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 de-

pletion 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):

$$U_{Mie}(r) = \epsilon \frac{\lambda_r}{\lambda_r - \lambda_a} \left(\frac{\lambda_r}{\lambda_a}\right)^{\left(\frac{\lambda_a}{\lambda_r - \lambda_a}\right)} \left[\left(\frac{\sigma}{r}\right)^{\lambda_r} - \left(\frac{\sigma}{r}\right)^{\lambda_a}\right]. \quad (1)$$

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, σ, ϵ, λ_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.

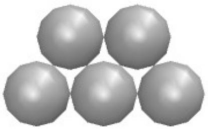


Figure 1: Representation of phenanthrene with the geometry proposed by Müller and Mejía.[25]

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

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

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.

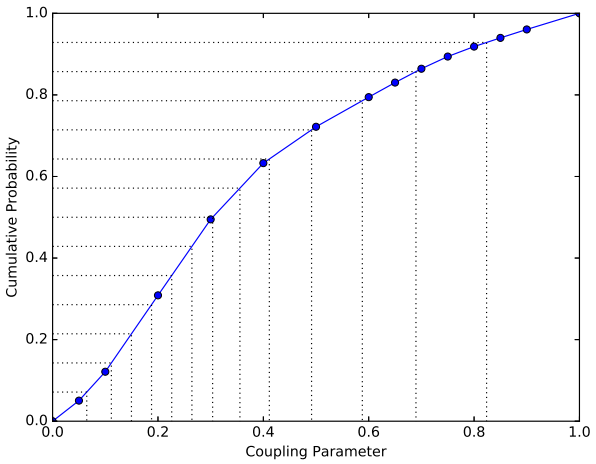
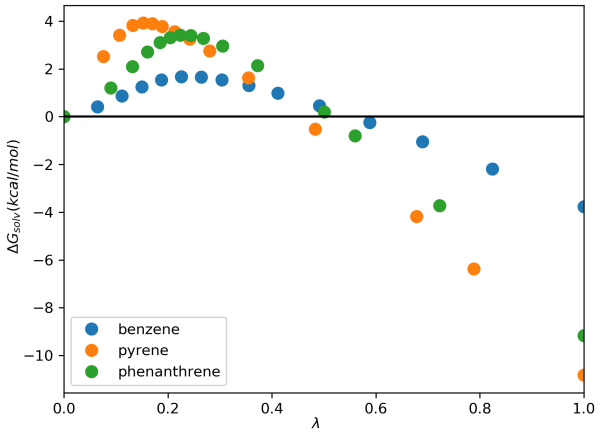
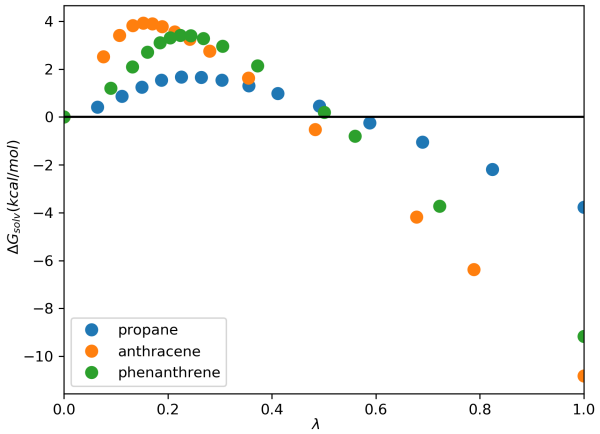


Figure 2: Cumulative probability used to obtain the optimized values of $\lambda's$ for the pair hexane+benzene.

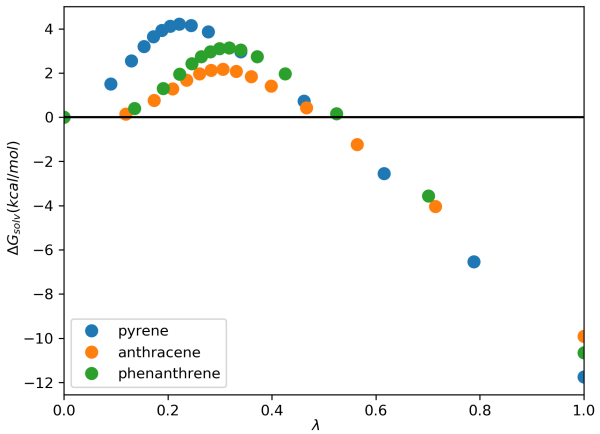
After the expanded ensemble simulations with the intermediate states and weights estimated, we calculated the solvation free energy differences with MBAR. These results and and the absolute deviations to experimental data[59] are available in Table 3. The numerical values for solvation free energies 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 non-polarity 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^{exp}_{solv}	ΔG^{Mie}_{solv}	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

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.

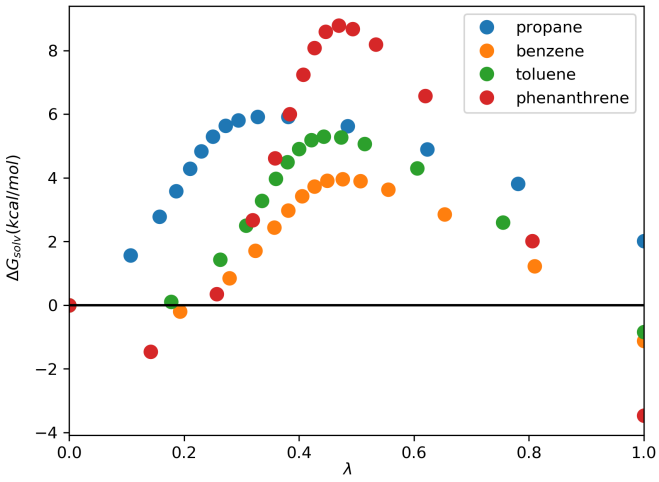


Figure 4: Hydration free energy profiles for different solutes.

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^{exp}_{solv}	ΔG^{Mie}_{solv} $k_{ij} = 0$	Absolute Deviation	ΔG^{Mie}_{solv} $k_{ij} \neq 0$	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.