v. 135, p. 134121, 2011.

SHELL, M. S. *Thermodynamics and Statistical Mechanics: a Integrated Approach*. 1. ed. [S.l.]: Cambridge University Press, 2015.

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. *Journal of Chemical Physics*, 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, O.; HAVRE, T. E.; SÆTHER, O.; 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, p. 399 – 473, 2003.

SJÖBLOM, J.; SIMON, S.; XU, Z. Model molecules mimicking asphaltenes. *Advances in Colloid and Interface Science*, v. 218, p. 1 – 16, 2015.

SMITH, J.; van Ness, H.; 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, p. 405 – 411, 2014.

STERPONE, F.; DERREUMAUX, P.; MELCHIONNA, S. Protein simulations in fluids: Coupling the opep coarse-grained force field with hydrodynamics. *Journal of Chemical Theory and Computation*, 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.

STRANG, G. On the construction and comparison of difference schemes. *SIAM Journal on Numerical Analysis*, v. 5, p. 506–517, 1968.

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, p. 1040–1051, 2014. PMID: 24397444.

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

TROTTER, H. F. On the product of semi-groups of operators. *Proceeding of the American Mathematical Society*, v. 10, p. 545–551, 1959.

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.

WESTMORELAND, P.; KOLLMAN, P. A.; CHAKA, A. M.; CUMMINGS, P. T.; MOROKUMA, K.; NEUROCK, M.; STECHEL, E. B.; VASHISHTA, P. *Applying Molecular and Materials Modeling*. 1. ed. [S.l.]: Springer Netherlands, 2002.

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 – Detailing of the SAFT-VR Mie Equation of State

The term a_1 in Eq. 3.62 is the first-order mean-attractive energy of the mixture (LAFITTE *et al.*, 2013), and is given by

$$a_1 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{s,i}x_{s,j}a_{1,ij}, \tag{A.1}$$

where $a_{1,ij}$ is equal to

$$\begin{aligned} a_{1,ij} =& \mathcal{C}_{ij}\{x_{0,ij}^{\lambda_{a,ij}}[a_{1,ij}^{s}(\rho_s;\lambda_{a,ij}) + B_{ij}(\rho_s;\lambda_{a,ij})] \\ &- x_{0,ij}^{\lambda_{r,ij}}[a_{1,ij}^{s}(\rho_s;\lambda_{r,ij}) + B_{ij}(\rho_s;\lambda_{r,ij})]\}, \end{aligned} \tag{A.2}$$

with $\mathcal{C}_{ij}$ equals to

$$\mathcal{C}_{ij} = \frac{\lambda_{r,ij}}{\lambda_{r,ij}-\lambda_{a,ij}}\left(\frac{\lambda_{r,ij}}{\lambda_{a,ij}}\right)^{\left(\frac{\lambda_{a,ij}}{\lambda_{a,ij}-\lambda_{a,ij}}\right)}. \tag{A.3}$$

Also in Eq. A.2, $B_{ij}(\rho_s;\lambda_{ij})$ is equal to

$$B_{ij}(\rho_s;\lambda_{ij}) = 2\pi\rho_s d_{ij}^3\epsilon_{ij}\left[\frac{1-\zeta_x/2}{(1-\zeta_x)^3}I_{\lambda,ij} - \frac{9\zeta_x(1+\zeta_x)}{2(1-\zeta_x)^3}J_{\lambda,ij}\right], \tag{A.4}$$

$$\zeta_x = \frac{\pi\rho_s}{6}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{s,i}x_{s,j}d_{ij}^3. \tag{A.5}$$

Here, $I_{\lambda,ij}$ and $J_{\lambda,ij}$ are given by

$$I_{\lambda,ij} = \frac{(x_{0,ij})^{3-\lambda_{ij}}-1}{\lambda_{ij}-3}, \tag{A.6}$$

$$J_{\lambda,ij} = \frac{(x_{0,ij})^{4-\lambda_{ij}}(\lambda_{ij}-3)-(x_{0,ij})^{3-\lambda_{ij}}(\lambda_{ij}-4)-1}{(\lambda_{ij}-3)(\lambda_{ij}-4)}. \tag{A.7}$$

The perturbation terms $a_{1,ij}^{s}$ are obtained with the following equation:

$$a_{1,ij}^{S}(\rho_s;\lambda_{ij}) = -2\rho_s\left(\frac{\pi\epsilon_{ij}d_{ij}^3}{\lambda_{ij}-3}\right)\frac{1-\zeta_x^{eff}(\lambda_{ij})/2}{[1-\zeta_x^{eff}(\lambda_{ij})]^3}, \tag{A.8}$$

where $\zeta_x^{eff}(\lambda_{ij})$ is the effective packing fraction obtained within a van der Waals one-fluid approximation (LAFITTE *et al.*, 2013). It is equal to

$$\zeta_x^{eff}(\lambda_{ij}) = c_1(\lambda_{ij})\zeta_x + c_2(\lambda_{ij})\zeta_x^2 + c_3(\lambda_{ij})\zeta_x^3 + c_4(\lambda_{ij})\zeta_x^4. \tag{A.9}$$

Here, the coefficients c_1, c_2, c_3 and c_4 are

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} 0.81096 & 1.7888 & -37.578 & 92.284 \\ 1.0205 & -19.341 & 151.26 & -463.50 \\ -1.9057 & 22.845 & -228.14 & 973.92 \\ 1.0885 & -6.1962 & 106.98 & -677.64 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1/\lambda_{ij} \\ 1/\lambda_{ij}^2 \\ 1/\lambda_{ij}^3 \end{bmatrix}. \tag{A.10}$$

The term a_2 in Eq. 3.62 has a similar formulation to a_1. It is given by

$$a_2 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{s,i}x_{s,j}a_{2,ij}, \tag{A.11}$$

where $a_{2,ij}$ is equal to

$$\begin{aligned} a_{2,ij} =& \frac{1}{2}K^{HS}(1+\chi_{ij})\epsilon_{ij}\mathcal{C}_{ij}^2\{x_{0,ij}^{2\lambda_{a,ij}}[a_{1,ij}^s(\rho_s;2\lambda_{a,ij}) + B_{ij}(\rho_s;2\lambda_{a,ij})] \\ &- 2x_{0,ij}^{2\lambda_{a,ij}+2\lambda_{r,ij}}[a_{1,ij}^s(\rho_s;\lambda_{a,ij}+\lambda_{r,ij}) + B_{ij}(\rho_s;\lambda_{a,ij}+\lambda_{r,ij})] \\ &+ x_{0,ij}^{2\lambda_{r,ij}}[a_{1,ij}^s(\rho_s;2\lambda_{r,ij}) + B_{ij}(\rho_s;2\lambda_{r,ij})]\}. \end{aligned} \tag{A.12}$$

In Eq. A.12, K^{HS} is the isothermal compressibility of the mixture of hard spheres. It is equal to

$$K^{HS} = \frac{(1-\zeta_x)^4}{1+4\zeta_x+4\zeta_x^2+4\zeta_x^3+\zeta_x^4}. \tag{A.13}$$

and χ_{ij} is given by

$$\chi_{ij} = f_i(\alpha_{ij})\bar{\zeta}_x + f_2(\alpha_{ij})\bar{\zeta}_x^5 + f_3(\alpha_{ij})\bar{\zeta}_x^8, \tag{A.14}$$

where $\bar{\zeta}_x$ is equal to

$$\bar{\zeta}_x = \frac{\pi\rho_s}{6}\sum_{j=1}^{n} x_{s,i}x_{s,j}\sigma_{ij}^3, \tag{A.15}$$

and α_{ij} is equal to

$$\alpha_{ij} = \mathcal{C}_{ij}\left(\frac{1}{\lambda_{a,ij}-3} - \frac{1}{\lambda_{r,ij}-3}\right). \tag{A.16}$$

Finally, a_3 in Eq. 3.62 is equal to

$$a_3 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{s,i}x_{s,j}a_{3,ij}, \tag{A.17}$$

where $a_{3,ij}$ is equal to

$$a_{3,ij} = -\epsilon_{ij}^3 f_4(\alpha_{ij})\bar{\zeta}_x \exp[f_5(\alpha_{ij})\bar{\zeta}_x + f_6(\alpha_{ij})\bar{\zeta}_x^2]. \tag{A.18}$$

The functions $f_k(k = 1, ..., 6)$ are obtained with

$$f_k(\alpha_{ij})\frac{\sum_{n=0}^{n=3}\phi_{k,n}\alpha_{ij}^n}{1+\sum_{n=4}^{n=6}\phi_{k,n}\alpha_{ij}^{n-3}}, \tag{A.19}$$

where $\phi_{k,n}$ are coefficients defined in the original paper by Lafitte *et al.* (2013).

The density dependent coefficients of Eq. 3.69 are given by the following equations

$$k_0 = -\ln(1-\zeta_x) + \frac{42\zeta_x - 39\zeta_x^2 + 9\zeta_x^3 - 2\zeta_x^4}{6(1-\zeta_x)^3}, \tag{A.20}$$

$$k_1 = \frac{\zeta_x^4 + 6\zeta_x^2 - 12\zeta_x}{2(1-\zeta_x)^3}, \tag{A.21}$$

$$k_2 = \frac{-3\zeta_x^2}{8(1-\zeta_x)^2}, \tag{A.22}$$

$$k_3 = \frac{-\zeta_x^4 + 3\zeta_x^2 + 3\zeta_x}{6(1-\zeta_x)^3}. \tag{A.23}$$

In Eq. 3.68, the term $g_{1,ij}(\sigma_{ij})$ is the first-order contribution to the radial distribution function. It has the following form:

$$\begin{aligned}g_{1,ij}(\sigma_{ij}) =&\frac{1}{2\pi\epsilon_{ij}d_{ij}^3}\left[3\frac{\partial a_{1,ij}}{\partial\rho_s}\right.\\&-\mathcal{C}_{ij}\lambda_{a,ij}x_{0,ij}^{\lambda_{a,ij}}\frac{a_{1,ij}^s(\rho_s;\lambda_{a,ij})+B_{ij}(\rho_s;\lambda_{a,ij})}{\rho_s}\\&\left.+\mathcal{C}_{ij}\lambda_{r,ij}x_{0,ij}^{\lambda_{r,ij}}\frac{a_{1,ij}^s(\rho_s;\lambda_{r,ij})+B_{ij}(\rho_s;\lambda_{r,ij})}{\rho_s}\right].\end{aligned} \tag{A.24}$$

Also in Eq. 3.68, the second-order contribution to the radial distribution function ($g_{2,ij}(\sigma_{ij})$) is equal to

$$g_{2,ij}(\sigma_{ij}) = (1+\gamma_{c,ij})g_{2,ij}^{MCA}(\sigma_{ij}), \tag{A.25}$$

where $g_{2,ij}^{MCA}(\sigma_{ij})$ is equal to

$$\begin{aligned}g_{2,ij}(\sigma_{ij}) =&\frac{1}{2\pi\epsilon_{ij}d_{ij}^3}\left[3\frac{\partial\frac{a_2}{1+\gamma_{ij}}}{\partial\rho_s}\right.\\&-\epsilon_{ij}K^{HS}\mathcal{C}_{ij}^2\lambda_{r,ij}x_{0,ij}^{2\lambda_{r,ij}}\frac{a_{1,ij}^s(\rho_s;2\lambda_{r,ij})+B_{ij}(\rho_s;2\lambda_{r,ij})}{\rho_s}\\&+\epsilon_{ij}K^{HS}\mathcal{C}_{ij}^2(\lambda_{r,ij}+\lambda_{a,ij})(x_{0,ij})^{\lambda_{r,ij}+\lambda_{a,ij}}\frac{a_{1,ij}^s(\rho_s;\lambda_{r,ij}+\lambda_{a,ij})+B_{ij}(\rho_s;\lambda_{r,ij}+\lambda_{a,ij})}{\rho_s}\\&\left.-\epsilon_{ij}K^{HS}\mathcal{C}_{ij}^2\lambda_{a,ij}x_{0,ij}^{2\lambda_{a,ij}}\frac{a_{1,ij}^s(\rho_s;2\lambda_{a,ij})+B_{ij}(\rho_s;2\lambda_{a,ij})}{\rho_s}\right],\end{aligned} \tag{A.26}$$

and $\gamma_{c,ij}$ is a correction factor obtained from the equation bellow:

$$\gamma_{c,ij} = \phi_{7,0}\bar{\zeta}_x\theta_{ij}\exp(\phi_{7,3}\bar{\zeta}_x+\phi_{7,4}\bar{\zeta}_x^2)\{1-\tanh[\phi_{7,1}(\phi_{7,2}-\alpha_{ij})]\}, \tag{A.27}$$

where $\theta_{ij} = \exp(\beta\epsilon_{ij})-1$.

APPENDIX B – Overlap Matrices

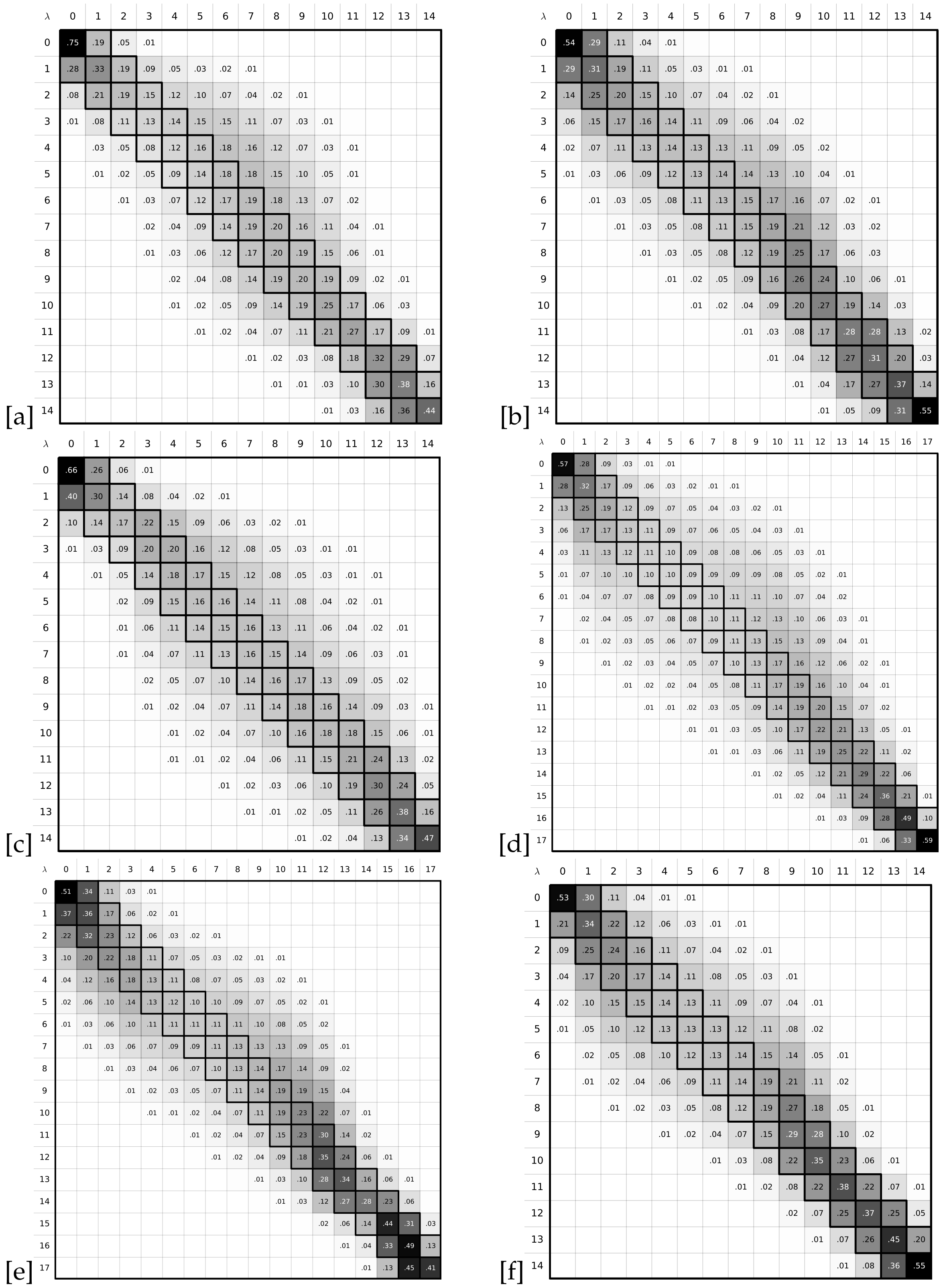


Figure B.0.1 – Overlap matrix for hexane+pyrene [a], hexane+phenanthrene [b], 1-octanol+propane [c], 1-octanol+anthracene [d], 1-octanol+phenanthrene [e], and toluene+pyrene [f].

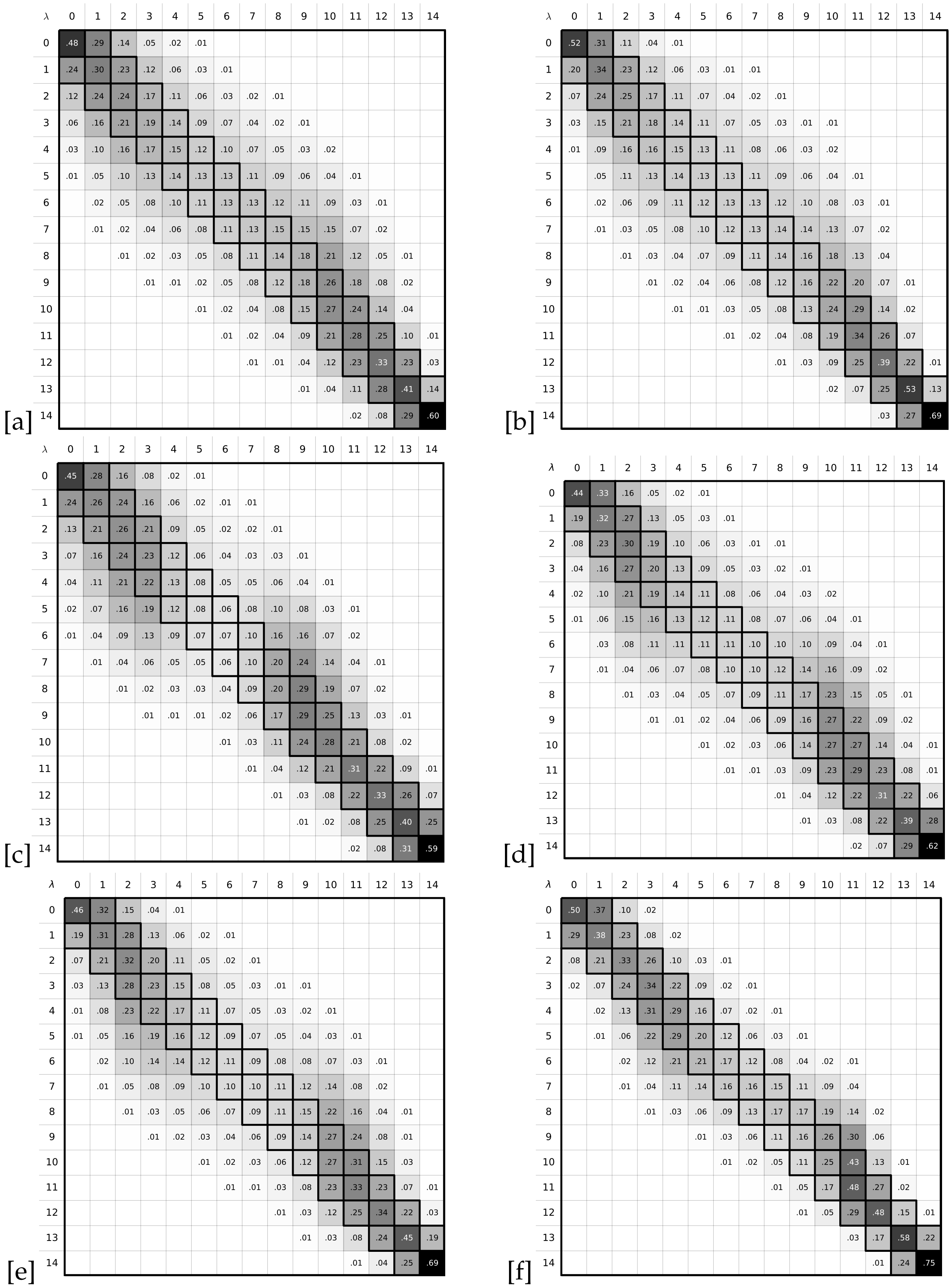


Figure B.0.2 – Overlap matrix for toluene+anthracene [a], toluene+phenanthrene [b], water+propane [c], water+benzene [d], water+toluene [e], and water+phenanthrene [f] .

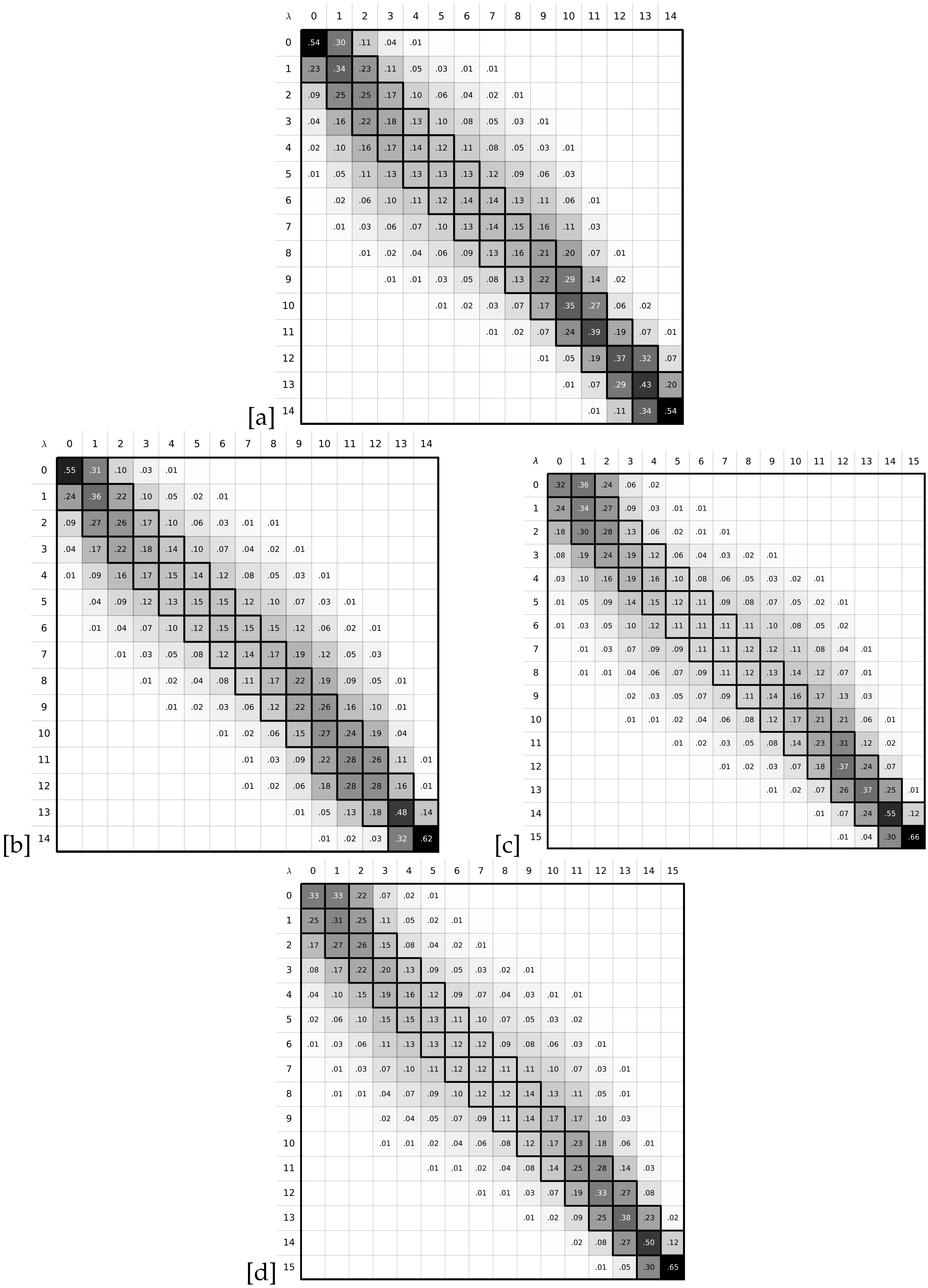


Figure B.0.3 – Overlap matrix for the different w_{CO_2} of the mixture toluene+CO_2+phenanthrene. $w_{CO_2} = 0.087$ [a], $w_{CO_2} = 0.119$ [b], $w_{CO_2} = 0.169$ [c], and $w_{CO_2} = 0.289$ [d].

APPENDIX C – Work Published in Scientific Conference

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.

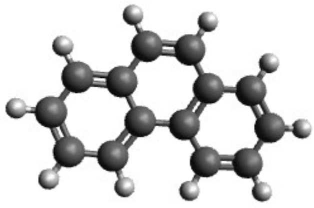


Figura1. Estrutura molecular do fenantreno.

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

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

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:

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

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:

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

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:

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

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

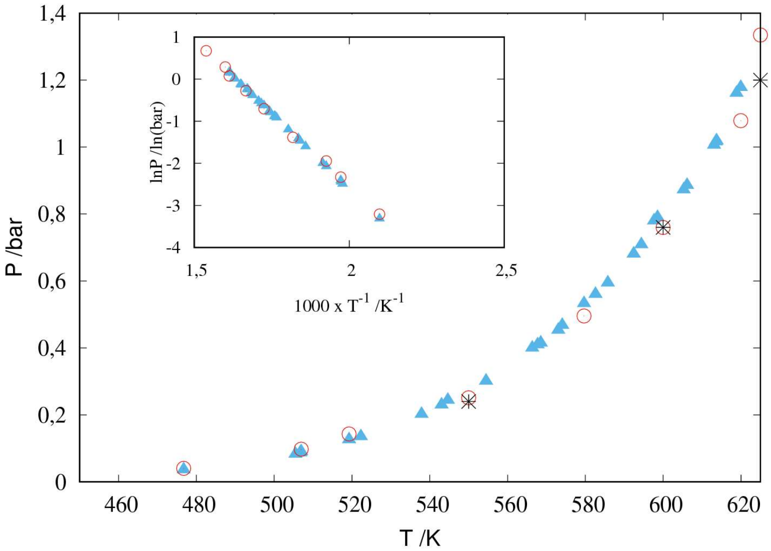


Figura 2. Pressão de vapor em função da temperatura calculada com o campo de força SAFT-γ CG (círculos vermelhos), dados experimentais (triângulos azuis) e resultados do campo de força TraPPE-EH (asteriscos pretos). A figura inserida é o mesmo gráfico, porém no formato de Clausius-Clapeyron.

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.

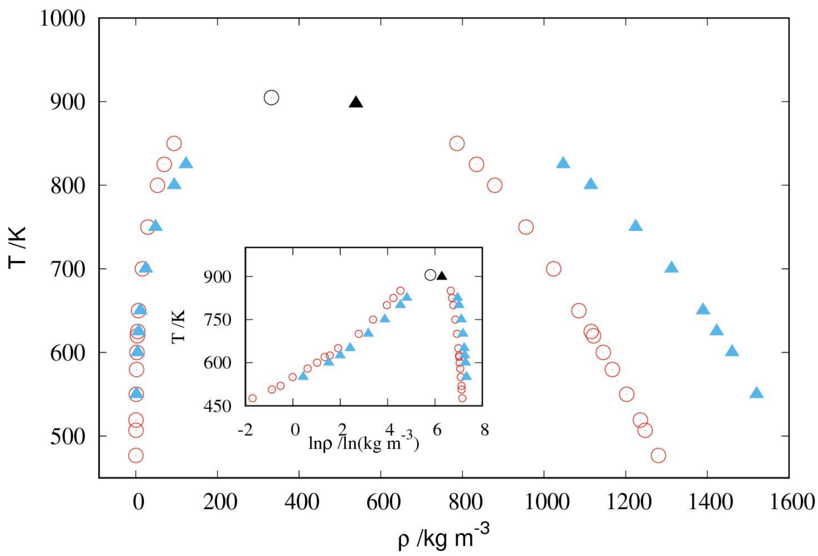


Figura 3. Curva de coexistência líquido vapor calculada com campo de força SAFT-γ CG (círculos vermelhos) e com o campo de força TraPPE-EH (triângulos azuis). O círculo e o triângulo pretos representam as propriedades críticas calculadas com o ajuste.

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

Evaluation of the SAFT-γ Mie force field with solvation free energy calculations

Isabela Q. Matos and Charlles R. A. Abreu*

School of Chemistry, Federal University of Rio de Janeiro, Av. Athos da Silveira Ramos 149, Rio de Janeiro,RJ 21941-909, Brazil

E-mail: *abreu@eq.ufrj.br

Abstract

We, at this work, studied the solvation free energy differences of molecules mimicking asphaltenes in different solvents with the SAFT-γ Mie force field. We obtained solvation free energy differences 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 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 this coarse-grained model could represent the free energy differences of the studied sets of solute-solvent.

Introduction

Solvation free energy calculations with molecular dynamics (MD) have a variety of applications ranging from drug design in the pharmaceutical industry to the development of separation technologies in the chemical industry. Solvation free energy is, more specifically, the difference in free energy related to the process of transferring the solute from the ideal gas phase condition to the liquid solvent phase condition[1]. 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 and partition coefficients. Additionally, solvation free energy calculations can be part of the process of calculating solubility with molecular dynamics[2].

The solvation phenomenon described above 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 low energy regions. Various simulation methodologies were developed to enable estimations of free energy differences such as the expanded ensemble,[3] thermodynamic integration,[4] free energy perturbation (FEP)[5–7] and umbrella sampling.[8] Utilizing FEP methodologies, recent papers[9,10] made available a big database of hydration free energy of small molecules using the GAFF force field. Beckstein et al.[11] also calculated the hydration free energies for fifty-two compounds with the OPLS-AA force field. They obtained an overall root mean square deviation to the experimental data of 1.75 kcal/mol and concluded that the reproducibility of the Lennard-Jones parameters is the main limiter of the precision of their results. Izairi and Kamberaj[12] also studied hydration free energies but with the intention of comparing the polar and nonpolar contributions. Garrido et al.[13,14] 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, ethylbenzene, and acetone with the force fields TraPPE-UA and TraPPE-AA. Roy et al.[15] addressed the choice of the Lennard-Jones parameters for predicting solvation free energy in 1-octanol. They calculated the solvation free energy of a set of 205 small organic molecules in 1-octanol and found that the force field parametrization of n-octanol proposed by Kobryn and Kovalenko[16] provided the best agreement. Gonçalves and Stassen[17] calculated the free energy of solvation using the polarizable continuum model coupled to molecular dynamics computer simulation with the GROMOS force field. These calculations were done with a representative set of solutes and with the solvents tetrachloride, chloroform, and benzene. Using the GAFF and the polarizable AMOEBA force fields, Mohamed et al.[18] 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. To define the role of solvent water in the docking structure determination of proteins, Matubayasi[19] developed a method to compute the solvation free energy of proteins while using OPLS-AA force field for the solutes and TIP3P for water. Genheden[20] expanded the Elba force field to calculate solvation free energies of more than 150 solutes taken from the Minnesota solvation database in polar (water, hexanol, octanol and nonanol) and apolar (hexane, octane, and nonane) solvents. They obtained mean absolute deviations of 1 kcal/mol for water and 1.5 kcal/mol for hexane. In this model, three carbons are represented by a single bead and water is also represented by a single bead.

As can be seem in the previous paragraph, solvation free energy simulation are performed in the literature using a variety force field since the choice of force field can be another influencing factor in the output of these calculations. Hence, we, in this study, assess the efficiencies and shortcomings of the SAFT-γ Mie coarse-grained force field[21] with free energy calculations for a variety of pairs solute-solvent. We chose a coarse-grained force fields because they generally reproduce free energy differences since the effects of reducing degrees of freedom in the entropy are counterbalanced by the reduction of enthalpic terms.[22] Additionally, the success of a coarse-grained force field is essential to decrease the computational time of solvation free energy calculations and to reveal deficiencies in the description of small molecules by these models.[1,23] The SAFT-γ Mie coarse-grained force field was specifically picked because it uses, unlike the majority of the force fields, the Mie potential and because its method of obtaining parameters is more straightforward than other models. It was initially parameterized with pure component equilibrium and interfacial tension data,[21] and this strategy has provided satisfactory results. Examples include the prediction of phase equilibrium of aromatic compounds, alkanes, light gases, and water,[24–26] thermodynamic properties of carbon dioxide and methane,[27] multiphase equilibrium of mixtures of water, carbon dioxide, and n-alkanes,[28] and water/oil interfacial tension.[29]

The solvents and solutes in our free energy calculations were picked to test the force field with standard sets used as a benchmark in solvation free energy calculations and with aromatic substances used as a model to asphaltenes. Asphaltenes are complicated to characterize by determining their composition on a molecular basis, but the literature broadly accepts that they can be described as a fraction of crude oil soluble in toluene and insoluble in n-alkenes (pentane, hexane, heptane).[30] They have motivated many studies with interest in developing models for their structure and behavior due to all the problems they can cause during their transportation and refining such as precipitation during the oil processing.[31] This precipitation issue is a recurrent problem due to the growing market of the production of crude oil in deep waters, which have conditions favorable to precipitation, such as pressure depletion and acid stimulation.[32] As an example, asphaltene precipitation due to pressure drop can clog oil production equipment and cause an almost exponential increase in the cost of production.[33] All these factors make the understanding of the behavior of asphaltenes in different chemical and physical environments relevant to the oil industry. As we said, asphaltene characterization still faces some issues. Hence, we choose to use polycyclic aromatic hydrocarbons (PAH'S), which have well-defined characteristics, to initially test the efficiency of the SAFT-γ Mie force field in describing the solvation phenomenon. The ones utilized in this work were phenanthrene, anthracene, and pyrene since they have similarities with asphaltenes regarding their solubility. Meanwhile, we selected compounds that are used to characterize asphaltenes (toluene, hexane) as solvents in our free energy calculations. We also tested the anti-solvent/solvent effect of carbon dioxide due to its influence in asphaltene precipitation during the oil processing.[34] With this study of solvation free energies with the SAFT-γ Mie model, we then intend to improve this force field and provide accurate free energy calculations of PAH's with a coarse-grained model. The correct description of these smaller asphaltene-like compounds by this force field opens up the possibility of obtaining satisfactory results for more complex asphaltene models with a less computational expensive force filed.

Computational Methods

SAFT-γ Mie Force Field

The SAFT-γ Mie Force Field uses a top-down coarse-graining 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 in molecular simulations since both the equation of state and the force field use the same explicit intermolecular potential model (Mie potential):

$$\begin{aligned} U_{Mie}(r) = &\epsilon \frac{\lambda_r}{\lambda_r - \lambda_a} \left(\frac{\lambda_r}{\lambda_a}\right)^{\left(\frac{\lambda_a}{\lambda_r - \lambda_a}\right)} \\ &\left[\left(\frac{\sigma}{r}\right)^{\lambda_r} - \left(\frac{\sigma}{r}\right)^{\lambda_a}\right]. \end{aligned} \tag{1}$$

The parameter ϵ is the potential well depth, σ is the segment diameter, r is the distance between the spherical segments, λ_r is the repulsive exponent and λ_a is the attractive exponent. his correspondence between models has been used to parametrize a variety of fluids.[35] This force field has the advantage of incorporating the degrees of freedom provided by the use of the Mie Potential.[24] This flexibility offers the exploration of a vast parameter space without using an iterative simulation scheme.[21] Despite these advantages, the force field can be restricted by the shortcomings of the equation of state. As an example, the lack of an association term in the equation can cause an inadequate representation of the properties of hydrogen bonding compounds.

Each substance has initially five parameters to be estimated ($m_s,\ \sigma,\ \epsilon,\ \lambda_r,\ and\, \lambda_a$) according to Eq. (1). 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 chosen value for this parameter is 6, corresponding to the London model, which is a good representation of the dispersion scale of most simple fluids that do not have strong polar interactions.[24,36] There are two strategies to obtain these parameters: one is by fitting the SAFT-Vr Mie EoS to experimental data as vapor pressure and liquid density,[37] and the other one is by using correspondent state parametrization.[38] In the present work, the first strategy was used to find the parameters for phenanthrene with vapor-liquid equilibrium data[39,40] following the methodology proposed by Müller and Mejía.[25] The parameterization was carried out with the number of segments equal to five and with a geometry such as that in Figure 1, since this level of coarse-graining was also used for a similar molecule (anthracene) in the original paper.

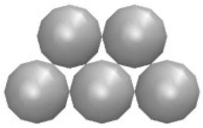


Figure 1: Representation of phenanthrene with the geometry proposed by Müller and Mejía.[25]

The parameter for the other compounds were retrieved from the literature, and all these parameters are exposed in Table 1. For a mixture, the mixing rules used on can be seen on Eqs. (2) to (4)[41] .

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

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

$$\epsilon_{ij} = (1 - k_{ij})\frac{\sqrt{\sigma_{ii}^3 \sigma_{jj}^3}}{\sigma_{ij}^3}\sqrt{\epsilon_{ii}\epsilon_{jj}}, \tag{4}$$

After the first estimations, we realized the need to estimate the binary interaction parameter of Eq. 4 for pairs with water as a solvent. Hence, we estimated k_{ij} for these pairs and, for all the other pairs, we set k_{ij} to zero. The estimation was done by performing trial expanded ensemble simulations in three values of the parameter, as suggested by Ervik et al..[42] With the ΔG_{solv} obtained with these simulations, we did a linear fit to acquire the refined value of the parameter. We used this strategy because the estimation with SAFT VR Mie EoS gave poor results for the solvation free energies.

Expanded Ensemble

The strategy chosen in this work to calculate the solvation free energy differences was to use an alchemical method in which the solute molecule is gradually inserted in the solvent using a thermodynamic path.[43] Each insertion or alchemical state is represented by a coupling parameter, λ, that ranges from 0 to 1. When $\lambda = 0$, there is no interaction with the solvent and, when $\lambda = 1$, the interactions are fully activated. Since the force field used does not explicitly take in consideration the charges, the interactions are only due to the Mie potential. For the coupling of the Mie Potential, we propose generalized softcore Mie potential based on the softcore potential of Beutler et al.[44] :

Table 1: SAFT-γ Mie Force Field for each substance used in this work.

	m_s	ϵ/κ_b (K)	$\sigma(\dot{A})$	λ_r
Water[28]	1	305.21	2.902	8.0
Propane[24]	1	426.08	4.871	34.29
Carbon dioxide[24]	2	194.94	2.848	14.65
Hexane[24]	2	376.35	4.508	19.57
Octanol[35]	3	495.71	4.341	28.79
Toluene[25]	3	268.24	3.685	11.80
Benzene[25]	3	230.30	3.441	10.45
Pyrene[25]	4	459.04	4.134	15.79
Anthracene[25]	5	259.68	3.631	9.55
Phenanthrene	5	262.74	4.077	9.55

$$\begin{aligned} U_{Mie}^{sc}(r) =& \lambda\epsilon\frac{\lambda_r}{\lambda_r - \lambda_a}\left(\frac{\lambda_r}{\lambda_a}\right)^{\left(\frac{\lambda_a}{\lambda_r - \lambda_a}\right)} \\ & \left\{\frac{1}{[\alpha(1-\lambda) + (r/\sigma)^{\lambda_a}]^{\lambda_r/\lambda_a}} - \right. \\ & \left.\frac{1}{\alpha(1-\lambda) + (r/\sigma)^{\lambda_a}}\right\}. \end{aligned} \tag{5}$$

where α is a constant whose value is normally assumed to be 0.5. We decided to use the Expanded Ensemble method[3] in our solvation free energy simulations since it allows a non-Boltzmann sampling scheme of different states in a single simulation. In this scheme, the sampling is done by biasing the phase space exploration process with weights not related to the statistical ensemble. The partition function of 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 subensembles Z_i in different values of λ, that is,

$$Z = \sum_{i=1}^{N} Z_i exp(\eta_i), \tag{6}$$

where N is the number of alchemical states, η_i is the arbitrary weight of the subensemble at each state, and Z_i is the configurational partition function of state i. Here, we followed the flat-histogram approach[45–47] to calculate the weights. This strategy aims to obtain adequate sampling by ensuring that all the states have an equal number of visits, i.e. the ratio of the probability of sampling state i (π_i) to the probability of sampling state j (π_j) is equal to one. Using this relation, the following equation can be obtained:

$$(\eta_i - \eta_j)_{k+1} = \beta(G_i - G_j)_k. \tag{7}$$

Eq. (7) proposes that the choice of the new weights is dependent on the free energies that we are attempting to obtain. This equation is then 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. (7), the next values of η can be calculated. This iteration goes on until a uniform distribution is attained. 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 free energy differences. In this work, the method chosen to obtain the optimal staging of the λ domain is the one developed by Escobedo and Martinez-Veracoechea[48] with a basis in the study of Katzgraber et al..[49] 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 (0) and highest (1) values of λ. This is specifically done by maximizing the steady-state stream ϕ of the simulation, which "walks" among the values of λ. This flow is estimated from a Fick's diffusion type of law:

$$\phi = D(\Lambda)\Pi(\Lambda)\frac{dx(\Lambda)}{d\Lambda}. \tag{8}$$

In the equation above, Λ is the actual continuous value of the coupling parameter. This continuous function of $\lambda' s$ 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:

$$\Pi(\Lambda) = \frac{C'\bar{\Pi}(\lambda)}{\Lambda_{i+1} - \Lambda_i}. \tag{9}$$

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

$$\bar{\Pi}_i(\lambda) = \frac{\pi_{i+1} - \pi_i}{2}. \tag{10}$$

The ϕ is maximum when the optimal probability $\Pi'(\Lambda_i)$ of visiting state Λ_i is proportional to $1/\sqrt{D(\Lambda)}$.[50] With that information, it is possible to estimate the diffusivity using one trial simulation with the following equation:

$$D(\Lambda) = \frac{\Lambda_{i+1} - \Lambda_i}{\bar{\Pi}(\lambda)dx(\Lambda)/d\Lambda}. \tag{11}$$

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

$$\Phi = \int_{\lambda=0}^{\lambda=1} \Pi'(\Lambda_i)d\Lambda = \frac{i}{K}, \tag{12}$$

where K is the total number of λ states. We obtained these cumulative probabilities for every λ set we estimated in order to carry out our solvation free energy simulations.

Molecular Dynamic Simulations

Using the parameters of Table 1, we carried out molecular dynamic simulations to estimate solvation free energy differences. The chosen software package to perform the simulations was the LAMMPS.[51] In this package, the equations of motion were integrated with the velocity-Verlet algorithm[52] with a time step of 2 fs. As required by the coarse-grained model, molecules with more than one bead were treated as rigid bodies. The thermostat and the barostat were the Nosé Hoover chains as described originally in Hoover[53] and Martyna et al.[54] with damping factors of 100 and 1000 time steps, respectively. For the rigid bodies in our simulations, we used the rigid-body algorithm of Kamberaj et al..[55] The potential cutoff was equal to 20 Å[25] with a neighbor list skin of 2 Å. The initial configurations of the solvated systems were generated using the Playmol package,[56] which is integrated with the Packmol package.[57] For the binary mixtures, one molecule of solute and a varying number of solvent molecules- 700 molecules of toluene, 700 molecules of octanol, 1024 molecules of hexane, 3000 molecules of water - were randomly added to a cubic box. Besides the systems with pure substances acting as solvents, we performed simulations to study solvation free energy of phenanthrene in a mixture of toluene and carbon dioxide with different weight fractions (w_{CO_2}). The system consisted of one molecule of phenanthrene for all the cases and 123 molecules of CO_2 and 618 molecules of toluene ($w_{CO_2} = 0.087$); 166 molecules of CO_2 and 589 molecules of toluene ($w_{CO_2} = 0.119$); 232 molecules of CO_2 and 545 molecules of toluene ($w_{CO_2} = 0.169$); 380 molecules of CO_2 and 446 molecules of toluene ($w_{CO_2} = 0.289$). As we commented in the Introduction, the solvents and solutes used in this study were selected with the intention of testing the force field with standard sets used as a benchmark in solvation free energy calculations and with aromatic substances used as models to asphaltenes.

All simulations were performed with the constant temperature and pressure values of 298 K and 1 bar, except the ones containing carbon dioxide. These had the temperature of 298 K and the pressure of the experimental liquid-phase equilibrium correspondent to each composition of the system CO_2+toluene.[58] For all simulations, the initial box was equilibrated at the NPT ensemble for 2 ns, and 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 research group, available at https://github.com/atoms-ufrj/USER-ALCHEMICAL.

During these expanded ensemble simulations, the sampling of a new alchemical state was tried at every 10 MD steps. To define the optimal values of λ and η corresponding to each state, trial simulations, having around 9 ns of production time, were carried out. In the first simulation, we chose the group of λ values arbitrarily, and we either set all $\eta's$ to zero or assigned values previously found for similar solute-solvent pairs. The subsequent group of $\eta's$ were estimated with the flat histogram approach (Eq. (7)). We then did another trial simulation with the new weights. The results of this simulation were used to optimize the group of $\lambda's$ by minimizing the number of round trips, as described in the previous section. The $\eta's$ corresponding to the newest group of $\lambda's$ 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 employed force field considers that the beads do not have charges, there are no Coulombic interactions and the the only contribution to the total potential energy is due to the softcore potential of Eq. 5. The post-processing method used to effectively calculate free energy differences with the potential energies obtained from the expanded ensemble simulations was the Multistate Bennett Acceptance Ratio (MBAR) method.[7] The software alchemical-analysis[43] was utilized to obtain the ΔG_{solv} with MBAR and to assess the quality of the results. After the first estimations, we realized that the binary interaction parameter of Eq. (4) was necessary for systems containing water. Hence, we estimated k_{ij} for these pairs and, for all the other pairs, we set k_{ij} to zero. The estimation was done by performing trial expanded ensemble simulations in three values of k_{ij}, as suggested by Ervik et al..[42] With the ΔG_{solv} obtained with these simulations, we did a linear fit to obtain the refined value of the parameter. We used this strategy because the estimation with SAFT VR Mie EoS gave poor results for the solvation free energies.

Results and discussion

Solvation free energies

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. A total of 15 to 18 $\lambda's$, depending on the solute-solvent pairs, and their respective $\eta's$ were estimated. The final λ set was found using the cumulative probability distribution (Eq. (12)) for all pairs. The distribution for the hexane(solvent)+benzene(solute) pair can be seen in Figure 2. The optimized values of λ and η for this pair is available in Table 2 and values for al the other pairs are available at the supporting information. Observing the coupling parameters found for all the pairs, we can see that they are concentrated on the region with a steeper slope as it is expected in this method.

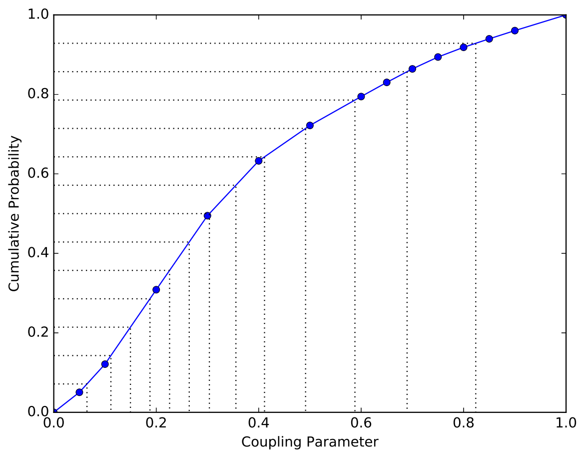
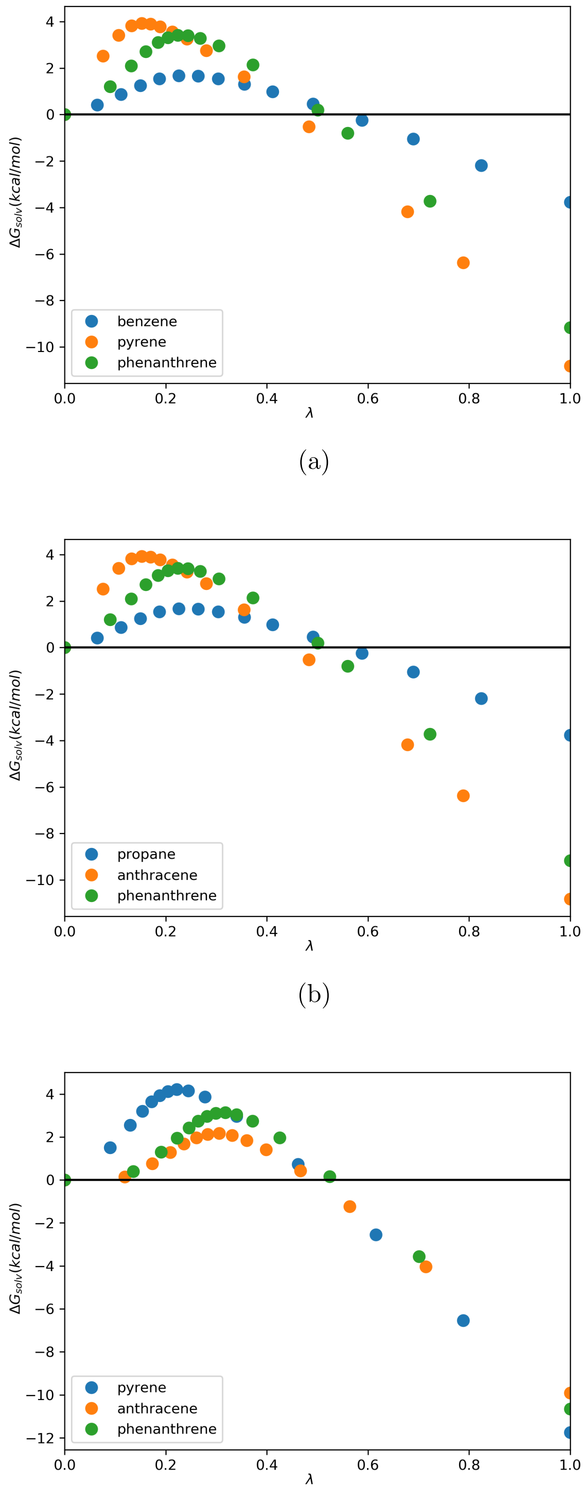


Figure 2: Cumulative probability used to obtain the optimized values of $\lambda's$ for the pair hexane+benzene.

After the expanded ensemble simulations with the intermediate states and weights estimated, we calculated the solvation free energy differences with MBAR. These results and and the absolute deviations to experimental data[59] are available in Table 3. The numerical values for solvation free energies in hexane had smaller absolute deviations to experimental data, what shows that the SAFT-γ Mie force field performs better for a non-polar solvent.



(a)

(b)

(c)

Figure 3: Solvation free energy profiles of different solutes in hexane (a), 1-octanol (b), and toluene (c).

Table 2: Optimized values of λ and η for the pair hexane + benzene.

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

Additionally, this force field presented better results for the pair hexane+benzene than the Trappe force field (- 4.35 ± 0.05 kcal/mol)[13] and the ELBA coarse-grained force field (-2.92 ± 0.01 kcal/mol).[20] We also observed the effect of molecule's size on the entropic region of the free energy curve in Figure 3. It was expected that a force field based on an EoS that does not 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.[60] The solvation free energy absolute deviation for propane was much smaller when compared to the other solutes, what can be attributed to propane's nonpolarity and smoother free energy curve (Figure 3). This solvation free energy of propane in 1-octanol also had a smaller deviation than the prediction of the ELBA force field (-0.92 ± 0.01).[20] 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 high deviation may indicate a problem in the parameterization of anthracene. 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 the free energy curves was the same as the one observed for other solvents (Figure 3). ΔG_{solv} was also calculated for phenanthrene in toluene and in toluene+CO_2. To the best of our knowledge, there were 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 in Table 4.

Table 4: 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 ± 0.02
0.087	-10.73 ± 0.02
0.119	-10.78 ± 0.02
0.169	-10.71 ± 0.02
0.289	-10.69 ± 0.02

The increase 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}. After the 0.119 fraction, the effect was reversed and carbon dioxide became an anti-solvent. Soroush et al.[34] reported that asphaltene precipitation occurs when carbon dioxide mass fractions became higher than 0.10 in the system asphaltene+toluene+carbon dioxide, what is in agreement 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 may indicate the insignificance of CO_2 in the solvation of phenanthrene in toluene when using the SAFT-γ Mie force field. But, more studies need to be done to make a secure assertion about it since this is a qualitative study due to the lack of experimental data.

Table 3: 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

Hydration free energies

We also calculated the hydration free energies of widely studied solutes (propane, benzene) and aromatic solutes (toluene, phenanthrene) with a group of fifteen intermediate states. First, the binary interaction parameter was set to zero, but the preliminary results for hydration free energies, exposed in Table 6, had a high deviation from the experimental data.[61,62] After these results, the need for binary interaction parameters was clear. First, we estimated k_{ij} with the SAFT VR Mie EoS and experimental vapor pressure data, but this strategy also did not provide good results. Hence, we used the approach of estimating the k_{ij} with the output from solvation free energy calculations with molecular dynamics. We initially found individual values for the interaction parameter of each pair, but, since the parameters for aromatic solutes were very similar (0.148, 0.162, 0.152), we averaged these values. By doing that, we obtained a general parameter for the water+aromatic pairs:

Table 5: Binary interaction parameters employed.

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

The relatively large k_{ij} value of the aromatic solutes can be pinned on the lack of an explicit association term in the model and on the water model itself since the force field did not need a k_{ij} for mixtures with the other hydrogen bonding solvent (1-octanol). This SAFT-γ Mie model for water[28] has two different temperature-dependent sets of parameters. The parameters utilized in this work was the one estimated with experimental interfacial tension data. Hence, we tested the binary interaction parameter for water+toluene estimated with MD interfacial data by Herdes et al..[29] Nevertheless, the result was not satisfactory and this parameter could not be transferable to the solvation free energy of toluene in water.

These issues faced by SAFT-γ Mie model are related to the problems of modeling water with a coarse-grained force field. One of the main difficulties is the choice of which water molecules are going to be represented by which specific beads since water molecules move independently and are only bound by non bonded interactions.[63,64] The SAFT-γ Mie water considers that one water molecule corresponds to one bead. This strategy only saves small simulation time, but it can 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. Using these parameters, we then obtained the final hydration free energy differences presented in Table 6.

Hydration free energy differences calculated using the SAFT-γ Mie force field with $k_{ij} \neq 0$ had low absolute deviations to the experimental data, as expected since the parameters were adjusted to fit the experimental data. Comparing our results with other force fields, the root mean square error (RMSE) for the pairs tested with the SAFT-γ Mie model was 0.24, the RMSE for hydration free energy differences with the GAFF force field was 0.73,[9] and the RMSE for the ELBA coarse-grained force field was 0.44.[20] The difference in absolute deviations between the GAFF and SAFT-γ Mie 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. The results also indicated that the SAFT-γ Mie Model with the binary interaction parameter performed better than the ELBA force field in modeling the solvation phenomenon of the pairs studied in this work and worst with the binary parameter equal to zero. This fact occurred despite the fact that both models have the same level of coarse-graining (one bead represents one water molecule). Hence, the choice between the two coarse-grained models is dependent on the availability and transferability of binary interaction parameters for the Mie Model. We also present, for the SAFT-γ Mie force field, the hydration free energy profiles in Figure 4. The geometry dependence on the free energy profiles is apparent as it was for the solvation free energy study in other solvents. We also observe that the hydration free energy for the first non zero λ is negative for benzene and toluene when a positive value is expected since energy is required to 'open space' in the solvent for the solute's insertion. This anomaly can be caused by numerical errors during the estimation or by another inconsistency in the force field.

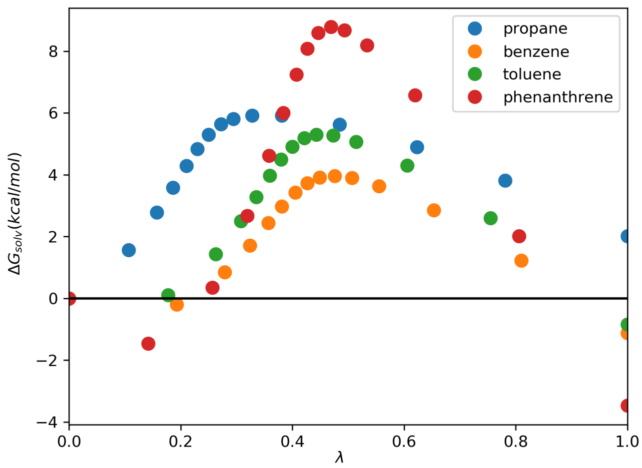


Figure 4: Hydration free energy profiles for different solutes.

Conclusions

This study consisted of solvation free energy calculations of aromatic solutes that can mimic asphaltenes in different solvents with the SAFT-γ Mie coarse-grained force field. Solvation free energy studies are mostly done using water as a solvent and with all-atom force fields based on the Lennard-Jones Potential, therefore, with this study, we provided data that were lacking in the literature. Additionally, the free energy estimations done can help improve the SAFT-γ Mie force field since these calculations are helpful in identifying errors in the modeling process. The SAFT-γ Mie uses the SAFT-VR Mie EoS in its parameterization, which results in a more straightforward method of obtaining parameters. Following this strategy, the phenanthrene parameters, which were not available in this force field database, were estimated using phase equilibrium data.

To obtain accurate solvation free energies, we carefully selected and optimized the coupling parameter and their respective simulation weights used in our Expanded Ensemble simulations. The resulting potential energies from these simulations were then served as input to estimate solvation free energy differences with 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. We also observed the geometry effect on the free energy curves - larger molecules had steeper curves and more substantial absolute deviations. The influence of carbon dioxide on the solvation free energy of

Table 6: Calculated and experimental hydration free energy differences (kcal/mol) of solutes in water.

Solute	ΔG_{solv}^{exp}	ΔG_{solv}^{Mie} $k_{ij} = 0$	Absolute Deviation	ΔG_{solv}^{Mie} $k_{ij} \neq 0$	Absolute Deviation
propane	2.00 ± 0.20	1.10 ± 0.01	0.90	2.01 ± 0.01	0.01
benzene	-0.86 ± 0.20	-4.45 ± 0.03	3.59	-1.12 ± 0.01	0.26
toluene	-0.83 ± 0.20	-10.98 ± 0.30	10.15	-0.84 ± 0.01	0.01
phenanthrene	-3.88 ± 0.60	-10.90 ± 0.04	7.12	-3.47 ± 0.02	0.41

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} to obtain satisfactory results. We chose to estimate the parameter with the output from molecular dynamics data since the strategy of using the SAFT-VR Mie EoS did not provide good results. This necessity of one additional parameter happens probably due to the lack of a term to account for the hydrogen bond on the EoS that the model is based and due to the problems associated with the coarse-graining of water molecules. 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 and ELBA force field.

Overall, the SAFT-γ Mie force field proved to be an excellent 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 more significant for this water model due to its coarse-graining level. Nevertheless, the SAFT-γ Mie force field for water used does not predict freezing at room temperature as other force fields, which is essential for our hydration free energy calculations.

Acknowledgement The authors thanks the financial support provided by Petrobras (project code:).

Supporting Information Available: This will usually read something like: "Experimental procedures and characterization data for all new compounds. The class will automatically add a sentence pointing to the information on-line: This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740.

(2) Schnieders, M. J.; Baltrusaitis, J.; Shi, Y.; Chattree, G.; Zheng, L.; Yang, W.; Ren, P. *J. Chem. Theory Comput.* **2012**, *8*, 1721–1736.

(3) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; VorontsovVelyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776–1783.

(4) Kirkwood, J. *J. Chem. Phys.* **1935**, *3*, 300313.

(5) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.

(6) Bennett, C. *J. Comput. Phys.* **1976**, *22*, 245268.

(7) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.

(8) Torrie, G.; Valleau, J. *J. Comput. Phys.* **1977**, *23*, 187 – 199.

(9) Mobley, D. L.; Guthrie, J. P. *J. Comput. Aided Mol. Des.* **2014**, *28*, 711720.

(10) Matos, G. D. R.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; ; Mobley, D. L. *J. Chem. Eng. Data* **2017**, *62*, 1559–1569.

(11) Beckstein, O.; Fourrier, A.; Iorga, B. I. *J. Comput. Aided Mol. Des.* **2014**, *28*, 265–276.

(12) Izairi, R.; Kamberaj, H. *J. Chem. Inf. Model.* **2017**, *57*, 25392553.

(13) Garrido, N. M.; Jorge, M.; Queimada, A. J.; Macedo, E. A.; Economou, I. G. *Phys. Chem. Chem. Phys* **2011**, *20*, 9155–9164.

(14) Garrido, N. M.; Queimada, A. J.; Jorge, M.; Macedo, E. A.; Economou, I. G. *J. Chem. Theory Comput.* **2009**, *5*, 2436–2446.

(15) Roy, D.; Blinov, N.; Kovalenko, A. *J. Phys. Chem. B,* **2017**, *121*, 92689273.

(16) Kobryn, A. E.; Kovalenko, A. *J. Chem. Phys.* **2008**, *129*, 134701.

(17) Gonçalves, P. F. B.; Stassen, H. *J. Chem. Phys.* **2005**, *123*, 214109.

(18) Mohamed, N. A.; Bradshaw, R. T.; Essex, J. W. *J. Comput. Chem.* **2016**, *37*, 2749–2758.

(19) Matubayasi, N. *Curr. Opin. Struct. Biol.* **2017**, *43*, 45 – 54.

(20) Genheden, S. *J. Chem. Theory Comput.* **2016**, *12*, 297–304.

(21) Avendaño, C.; Lafitte, T.; Galindo, A.; Adjiman, C. S.; Jackson, G.; Muller, E. A. *J. Phys. Chem. B* **2011**, *115*, 11154–11169.

(22) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. *Chem. Rev.* **2016**, *116*, 78987936.

(23) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Chem. Phys. B* **2007**, *111*, 2242–2254.

(24) Herdes, C.; Totton, T. S.; Müller, E. A. *Fluid Phase Equilib.* **2015**, *406*, 91–100.

(25) Müller, E. A.; Mejía, A. *Langmuir* **2017**, -, AL.

(26) Lobanova, O.; Avendaño, C.; Lafitte, T.; Müller, E. A.; Jackson, G. *Mol. Phys.* **2015**, *113*, 12281249.

(27) Aimoli, C. G.; Maginn, E. J.; Abreu, C. R. *Fluid Phase Equilib.* **2014**, *368*, 80–90.

(28) Lobanova, O.; Mejía, A.; Jackson, G.; Müller, E. A. *J. Chem. Thermodyn.* **2016**, *93*, 320–336.

(29) Herdes, C.; Ervik, A.; Mejía, A.; Müller, E. A. *Fluid Phase Equilib.* **2017**,

(30) Sjöblom, J.; Aske, N.; Auflem, I. H.; ystein Brandal,; Havre, T. E.; ystein Sther,; Westvik, A.; Johnsen, E. E.; Kallevik, H. *Adv. Colloid Interface Sci.* **2003**, *100-102*, 399 – 473.

(31) Sjöblom, J.; Simon, S.; Xu, Z. *Adv. Colloid Interface Sci.* **2015**, *218*, 1 – 16.

(32) Buenrostro-Gonzalez, E.; Lira-Galeana, C.; Gil-Villegas, A.; Wu, J. *AIChE J.* **2004**, *50*, 2552–2570.

(33) Joshi, N. B.; Mullins, O. C.; Jamaluddin, A.; Creek, J.; McFadden, J. *Energy Fuels* **2001**, *15*, 979–986.

(34) Soroush, S.; Straver, E. J.; Rudolph, E. S. J.; Peters, C. J.; de Loos, T. W.; Zitha, P. L.; Vafaie-Sefti, M. *Fuel* **2014**, *137*, 405 – 411.

(35) Ervik, A.; Mejía, A.; Müller, E. A. *J. Chem. Inf. Model.* **2016**, *56*, 1609–1614.

(36) Ramrattan, N.; Avendaño, C.; Müller, E.; Galindo, A. *Mol. Phys.* **2015**, *113*, 1–16.

(37) Avendaño, C.; Lafitte, T.; Adjiman, C. S.; Galindo, A.; Muller, E. A.; Jackson, G. *J. Phys. Chem. B* **2013**, *117*, 2717–2733.

(38) Mejía, A.; Herdes, C.; Müller, E. A. *Ind. Eng. Chem. Res.* **2014**, *53*, 4131–4141.

(39) Mortimer, S.; Murphy, R. *Ind. Eng. Chem. Res.* **1923**, *14*, 1140–1142.

(40) Osborn, A. G.; Douslin, D. R. *J. Chem. Eng. Data* **1975**, *20*, 229–231.

(41) Lafitte, T.; Apostolakou, A.; Avendano, C.; Galindo, A.; Adjiman, C. S.; Muller, E. A.; Jackson, G. *J. Chem. Phys.* **2013**, *139*, 154504.

(42) Ervik, A.; Lysgaard, M. O.; Herdes, C.; Jiménez-Serratos, G.; Mller, E. A.; Munkejord, S. T.; Müller, B. *J. Comput. Phys* **2016**, *327*, 576–611.

(43) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. *J. Comput. Aided Mol. Des.* **2015**, *29*, 397–411.

(44) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett* **1994**, *6*, 529–539.

(45) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9–12.

(46) Lee, J. *Phys. Rev. Lett.* **1993**, *71*, 211–214.

(47) Dayal, P.; Trebst, S.; Wessel, S.; Würtz, D.; Troyer, M.; Sabhapandit, S.; Coppersmith, S. N. *Phys. Rev. Lett.* **2004**, *92*, 097201.

(48) Escobedo, F. A.; Martinez-Veracoechea, F. J. *J. Chem. Phys* **2007**, *127*, 174103.

(49) Katzgraber, H. G.; Trebst, S.; Huse, D. A.; Troyer, M. *J. Stat. Mech. Theory Exp.* **2006**, *2006*, P03018.

(50) Trebst, S.; Huse, D. A.; Troyer, M. *Phys. Rev. E* **2004**, *70*, 046701.

(51) Plimpton, S. *J. Comp. Phys.* **1995**, *117*, 1–19.

(52) Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.

(53) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(54) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635.

(55) Kamberaj, H.; Low, R.; Neal, M. *J. Chem. Phys.* **2005**, *122*, 224114.

(56) ABREU, C. R. A. Playmol. `http://atoms.peq.coppe.ufrj.br/playmol/index.html`, 2017; Accessed: 2017-03-10.

(57) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

(58) J.Chang, C. *Fluid Phase Equilib.* **1992**, *15*, 235–242.

(59) Katritzky, A. R.; Oliferenko, A. A.; Oliferenko, P. V.; Petrukhin, R.; Tatham, D. B.; Maran, U.; Lomaka, A.; Acree, W. E. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1794–1805.

(60) Mobley, D. L.; Gilson, M. K. *Annu. Rev. Biophys.* **2017**, *46*, 531–558.

(61) Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. *J. Chem. Soc., Perkin Trans. 2* **1990**, 291–300.

(62) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. *J. Chem. Theory Comput.* **2006**, *2*, 128–139.

(63) Hadley, K. R.; McCabe, C. *J. Phys. Chem. B* **2010**, *114*, 4590–4599.

(64) Hadley, K. R.; McCabe, C. *Mol. Simul.* **2012**, *38*, 671681.