Solubilities of pyrene in organic solvents: Comparison between chemical potential calculations using a cavity-based method and direct coexistence simulations

C. R. Wand[a,b], M. Fayaz-Torshizi[c], G. Jimenez-Serratos[c], E. A. Müller[c], D. Frenkel[a,1]

[a]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom*
[b]*School of Chemical Engineering and Analytical Science, University of Manchester, Manchester, M13 9PL, United Kingdom*
[c]*Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom*

Abstract

In this paper, we benchmark a cavity-based simulation method for calculating the relative solubility of large molecules in explicit solvents. The essence of the procedure is the accounting of the free energy change associated with an alchemical thermodynamic cycle where, in sequence, a cavity is created in a solvent, a solute is inserted in the cavity and the cavity is annihilated. The free energy change is equated to the excess chemical potential allowing the comparison of solubilities in different solvents. The results obtained using the cavity-based method are compared to direct large-scale molecular dynamics simulations performed using coarse-grained models for calculating the partition coefficient of pyrene between heptane and toluene. We demonstrate the applicability of this cavity-based technique under high pressure/temperature conditions.

Keywords: Solubility, Pyrene, Simulation, heptane, toluene, PAH, SAFT, Thermodynamic Integration

2018 MSC: 00-01, 99-00

---

*Corresponding author: d246@cam.ac.uk

1. Introduction

The understanding and impact of solubility is ubiquitous throughout many scientific areas of study, and is of significant interest to the pharmaceutical [1, 2, 3, 4], oil and gas [5], and food [6, 7] industries. However, currently there is no single prescribed technique to measure the solubility. Experimentally, solubility can be measured by the preparation of saturated solutions from which the concentration of the supernatant fluid is then determined by spectroscopic, chromatographic or volumetric methods. These methods can be both time-consuming and expensive and may suffer from accuracy issues for sparingly soluble solutes where the saturated concentration is low. To address this there has been significant interest in computational calculations that can provide a means to scan multiple targets with minimal experimental input. Methods based on analytical models [8, 9, 10] and machine learning [11] have previously been employed, although these provide no physical insight into the trends in solubility. Molecular level insight can be provided by particle-based simulations and as such would be the preferred *in silico* option. Of these there are many different methods reported in the literature, including osmotic ensemble simulations [12, 13, 14, 15], direct coexistence simulations [16, 17, 18], and chemical potential calculations [19, 20, 21]. The osmotic ensemble uses a grand canonical approach of trial insertions to impose the chemical potential of a solid on a fluid phase in order to calculate the saturated solution concentration. This relies on knowledge of the initial (solid) phase and a non-negligible insertion probability rendering it unsuitable for large solutes in dense solvents. As such it has mainly been employed for small molecules in similar states, *e.g.* in solution and to study adsorption of small molecules in metal organic frameworks [22, 23, 24]. A recently developed cavity-based method [25] where a cavity is created in the solvent and the insertion of the solute molecule is used herein to compute the partition coefficient ($P_{TOL/HEP}$), or solubility ratio, of pyrene between toluene and heptane at a range of conditions. We compare it with a conceptually simpler but computationally more demanding direct simulation method.

2

Pyrene is used as a test case for this work and is a polyaromatic hydrocarbon (PAH) formed of four fused rings (fig 1) commonly found in a wide range of combustion products and in naturally occurring crude oil mixtures [26]. Pyrene shares many characteristics with asphaltene molecules, which are large molecules with pericondensed cores found in crude oil and defined operationally as those who are soluble in toluene and insoluble in n-heptane [27, 28, 29, 30, 5, 31]. Asphaltenes are of primary interest to the oil and gas industry due to their tendency to precipitate in oil wells and pipelines upon by changes in composition, temperature and pressure [32].
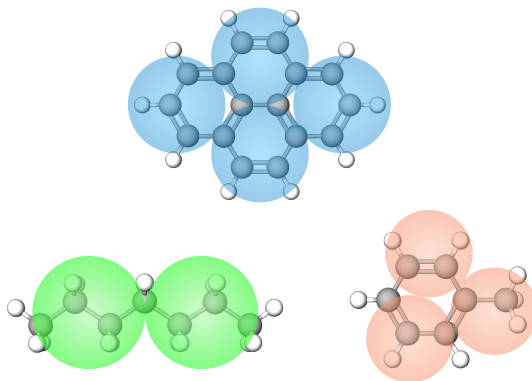

Figure 1: Coarse-grained SAFT models pyrene (blue), heptane (green), toluene (red). Beads are not shown to scale and are overlayed on an atomistic model as a guide to the eye. SAFT models are not fitted to atomistic models, rather to macroscopic thermophysical properties.

The logarithm of the partition coefficient, $P_{I/II}$, of a solute A between solvents I and II is defined as

$$\log P_{I/II} = \frac{[A]_I}{[A]_{II}} \tag{1}$$

Where $[A]_x$ is the saturation concentration of A in solvent x. For two immiscible solvents, *e.g.* water and octanol, this is typically measured experimentally using a shake-flask containing both solvents. However, as with measuring the absolute solubility, the degree of accuracy and expense must be taken into consideration,

particularly when considering sparingly soluble systems [33, 34].

The rest of the paper is organised as follows; Section 2 describes the molecular model, theoretical background and simulation set-up, Section 3 presents the findings of both the direct coexistence simulations and chemical potential calculations. Section 4 provides a comparison and discussion between the two simulation methods and in Section 5 we include the concluding remarks.

2. Model and simulation details

2.1. SAFT Model

We employ the SAFT coarse-grained [35, 36, 37] which represents groups of atoms as single isotropic beads described by the Mie potential;

$$U^{\text{Mie}}(r) = \mathcal{C} \; \epsilon \left[\left(\frac{\sigma}{r} \right)^n - \left(\frac{\sigma}{r} \right)^6 \right] \qquad \text{where} \qquad \mathcal{C} = \frac{n}{n-6} \left(\frac{n}{6} \right)^{\frac{6}{n-6}}, \quad (2)$$

where σ is the size parameter, roughly representative of the bead diameter, ϵ the energy parameter corresponding to the potential energy well depth, n is the repulsive exponent, which determines the range of the potential and r the distance between two beads. Cross interactions are parametrised using the Lafitte mixing rules for a Mie fluid [38], with the $k_{ij} = 0$, i.e. with no adjustable parameters employed.

$$\sigma_{ij} = 0.5(\sigma_{ii} + \sigma_{jj})$$

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

$$n_{ij} = 3 + \sqrt{(n_{ii} - 3)(n_{jj} - 3)}.$$

The SAFT force field is parametrized using a top-down approach, where the equation of state is used to fit the thermophysical of real fluids, *e.g.* vapour-liquid equilibria, and directly inform the corresponding molecular parameters.

4

The SAFT molecular models are built by combining spherical beads ideally resembling the geometry of the parent molecule. (c.f. fig 1).

Pyrene is modelled as four tangentially bonded segments where the angle between every three touching segments is $\frac{\pi}{3}$ radians. Heptane is modelled as a dimer, whilst toluene is modelled using three beads bonded in an equilateral triangle configuration. All the molecules are homonuclear where all beads have the same non-bonded interactions given in Table 1. All bonds are rigid with the bond length connecting two adjacent segments fixed to the value of σ reported in Table 1.

Table 1: Non-bonded interactions for pyrene, heptane and toluene. m_s is the number of beads in each molecule.

Name	m_s	$\frac{\epsilon}{k_\mathrm{B}}[\mathrm{K}]$	$\sigma[\mathrm{\AA}]$	n	ref
Pyrene	4	459.04	4.134	15.79	[37]
Heptane	2	436.13	4.766	23.81	[36]
Toluene	3	267.12	3.684	11.74	[37]

To assess the quality of the force fields used the bulk properties of pyrene and the solvents are investigated. The radial distribution function, $g\left(r\right)$, for the centre of mass (CoM) was calculated for toluene and pyrene and compared to atomistic simulations performed using OPLS-AA [39]. The resulting $g\left(r\right)$ are shown in figs 2 and 3 and good agreement is observed between the coarse-grained and atomistic models.

The vapour-liquid equilibrium (VLE) curves were also calculated for the pure substances and compared to experimental literature values [40, 41]. Simulations are set-up using cubic simulation boxes containing 1000 molecules in a liquid phase at a temperature equal to 70 percent of the critical temperature of the liquid. Once the liquid is allowed to relax at 1 bar, the simulation box is extended in the z-direction by a factor of 4. The system is allowed to equilibrate and canonical NVT simulations are run to calculate saturation liquid densities and vapour pressures. Vapour pressure is estimated using the virial theorem and equated to the component of the pressure tensor in the z-direction.
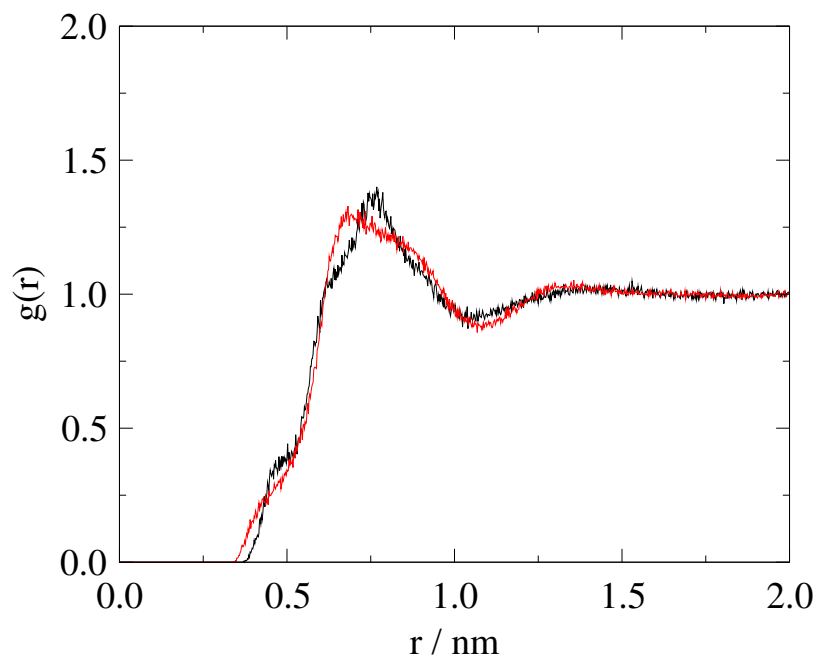
Figure 2: Radial distribution function of CoM-CoM of pyrene at 550 K, where the red line is from atomistic simulations and the black line the SAFT-γ model.
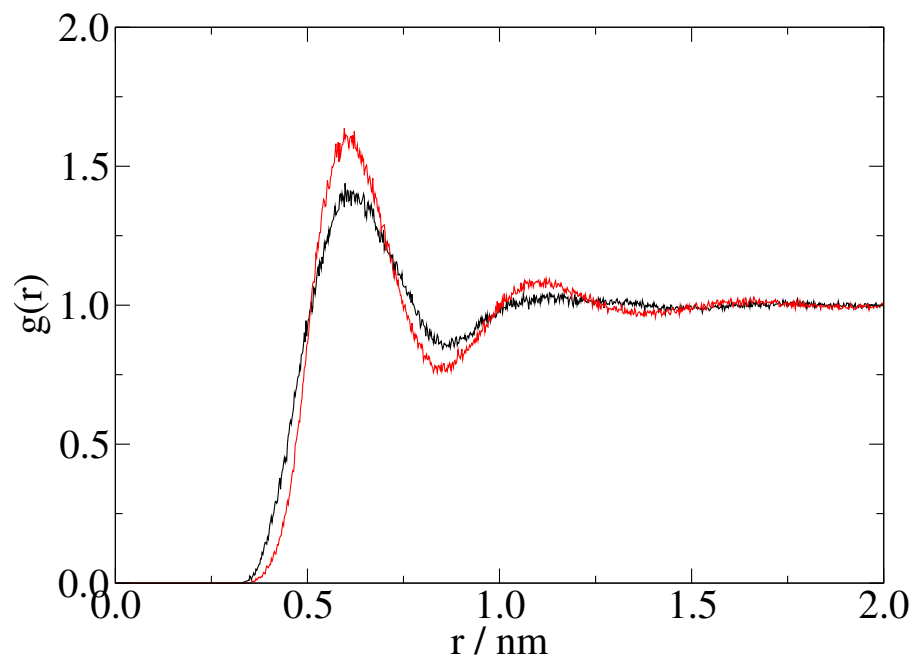

Figure 3: CoM-CoM RDF of toluene at 400 K as calculated using both SAFT and OPLS-AA force fields, colours are equivalent to fig 2.

The results summarized in figs 4 and 5 for the solvents show a good agreement between the simulations and vapour-liquid equilibrium properties for regions in the phase diagram where $T/T_{\text{crit}} \leq 0.9$, the pressures and densities

90 agree well with experimental results. For toluene, it can be clearly seen from fig 5 that the three bead representation captures the essential conformational information of a toluene molecule, thus justifying the usage of such a model in the systems studied. The vapour pressure is almost always overestimated at higher temperatures.

95 For pyrene, the comparison to available data requires the caveat that organic components will decompose much before the critical point is reached. Similarly, the data available in accepted databases are correlations based on extrapolations. Without prejudice, the data from DIPPR [41] is employed in the SAFT parameter fitting procedure which is in conflict with the data from NIST [40]
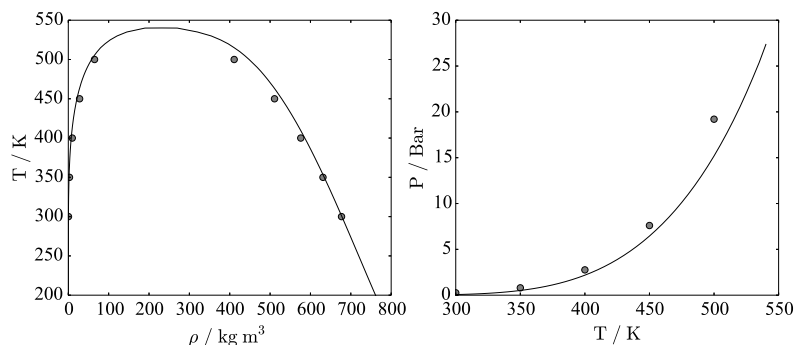
100 (fig 6).

Figure 4: Heptane pure component phase behaviour: **Left** temperature vs saturation densities, **Right** vapour pressure vs temperature. Symbols are simulation results with the SAFT-γ force field, lines are the smoothed experimental results taken from [40].

2.2. Excess chemical potential calculations

We consider the partition coefficient of a solute between two phases, I and II, in contact with each other. Apart from the thermal and mechanical equilibrium conditions, a diffusive equilibrium is required, *i.e.* the chemical potential of the
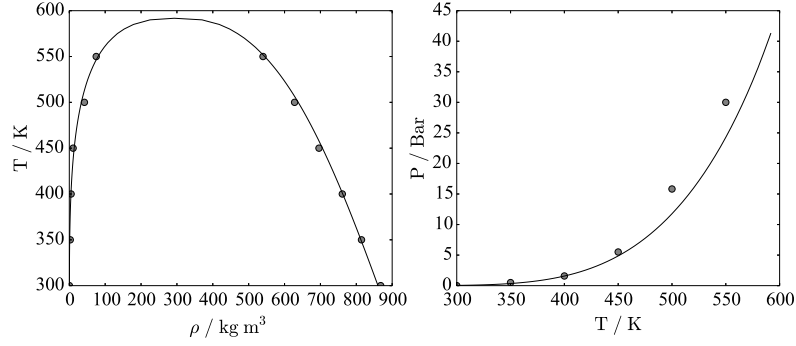
7

Figure 5: Toluene pure component phase behaviour: **Left** temperature vs saturation densities, **Right** vapour pressure vs temperature. Symbols are simulation results with the SAFT-γ force field, lines are the smoothed experimental results taken from [40].
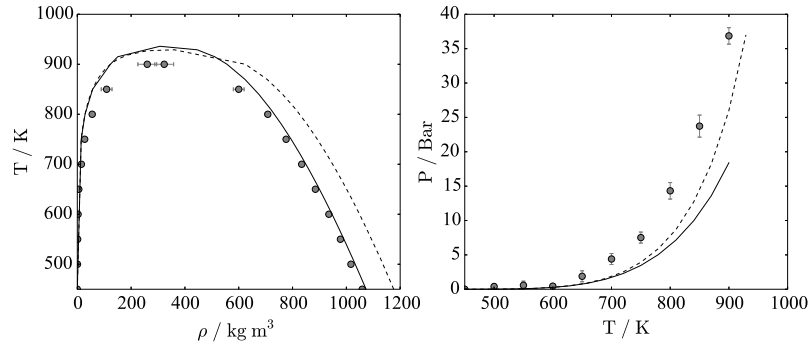

Figure 6: Pyrene pure component phase behaviour: **Left** saturation densities vs temperature, **Right** vapour pressure vs temperature. Symbols are simulation results with the modified SAFT-VR [37] force field, solid lines are estimates from the DIPPR database [41], dashed lines are from NIST [40].

solute should be the same in both phases

$$\mu^I\left(T, p\right) = \mu^{II}\left(T, p\right) \tag{4}$$

where the superscript denotes the solvent phase. The chemical potential is defined as the change in free energy upon the addition of a solute molecule [42, 43] and has three contributions, an intramolecular contribution, μ_{ideal},

$$\mu_{\text{ideal}} = -k_{\text{B}}T \ln \left\{ \frac{q_{\text{solute}}}{\Lambda_{\text{solute}}^3} \right\} \tag{5}$$

where q_{solute} is the solute intramolecular partition function and Λ the solute's de Broglie wavelength; a configurational contribution, μ_{conf},

$$\mu_{\text{conf}} = k_{\text{B}}T \ln \rho, \tag{6}$$

that is density (ρ) or concentration dependent; and the excess chemical potential, μ_{excess},

$$\mu_{\text{excess}} = -k_{\text{B}}T \ln \left\langle \exp\left[-\beta U_{\text{solute}-\text{solvent}}\left(\mathbf{R}_{\text{solute}}\right)\right]\right\rangle_0 \tag{7}$$

where $\beta = 1/k_{\text{B}}T$, $U_{\text{solute}-\text{solvent}}\left(\mathbf{R}_{\text{solute}}\right)$ is the solute-solvent interaction as a function of the position of the solute molecule in the solution $\mathbf{R}_{\text{solute}}$, and $\langle\ldots\rangle_0$ indicates an ensemble average at state 0, that is, the state before the additional solute particle. Thus, equation (4) can be recast as

$$k_{\text{B}}T \ln \rho^I + \mu_{\text{excess}}^I = k_{\text{B}}T \ln \rho^{II} + \mu_{\text{excess}}^{II}, \tag{8}$$

where the intramolecular contributions have cancelled as they are independent of the solvent. When rearranged, equation (8) gives

$$\ln\left(\frac{\rho^I}{\rho^{II}}\right) = \beta\left(\mu_{\text{excess}}^{II} - \mu_{\text{excess}}^I\right) \tag{9}$$

where the left-hand side is $\ln P_{I/II}$ which can then be converted to the more

9

usual $\log P_{I/II}$ using simple logarithmic relationships. Here we consider the case at infinite dilution, *i.e.* the condition where one solute molecule is surrounded entirely by solvent molecules, with the assumption that the solute molecules do not interact with each other. This assumption is valid for sparingly soluble solutes. Commonly μ_{excess} is calculated using Widom insertion [42] and direct growth of the molecule into the bulk solvent [44, 45]. However, these methods are limited to small solutes due to an integrable end-point singularity in the derivative that is caused by the hard-core nature of the solvent potentials *e.g.* Lennard-Jones and the more general Mie potential used here. This singularity has been addressed previously by non-linear coupling schemes [44] and soft-core potentials [46, 47]. Here we seek to remove the particle-particle overlaps responsible for the end-point singularity by the use of a reversible cavity. We calculate the free energy change using standard thermodynamic integration (TI) [48]

$$\Delta G = \int \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda \qquad (10)$$

where λ is the thermodynamic coupling parameter. The cavity-based method employed here avoids this end-point singularity at $\lambda \to 0$ by first growing a cavity large enough to encompass the solute and thus avoid particle-particle overlaps between the solute and solvent (ΔG_{grow}) in the bulk solvent, then inserting the solute into the cavity (ΔG_{insert}) before shrinking and annihilating the cavity (ΔG_{shrink}, see fig 7). $\Delta G_{\text{solvation}}$ is then the sum of the three contributing parts, *i.e.*

$$\Delta G_{\text{solvation}} = \Delta G_{\text{grow}} + \Delta G_{\text{insert}} + \Delta G_{\text{shrink}} \qquad (11)$$

In principle μ_{excess} is independent of the cavity's attributes. However, previous work [25] has shown that the functional form is important for the reversibility of work carried out. A poor choice of cavity potential can result in a hysteresis due to a non-reversible nucleation event. We have found the following cavity potential based on the Weeks-Chandler-Andersen (WCA) potential [49]
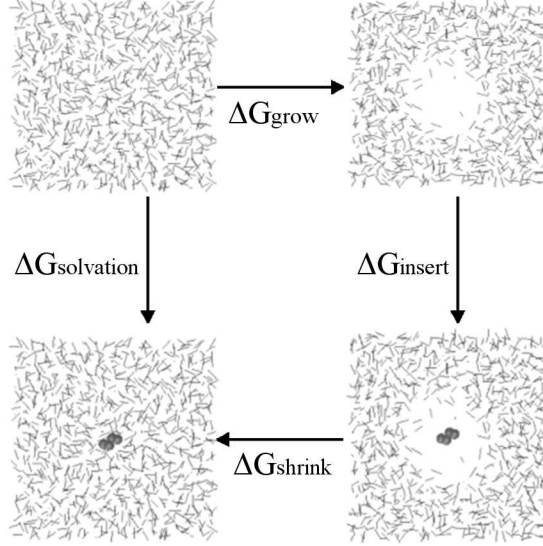
Figure 7: Schematic represeentation of the alchemical thermodynamic cycle used to calculate the excess chemical potential showing $\Delta G_{\text{grow}}, \Delta G_{\text{insert}}$ and ΔG_{shrink}.

suitable for calculating the reversible work in this case;

$$
U^{\text{cavity}}(r, \lambda) = \begin{cases} 4\epsilon_c \lambda \left\{ \left[\alpha \left(1 - \lambda\right) + \left(\frac{r}{\sigma_c}\right)^m \right]^{-2} - \left[\alpha \left(1 - \lambda\right) + \left(\frac{r}{\sigma_c}\right)^m \right]^{-1} \right\} + \epsilon_c \lambda, \\ \qquad\qquad\qquad\qquad \text{if } r < \left[2 - \alpha \left(1 - \lambda\right)\right]^{1/m} \sigma_c \\ 0, \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}
$$

$$(12)$$

where σ_c is the cavity radius, r the distance between the centre of the cavity and solvent bead and ϵ_c the modified cavity well-depth. α and m are constants taken as 0.5 and 2 respectively. Note that for the second step, inserting the solute, $\lambda = 0$ in equation (12), which reverts to a generalised WCA potential. For the insertion, a simple linear coupling scheme is used,

$$
U(r, \lambda) = \lambda U_1 - (1 - \lambda) U_0 \tag{13}
$$

where U_0 and U_1 are the values of U at $\lambda = 0, 1$ respectively, *i.e.* without

11

and with the solute-solvent interactions at full strength. The calculated μ_{excess} is independent of cavity size as the cavity is both created and annihilated in the thermodynamic cycle (if the cavity is large enough to remove any particle-particle overlaps between the solute and solvent) and we calculate μ_{excess} for at least three σ_c values per condition, shown in Table 2. In all cases, $\Delta G_{\mathrm{insert}}$ is calculated using 10 uniformly spaced points, whilst 56 points were used to calculate ΔG_{grow} and $\Delta G_{\mathrm{shrink}}$ using Gaussian quadrature (GQ) and simple trapezium (Trap) numerical integration methods.

Table 2: Cavity sizes employed in calculating μ_{excess} for pyrene in toluene and heptane.

Temperature /K	pressure/bar	σ_c/nm
298.15†	1	1.0, 1.3, 1.5
366.00	1	1.0, 1.2, 1.3, 1.5
366.00	500	1.0, 1.2, 1.3
366.00	1000	1.0, 1.2, 1.3
422.00	500	1.0, 1.2, 1.3
422.00	1000	1.0, 1.2, 1.3

† from reference [25]

In addition to using TI to calculate ΔG, we have used a free energy perturbation method which explicitly calculates μ_{excess} as given in equation (7). Derivations of the free energy perturbation technique can be found in standard textbooks on the subject [48, 50].

The μ_{excess} simulations were performed using an in-house Monte Carlo (MC) program at the conditions given in Table 2. In the MC simulations, particles were treated as rigid bodies with trial moves consisting of a random centre of mass translation, random rotation around the centre of mass or a combination of the two. Volume varying moves were considered on average once per MC cycle (where a cycle is defined as N trial attempts plus on average one volume change trial). Ensemble averages were taken over at least 3.2×10^5 configurations. The errors were estimated using block averages [48, 50].

2.3. Coexistence studies

Direct coexistence simulations are conceptually simple to understand. A system is set up in a way such that the two phases of interest, *i.e.* solid and solution, are at equilibrium. In practice however, this often requires long simulation times and large simulation boxes to gain the statistical accuracy required [51]. Here molecular dynamics (MD) simulations are run at temperatures ranging from 290 K to 420 K and pressures of 1 bar up to 1000 bar for pyrene in toluene and heptane. MD simulations for the direct coexistence studies were run using GROMACS 4.5.5 and 4.6.5 [52] simulation packages with Nose-Hoover thermostat [53] and Parrinello-Rahman barostat [54]. Initial simulation configurations are set up so that there are two unmixed phases in the simulation box. One of the phases is rich in pyrene, and the other rich in solvent. Bonds are constrained using the SHAKE algorithm. [55] Each simulation box contains more than 800 pyrene molecules (3200 particles). Solvents are included to ensure an overall pyrene mole fraction of 30%. This is considerably higher than the solubility limit of pyrene in all the solvents mentioned at the temperatures investigated at 1 bar and allows the spontaneous formation of a solid phase and a liquid phase which are in coexistence with each other. The width of the box is around $10\sigma_{Pyrene}$, where σ_{Pyrene} is the pyrene bead diameter, and the box length is more than $80\sigma_{Pyrene}$. Each simulation is run for 3μs, with a time-step of 0.01ps. The molar fraction of pyrene in the solvent-rich phase is calculated from the density profiles obtained from the last 2.4μs in the liquid phase and compared the experiments. The concentration of pyrene in the solvent-rich phase is calculated directly from the fluid phase;

$$x_{\text{solute}} = \frac{N_{\text{pyrene}}}{N_{\text{pyrene}} + N_{\text{solvent}}} \tag{14}$$

The direct procedure provides the absolute solubility. However, the results are strongly dependent on the capability of the force field to accurately describe the morphology and energetics of the solid phase. SAFT force fields are not parametrized for this purpose hence are expected to be only qualitatively

13

correct. Notwithstanding, the focus of the calculations here are the ratio of solubilities (or partition coefficients), which we presume to be independent of the coexisting solid phase.

3. Results

3.1. Excess chemical potential calculations

We calculated the partition coefficient utilizing the fact that, at equilibrium, the chemical potential of the solute in the two phases is equal (eqn 4). The excess chemical potential was calculated for several cavity sizes (see Table 2 for each condition). An example of the results is presented in fig 8, where we show that μ_{excess} was independent of the cavity radius in all cases, as the cavity is both created and destroyed in the thermodynamic cycle. Analogous plots for the other conditions can be found in the Supporting Information (SI). The results for pyrene in toluene and heptane are given in columns 4 to 6 of Table 3.

Table 3: Results for μ_{excess} (kJ mol^{-1}) for pyrene in toluene and heptane, and $\log P_{TOL/HEP}$ from free energy calculations (columns 3 to 5). The molar compositions of pyrene in toluene and heptane, and $\log P_{TOL/HEP}$ from direct coexistence simulations are given in columns 6 to 8.

T	p	Cavity method			Direct coexistence		
		$\mu_{\text{excess}}^{HEP}$	$\mu_{\text{excess}}^{TOL}$	$\log P_{TOL/HEP}$	$x^{HEP}10^{-2}$	$x^{TOL}10^{-2}$	$\log P_{TOL/HEP}$
298	1	-45.01	-48.51	0.61	1.73	6.91	0.61 [25]
366	1	-36.67	-40.77	0.59	3.67	17.44	0.68
366	500	-28.97	-33.25	0.61	2.81	12.04	0.63
422	500	-24.84	-28.28	0.43	10.11	soluble*	-
366	1000	-22.25	-25.45	0.46	2.72	8.92	0.52
422	1000	-17.37	-20.31	0.36	7.69	23.3	0.48

In both solvents μ_{excess} becomes less negative with increasing temperature. The relationship between μ_{excess} and T can be expressed as

$$\left(\frac{\partial \Delta \mu_{\text{excess}}}{\partial T}\right)_p = -\Delta s_{\text{solute}} \tag{15}$$
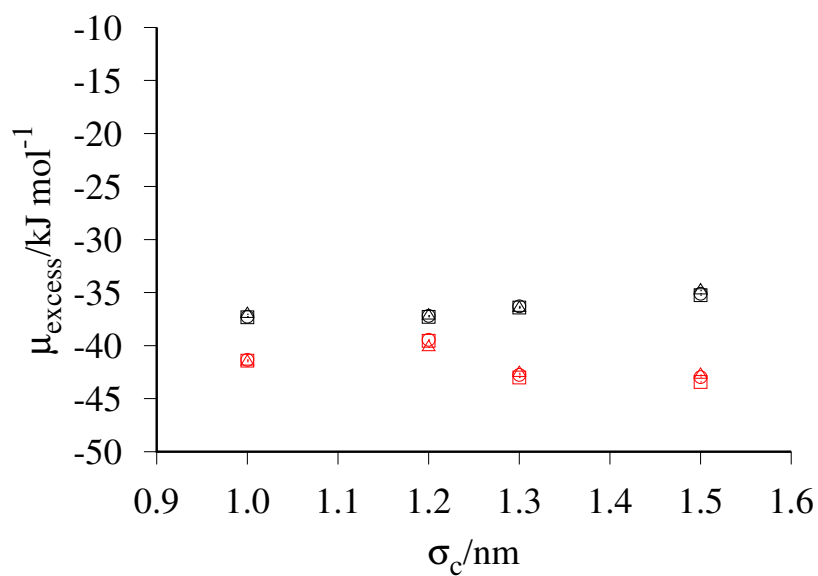
14

Figure 8: Calculated μ_{excess} for each cavity radius (σ_c) $T = 366$ K, $p = 1$ bar. The red symbols are for toluene whilst the black symbols are for heptane. Results obtained by TI(GQ), TI(Trapezium) and free energy perturbation methods are represented by circles, squares and triangles respectively. Analogous plots for the other conditions can be found in the SI.

where Δs_solute is the partial molar entropy of pyrene.

Table 4: Δs_solute (in J K^{-1} mol^{-1}) for pyrene in heptane and toluene.

	Δs_solute	
p/bar	Heptane	Toluene
1	-121.4	-117.5
500	-73.7	-88.7
1000	-87.1	-89.2

Table 4 shows that in both solvents Δs_solute is negative. The magnitude of Δs_solute decreases with increasing pressure due to the reduction in accessible low energy configurations. We were unable to find any values in the literature for Δs_solute of pyrene in any solvent. However, it is comparable to that found by Li *et al* [21] for naphthalene (2 fused rings) in water using an atomistic model, $\Delta s_\text{solute} = -117.15$J K^{-1}mol^{-1}.

Analogously, the relationship between μ_excess and p gives the partial molar volume, v_solute, by

$$\left(\frac{\partial \Delta \mu_\text{excess}}{\partial p}\right)_T = v_\text{solute} \tag{16}$$

and given in Table 5.

Table 5: v_solute (in cm^3 mol^{-1}) for pyrene in heptane and toluene.

	v_molar	
T/K	Heptane	Toluene
366	144.4	154.8
422	149.5	159.6

Whilst there are no literature values available for the partial molar volume of pyrene in heptane or toluene, v_solute for pyrene in CCl$_4$ has been experimentally determined [56] with $v_\text{solute} = 166.5$ cm^3/mol at ambient conditions. Note that the solubility of pyrene in CCl$_4$ ($x^\text{sat} = 0.04229$) is between that of toluene

16

($x^{\text{sat}} = 0.06785$) and heptane ($x_A^{\text{sat}} = 0.01101$) [57]. Whilst the calculated v_{solute} here is slightly smaller, it is of the expected order of magnitude. The deviation is likely to be due to the coarse-grained nature of the model employed. The trends shown in v_{solute} are as expected, with larger v_{solute} for toluene, which is a better solvent for pyrene. In both cases the estimated v_{solute} increases with increasing temperature, consistent with experimental results that find that for many liquid organic molecules expand by about 1% for every 5-10K increase in temperature [58]. The expansion observed in v_{solute} 3.5% and 3.1% for heptane and toluene respectively over a temperature range of 56K, is again consistent with experimental findings.

With increasing pressure, $\Delta G_{\text{solvation}}$ becomes less favourable as the pV term included in the reversible work becomes more significant. However, the effect of pressure on $\log P_{TOL/HEP}$ is less clear and no firm conclusions can be drawn.

The integrand, $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$, calculated in the TI method is shown in fig 9 for $T = 366$ K at 1 bar and 1000 bar. Here one can see that the pressure affects the location of the maximum (minimum) for calculating ΔG_{grow} (ΔG_{shrink}). This indicates that the largest contribution to the reversible work occurs at a lower λ value, $i.e.$ a smaller cavity at higher pressure. Again this is due to the growing importance of the pV term in the free energy calculation. The temperature has much less influence on the shape of $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$. Similar plots for toluene as a solvent and other conditions investigated can be found in the SI for $\sigma = 1.0$nm.

3.2. Coexistence studies

The solubility results from direct coexistence simulations are shown in fig 10 for pyrene in heptane and toluene at 1 bar, along with experimental results. The experimental values of solubilities of pyrene in n-heptane and toluene are taken from Hansen et al [8]. Pyrene is clearly more soluble in toluene than in heptane by at least a factor of 4. For the case of toluene in particular, solubility is slightly underestimated by the direct coexistence simulations, while for heptane, solubilities are overestimated. The temperature dependence of the solubility is captured in both cases. An improvement could be obtained
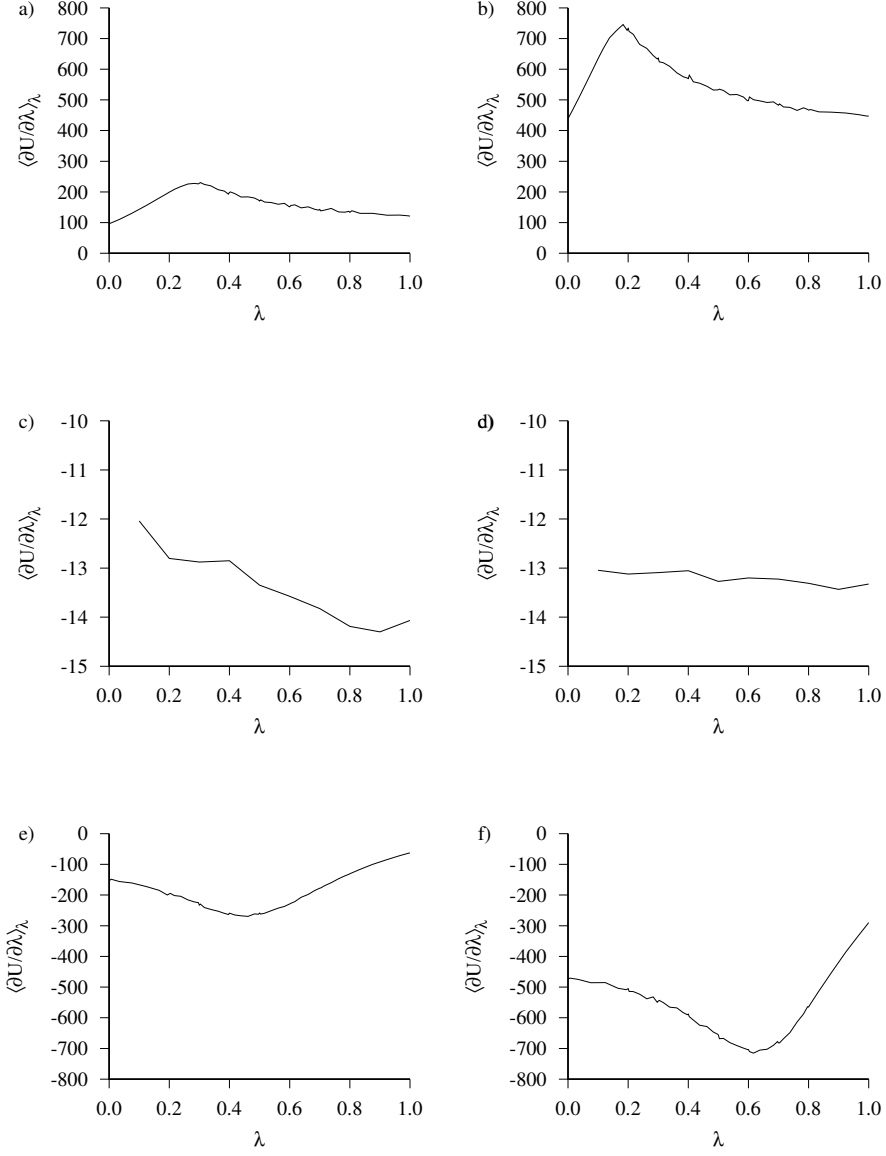
17

Figure 9: $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ used in TI for ΔG_{grow} (top row), ΔG_{insert} (middle row) and ΔG_{shrink} (bottom) for $T = 366$ K, $p = 1$ bar (left) and $T = 366$ K, $p = 1000$ bar (right). In both cases $\sigma_c = 1.0$ nm.
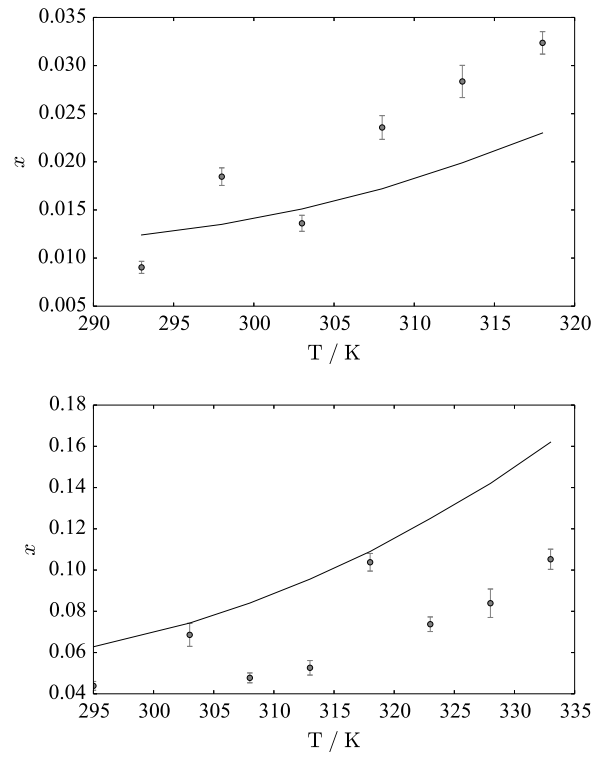
18

Figure 10: Solubility of pyrene in different solvents, symbols are direct coexistence simulation results, lines are the smoothed experimental results [59]: **Top** n-heptane as solvent, **Bottom** toluene as solvent.

by incorporating a binary interaction coefficient to moderate the solute-solvent interaction. However, it is not the objective of the work to produce refined solubility results, but rather to benchmark simulation methods using identical (and relevant) force fields.

We investigate if the pyrene-rich phase is a solid phase by calculating the radial distribution function of the pyrene-rich region, as shown in fig 11. The figure clearly shows that there is short and long range ordering in the pyrene-rich phase, indicating solid formation. However, as the snapshot in fig 11 shows, there is no molecular order in the solid phase, indicating an amorphous glassy state.

Experimentally pyrene is found to form two crystalline polymorphs, both with a sandwich-herringbone structure at ambient pressure [60, 61] and an additional polymorph at higher pressure with a lower interstitial free volume [62]. The lack of spontaneous formation of a crystalline solid phase could be explained by the very slow kinetics expected for solid formation, nevertheless we presume it is an inherent inability of the coarse-grained model. SAFT coarse-grained models are designed to produce, on average, correct volumetric properties of fluids and are not expected to be accurate to represent crystalline phases. The absolute solubility depends on the solid structure, when calculating the partition coefficient as a ratio of solubilities one can make the assumption that the difference in solid structure does not affect the relative solubility ratios.

Further results for higher pressure systems are given in Table 3. It can be seen that at all pressures investigated the solubility increases with increasing temperature. For pyrene in toluene at 422 K, 500 bar no solid phase was observed. We do not currently have an explanation for this, however, it is most likely a result of the finite size of the simulation box along with the finite time explored by the simulation.
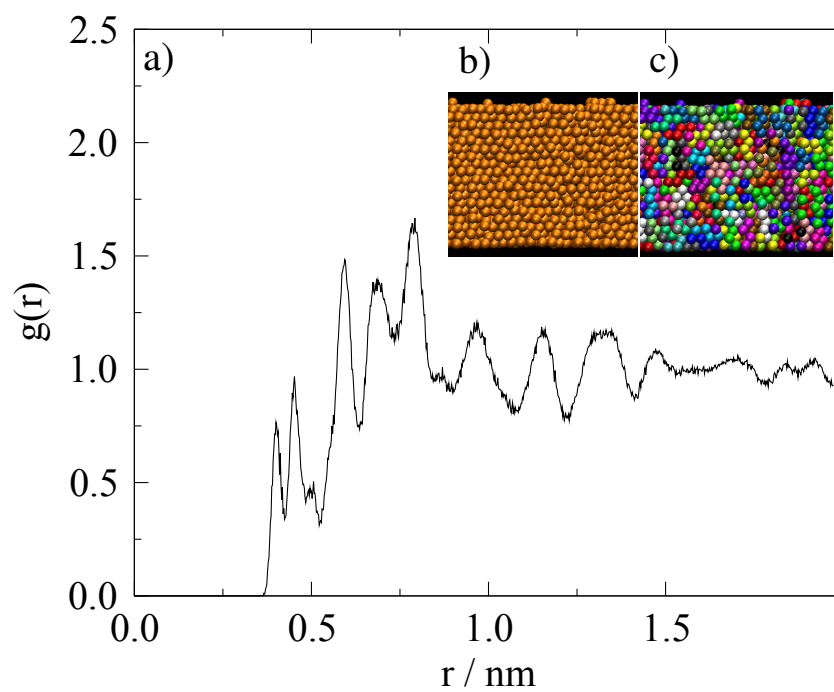
20

Figure 11: **a)** CoM-CoM radial distribution function of CG pyrene in the "pyrene-rich phase" in equilibrium with toluene at 323 K. **b and c)** Two representation of molecules from the same snapshot for the solid phase of pyrene, **b** the ordering of the solid phase with all pyrene molecules with the same colour, **c** same as the left figure with each molecule coloured differently to the neighbouring molecules, indicating an amorphous solid structure.

4. Comparison of methods

We have used computer simulations to investigate the solubility of pyrene in toluene and heptane. We employ two complementary methodologies, direct coexistence simulations and excess chemical potential calculations and compare the partition coefficient, $\log P_{TOL/HEP}$. As can be observed in columns 5 and 8 of Table 3, good agreement was found at 1 bar, however at higher pressures the results diverge, indeed direct coexistence simulations found no phase separation for pyrene in toluene at $T = 422$ K, $p = 500$ bar. The disagreement at high pressure may be down to two reasons. Firstly, the direct coexistence simulations may not be fully equilibrated as they depend on the nucleation and stability of a nascent solid phase, which may require simulation times and sizes several orders of magnitude larger than those employed here and possible with current available hardware.

On the other hand, for systems at high pressure we found evidence of ordering around the cavity for heptane, as shown in fig 12 (top row) which clearly shows the order present in the radial density for heptane at 1000 bar (a) but not at 1 bar (b). Whilst this does not affect the reversibility of the work done to grow and shrink the cavity and the technique is still valid, it did increase the required simulation time required. The ordering was not found for toluene (fig 12 bottom row).

Secondly, the mismatch in the direct coexistence results could be due to the parametrization of the force field. As previously stated, the force field was parametrized for the liquid state and does not accurately represent the solid phase, as evidenced by the amorphous solid pyrene structure found in the simulation in comparison to the expected FCC crystalline form. The excess chemical potential calculations circumvent this problem when calculating the partition coefficient by not considering the solid phase. Of course, if one was interested in the absolute solubility, rather than the relative solubility given by $\log P_{TOL/HEP}$, one would need to equate the chemical potential in the solution to that in the solid phase. This requires knowledge of both the solid structure
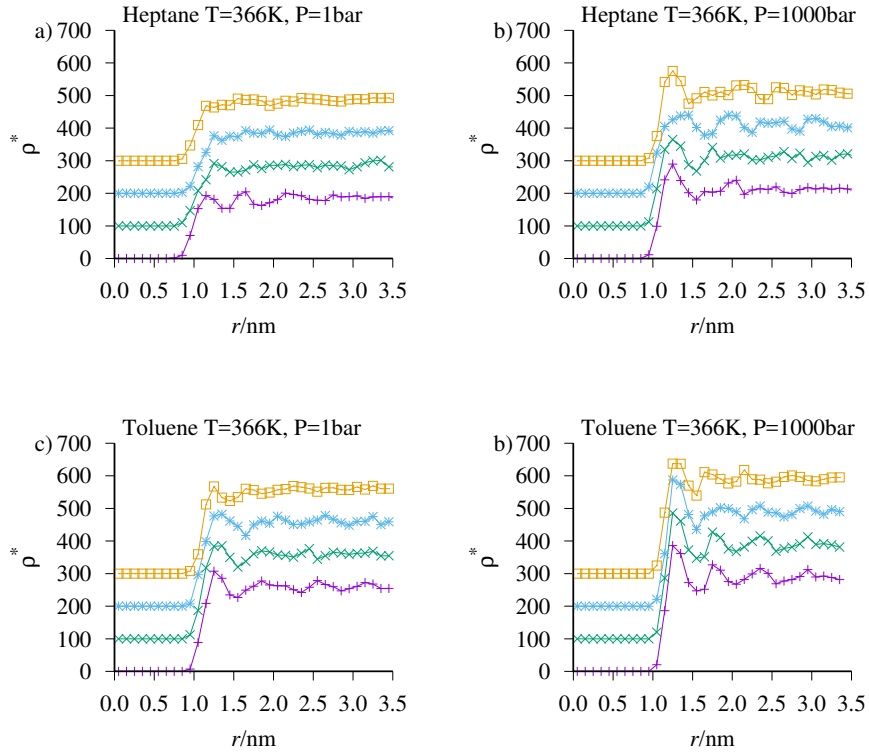
Figure 12: Radial solvent density plots for a) Heptane, $T = 366$ K, $p = 1000$ bar; b) Heptane, $T = 366$ K, $p = 1$ bar; c) Toluene, $T = 366$ K, $p = 1000$ bar; and d) Toluene, $T = 366$ K, $p = 1$ bar. In all cases $\sigma_c = 1.0$ nm. The densities have been averaged over 5000 configurations. The different colour plots show different averages throughout the simulations. They have been shifted upwards in the y-direction for ease of visibility.

and an accurate force field for both the solid and solution phases, with the associated difficulties found with the direct coexistence simulations.

5. Conclusion

In this paper, we have compared two methods to calculate the partition coefficient at a wide range of conditions. The partition coefficient is of importance in both the pharmaceutical industry, where the water-octanol partition coefficient is used to provide information on the blood-cell barrier, and also the oil and gas industry, where oil fractions may be defined by their relative solubilities. We consider the partition coefficient of pyrene between toluene and heptane as an example of a large solute in a dense liquid phase. The two methods employed, that is, direct coexistence simulations and excess chemical potential calculations show good agreement at 1 bar. Here the direct coexistence simulations can provide insight into absolute solubilities, whilst in order to calculate absolute solubilities using thermodynamic relations would require knowledge of the solid phase. However, at high pressure the agreement between two methods diverges with the direct coexistence simulations suffering from long equilibration times. It is at these high-pressure conditions that the cavity-based excess chemical potential methodology becomes superior to both the direct coexistence methodology and other methods for calculating the excess chemical potential. By utilising the creation and annihilation of a cavity in the thermodynamic cycle, the cavity-based method employed here avoids the low insertion probability issue that the well-known trial insertion method [42] can suffer from in dense systems, particularly with large solutes. We find that the cavity-based method for calculating the excess chemical potential is able to accurately and robustly calculate the partition coefficient at all pressures. The cavity-based method can be easily implemented in both bespoke and existing software. An attractive feature of this method is that the initial step, *i.e.* growing the cavity in the solvent, is independent of the solute identity, thus only needs to be calculated once (per condition). The second step, *i.e.* inserting the solute, can also be

performed over relatively few λ intervals These two factors can that the cavity-based method can lend itself to high-throughput simulations in drug-discovery screening applications. The applicability at high temperature/high pressure in particularly attractive to the oil and gas industries, which operate under these conditions.

6. Acknowledgements

7. References

[1] G. L. Amidon, H. Lennernäs, V. P. Shah, J. R. Crison, A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability, Pharmaceutical Research 12 (3) (1995) 413–420.

[2] M. H. Abraham, J. Le, The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship, Journal of Pharmaceutical Sciences 88 (9) (1999) 868–880.

[3] W. L. Jorgensen, E. M. Duffy, Prediction of drug solubility from Monte Carlo simulations, Bioorganic & Medicinal Chemistry Letters 10 (11) (2000) 1155–1158.

[4] W. L. Jorgensen, E. M. Duffy, Prediction of drug solubility from structure, Advanced Drug Delivery Reviews 54 (3) (2002) 355–366.

[5] O. C. Mullins, A. E. Pomerantz, J. Y. Zuo, C. Dong, Downhole fluid analysis and asphaltene science for petroleum reservoir evaluation, Annual Review of Chemical and Biomolecular Engineering 5 (2014) 325–345.

[6] M. Esmaili, S. M. Ghaffari, Z. Moosavi-Movahedi, M. S. Atri, A. Shari-
fizadeh, M. Farhadi, R. Yousefi, J.-M. Chobert, T. Haertlé, A. A. Moosavi-
Movahedi, Beta casein-micelle as a nano vehicle for solubility enhancement
of curcumin; food industry application, LWT-Food Science and Technology
44 (10) (2011) 2166–2172.

[7] J. Gajdoš, K. Galić, Ž. Kurtanjek, N. Ciković, Gas permeability and DSC
characteristics of polymers used in food packaging, Polymer Testing 20 (1)
(2000) 49–57.

[8] H. K. Hansen, C. Riverol, W. E. Acree, Solubilities of anthracene, fluo-
ranthene and pyrene in organic solvents: comparison of calculated values
using UNIFAC and modified UNIFAC (Dortmund) models with experimen-
tal data and values using the mobile order theory, The Canadian Journal
of Chemical Engineering 78 (6) (2000) 1168–1174.

[9] A. Jouyban, A. Shayanfar, W. E. Acree, Solubility prediction of polycyclic
aromatic hydrocarbons in non-aqueous solvent mixtures, Fluid Phase Equi-
libria 293 (1) (2010) 47–58.

[10] A. Shayanfar, S. H. Eghrary, F. Sardari, W. E. Acree Jr, A. Jouyban, Solu-
bility of anthracene and phenanthrene in ethanol + 2, 2, 4-trimethylpentane
mixtures at different temperatures, Journal of Chemical & Engineering
Data 56 (5) (2011) 2290–2294.

[11] N. Habibi, S. Z. M. Hashim, A. Norouzi, M. R. Samian, A review of machine
learning methods to predict the solubility of overexpressed recombinant
proteins in Escherichia coli, BMC Bioinformatics 15 (1) (2014) 134.

[12] M. Lísal, W. R. Smith, J. Kolafa, Molecular simulations of aqueous elec-
trolyte solubility: 1. The expanded-ensemble osmotic molecular dynam-
ics method for the solution phase, The Journal of Physical Chemistry B
109 (26) (2005) 12956–12965.

[13] F. Moucka, M. Lísal, J. Skvor, J. Jirsák, I. Nezbeda, W. R. Smith, Molecular simulation of aqueous electrolyte solubility. 2. Osmotic ensemble Monte Carlo methodology for free energy and solubility calculations and application to NaCl, The Journal of Physical Chemistry B 115 (24) (2011) 7849–7861.

[14] I. Nezbeda, F. Moučka, W. R. Smith, Recent progress in molecular simulation of aqueous electrolytes: force fields, chemical potentials and solubility, Molecular Physics 114 (11) (2016) 1665–1690.

[15] F. Moučka, J. Kolafa, M. Lísal, W. R. Smith, Chemical potentials of alkaline earth metal halide aqueous electrolytes and solubility of their hydrates by molecular simulation: Application to $CaCl_2$, antarcticite, and sinjarite, The Journal of Chemical Physics 148 (22) (2018) 222832.

[16] J. R. Espinosa, J. M. Young, H. Jiang, D. Gupta, C. Vega, E. Sanz, P. G. Debenedetti, A. Z. Panagiotopoulos, On the calculation of solubilities via direct coexistence simulations: Investigation of NaCl aqueous solutions and Lennard-Jones binary mixtures, The Journal of Chemical Physics 145 (15) (2016) 154111.

[17] H. M. Manzanilla-Granados, H. Saint-Martín, R. Fuentes-Azcatl, J. Alejandre, Direct coexistence methods to determine the solubility of salts in water from numerical simulations. test case NaCl, The Journal of Physical Chemistry B 119 (26) (2015) 8389–8396.

[18] A. L. Benavides, J. L. Aragones, C. Vega, Consensus on the solubility of NaCl in water from computer simulations using the chemical potential route, The Journal of Chemical Physics 144 (12) (2016) 124504.

[19] Z. Mester, A. Z. Panagiotopoulos, Temperature-dependent solubilities and mean ionic activity coefficients of alkali halides in water from molecular dynamics simulations, The Journal of Chemical Physics 143 (4) (2015) 044505.

27

[20] M. Ferrario, G. Ciccotti, E. Spohr, T. Cartailler, P. Turq, Solubility of KF in water by molecular dynamics using the Kirkwood integration method, The Journal of Chemical Physics 117 (10) (2002) 4947–4953.

[21] L. Li, T. S. Totton, D. Frenkel, Computational methodology for solubility prediction: Application to the sparingly soluble solutes, The Journal of Chemical Physics 146 (21) (2017) 214110.

[22] A. Ghoufi, G. Maurin, Hybrid Monte Carlo simulations combined with a phase mixture model to predict the structural transitions of a porous metal- organic framework material upon adsorption of guest molecules, The Journal of Physical Chemistry C 114 (14) (2010) 6496–6502.

[23] F.-X. Coudert, The osmotic framework adsorbed solution theory: predicting mixture coadsorption in flexible nanoporous materials, Physical Chemistry Chemical Physics 12 (36) (2010) 10904–10913.

[24] L. J. Dunne, G. Manos, Statistical mechanics of binary mixture adsorption in metal-organic frameworks in the osmotic ensemble., Philosophical Transactions A: Mathematical, Physical and Engineering Sciences.

[25] C. Wand, T. Totton, D. Frenkel, Addressing hysteresis and slow equilibration issues in cavity-based calculation of chemical potentials, The Journal of Chemical Physics 149 (1) (2018) 014105.

[26] S. Senkan, M. Castaldi, Combustion in Ullmanns encyclopedia of industrial chemistry, Wiley-VCH, Weinheim. Polycyclic aromatic hydrocarbons in urine from two psoriatic patients, Acta Dermato Venereologica 73 (2003) 188–190.

[27] A. Hirschberg, L. N. J. DeJong, B. A. Schipper, J. G. Meijer, et al., Influence of temperature and pressure on asphaltene flocculation, Society of Petroleum Engineers Journal 24 (03) (1984) 283–293.

28

[28] E. Rogel, Studies on asphaltene aggregation via computational chemistry, Colloids and Surfaces A: Physicochemical and Engineering Aspects 104 (1) (1995) 85–93.

[29] O. C. Mullins, The asphaltenes, Annual Review of Analytical Chemistry 4 (2011) 393–418.

[30] Q. Wu, A. E. Pomerantz, O. C. Mullins, R. N. Zare, Laser-based mass spectrometric determination of aggregation numbers for petroleum-and coal-derived asphaltenes, Energy & Fuels 28 (1) (2013) 475–482.

[31] T. F. Headen, E. S. Boek, G. Jackson, T. S. Totton, E. A. Müller, Simulation of asphaltene aggregation through molecular dynamics: Insights and limitations, Energy & Fuels 31 (2) (2017) 1108–1125.

[32] J. Ancheyta, F. Trejo, M. S. Rana, Asphaltenes: chemical transformation during hydroprocessing of heavy oils, CRC press, 2010.

[33] E. Baka, J. E. A. Comer, K. Takács-Novák, Study of equilibrium solubility measurement by saturation shake-flask method using hydrochlorothiazide as model compound, Journal of Pharmaceutical and Biomedical Analysis 46 (2) (2008) 335–341.

[34] J. Alsenz, M. Kansy, High throughput solubility measurement in drug discovery and development, Advanced Drug Delivery Reviews 59 (7) (2007) 546–567.

[35] E. A. Müller, G. Jackson, Force-field parameters from the SAFT-γ equation of state for use in coarse-grained molecular simulations, Annual Review of Chemical and Biomolecular Engineering 5 (2014) 405–427.

[36] C. Herdes, T. S. Totton, E. A. Müller, Coarse grained force field for the molecular simulation of natural gases and condensates, Fluid Phase Equilibria 406 (2015) 91–100.

29

[37] E. A. Müller, A. Mejía, Extension of the SAFT-VR Mie EoS to model homonuclear rings and its parametrization based on the principle of corresponding states, Langmuir 33 (42) (2017) 11518–11529.

[38] T. Lafitte, A. Apostolakou, C. Avendano, A. Galindo, C. S. Adjiman, E. A. Müller, G. Jackson, Accurate statistical associating fluid theory for chain molecules formed from Mie segments, Journal of Chemical Physics 139 (15).

[39] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, Journal of the American Chemical Society 118 (45) (1996) 11225–11236.

[40] P. J. Linstrom, W. G. Mallard, The NIST Chemistry WebBook: A chemical data resource on the internet, Journal of Chemical & Engineering Data 46 (5) (2001) 1059–1063.

[41] S. b. A. Design Institute for Physical Properties, Design Institute for Physical Property Research/AIChE, 2005; 2008; 2009; 2010; 2011; 2012; 2015; 2016; 2017. [link].
URL https://app.knovel.com/hotlink/toc/id:kpDIPPRPF7/dippr-project-801-full/dippr-project-801-full

[42] B. Widom, Some topics in the theory of fluids, The Journal of Chemical Physics 39 (11) (1963) 2808–2812.

[43] A. Ben-Naim, Standard thermodynamics of transfer. uses and misuses, The Journal of Physical Chemistry 82 (7) (1978) 792–803.

[44] T. Steinbrecher, D. L. Mobley, D. A. Case, Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations, The Journal of Chemical Physics 127 (21) (2007) 214108.

[45] T. Simonson, Free energy of particle insertion: an exact analysis of the origin singularity for simple liquids, Molecular Physics 80 (2) (1993) 441–447.

[46] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, W. F. van Gunsteren, Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations, Chemical Physics Letters 222 (6) (1994) 529–539.

[47] M. Zacharias, T. P. Straatsma, J. A. McCammon, Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration, The Journal of Chemical Physics 100 (12) (1994) 9025–9031.

[48] D. Frenkel, B. Smit, Understanding molecular simulation: from algorithms to applications, 2nd Edition, Academic Press, 2001.

[49] J. D. Weeks, D. Chandler, H. C. Andersen, Role of repulsive forces in determining the equilibrium structure of simple liquids, The Journal of chemical physics 54 (12) (1971) 5237–5247.

[50] M. P. Allen, D. J. Tildesley, Computer simulation of liquids, 2nd Edition, Oxford University Press, 2017.

[51] Y. Zhang, E. J. Maginn, A comparison of methods for melting point calculation using molecular dynamics simulations, The Journal of chemical physics 136 (14) (2012) 144116.

[52] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, Journal of Chemical Theory and Computation 4 (3) (2008) 435–447.

[53] W. G. Hoover, Canonical dynamics: equilibrium phase-space distributions, Physical Review A 31 (3) (1985) 1695.

[54] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, Journal of Applied Physics 52 (12) (1981) 7182–7190.

[55] J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dy-

namics of n-alkanes, Journal of Computational Physics 23 (3) (1977) 327–341.

[56] F. Shahidi, P. G. Farrell, J. T. Edward, P. Canonne, Partial molal volumes of organic compounds in carbon tetrachloride. 3. Aromatic hydrocarbons: steric effects, The Journal of Organic Chemistry 44 (6) (1979) 950–953.

[57] J. R. Powell, D. Voisinet, A. Salazar, W. E. Acree Jr, Solubility of pyrene in organic nonelectrolyte solvents. comparison of observed versus predicted values based upon mobile order theory, Physics and Chemistry of Liquids 28 (4) (1994) 269–276.

[58] P. Hynčica, L. Hnědkovskỳ, I. Cibulka, Partial molar volumes of organic solutes in water. X. Benzene and toluene at temperatures from (298 to 573) K and at pressures up to 30 MPa, The Journal of Chemical Thermodynamics 35 (12) (2003) 1905–1915.

[59] W. E. Acree Jr, IUPAC-NIST solubility data series. 98. Solubility of polycyclic aromatic hydrocarbons in pure and organic solvent mixturesrevised and updated. Part 3. Neat organic solvents, Journal of Physical and Chemical Reference Data 42 (1) (2013) 013105.

[60] Y. Kai, F. Hama, N. Yasuoka, N. Kasai, Structural chemistry of layered cyclophanes. III. Molecular structures of [2.2](2, 7) pyrenophane-1, 1'-diene and pyrene (redetermined) at–160 c, Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry 34 (4) (1978) 1263–1270.

[61] C. Frampton, K. Knight, N. Shankland, K. Shankland, Single-crystal X-ray diffraction analysis of pyrene II at 93k, Journal of Molecular Structure 520 (1) (2000) 29–32.

[62] F. P. A. Fabbiani, D. R. Allan, S. Parsons, C. R. Pulham, Exploration of the high-pressure behaviour of polycyclic aromatic hydrocarbons: naphthalene, phenanthrene and pyrene, Acta Crystallographica Section B: Structural Science 62 (5) (2006) 826–842.