

Isabela Quintela Matos

**Solvation Free Energy Calculations of
Molecules Mimicking Asphaltenes Through
Molecular Dynamics Using The
Coarse-Grained SAFT- γ Mie Force Field**

Rio de Janeiro

2018

Isabela Quintela Matos

**Solvation Free Energy Calculations of Molecules
Mimicking Asphaltenes Through Molecular Dynamics
Using The Coarse-Grained SAFT- γ Mie Force Field**

Dissertação de Mestrado apresentada ao
Programa de Pós-Graduação em Tecnologia
de Processos Químicos e Bioquímicos,
Escola de Química, Universidade Federal
do Rio de Janeiro, como requisitos parcial à
obtenção do título de Mestre em Engenharia
Química.

Universidade Federal do Rio de Janeiro

Escola de Química

Programa de Pós-Graduação em Tecnologia de Processos Químicos e Bioquímicos

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 Through Molecular Dynamics Using The Coarse-Grained SAFT- γ Mie Force Field/ Isabela Quintela Matos. – Rio de Janeiro, 2018-

54 p. : il. (algumas color.) ; 30 cm.

Supervisor: Charlles Rubber de Almeida Abreu

Dissertação (Mestrado) – Universidade Federal do Rio de Janeiro

Escola de Química

Programa de Pós-Graduação em Tecnologia de Processos Químicos e Bioquímicos, 2018.

1. Palavra-chave1. 2. Palavra-chave2. 2. Palavra-chave3. I. Orientador. II. Universidade xxx. III. Faculdade de xxx. IV. Título

Errata sheet

Elemento opcional da ??, 4.2.1.2). Exemplo:

FERRIGNO, C. R. A. **Tratamento de neoplasias ósseas apendiculares com reimplantação de enxerto ósseo autólogo autoclavado associado ao plasma rico em plaquetas**: estudo crítico na cirurgia de preservação de membro em cães. 2011. 128 f. Tese (Livre-Docência) - Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2011.

Folha	Linha	Onde se lê	Leia-se
1	10	auto-conclavo	autoconclavo

Isabela Quintela Matos

Solvation Free Energy Calculations of Molecules Mimicking Asphaltenes Through Molecular Dynamics Using The Coarse-Grained SAFT- γ Mie Force Field

Dissertação de Mestrado apresentada ao
Programa de Pós-Graduação em Tecnologia de Processos Químicos e Bioquímicos,
Escola de Química, Universidade Federal
do Rio de Janeiro, como requisitos parcial à
obtenção do título de Mestre em Engenharia
Química.

Trabalho aprovado. Rio de Janeiro, 24 de novembro de 2012:

Charles Rubber de Almeida Abreu
Orientador

Professor
Convidado 1

Professor
Convidado 2

Rio de Janeiro
2018

*Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.*

Acknowledgements

Os agradecimentos principais são direcionados à Gerald Weber, Miguel Frasson, Leslie H. Watter, Bruno Parente Lima, Flávio de Vasconcellos Corrêa, Otavio Real Salvador, Renato Machnievscz¹ e todos aqueles que contribuíram para que a produção de trabalhos acadêmicos conforme as normas ABNT com L^AT_EX fosse possível.

Agradecimentos especiais são direcionados ao Centro de Pesquisa em Arquitetura da Informação² da Universidade de Brasília (CPAI), ao grupo de usuários *latex-br*³ e aos novos voluntários do grupo *abnT_EX2*⁴ que contribuíram e que ainda contribuirão para a evolução do *abnT_EX2*.

¹ Os nomes dos integrantes do primeiro projeto *abnT_EX* foram extraídos de <<http://codigolivre.org.br/projects/abntex/>>

² <<http://www.cpai.unb.br/>>

³ <<http://groups.google.com/group/latex-br>>

⁴ <<http://groups.google.com/group/abntex2>> e <<http://www.abntex.net.br/>>

*“Não vos amoldeis às estruturas deste mundo,
mas transformai-vos pela renovação da mente,
a fim de distinguir qual é a vontade de Deus:
o que é bom, o que Lhe é agradável, o que é perfeito.
(Bíblia Sagrada, Romanos 12, 2)*

Abstract

Segundo a ??, 3.1-3.2), o resumo deve ressaltar o objetivo, o método, os resultados e as conclusões do documento. A ordem e a extensão destes itens dependem do tipo de resumo (informativo ou indicativo) e do tratamento que cada item recebe no documento original. O resumo deve ser precedido da referência do documento, com exceção do resumo inserido no próprio documento. (...) As palavras-chave devem figurar logo abaixo do resumo, antecidas da expressão Palavras-chave:, separadas entre si por ponto e finalizadas também por ponto.

Palavras-chave: latex. abntex. editoração de texto.

Abstract

This is the english abstract.

Keywords: latex. abntex. text editoration.

List of Figures

Figure 2.2.1–Thermodynamic cycle for solvation free energy calculations with molecular dynamics (Adapted from Klimovich, Shirts e Mobley (2015))	21
Figure 3.1.1–Values for parameter χ according to the ring geometry (MÜLLER; MEJÍA, 2017)	28
Figure 4.1.1–Geometry for $m_s = 5$	36
Figure 4.1.2–Geometry for $m_s = 3$	36
Figure 5.1.1–Solvation free energy profiles for the hexane solvent	41
Figure 5.1.2–Solvation free energy profiles for the 1-octanol solvent	41
Figure 5.1.3–Solvation free energy profiles for the toluene solvent	42
Figure 5.1.4–Solvation free energy profiles of phenanthrene in toluene+ CO_2	43
Figure 5.2.1–Hydration free energy profiles	44

List of Tables

Table 1 – Estimated SAFT- γ Mie Force Field parameters for phenanthrene	39
Table 2 – SAFT- γ Mie Force Field for each substance used in this work	39
Table 3 – Calculated and experimental values for solvation free energies (kcal/mol) of solutes in non aqueous solvents	40
Table 4 – Calculated values for the solvation free energies (kcal/mol) of phenan- threne in toluene+ CO_2	42
Table 5 – Calculated values for the Gibbs energy of solvation (kcal/mol) of solutes in water for $k_{ij} = 0$	43
Table 6 – Binary interaction parameters employed	43
Table 7 – Calculated and experimental values for the Gibbs energy of solvation (kcal/mol) of solutes in water	44

List of symbols

Γ	Letra grega Gama
Λ	Lambda
ζ	Letra grega minúscula zeta
\in	Pertence

Contents

1	INTRODUCTION	15
2	LITERATURE REVIEW	17
2.1	Coarse Grained Force Fields	17
2.2	Solvation Free Energies Based on Molecular Dynamics	19
2.3	Post simulation methods	22
2.3.1	Thermodynamic integration	22
2.3.2	Free energy of Pertubation (FEP)	22
2.3.3	Bennet Acceptance Ratio (BAR)	23
2.3.4	Multistate Bennet Acceptance Ratio (MBAR)	24
3	FUNDAMENTALS OF THE COMPUTATIONAL METHODS	25
3.1	SAFT-γ Mie Force Field	25
3.1.1	SAFT-VR Mie EoS	25
3.1.1.1	Ideal Contribution	25
3.1.1.2	Monomer Contribution	26
3.1.1.3	Chain Contribution	26
3.1.1.4	Ring Contribution	27
3.1.1.5	Combining rules for the intermolecular potential parameters	27
3.1.2	Parameter Estimation for the SAFT- γ Mie Force Field	28
3.2	Expanded Ensemble Method	31
3.3	Gibbs Ensemble Monte Carlo (GEMC)	34
4	METHODOLOGY	36
4.1	Phenanthrene Parameterization	36
4.2	Solvation Free energy Calculations	37
5	RESULTS AND DISCUSSION	39
5.1	Solvation free energies	39
5.2	Hydration free energies	43
6	CONCLUSIONS	45
	BIBLIOGRAPHY	46

APPENDIX	52
APPENDIX A – FINAL VALUES OF λ AND η	53
APPENDIX B – OVERLAPPING MATRICES	54

1 Introduction

Solvation free energy differences 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. The solvation free energy difference is, more specifically, the Gibbs free energy difference between the solute alone in the gas phase and the solute interacting with the solvent. Through the study of this solvation phenomena, 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 like the activity coefficient at infinite dilution, the Henry constant, partition coefficients and solubility with the solvation free energy.

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

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

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

The sets of solvents and solutes chosen in this study range from the standard sets used as benchmark in solvation free energy calculations to ring substances used as a model to asphaltenes. Asphaltenes are complicated to characterize by determining its composition on a molecular basis, but it is generally accepted that they can be characterized as a fraction of crude oil soluble in toluene and insoluble in n-alkenes (pentane, hexane, heptane) ([SJÖBLOM *et al.*, 2003](#)). They've attracted a lot of studies to develop models for its structure for the last ten years due all the problems they can cause during its transportation and refining ([SJÖBLOM; SIMON; XU, 2015](#)). The strategy for this work was to use molecules that appears in the asphaltenes models and that have similarities in terms of solubilities with it (phenanthrene, anthracene, pyrene) and solvents that characterize asphaltenes (toluene, hexane) to test the efficiency of the SAFT- γ Mie force field in describing the solvation phenomena. The anti solvent/solvent effect of carbon dioxide was also tested due to the CO_2 influence in the asphaltene precipitation during the oil processing. The asphaltenes are described by its solubility, hence if this coarse grained model can correctly describe solvation free energy differences of simple molecules mimicking them, the simulations scale can be increased and studies with more complete asphaltenes models can be carried out.

2 Literature Review

2.1 Coarse Grained Force Fields

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

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

One of the first applications of coarse grained models is the study of protein folding (LEVITT; WARSHE, 1975; LEVITT, 1976). These earlier protein CG models were based on the structure of the molecule and they contributed for the knowledge of the physicochemical forces associated with protein folding and protein interactions (KOGA; TAKADA, 2001). More recent, models focused on retaining the protein's chem-

ical specificity. The Bereau and Deresmo model (BEREAU; DESERNO, 2009) has a up to four-bead representation and was used in studies of protein folding and aggregation. However, this model still needs tuning to improve protein stability (BEREAU; BACHMANN; DESERNO, 2010). The OPEP (Optimized Potential for Efficient Protein Structure Prediction) model (STERPONE *et al.*, 2014; STERPONE; DERREUMAUX; MELCHIONNA, 2015) has up to six-bead representation. It was used to investigate a variety of phenomena, ranging from protein folding to *ab initio* peptide structure prediction (BARDUCCI; BONOMI; DERREUMAUX, 2011; CHEBARO *et al.*, 2009b; CHEBARO *et al.*, 2009a). 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 later one is the most popular model for the CG modeling of membrane proteins (MARRINK; TIELEMAN, 2013). The MARTINI model is also extensively used as CG model for water. This model represents four water molecules as one bead using a shifted Lennard Jones potential for the non bonded interactions. Though its extensive use, the MARTINI water model doesn't properly represent properties as interfacial tension and compressibility (HE *et al.*, 2010) and can freeze at room temperature (WINGER *et al.*, 2009; MARRINK *et al.*, 2007), what makes necessary 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 the 12-6 potential. With the idea of improving the MARTINI model, Chiu, Scott e Jakobsson (2010) used the Morse Potential, which is softer than the LJ potential. Meanwhile, Shinoda, Devane e Klein (2007) used different forms of the Mie potential. They concluded that a 12-4 Mie Potential was ideal for the all water cross interactions and a 9-6 Mie Potential was suited for all the solute-solute interactions.

Outside of the Martini framework, He *et al.* (2010) studied different levels of coarse-graining for water ranging for one to 4 molecules per bead using different Mie and Morse potentials. Other works also assessed the use of Soft-core potentials to study aqueous solutions of surfactants (SHINODA; DEVANE; KLEIN, 2007), ionic liquids (BHARGAVA; KLEIN, 2009), lipids (SHINODA; DEVANE; KLEIN, 2010) and membranes (PANTANO; KLEIN, 2009). Other CG force field for water based on the Mie Potential is the SAFT- γ Mie (LOBANOVA *et al.*, 2015). In this model, the water molecule can be represented by two different one isotropic bead interacting via a 8-6 Mie Potential models. The CGW1-vle model was parametrized using saturated-liquid density and vapor pressure data, and should be used for simulations of aqueous systems' fluid-phase equilibrium at high temperatures and pressures. This model still suffers from premature freezing with a triple point at 343 K. The other model, CGW1-ift, was parametrized using saturated-liquid density and vapor-liquid interfacial tension, hence it is best suited for interfacial properties calculations. Both models have temperature-dependent

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

2.2 Solvation Free Energies Based on Molecular Dynamics

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

$$F(N, V, T) = -\kappa_b T \ln Q(N, V, T) \quad (2.1)$$

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

$$Q(N, V, T) = \int d^n p d^n r \exp \left[-\beta \left(\sum_{i=1}^N \frac{p_i^2}{2m_i} + U(r_1, \dots, r_n) \right) \right] \quad (2.2)$$

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

$$G(N, P, T) = -\kappa_b T \ln \Delta(N, P, T) \quad (2.3)$$

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

$$\Delta(N, P, T) = \frac{1}{V_0} \int_0^\infty dV \int d^n p d^n r \exp \left[-\beta \left(\sum_{i=1}^N \frac{p_i^2}{2m_i} + U(r_1, \dots, r_n) + PV(r_1, \dots, r_n) \right) \right] \quad (2.4)$$

Since it is only possible to obtain free energy differences, solvation free energy calculations based on molecular dynamics estimate the difference between the Gibbs

free energies of end states:

$$\Delta G_{AB} = G_B - G_A = -\kappa_b T \ln \left(\frac{\Delta_B}{\Delta_A} \right) = -\kappa_b T \ln \left(\frac{Z_B}{Z_A} \right) \quad (2.5)$$

Since both states are at the same temperature, the moment integrals in the ratio Δ_B/Δ_A can be simplified. Hence this ratio becomes equal to the ratio of configuration partition functions:

$$\frac{Z_B}{Z_A} = \frac{\int_0^\infty dV \int d^n r \exp [-\beta (U(r_1, \dots, r_n) + PV(r_1, \dots, r_n))]_B}{\int_0^\infty dV \int d^n r \exp [-\beta (U(r_1, \dots, r_n) + PV(r_1, \dots, r_n))]_A} \quad (2.6)$$

The Gibbs free energy difference between end states A and B are, more specifically, the difference between the solute alone in the gas phase and the solute interacting with the solvent. In order to these differences be accurate, the states' phase integral must have sufficient overlap (KLIMOVICH; SHIRTS; MOBLEY, 2015). This can be achieved by calculating the free energy difference between a series of intermediates states. The result of these differences are independent of the path chosen since the free energy is a state function. That's why the states used typically don't have a physical sense, they are alchemical states which are only linking the physical states of interest. The strategy for solvation free energy calculations follows a thermodynamic cycle to gradually insert the solute molecule into the solvent as illustrated in the [Figure 2.2.1](#). According to this cycle, the free energy of solvation can be expressed as:

$$\Delta G_{solv} = \Delta G_{1 \rightarrow 4} = \Delta G_{1 \rightarrow 2} + \Delta G_{2 \rightarrow 3} + \Delta G_{3 \rightarrow 4} \quad (2.7)$$

The solvation free energy between states 1 and 2 in the cycle is the one associated with turning off the molecule's non bonded interactions in the gas phase. The following transformation, $\Delta G_{2 \rightarrow 3}$, is the free energy of moving the non-interacting molecule in the gas phase to the solvent and is equal to zero since the transformation of a non interacting molecule doesn't depend on the environment. Lastly, $\Delta G_{3 \rightarrow 4}$ is the free energy required to the non-interaction molecule in the aqueous phase regain its non-bonded interactions. The solvation free energy calculation can be classified according to the types of the non bonded interactions that are turned of in the $1 \rightarrow 2$ and $3 \rightarrow 4$ parts of the cycle. If both the non-bonded interactions with the environment and the internal interactions are turned of, this is the annihilation free energy. Meanwhile, if only the non-bonded interactions with the environment are turned off, this is the decoupling free energy. In this later case, $\Delta G_{1 \rightarrow 2} = 0$ and the $\Delta G_{solv} = \Delta G_{3 \rightarrow 4}$. The methods used to carry out theses transformations scale the solute charges to zero and then turn of the interactions corresponding to the Lennard Jones potential. In order to carry out the later process, a modified potential with a coupling parameter λ is used. Each λ represent an

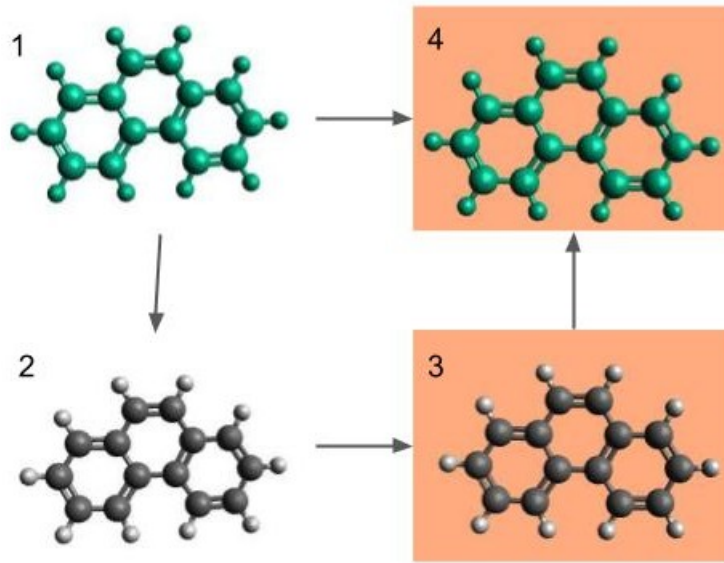


Figure 2.2.1 – Thermodynamic cycle for solvation free energy calculations with molecular dynamics (Adapted from Klimovich, Shirts e Mobley (2015))

alchemical state and, when $\lambda = 0$, there is no interaction with the solvent and, when $\lambda = 1$, the interactions are fully activated. The coupling of the λ parameter could be linear, but it could generate numerical problems related to the exponential part of the Potential. That's why the non-linear soft-core scheme (BEUTLER *et al.*, 1994) is usually used, the generalized soft core potential is given by:

$$U^{sc}(r) = \lambda \epsilon \frac{\lambda_r}{\lambda_r - \lambda_a} \left(\frac{\lambda_r}{\lambda_a} \right)^{\left(\frac{\lambda_a}{\lambda_r - \lambda_a} \right)} \left\{ \frac{1}{[\alpha(1 - \lambda) + (r/\sigma)^{\lambda_a}]^{\lambda_r/\lambda_a}} - \frac{1}{\alpha(1 - \lambda) + (r/\sigma)^{\lambda_a}} \right\} \quad (2.8)$$

Using the Lennard Jones exponents, Eq. (2.8) becomes:

$$U_{LJ}^{sc}(r) = 4\lambda \epsilon \left\{ \frac{1}{[\alpha(1 - \lambda) + (r/\sigma)^6]^2} - \frac{1}{\alpha(1 - \lambda) + (r/\sigma)^6} \right\} \quad (2.9)$$

where α is a constant in which the value of 0.5 is normally assumed to it. The $\Delta G_{3 \rightarrow 4}$ can be then obtained by doing independent simulations in different values of λ or by doing expanded ensemble simulations (LYUBARTSEV *et al.*, 1992) which samples all state in a single simulation. This method allows a faster sampling across the alchemical states considering that the kinetic barriers are not substantial. The free energies of solvation obtained can be used to calculate the partition coefficients, that are a measure of the partitioning of one solute in two solvents (a and b) at a temperature T:

$$\log P^{a/b} = \frac{\Delta G_{solv}^a - \Delta G_{solv}^b}{2.303RT} \quad (2.10)$$

2.3 Post simulation methods

The data from molecular dynamics simulations method explained in the section above provide the potential energies correspondent to each λ . These potential energies obtained needs to be post processed and analyzed in order to calculate the solvation free energies. Some of the widely used method for these calculations are going to be briefly describe below.

2.3.1 Thermodynamic integration

The thermodynamic integration method (KIRKWOOD, 1935) uses equilibrium averages to evaluate the derivative of the potential energy with respect to the coupling parameter. Then, the free energies are obtained as the integration of the derivatives of the initial and final state:

$$\Delta G_{solv} = \int_0^1 \frac{\partial U}{\partial \lambda} d\lambda \quad (2.11)$$

The integration in Eq. (2.11) is obtained by interpolating the output data form the simulations in different ways. Some examples of methods for the interpolations are the trapezoidal rule or natural cubic spline (PALIWAL; SHIRTS, 2011). There are also other more complex schemes that are usually system specific as the works of Jorge *et al.* (2010) and Shyu e Ytreberg (2010) and that use fitting functions to interpolate the data.

2.3.2 Free energy of Pertubation (FEP)

The free energy of perturbation method (ZWANZIG, 1954) is the oldest and one of the most general purpose strategy to calculate free energy differences. In this method, the difference between two thermodynamic states A and B is given by:

$$\Delta G_{AB} = -\frac{1}{\beta} \ln \langle e^{-\beta(U_B - U_A)} \rangle_A \quad (2.12)$$

According to the equation above, the free energy difference is calculated by doing an average over the potential energies of state A and B obtained during the simulation of state A. This method requires a great overlap between states, the state B needs to represent a small perturbation in state A, in order to obtain a rapid convergence of the free energy difference. To assure the overlap, it is possible to carry out simulations in N intermediate sates between A and B, so Eq. (2.12) becomes:

$$\Delta G_{AB} = -\frac{1}{\beta} \ln \left(\frac{1}{N} \sum_{i=0}^{N+1} e^{-\beta(U_{i+1} - U_i)} \right)_i \quad (2.13)$$

The way of calculating ΔG of Eq. (2.13) is called Exponential Averaging (EXP) (ZWANZIG, 1955; PALIWAL; SHIRTS, 2011). The direction of the transformation is also important in this method. If the direction is of decreasing entropy, the step is of insertion (ΔG_{AB}) and method is called insertion exponential averaging (IEXP). The direction of increasing entropy is called a deletion step (ΔG_{BA}) and the method is labeled as deletion exponential averaging (DEXP). These directions can yield different values of free energy differences due the under sampling in the tail regions of the ΔG_{AB} distribution (KLIMOVICH; SHIRTS; MOBLEY, 2015; POHORILLE; JARZYNSKI; CHIPOT, 2010). These problems makes the EXP methods not suited to calculate free energy differences when the system hasn't a sufficient a overlap. For these cases, the Bennet Acceptance Ratio or the Multi-State Bennet Acceptance Ratio is more indicated.

2.3.3 Bennet Acceptance Ratio (BAR)

The BAR method (BENNETT, 1976) was developed with the intent of eliminating the bias in the free energy estimation. It uses the uncorrelated samples of the potential energy in both directions ($A \rightarrow B$ 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 (i and j) is calculated by the self-consistence solution of the following equations:

$$\Delta G_{ij} = \frac{1}{\beta} \ln \left(\frac{\sum_{k=1}^{N_j} \frac{1}{1 + \exp[-\beta(\Delta U_k^j - C)]}}{\sum_{l=1}^{N_i} \frac{1}{1 + \exp[-\beta(\Delta U_l^i - C)]}} \right) + C - \frac{1}{\beta} \ln \left(\frac{N_j}{N_i} \right) \quad (2.14)$$

$$C = \Delta G_{ij} + \frac{1}{\beta} \ln \left(\frac{N_j}{N_i} \right) \quad (2.15)$$

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

$$s_{ij}^2 = \frac{1}{\beta^2 N_i} \left[\frac{\langle f^2(x) \rangle_i}{\langle f(x) \rangle_i^2} - 1 \right] + \frac{1}{\beta^2 N_j} \left[\frac{\langle f^2(x) \rangle_j}{\langle f(x) \rangle_j^2} - 1 \right] \quad (2.16)$$

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

unbiased estimate of an entire transformation with the BAR method (PALIWAL; SHIRTS, 2011).

There are two other methods related to the BAR method that don't solve the Eqs. (2.14) and (2.15) self consistently. By doing that, the free energy difference will not have minimum variance, but significant space and disk memory can be saved since the averages of Eqs. (2.14) - (2.16) are accumulated. The two methods are the Unoptimized Bennett Acceptance Ratio (UBAR) and Range-Based Bennett Acceptance Ratio (RBAR). The first one avoids the self consistently resolution of the BAR equations by defining $C = \beta^{-1} \ln(N_j/N_i)$. The UBAR method requires that the intermediate free energies differences are approximately equal to zero to obtain optimal estimations. Meanwhile, the RBAR method selects a range initial guesses of the constant C in order to calculate a range of ΔG_{ij} . The value of free energy difference correspondent to the minimum variance is then used as input in Eq. (2.15) to calculate the value of C . Hence, this method requires a good estimation of the initial range of the values of C . In terms of accuracy, the UBAR method can be as accurate as the BAR method, but it may end up being as computational costly (PALIWAL; SHIRTS, 2011).

2.3.4 Multistate Bennet Acceptance Ratio (MBAR)

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

$$G_i = \frac{1}{\beta} \ln \sum_{k=1}^K \sum_{n=1}^{N_k} \frac{\exp[-\beta U_i(x_{kn})]}{\sum_{l=1}^K N_l \exp[\beta(G_l - U_l(x_{kn}))]} \quad (2.17)$$

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

$$s_{ij}^2 \Delta G_{ij} \equiv \text{cov}(-\ln \hat{Z}_j / \hat{Z}_i, -\ln \hat{Z}_j / \hat{Z}_i) \quad (2.18)$$

where \hat{Z}_j and \hat{Z}_i are the partition functions of states i and j . The MBAR method can be considered as limit case of the weighted histogram analysis method (WHAM) (KUMAR *et al.*, 1992) for computing free energies. The Eq. (2.17) becomes equal to the WHAM equations if the histogram width tend to zero. Despite this, the MBAR is still more suited because it doesn't have the bias associated with the discretization and it allows to calculate an error estimate.

3 Fundamentals of the Computational Methods

3.1 SAFT- γ Mie Force Field

3.1.1 SAFT-VR Mie EoS

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

$$U_{Mie}(r) = \epsilon \frac{\lambda_r}{\lambda_r - \lambda_a} \left(\frac{\lambda_r}{\lambda_a} \right)^{\left(\frac{\lambda_a}{\lambda_r - \lambda_a} \right)} \left[\left(\frac{\sigma}{r} \right)^{\lambda_r} - \left(\frac{\sigma}{r} \right)^{\lambda_a} \right] \quad (3.1)$$

where 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. This equation uses the Barker e 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 a 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:

$$\frac{A}{N\kappa_b T} = a = a^{IDEAL} + a^{MONO} + a^{CHAIN} \quad (3.2)$$

3.1.1.1 Ideal Contribution

The ideal contribution for a mixture is given by:

$$a^{IDEAL} = \sum_{i=1}^{N_c} x_i \ln (\rho_i \Lambda_i^3) - 1 \quad (3.3)$$

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

3.1.1.2 Monomer Contribution

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

$$a^{MONO} = \left(\sum_{i=1}^{N_c} x_i m_{s,i} \right) a^M \quad (3.4)$$

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):

$$a^M = a^{HS} + \beta a_1 + \beta a_2^2 + \beta a_3^3 \quad (3.5)$$

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

$$a^{HS} = \frac{6}{\pi \rho_s} \left[\left(\frac{\zeta_2^3}{\zeta_3^2} - \zeta_0 \right) \ln(1 - \zeta_3) + \frac{3\zeta_1\zeta_2}{1 - \zeta_3} + \frac{\zeta_2^3}{\zeta_3(1 - \zeta_3)^2} \right] \quad (3.6)$$

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

$$\zeta_l = \frac{\pi \rho_s}{6} \left(\sum_{i=1}^{N_c} x_{s,i} d_{ii}^l \right), l = 0, 1, 2, 3 \quad (3.7)$$

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

$$x_{s,i} = \frac{m_{s,i} x_i}{\sum_{k=1}^{N_c} m_{s,k} x_k} \quad (3.8)$$

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

$$d_{ii} = \int_0^{\sigma_{ii}} (1 - \exp(-\beta U_{ii}^{Mie}(r))) dr \quad (3.9)$$

The integral in Eq. (3.9) is normally obtained by means of Gauss-Legendre with a 5-point quadrature (PAPAIIOANNOU *et al.*, 2014). The detailing of terms of Eq. (3.4) can be found in Lafitte *et al.* (2013).

3.1.1.3 Chain Contribution

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

$$a^{CHAIN} = - \sum_{i=1}^{N_c} x_i (m_{s,i} - 1) \ln(g_{ii}^{Mie}(\sigma_{ii})) \quad (3.10)$$

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

$$g_{ij}^{Mie}(\sigma_{ij}) = g_{d,ij}^{HS}(\sigma_{ij}) \exp[\beta \epsilon g_{1,ij}(\sigma_{ij})/g_{d,ij}^{HS}(\sigma_{ij}) + (\beta \epsilon)^2 g_{2,ij}(\sigma_{ij})/g_{d,ij}^{HS}(\sigma_{ij})] \quad (3.11)$$

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

3.1.1.4 Ring Contribution

There are two forms for the Helmholtz free energy for rings formed from m_s tangentially bonded segments in the literature. The first one (LAFITTE *et al.*, 2012) considered that the difference between a chain and a ring molecule is that the latter has one more bond that is connecting the first segment to the last. With this assumption, the Eq. (3.10) can be adapted to rings by:

$$a^{RING} = - \sum_{i=1}^{N_c} x_i m_{s,i} \ln(g_{ii}^{Mie}(\sigma_{ii})) \quad (3.12)$$

According to Lafitte *et al.* (2012), Eq. (3.12) needs an additional parametrization with molecular simulation data so the EoS can be used in molecular simulations, but this procedure is not the necessary for chain molecules. Recently Müller e Mejía (2017) tried to correct this inconsistency by means of developing the ring free energy based on the work of Müller e Gubbins (1993) who obtained rigorous expressions for hard fluids with molecular geometries of rings of $m_s = 3$. The final expression developed for the ring dimensionless Helmholtz free energy is:

$$a^{RING} = - \sum_{i=1}^{N_c} x_i (m_{s,i} - 1 + \chi_i \eta_i) \ln(g_{ii}^{Mie}(\sigma_{ii})) \quad (3.13)$$

$\eta_i = m_{s,i} \rho_i \sigma_{ii}^3 / 6$ is the packing fraction and χ_i is a parameter which depends on $m_{s,i}$ and on the geometry of the ring of each component i . For a value of $\chi = 0$ Eq. (3.13) is equal to Eq. (3.10). Meanwhile, the equation corresponds to a hard sphere system of triangles when $\chi = 1.3827$. Müller e Mejía (2017) also calculated values of ζ for $m_s = 3, m_s = 4, m_s = 5, m_s = 7$ with pseudo-experimental data from molecular dynamics (MD) for a defined pure fluid. The values of χ for each geometry estimated can be seen in the Figure 3.1.1.

3.1.1.5 Combining rules for the intermolecular potential parameters

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

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \quad (3.14)$$

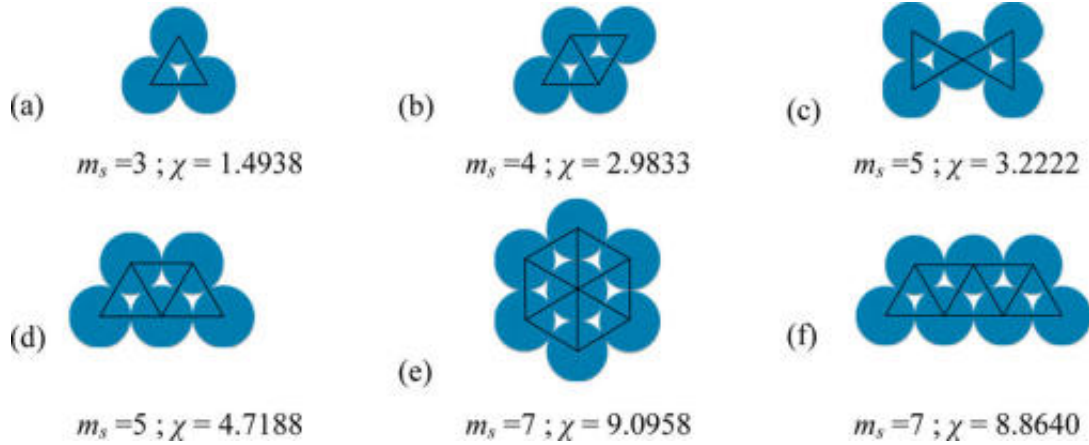


Figure 3.1.1 – Values for parameter χ according to the ring geometry (MÜLLER; MEJÍA, 2017)

$$\lambda_{k,ij} - 3 = \sqrt{(\lambda_{k,ii} - 3)(\lambda_{k,jj} - 3)}, k = r, a \quad (3.15)$$

$$\epsilon_{ij} = (1 - k_{ij}) \frac{\sqrt{\sigma_{ii}^3 \sigma_{jj}^3}}{\sigma_{ij}^3} \sqrt{\epsilon_{ii} \epsilon_{jj}} \quad (3.16)$$

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

3.1.2 Parameter Estimation for the SAFT- γ Mie Force Field

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

Each substance has initially five parameters to be estimated ($m_s, \sigma, \epsilon, \lambda_r$ and λ_a) according to Eq. (3.1). The number of segments are usually fixed in an integer value so it can be used in the coarse grained simulations. The attractive parameter can also be

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

$$\min_{\sigma, \epsilon, \lambda_r} F_{obj}(\sigma, \epsilon, \lambda_r) = \sum_{i=1}^{N_p} \left(\frac{P_v^{SAFT}(T_i, \sigma, \epsilon, \lambda_r) - P_v^{exp}(T_i)}{P_v^{exp}(T_i)} \right)^2 + \sum_{i=1}^{N_p} \left(\frac{\rho_l^{SAFT}(T_i, \sigma, \epsilon, \lambda_r) - \rho_l^{exp}(T_i)}{\rho_l^{exp}(T_i)} \right)^2 \quad (3.17)$$

where N_p is the number of experimental points, P_v is the vapor pressure and ρ_l is the saturated liquid density. The minimized properties can also change and other possible ones as superficial tension and speed of sound can also be taken into account. These multiple parameters make it necessary the use of a wide range of experimental data since multiple solutions can be found for the fit. So one need to be careful in deciding the level of coarse graining (i.e. the parameter m_s) and subsequent parameter space that will not result in some physical inconsistencies like a premature freezing fluid.

Lafitte *et al.* (2012) suggested that the two corrections factors (c_σ and c_ϵ) should be estimated with simulation data when using Eq. (3.12) for the ring contribution. They are related to the EoS parameters by scaled parameters:

$$\sigma^{scaled} = c_\sigma \sigma^{SAFT} \quad (3.18)$$

$$\epsilon^{scaled} = c_\epsilon \epsilon^{SAFT} \quad (3.19)$$

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 results for ring molecules modeled with Eq. (3.12). The objective function for this second estimation is given by:

$$\min_{c_\sigma, c_\epsilon} F_{obj}(c_\sigma, c_\epsilon) = \sum_{i=1}^{N_p} \left(\frac{P_v^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT}) - P_v^{SAFT}(T_i, \sigma^{scaled}, \epsilon^{scaled})}{P_v^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT})} \right)^2 + \sum_{i=1}^{N_p} \left(\frac{\rho_{liq}^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT}) - \rho_{liq}^{SAFT}(T_i, \sigma^{scaled}, \epsilon^{scaled})}{\rho_{liq}^{sim}(T_i, \sigma^{SAFT}, \epsilon^{SAFT})} \right)^2 \quad (3.20)$$

The repulsive parameter is maintained in the value found on the minimization of Eq. (3.17), so the refined values for the force field are:

$$\sigma^{sim} = \sigma^{SAFT} / c_\sigma \quad (3.21)$$

$$\epsilon^{sim} = \epsilon^{SAFT} / c_\epsilon \quad (3.22)$$

It is interesting to point out that this new parametrization is not necessary when using Eq. (3.13) as the ring contribution. The other method to obtain the force field parameters is the correspondent state parametrization for the EoS SAFT-VR Mie (MEJÍA; HERDES; MÜLLER, 2014). This method considers that the unweighted volume average of the attractive contribution to the Mie intermolecular potential, a_1 , can be given a mean field approximation:

$$a_1 = 2\pi\rho\sigma^3\epsilon\alpha \quad (3.23)$$

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

$$\alpha = \frac{1}{\epsilon\sigma^3} \int_{\sigma}^{\infty} \phi(r)r^2 dr = \frac{\lambda_r}{3(\lambda_r - 3)} \left(\frac{\lambda_r}{6} \right)^{6/(\lambda_r - 6)} \quad (3.24)$$

The parametrization 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:

$$\lambda_r = \frac{\sum_{i=0} a_i \omega^i}{1 + \sum_{i=1} b_i \omega^i} \quad (3.25)$$

a_i and b_i are dependent parameters of the number of segments and a table with its values is presented in the original paper (MEJÍA; HERDES; MÜLLER, 2014). Substituting λ_r into Eq. (3.24), the van der Waals constant can be found. The reduced critical potential T_c^* can also be related to α by a Padé series:

$$T_c^* = \frac{\sum_{i=0} c_i \alpha^i}{1 + \sum_{i=1} d_i \alpha^i} \quad (3.26)$$

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:

$$T_c^* = \frac{\kappa_b T_c}{\epsilon} \quad (3.27)$$

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

$$\rho_{T_r=0.7}^* = \frac{\sum_{i=0} j_i \alpha^i}{1 + \sum_{i=1} k_i \alpha^i} \quad (3.28)$$

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

$$\rho_{T_r=0.7}^* = \rho_{T_r=0.7} \sigma^3 N_{av} \quad (3.29)$$

where N_{av} is The Avogadro number. This correspondent state method has the advantage of only requiring critical data, that it is available for a great range of fluids, and one liquid density point. In addition to that, there is an on line parameter database obtained with this strategy ([ERVIK; MEJÍA; MÜLLER, 2016](#)).

The binary interaction parameter k_{ij} of Eq. (3.16) is necessary to adjust the mixture behavior of chemically distinct components. Normally, it is estimated minimizing the difference between experimental binary vapor liquid equilibrium or superficial tension data and the SAFT-VR Mie EoS output data ([MÜLLER; MEJÍA, 2017; LOBANOVA et al., 2016](#)). The objective function is similar to:

$$\min_{k_{ij}} F_{obj}(k_{ij}) = \sum_{k=1}^{N_p} \left(\frac{P_v^{SAFT}(T_k, x, k_{ij}) - P_v^{exp}(T_k, x)}{P_v^{exp}(T_k, x)} \right)^2 + \sum_{k=1}^{N_p} \left(\frac{\rho_l^{SAFT}(T_k, x, k_{ij}) - \rho_l^{exp}(T_i)}{\rho_l^{exp}(T_i)} \right)^2 \quad (3.30)$$

However, [Ervik et al. \(2016\)](#) used molecular simulation results to fit the parameter to the superficial tension data of the mixture water-toluene. The strategy followed by them was to do simulations in three values of k_{ij} and then refine the parameter until a value in good agreement with the experimental data was found.

3.2 Expanded Ensemble Method

Instead of doing various simulations in different values of λ , expanded ensemble simulations ([LYUBARTSEV et al., 1992](#)) were developed to allow a non-Boltzmann

sampling scheme of different states in only one simulation. The statistical expanded ensemble, Z^{EE} can be defined as a sum of sub ensembles Z_i in different values of λ :

$$Z^{EE} = \sum_{i=1}^N Z_i(\lambda_i) \exp(\eta_i) \quad (3.31)$$

where N is the number of alchemical states and η_i is the arbitrary weight of the sub ensemble Z_i at each state. In the application of this method for solvation energy calculations with molecular dynamics, λ corresponds to the coupling parameter of the soft-core potential (Eq. 2.8) and the expanded ensemble is sampled in the MD simulations by performing an arbitrary number of steps followed by a state transition. Chodera e Shirts (2011) proved that the sampling of the expanded ensemble are 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 using the conditional distribution:

$$\pi(x|\lambda_k) = \frac{\exp[-\beta u(x, \lambda_k)]}{\int dx \exp[-\beta u(x, \lambda_k)]} \quad (3.32)$$

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

$$\pi(\lambda_k|x) = \frac{\exp[-\beta u(x, \lambda_k) + \eta_k]}{\sum_{k=1}^K \exp[-\beta u(x, \lambda_k) + \eta_k]} \quad (3.33)$$

where $u(x, \lambda_k) = U(x, \lambda_k) + PV(x, \lambda_k)$ is the reduced potential function for the NPT ensemble. There are a variety of proposal or acceptance schemes to do the expanded sampling using the Eq. (3.33), but Chodera e Shirts (2011) suggested that the independence sampling (LIU, 2002) is the best strategy to increase the number of uncorrelated configurations. The implementation suggested by them updates the state index from i to j by first generating a uniform random number R on the interval $[0, 1)$ and then selecting the smallest new value of j that satisfies the relation below:

$$R < \sum_{i=1}^j \pi(\lambda_i|x) \quad (3.34)$$

The sampling strategy above depends on the weights above in order to assure an adequate sampling of the states. If there isn't a sufficient number of states sampled, the expanded ensemble becomes deficient in obtaining input data to estimate free energy differences with the methods exposed in Chapter 2. Here, the weights can be calculated following the flat-histogram approach (BERG; NEUHAUS, 1992; LEE, 1993; DAYAL *et al.*, 2004). This strategy aims to obtain adequate sampling by sampling all the states

in a equal number of ways, i. e. the ratio of the probability of sampling state (π_i) to the probability of sampling state j (π_j) is equal to one. Given that π_i has the following equation:

$$\pi_i = \frac{Z_i(\lambda_i) \exp(\eta_i)}{Z^{EE}} \quad (3.35)$$

and using Eqs. 2.5 and 2.6, the following relation can be obtained for $\pi_i/\pi_j = 1$:

$$(\eta_i - \eta_j) = \beta(G_i - G_j) \quad (3.36)$$

The Eq. (3.36) is solved iteratively with trial simulations. For the first simulation, the values of η are chosen or set to zero and the histogram of the states visited is obtained. With this histogram is possible to estimate the free energy differences and, since the weights are related to the free energies by Eq. (3.36), the next values of η can be calculated from the previous result. This iteration goes on until a uniform distribution is secured. The weights found can then be used in a longer simulation to obtain the final solvation free energy differences.

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

$$\phi = D(\Lambda) \Pi(\Lambda) \frac{dx(\Lambda)}{d\Lambda} \quad (3.37)$$

In the equation above, Λ is the actual continue value of the coupling parameter. This continue function of λ s may be obtained by interpolating them linearly. $D(\Lambda)$ is the diffusivity at the 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 $dx(\Lambda)/d\Lambda$ can be approximated with the central finite differences. Finally, $\Pi(\Lambda)$ represents the probability of visiting Λ .

$$\Pi(\Lambda) = \frac{C' \bar{\Pi}(\lambda)}{\Lambda_{i+1} - \Lambda_i} \quad (3.38)$$

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

$$\bar{\Pi}(\lambda) = \frac{\pi_{i+1} - \pi_i}{2} \quad (3.39)$$

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

$$D(\Lambda) = \frac{\Lambda_{i+1} - \Lambda_i}{\bar{\Pi}(\lambda) dx(\Lambda)/d\Lambda} \quad (3.40)$$

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

$$\Phi = \int_{\lambda=0}^{\lambda=1} \Pi'(\Lambda_i) d\Lambda = \frac{i}{K} \quad (3.41)$$

where, K is the total number of λ states.

3.3 Gibbs Ensemble Monte Carlo (GEMC)

The Gibbs Ensemble Method ([PANAGIOTOPOULOS, 1987](#)) is used to study phase coexistence with simultaneous Monte Carlo simulations of two boxes, representing a two phase system, with periodic conditions. The boxes exchange molecules, energy and volume between them. The boxes equilibrium is obtained through MC steps that consist of translation and rotation moves, volume exchange and randomly exchanges of molecules between the boxes. For the phase equilibrium of multi-component systems, the GEMC simulations should be carried out at the NPT (constant number of particles, pressure and temperature) ensemble to obey the requirement of an additional degree of freedom for mixtures. Meanwhile, the simulation of single component systems is carried out at constant number of particles, temperature and volume (NVT) since the two phase region would be a line for a system at constant pressure and temperature. The constant volume GEMC ensemble is rigorously equivalent to the canonical ensemble in the thermodynamic limit as demonstrated by [Frenkel e Smit \(2001\)](#). The partition function of the GEMC-NVT ensemble is obtained considering that the particles in both boxes are subjected to the same intermolecular interactions and that the boxes volumes and number of particles (N_1, N_2, V_1 and V_2) can vary while the total volume (V) and total

number of particles (N) remain constant ($N = N_1 + N_2, V = V_1 + V_2$):

$$Q(NVT) \equiv \sum_{N_1}^N \frac{1}{V \Lambda^{3N} N_1! (N - N_1)!} \int_0^V dV_1 V_1^{N_1} V_2^{N_2} \int dx_1^{N_1} \exp[-\beta U(x_1^{N_1})] \int dx_2^{N_2} \exp[-\beta U(x_2^{N_2})] \quad (3.42)$$

In order to define the acceptance rules for the MC moves, 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:

$$\pi(x_1^{N_1}, x_2^{N_2}, N_1, N_2, V_1, V_2) \propto \frac{V_1^{N_1} V_2^{N_2}}{N_1! N_2!} \exp[-\beta U(x_1^{N_1}) - \beta U(x_2^{N_2})] \quad (3.43)$$

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

$$acc_{A \rightarrow B} = \min(1, \exp[-\beta U(x_A^{N_1}) - \beta U(x_B^{N_1})]) \quad (3.44)$$

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

$$acc_{A \rightarrow B} = \min \left(1, \left(\frac{V_1^B}{V_1^A} \right)^{N_1-1} \left(\frac{V_2^B}{V_2^A} \right)^{N_2+1} \exp[-\beta U(x_A^N) - \beta U(x_B^N)] \right) \quad (3.45)$$

Particle exchange moves are carried out to obtain the equality of chemical potential the two 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:

$$acc_{A \rightarrow B} = \min \left(1, \frac{N_1 V_2}{N_2 V_1} \exp[-\beta U(x_A^N) - \beta U(x_B^N)] \right) \quad (3.46)$$

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

4 Methodology

4.1 Phenanthrene Parameterization

The two parameterization strategies for ring molecules described in section 3.1.2 were implemented for phenanthrene. For both of them, only vapor pressure data (MORTIMER; MURPHY, 1923) was used due to the unavailability of saturated liquid density. The attractive parameter λ_a was set to six to avoid correlation with the repulsive parameter. The parameterization with the ring equation of Müller e Mejía (2017) was done with the number of segments fixed in 5 since this level of coarse graining was also used for the similar molecule anthracene:

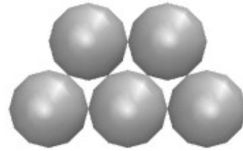


Figure 4.1.1 – Geometry for $m_s = 5$

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

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

The P_v^{exp} is the experimental vapor pressure and P_v^{SAFT} is the vapor pressure obtained with SAFT-VR Mie EoS. The method used to calculate the bubble point with the EoS was the one proposed by Smith, Ness e Abbot (2007). The resulting parameters σ , ϵ and λ_r from this minimization are the final force field parameters to be used in molecular simulations. The parameterization with the Lafitte *et al.* (2012) ring equation was done with $m_s = 3$ so every bead would represent one ring:

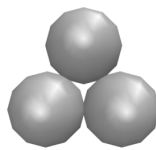


Figure 4.1.2 – Geometry for $m_s = 3$

The first part of the estimation followed the same procedure of the one described above for the Müller e Mejía (2017) equation. As explained in section 3.1.2, the Lafitte

et al. (2012) equation requires the estimation of the correction factors c_σ and c_ϵ (Eqs. (3.18) and (3.19)). The PSO method was also used and the objective function is the one in Eq. (3.20). In this equation, the vapor pressures and saturated liquid densities from molecular simulations was obtained using the Gibbs Ensemble Monte Carlo method on the NVT ensemble (section 3.3).

The boxes for the simulations on the GEMC-NVT were generated inserting 400 molecules on the liquid box and 100 on the vapor one using the Packmol package (MARTÍNEZ *et al.*, 2009). The initial densities of each box were given the value of the saturated densities found with the SAFT-VR Mie Eos, to avoid the migration of all molecules to a single phase throughout the simulation. The equilibration and production time consisted of 10^4 and $5 \cdot 10^4$ 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 \AA with no long range interactions. The saturated vapor density (ρ_{vap}), the saturated liquid density (ρ_{liq}) and the vapor pressure (P_v) were sampled at each 100 MC cycles and this data were divided in five blocks to calculate the averages and standard deviations. The corrections found after the estimations was then used to calculate the of σ and ϵ parameters to use in molecular simulations with Eqs. (3.21) and (3.22).

4.2 Solvation Free energy Calculations

The molecular dynamic simulations to estimate the free energy differences with the SAFT- γ Mie force field were performed using the Lammmps package (PLIMPTON, 1995). The motion's equations were integrated with the velocity-Verlet algorithm (VERLET, 1967) with a time step of 1 fs. The molecules were treated as rigid bodies, as required by the coarse grained model. The thermostat and the barostat were the Nose/Hoover with chains with a damping factors of 100 and 1000 fs respectively. The SAFT- γ Mie model doesn't consider electrostatics interactions explicitly, hence there was no shifting of forces or long range corrections and the potential cutoff was equal to 20 \AA (MÜLLER; MEJÍA, 2017) with a neighbor skin of 2 \AA . The initial configuration of the solvated systems were generated with the Packmol package (MARTÍNEZ *et al.*, 2009). For the binary mixtures, one molecule of solute and a varying number of solvent molecules- 700 molecules for toluene and octanol, 1024 for hexane, 3000 for water - were randomly added to a cubic box. The simulations to study solvation free energy of phenanthrene in a mixture of toluene and carbon dioxide were done with different fractions of carbon dioxide. The system consisted of one molecule of phenanthrene for all the fractions and 123 molecules of CO_2 and 618 molecules of toluene for $w_{CO_2} = 0.087$; 166 molecules of CO_2 ; 589 molecules of toluene for $w_{CO_2} = 0.119$; 232 molecules of CO_2 and 545 molecules of toluene for $w_{CO_2} = 0.169$ and 380 molecules of CO_2 and 446

molecules of toluene for $w_{CO_2} = 0.289$.

All simulations were carried out maintaining the temperature and pressure constant at 298 K and 1 bar, except the ones containing carbon dioxide. These were performed at 298 K and at the pressure of the liquid phase equilibrium correspondent to the CO_2 fraction (CHANG, 1992). Primarily, the initial box was equilibrated at the NPT ensemble for 2 ns and then the resulting configurations were used on the alchemical simulations. These were carried out with the LAMMPS user package for alchemical simulations with the Mie Potential developed by our group, available at <https://github.com/atoms-ufrj/USER-ALCHEMICAL>. In the simulations, the sampling of a new state was tried at every 10 MD steps. In order to define the optimal values of λ and η related to each state, short trials simulations, lasting around 10 ns, were carried out. In the first simulation, the group of λ for all the pairs solvent-solute was: (0.0,0.15,0.2,0.25,0.3,0.4,0.45,0.5,0.55,0.7,0.9,1.0) and the η s were set to zero or were given the values of the ones found for similar mixtures. The subsequent groups of η were estimated with the flat histogram approach (Eq. (3.36)) using the solvation free energy values stemming from the previous simulations. The results with the new weights were then utilized to optimize the group of λ s by minimizing the number of round trips as described in section 3.2. The η s correspondent to the newest group of λ s were interpolated from the free energy differences. With the final values of the η and λ defined for each mixture, larger simulations with a time of 20 ns were performed.

Since the force field considers that the beads don't have charges, there is no coulombic interaction and the Eq. (2.7) becomes equal to $\Delta G_{3 \rightarrow 4}$. The post processing method used to calculate the free energies was the Multistate Bennett Acceptance Ratio (MBAR) described in section 2.3.4. The software alchemical-analysis (KLIMOVICH; SHIRTS; MOBLEY, 2015) were used to obtain the ΔG_{solv} with MBAR and to assess the results quality. The final values of ΔG_{solv} was then used to calculate the partition coefficients in different solvents with Eq. (2.10). The binary interaction parameter of Eq. (3.16) was only estimated for pairs with water as a solvent. The strategy picked was to carry out the simulations described above in three values of k_{ij} and then refine the parameter until a value in good agreement with the experimental solvation free energy was found. The estimation didn't follow the methodology of section 4.1 because the k_{ij} obtained with the EoS gave poor results for the free energy.

5 Results and Discussion

5.1 Solvation free energies

The force field parameters for phenanthrene were not available for the ring geometry on the force field database, hence they were estimated as described in section 4.1. The parameters obtained and the mean percentage error (MPE) of the estimation were:

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

m_s	ϵ/k_B (K)	σ (Å)	λ_r	MPE(%)
3 (LAFITTE <i>et al.</i> , 2012)	485.55	4.197	14.34	1.64 9.74
5 (MÜLLER; MEJÍA, 2017)	262.74	4.077	9.55	0.88

In the table above the first value of MPE for the Lafitte *et al.* (2012) strategy corresponds to the found for the estimation with experimental data and the second corresponds to the the estimation of the corrections factors with uses molecular simulation data. This strategy has this inconsistency of requiring two estimations because the parameters solely estimated with the EoS aren't accurate for molecular simulation, hence the solvation free energy of phenanthrene was only studied with the set of parameters estimated with Müller e Mejía (2017) strategy. In fact, the (LAFITTE *et al.*, 2012) was only followed because it was the only one available when the study first started. 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):

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

	m_s	ϵ/k_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

The solvation free energies of aromatic solutes in non polar (hexane), aromatic (toluene) and hydrogen bonding (1-octanol) solvents were examined with binary in-

teraction parameter equal to zero. Since the force field doesn't account for charges, the solvation free energy is equal to the Mie contribution (Eq. (2.8)). A total of fifteen values of λ were used, except 1-octanol+phenanthrene and 1-octanol+anthracene which had eighteen intermediate states. The final values of λ were concentrated in the region with a steeper slope. This values and the final values of η are available at Appendix A. The overlapping of the intermediate states are an important measure of the reliability of the solvation free energy estimation. In order demonstrate that the group of λ found have a sufficient overlap, the overlapping matrices obtained with the software alchemical-analysis are available at the Appendix B. The solvation free energies of solvation estimated were:

Table 3 – Calculated and experimental values for solvation free energies (kcal/mol) of solutes in non aqueous solvents

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

The absolute deviations when the solvent is hexane are smaller, what shows that the Saft- γ Mie force field performs better for the non polar solvent. Additionally, this force field presented a better result for the pair hexane+benzene than the Trappe force field (GARRIDO *et al.*, 2011). Observing the free energy profile in Figure 5.1.1, the effect of molecule's size on the entropic region of the curve. It was expected that a force field based on a EoS that doesn't explicitly account for hydrogen bond would not perform well for 1-octanol. Despite this, the solvation free energies of propane and phenanthrene stayed in the desired range for applications of 1-2 kcal/mol (MOBLEY; GILSON, 2017). The deviation was much smaller for propane and, this can be attributed to its non polarity and the molecule geometry as in observed by the smoother free energy curve (Figure 5.1.2). The geometry for anthracene and phenanthrene are the same in the force field and they have similar properties, but the absolute deviation of the solvation free energy of anthracene in 1-octanol is much higher than the one of phenanthrene 1-octanol. This may indicate a problem in the parameterization of anthracene.

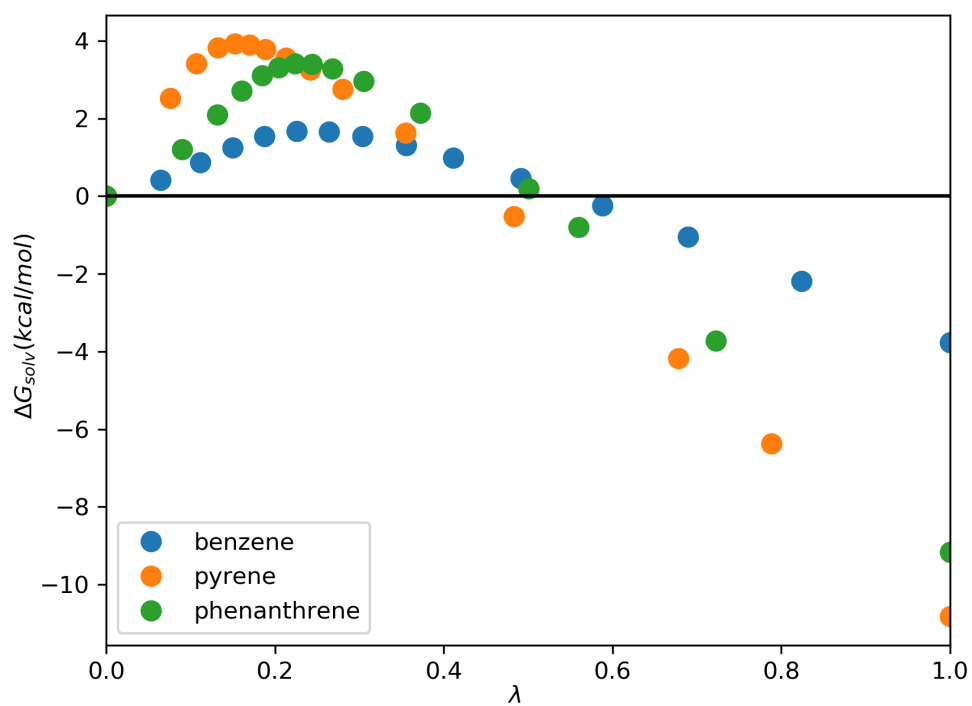


Figure 5.1.1 – Solvation free energy profiles for the hexane solvent

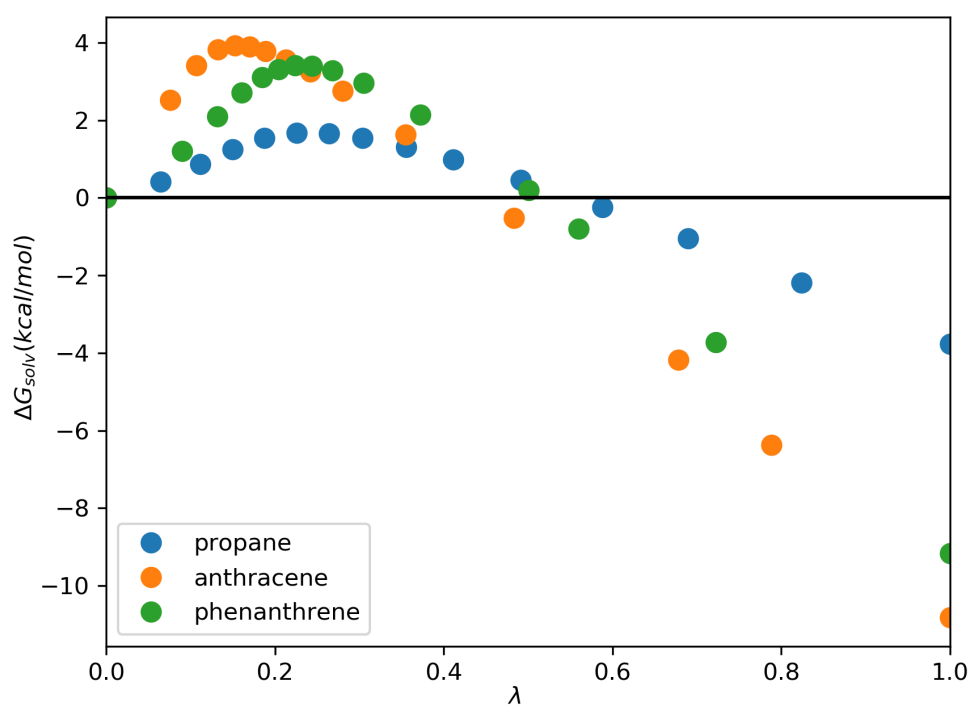


Figure 5.1.2 – Solvation free energy profiles for the 1-octanol solvent

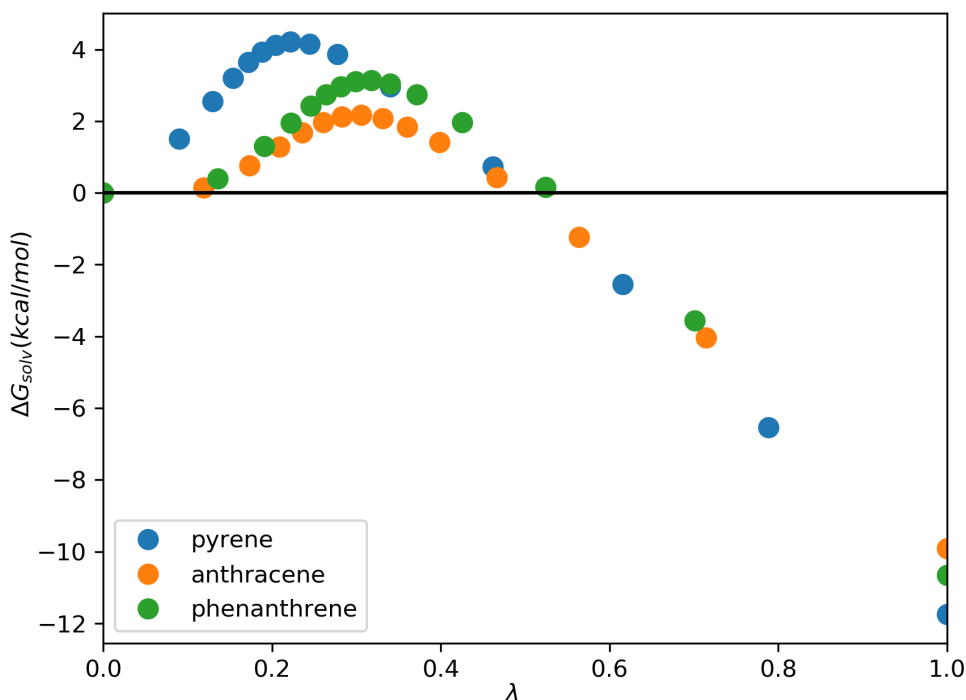


Figure 5.1.3 – Solvation free energy profiles for the toluene solvent

The results also indicated the prediction capability of the force field for the pairs of aromatic solute and solvent. The pattern of smaller deviations for smoother curves (Figure 5.1.3) was also observed for these pairs. The ΔG_{solv} was also calculated of phenanthrene in toluene and in toluene+ CO_2 . To the best of my knowledge, there was no available experimental data for these solvation free energies, but the previous results for phenanthrene in other solvents and for the pair anthracene+toluene indicated that the force field is adequate to describe the solvation phenomena of phenanthrene in an aromatic solvent. The results for these set are exposed bellow:

Table 4 – Calculated values for the solvation free energies (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 increasing of CO_2 mass fraction in toluene caused a slight effect on solvation free energies. First, the ΔG_{solv} decreased with the increase of w_{CO_2} , indicating a higher solubility. From the 0.169 fraction, the effected was reversed and carbon dioxide became

an anti solvent. It was observed that asphaltene precipitation occurs when carbon dioxide mass fractions became higher than 0.10 in the system asphaltene+toluene+carbon dioxide (SOROUSH *et al.*, 2014), what is in accordance with the anti solvent effect of carbon dioxide observed on the calculated values. It is also important to point out the small differences observed in the free energy profiles (Figure 5.1.4) may indicate that there is no influence of CO_2 in solvation of phenanthrene in toluene when using the Saft- γ Mie force field. But, since this is a qualitative study due the lack of this system experimental data, more studies need to be done in order to make a secure assertion about it.

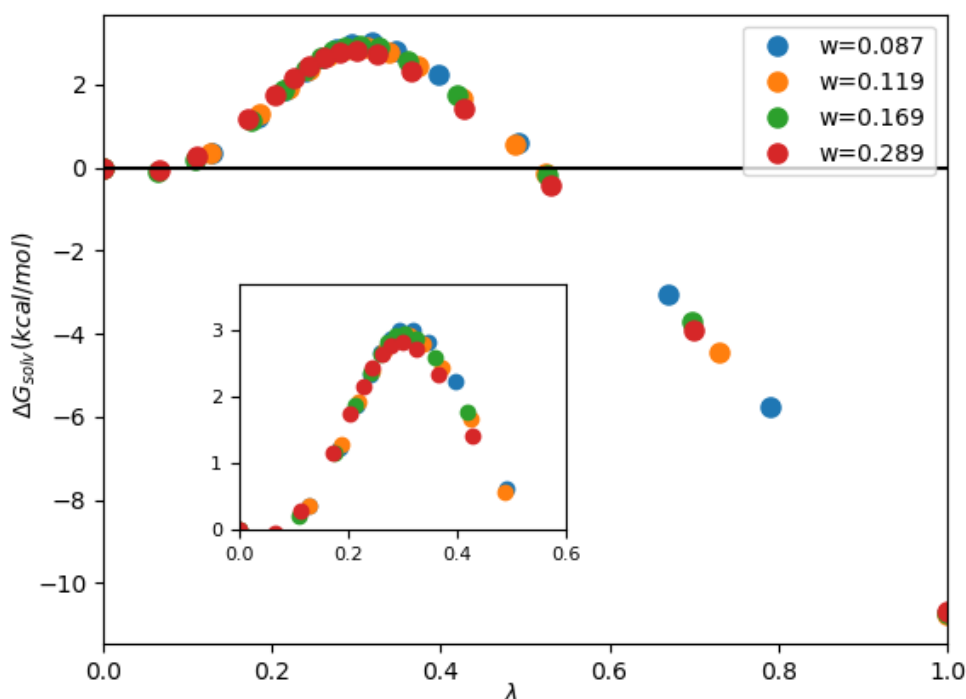


Figure 5.1.4 – Solvation free energy profiles of phenanthrene in toluene+ CO_2

5.2 Hydration free energies

Table 5 – Calculated values for the Gibbs energy of solvation (kcal/mol) of solutes in water for $k_{ij} = 0$

Solute	ΔG_{solv}^{Mie}
propane	1.10 ± 0.01
benzene	-4.45 ± 0.03
toluene	-15.80 ± 0.06
phenanthrene	-10.90 ± 0.04

Table 6 – Binary interaction parameters employed

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

Table 7 – Calculated and experimental values for the Gibbs energy of solvation (kcal/mol) of solutes in water

Solute	ΔG_{solv}^{exp}	ΔG_{solv}^{Mic}	Absolute Deviation	ΔG_{solv}^{GAFF}
propane	2.00 ± 0.20	2.01 ± 0.01	0.01	2.50 ± 0.02
benzene	-0.86 ± 0.20	-1.12 ± 0.01	0.26	-0.81 ± 0.02
toluene	-0.83 ± 0.20	-0.84 ± 0.01	0.01	-0.79 ± 0.03
phenanthrene	-3.88 ± 0.60	3.47 ± 0.02	0.41	-5.26 ± 0.03
RMSE			0.24	

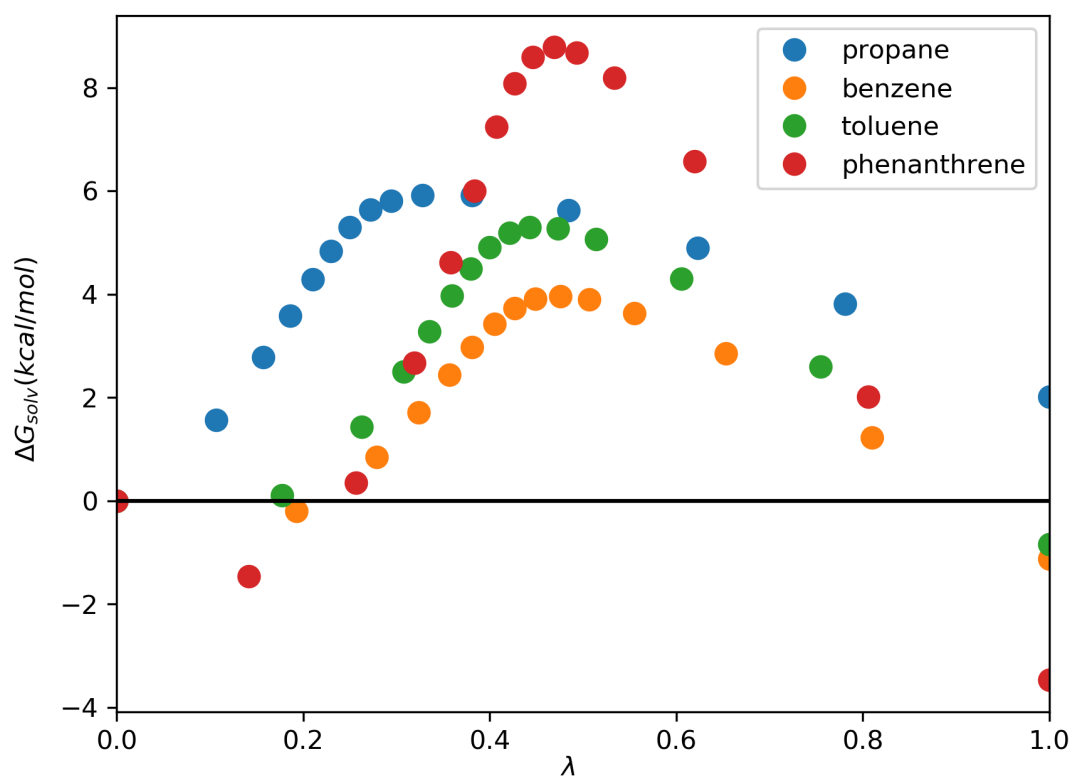


Figure 5.2.1 – Hydration free energy profiles

6 Conclusions

Bibliography

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

AIMOLI, C. G.; MAGINN, E. J.; ABREU, C. Force field comparison and thermodynamic property calculation of supercritical CO_2 and CH_4 using molecular dynamics simulations. *Fluid Phase Equilibria*, v. 368, p. 80–90, 2014.

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

AVENDAÑO, C.; LAFITTE, T.; GALINDO, A.; ADJIMAN, C. S.; JACKSON, G.; MULLER, E. A. Soft- 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.

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.

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

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

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

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

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

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

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

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

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

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

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.

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

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.

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.

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.

HERDES, C.; ERVIK, A.; MEJÍA, A.; MÜLLER, E. A. Prediction of the water/oil interfacial tension from molecular simulations using the coarse-grained soft- 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.

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.

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

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

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

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

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

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. Soft- γ force field for the simulation of molecular fluids: 3. coarse-grained models of benzene and hetero-group models of n-decylbenzene. *Molecular Physics*, v. 110, p. 1189–1203, 2012.

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

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

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

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

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. Soft- γ force field for the simulation of molecular fluids: 4. a single-site coarse-grained model of water applicable over a wide temperature range. *Mol. Phys.*, v. 113, p. 1228–1249, 2015.

LOBANOVA, O.; MEJÍA, A.; JACKSON, G.; MÜLLER, E. A. Soft- γ force field for the simulation of molecular fluids 6: Binary and 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.

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.

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

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

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

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

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

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

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.
- PAPAIIOANNOU, 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.
- 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.
- 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.
- ROWLINSON, J. S.; SWINTON, F. L. *Liquid and Liquid Mixtures*. 3. ed. London: Butterworth Scientific, 1982.
- 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.; PITERA, J. W.; SWOPE, W. C.; PANDE, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *Journal of Chemical Physics*, v. 119, p. 5740, 2003.
- SHYU, C.; YTREBERG, F. M. Reducing the bias and uncertainty of free energy estimates by using regression to fit thermodynamic integration data. *Journal of Computational Chemistry*, v. 6, p. 1018–1027, 2010.
- SJÖBLOM, J.; ASKE, N.; AUFLEM, I. H.; BRANDAL Øystein; HAVRE, T. E.; SæTHER Øystein; WESTVIK, A.; JOHNSEN, E. E.; KALLEVIK, H. Our current understanding of water-in-crude oil emulsions.: Recent characterization techniques and high pressure performance. *Advances in Colloid and Interface Science*, v. 100-102, n. Supplement C, p. 399 – 473, 2003. ISSN 0001-8686.
- SJÖBLOM, J.; SIMON, S.; XU, Z. Model molecules mimicking asphaltenes. *Advances in Colloid and Interface Science*, v. 218, n. Supplement C, p. 1 – 16, 2015. ISSN 0001-8686.

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

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

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

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

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

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

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

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

APPENDIX A – Final values of λ and η

APPENDIX B – Overlapping Matrices