Addressing hysteresis and slow equilibration issues in cavity-based calculation of chemical potentials

C. R. Wand,[1] T. S. Totton,[2] and D. Frenkel[1]

[1)] *Department of Chemistry, University of Cambridge, Cambridgeshire, CB2 1EW, United Kingdom* [a)]

[2)] *BP Exploration Operating Co. Ltd., Sunbury-on-Thames, TW16 7LN, United Kingdom*

(Dated: 5 June 2018)

In this paper we explore the strengths and weaknesses of a cavity-based method to calculate the excess chemical potential of a large molecular solute in a dense liquid solvent. Use of the cavity alleviates some technical problems associated with appearance of (integrable) divergences in the integrand during alchemical particle growth. The excess chemical potential calculated using the cavity-based method should be independent of the cavity attributes. However, the performance of the method (equilibration time and the robustness) does depend on the cavity attributes. To illustrate the importance of a suitable choice of the cavity attributes, we calculate the partition coefficient of pyrene in toluene and heptane using a coarse-grained model. We find that a poor choice for the functional form of the cavity may lead to hysteresis between growth and shrinkage of the cavity. Somewhat unexpectedly, we find that, by allowing the cavity to move as a pseudo-particle within the simulation box, the decay time of fluctuations in the integrand of the thermodynamic integration can be reduced by an order of magnitude, thereby increasing the statistical accuracy of the calculation.

PACS numbers: 61.20.Ja, 64.75.Bc, 05.70.Ce

[a)] Author to whom correspondence should be addressed: df246@cam.ac.uk

I. INTRODUCTION

Predicting the solubility of substances is of great importance across many scientific disciplines. In view of the importance of the topic, it is not surprising that many different methodologies to compute solubilities have been reported in the literature[1–9]. However, using molecular simulations to predict solubility, in particular for sparingly soluble substances, remains non-trivial. Examples of the importance of accurate solubility calculations include drug solubility in the pharmaceutical industry[10–13], and the solubility of sugars, food additives and unwanted contaminants in the food industry[14,15]. Not only aqueous solubility is of industrial importance: in the oil and gas industry crude oil fractions may be defined by their relative solubilities in hydrocarbon solvents. For example an asphaltene fraction is defined as soluble in toluene and insoluble in n-heptane[16]. Asphaltene phases typically consist of a mixture of different molecular species with a mean molecular weight of around 700u[17]. Asphaltenes are prone to aggregation that is driven by changes in condition, such as pressure, temperature or compositional changes by blending/incorporating with other elements such as other incompatible oils or gases.[16–21].

To explore some of the issues that may arise when computing the solubility of larger molecules, such as asphaltenes, we consider the relative solubility of pyrene (fig 2c) between toluene and heptane. Pyrene, which may be viewed as a mini-asphaltene, is a polycyclic aromatic hydrocarbon (PAH) formed of four fused rings. It is commonly found as a combustion product for a wide range of conditions[22]. There are a number of ways to calculate the absolute solubility *via* simulations, one of the simplest to understand is a 'brute force' simulation of the two coexisting phases, *i.e.* solid and solution or gas and solution. However, these brute force simulations require large simulations and long timescales to reach equilibrium which can be on the order of ns for even highly soluble solutes[6,9,23]. Other theoretical techniques employ standard thermodynamic relations to calculate the solubility, equating the chemical potentials of the solid and solution phases[7,24,25]. Such calculations require the direct calculation of the chemical potentials of the relevant species in the two phases or by employing the osmotic ensemble method/OEMD/OEMC[1,4,8,26]. The osmotic ensemble method calculates the composition of the saturated solution phase by imposing the chemical potential of the corresponding solid phase in a grand canonical approach. The osmotic ensemble has typically been applied to small solutes, both in solution and metal-

organic frameworks[27–29], as the initial and final states need to be similar. Moving away from particle-based methods, analytical models have been developed to investigate the solubility of large solutes in non-aqueous solvents and can accurately predict the solubility of PAHs in a range of solvents[30–32]. However, these analytical models do not provide insight into the molecular-level driving force for the solubility.

The relative solubility can be quantified by the partition coefficient, $P_{I/II}$, or solubility ratio[33], which gives the relative solubility of a substance A, between solvents I and II and is simply defined as

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

and is often reported as a logarithmic value, $\log P_{I/II}$. In principle, $\log P_{I/II}$ for two immiscible solvents (*e.g.* water and octanol) can be obtained from experiment by using a shake-flask. However this approach is less attractive when considering sparingly soluble solutes[34,35]. Moreover, the method cannot be applied to solvents that are miscible, *e.g.* toluene and heptane. Computer simulation based techniques allow us to estimate the partition coefficient of a solute between two miscible liquids. The advantage of considering relative solubilities, as opposed to absolute solubilities, is that the solid phase need not be considered. There are several advantages associated with bypassing the solid phase. Firstly, the structure of the solid phase may be unknown, particularly in the case of amorphous or polymorphic materials. Additionally, the same force field should be employed for both the solid and solution phases. Many pre-existing universal force fields are state specific, primarily developed for the solution[36,37]. Thus the force field may not be an adequate representation of the solute in the solid phase. By calculating the relative solubilities using the partition coefficient, we avoid these complications.

In this paper we show that the partition coefficient can be calculated using the cavity-based method previously employed by Li *et al*[25] with a coarse-grained model of pyrene between toluene and heptane as an example. The methodology employed is simple and can readily be implemented in available open-source software or in-house programs with a variety of force fields. We show that, although in principle the attributes of the softly repulsive cavity should not influence the calculation, in practice the cavity attributes may affect both the accuracy and efficiency of the calculation.

The remainder of the paper is organised as follows. In Section II we discuss the theoretical background of the partition coefficient and the cavity-based method for calculating the excess

chemical potential before discussing the model and simulation details. In Section III we present and discuss our findings.

II. METHODOLOGY

A. Theoretical background

1. Partition coefficient

We calculate the partition coefficient $\left(\log P_{I/II}\right)$ of pyrene between toluene and heptane as an example of the cavity-based method previously employed by Li *et al*[25], which is detailed below. The partition coefficient is defined as the ratio of solubilities of the solute of interest, in this case pyrene, between solvents I and II, equation (1). The derivation of the partition coefficient from the excess chemical potentials is given in Appendix A. However for the sake of clarity we briefly discuss it here.

At equilibrium the (total) chemical potential, $\mu\left(T,p\right)$, at a given temperature, T, and pressure, p, of the solute in both the solvents is equal;

$$\mu^{I}\left(T,p\right)=\mu^{II}\left(T,p\right) \tag{2}$$

where the superscript denotes the solvent of interest and the chemical potential is given by

$$\mu\left(T,p\right)=\mu_{ideal}\left(T,p\right)+k_{B}T\ln\rho-k_{B}T\ln\langle\exp\left[-\beta U_{solute-solvent}\left(\mathbf{R}_{solute}\right)\right]\rangle_{0} \tag{3}$$

where $\mu_{ideal}\left(T,p\right)$ is the ideal gas contribution to the chemical potential, k_B is the Boltzmann constant, ρ is the density (analogous to the concentration), $\beta=1/k_BT$ and $U_{solute-solvent}\left(\mathbf{R}_{solute}\right)$ is the interaction energy of the solute with the rest of the rest of the system. The final term in equation (3) is defined as the excess chemical potential, μ_{excess}. Given equations (2) and (3), it follows that when the two phases are at equilibrium

$$k_{B}T\ln\rho^{I}-\mu_{excess}^{I}=k_{B}T\ln\rho^{II}-\mu_{excess}^{II} \tag{4}$$

where the $\mu_{ideal}\left(T,p\right)$ contributions have cancelled as they are independent of the solvent. Rearranging equation (4) leads to an expression for the partition coefficient,

$$\ln P_{I/II}=\ln\left(\frac{\rho^{I}}{\rho^{II}}\right)=\beta\left(\mu_{excess}^{II}-\mu_{excess}^{I}\right)\ , \tag{5}$$

which can then be converted to $\log P_{I/II}$, the usual format for reporting the partition coefficient.

2. Excess Chemical potential calculation via the cavity method

The excess chemical potential, μ_{excess}, is defined as the change in Gibbs free energy, associated with adding a single solute into the system at constant $T, p, N_{solvent}, N_{solute}$, $\Delta G_{solvation}$, where $N_{solvent}$ and N_{solute} are the number of solvent and solute molecules respectively. In the present case we deal with a system at infinite dilution inserting a single solute molecule into a pure solvent system, *i.e.* $N_{solute} = 0$. However, the methodology is equally applicable to concentrated solutions. $\Delta G_{solvation}$ can be calculated from simulations utilising a variety of methods including test particle insertion[38], and growing the solute of interest directly into a solvent[39,40]. However, the former method fails at high densities and the latter method suffers from an (integrable) "end-point"singularity which can lead to inaccuracies due to particle-particle overlaps. This problem has previously been addressed by a variety of approaches including non-linear coupling schemes[39] and soft-core potentials[41,42]. In the systems that we study, the large size of the solute of interest and the high density of the solvent phase imply that such overlaps are highly likely in any methodology that relies on direct insertion. To eliminate the possibility of overlaps occurring when introducing the solute molecule, we initially grow a soft repulsive cavity within the solvent before inserting the solute into the cavity and finally shrinking the cavity. Thus

$$\Delta G_{solvation} = \Delta G_{grow} + \Delta G_{insert} + \Delta G_{shrink} \tag{6}$$

and is shown schematically in fig 1 .

An additional advantage of the use of a cavity to calculate $\Delta G_{solvation}$ is that the initial calculation for growing the cavity, ΔG_{grow} is independent of solute identity: thus it only has to be calculated once for each set of conditions, *i.e.* *T, p*, solvent and cavity size.

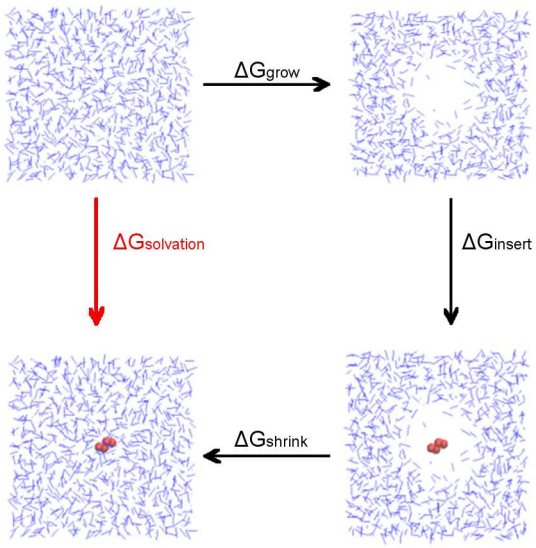

FIG. 1. A schematic showing the alchemical thermodynamic cycle employed to calculate μ_{excess}.

B. Model

We employed a coarse-grained model based on the SAFT-VR Mie potential[43,44] to model both pyrene and the solvents. The Mie potential has the form

$$U^{Mie}(r_{ij}) = C\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{\lambda_r} - \left(\frac{\sigma}{r_{ij}}\right)^{\lambda_a}\right] \tag{7}$$

$$C = \frac{\lambda_r}{\lambda_r - \lambda_a}\left(\frac{\lambda_r}{\lambda_a}\right)^{\frac{\lambda_a}{\lambda_r - \lambda_a}} \tag{8}$$

where r_{ij} is the distance between the centre of beads i and j, σ is the bead diameter, ϵ the potential depth and λ_a and λ_r are the attractive and repulsive exponents respectively. In all cases $\lambda_a = 6.00$ with the remaining model parameters listed in table I. Cross parameters are given by the Kong combining rules[45], that is

$$\sigma_{ij} = \frac{(\sigma_{ii} + \sigma_{jj})}{2} \tag{9}$$

$$\epsilon_{ij} = (1 - k_{ij})\frac{\sqrt{\sigma_{ii}^3 \sigma_{jj}^3}}{\sigma_{ij}^3}\sqrt{\epsilon_{ii}\epsilon_{jj}} \tag{10}$$

$$(\lambda_{ij} - 3) = \sqrt{(\lambda_{ii} - 3)(\lambda_{jj} - 3)} \tag{11}$$

where the subscripts denote the interactions between particles i and/or j. In all cases we take $k_{ij} = 0$.

TABLE I. Model parameters

Bead type	σ [Å]	ϵ/k_B [K]	λ_r
Heptane	4.770	436.13	23.81
Toluene	3.685	268.24	11.80
Pyrene	4.135	436.13	14.79

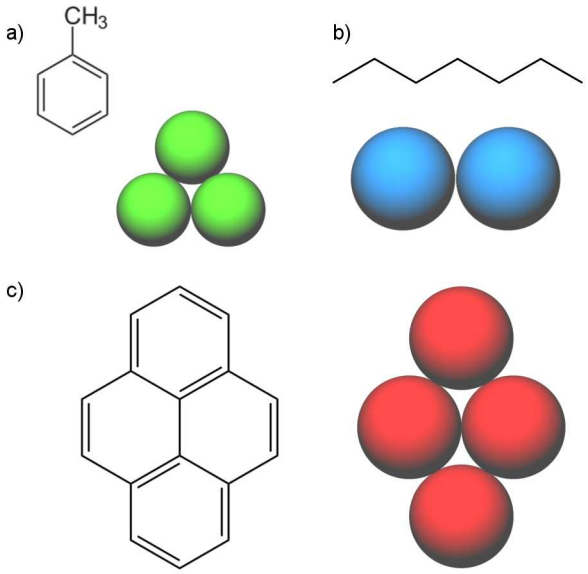

FIG. 2. A schematic showing coarse-grained representation of the three molecules present in the solvent. a) shows heptane represented as a dimer, b) toluene as an equilateral triangle trimer and c) the arrangement of four beads for pyrene.

A heptane molecule is represented by a dimer, whilst toluene is represented by an equilateral triangle and pyrene is comprised of four beads. All molecules have rigid bonds with length $(\sigma_i + \sigma_j)/2$. In molecules comprising three or more beads, the rigid-bond constraint also leads to fixed angles and dihedral angles with no additional constraints (fig 2). Preliminary bulk simulations of $N = 1950$ solvent molecules give densities of 682.92 kg m^{-3} and 869.63 kg m^{-3} for heptane and toluene respectively, both within 1.0% of the experimental values[46] of 679.60 kg m^{-3} and 862.24 kg m^{-3} .

As far as the choice of the cavity potential is concerned: any functional form that produces a soft cavity large enough for the solute to be inserted into without particle overlaps with the solvent molecules may be employed. For example, in the literature atomistic studies by

Li *et al*[25] employ a cavity potential in the form of

$$U_{(1)}^{cavity}(r,\lambda) = A\exp\left(-r/B+\lambda\right) \tag{12}$$

to investigate the solubility of sparingly soluble solutes such as naphthalene, whilst Postma *et al*[47] employ a cavity potential in the form of $U^{cavity}(r,\lambda) = \lambda\left(B/r\right)^{12}$ to study the formation of cavities in liquid water. Here we investigate two functional forms for the cavity potential: the exponential form employed by Li *et al*[25] given by equation (12) and referred to as $U_{(1)}^{cavity}$, and a second cavity potential modelled as a generalised soft Weeks-Chandler-Andersen (WCA) potential[48]

$$U_{(2)}^{cavity}(r) = \begin{cases} 4\epsilon_c\left[\left(\frac{\sigma_c}{r}\right)^{2n} - \left(\frac{\sigma_c}{r}\right)^{n}\right] + \epsilon_c, & \text{if } r < 2^{1/n}\sigma_c \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where σ_c and ϵ_c are the cavity radius and interaction strength respectively. In all cases we have taken a value of $n = 2$, however preliminary simulations have shown that the cavity produced is insensitive to the value of n used with $n = 6$ giving very similar results.

C. Simulation details

All simulations were run in the isobaric-isothermal (N, p, T) ensemble using an in-house, but standard, MC program. Systems were run for at least 9.6×10^5 MC cycles (where we define a MC cycle as N trial moves and, on average, one volume trial move) in the case of a fixed cavity and 3.2×10^5 MC cycles in the case of a moving cavity after an initial equilibration period. Trial MC moves consisted of centre of mass translations and/or a rotation around the centre of mass. Trial volume varying moves were attempted on average once per MC cycle. In the case of the moving cavity/solute simulations, a cavity/solute translation was attempted on average once per cycle. In the case of Stages 2 and 3 (inserting the solute and shrinking the cavity respectively), the solute molecule was also rotated about the centre of mass on cavity moves. An acceptance ratio of between 40% to 60% was maintained. Simulations were performed at 298.15K and 1bar using 1950 solvent molecules in a cubic box with sides l and periodic boundary conditions in three dimensions.

The change in free energy was calculated by standard thermodynamic integration (TI) where

$$\Delta G = \int \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda} d\lambda \tag{14}$$

and the integral computed numerically using either Gaussian Quadrature with six points, TI(GQ), or trapezium rule fitting, TI(Trap). For ΔG_{grow} and ΔG_{shrink} the cavity potential for $U_{(1)}^{cavity}$ is given explicitly by equation (12), whilst $U_{(2)}^{cavity}$ given in equation (13) is modified to avoid the end-point singularity by the introduction of an alchemical thermodynamic coupling parameter, λ, leading to

$$U_{(2)}^{cavity}(r,\lambda) = \begin{cases} 4\epsilon_c\lambda\left\{ \left[\alpha(1-\lambda) + \left(\frac{r}{\sigma_c}\right)^n\right]^{-2} - \left[\alpha(1-\lambda) + \left(\frac{r}{\sigma_c}\right)^n\right]^{-1} \right\} + \epsilon_c\lambda, \\ \qquad\qquad \text{if } r < \left[2-\alpha(1-\lambda)\right]^{1/n}\sigma_c \\ 0, \qquad\qquad \text{otherwise} \end{cases} \tag{15}$$

where α is a constant that we take to be 0.5. This coupling λ is analogous to the soft-core potentials utilised for inserting Lennard-Jones centres[41,42]. It acts to smooth the end-point singularity, gradually increasing the interaction potential from $U^{cavity}(r,\lambda) = 0$ at $\lambda = 0$ to the full cavity potential given in equation (13) at $\lambda = 1$. Note that for ΔG_{shrink} the λ and $(1-\lambda)$ terms in equation (15) are exchanged, leading to the full cavity potential at $\lambda = 0$ and 0 at $\lambda = 1$.

ΔG_{insert} was calculated using a linear alchemical potential regardless of the form of cavity potential, namely

$$U^{insert}(\lambda) = \lambda U_1 + (1-\lambda)U_0 \tag{16}$$

where U_0 and U_1 are the solute-solvent interactions before and after insertion respectively.

The change in free energy associated with steps 1 to 3 was also calculated using free energy perturbation (FEP) method with

$$\Delta G(0 \to 1) = -k_B T \ln\left\langle e^{-\beta\Delta U}\right\rangle_0 \tag{17}$$

where $\beta = \frac{1}{k_B T}$ and $\Delta U = U_1 - U_0$. The $\langle \ldots \rangle_0$ indicates the ensemble average at state 0.

The efficiency of the simulations can be estimated using the standard block average based method[49,50] to calculate the correlation time, t_c, that is, the average number of MC cycles required to produce an independent configuration, see Appendix B.

III. RESULTS AND DISCUSSIONS

A. Exponential cavity results, $U_{(1)}^{cavity}$

Simulations were run using $U_{(1)}^{cavity}$ previously utilised by Li *et al*[25], given by equation (12), with $A = 300\text{kJ mol}^{-1}, B = 1.0\text{nm}$. $U_{(1)}^{cavity}$ has the advantage that $\frac{\partial U}{\partial \lambda} = U_{(1)}^{cavity}$, needed for the TI procedure outlined in equation (14) making it trivial to implement within pre-existing software. In these simulations the cavity is fixed at the centre of the simulation cell. The integrand, $\langle \frac{\partial U}{\partial \lambda} \rangle_\lambda$, for growing the cavity in heptane is shown in fig 3a). Note that the same trend in behaviour is seen for both solvents and the corresponding graph for toluene can be found in the supplementary information (SI Figure 1). There is a clear discontinuity in $\langle \frac{\partial U}{\partial \lambda} \rangle_\lambda$ at a threshold λ, λ_t. To investigate the cause of this sudden change of behaviour in $\langle \frac{\partial U}{\partial \lambda} \rangle_\lambda$ we calculated the solvent radial density which allows for the cavity radius to be calculated. This was achieved by splitting the simulation cell into M concentric spherical shells of width $d_{hist} = 0.01\text{nm}$ from the centre of the cavity. The number density, $\rho_{hist} = N_{hist}/V_{hist}$ where N_{hist} is the number of solvent molecules whose centre of mass is found in the shell and V_{hist} is the volume of the shell. ρ_{hist} is calculated and averaged over 5000 MC cycles. The resulting histograms show a steep increase to a uniform density outside the cavity and the cavity radius can be estimated using a threshold ρ_{hist}, shown in fig 3b for $\lambda = -2.40, -2.50$ and -2.60. The cavity radius is not sensitive to the threshold value of ρ_{hist} used as long as it is significantly below the uniform density.

The cavity radius versus r is shown in fig 3b for three λ values, whilst the cavity radius versus λ is shown in 3c. From these plots it is clear that below λ_t there is no cavity is present in the system with the intermolecular attractions overcoming the repulsive cavity potential. At λ_t, the repulsive cavity-solvent interactions overcomes these intermolecular interactions and a cavity irreversibly nucleates, as evidenced by a sudden jump in the cavity radius from approximately 0.0nm to 1.2nm at λ_t before continuing to grow much more slowly, see figure 3a. Values of λ_t are given in table II. The coarse-grained nature of the model employed in these simulations is much softer than that of the atomistic SPC water[51] utilised by Li *et al*[25] previously. It is likely that the deeper potential well of the coarse-grained model employed here means that the attractive intermolecular interactions are much harder to overcome than for the smaller, atomistic model. Whilst the same behaviour is observed in both growing

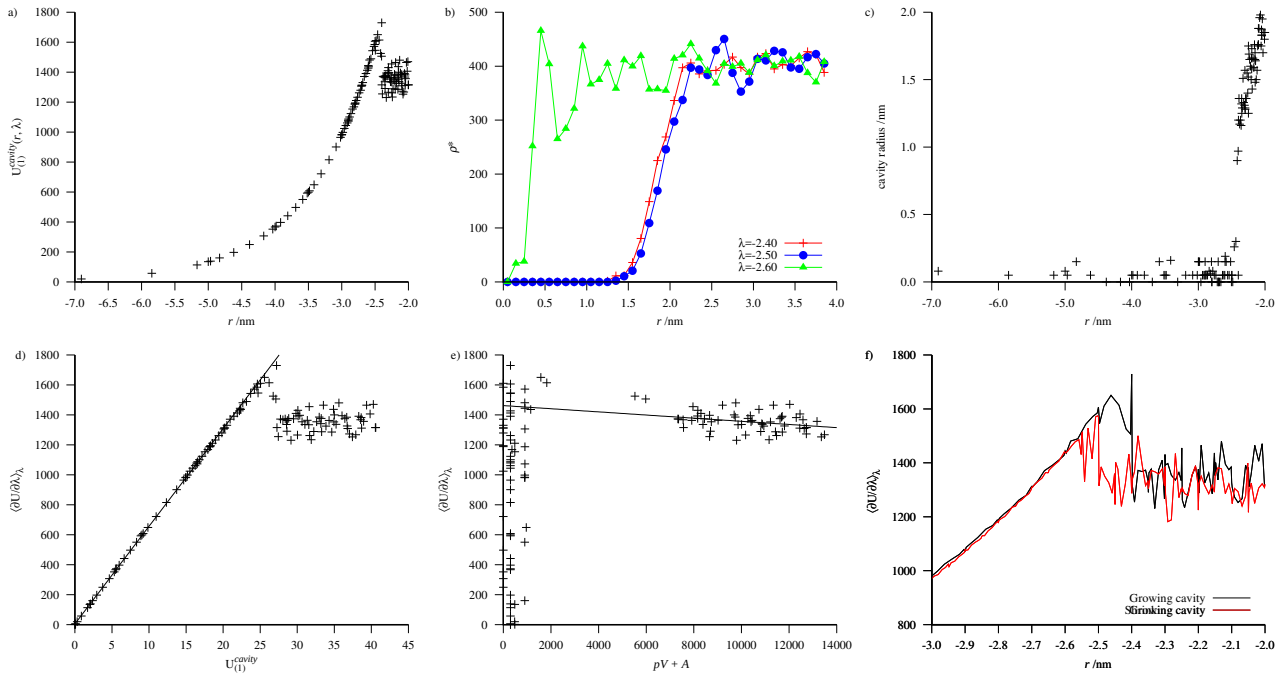

FIG. 3. Results for growing a cavity in heptane for $U_{(1)}^{cavity}$. a) $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for showing the discontinuity at $\lambda_t \approx -2.50$, b) for $\lambda = -2.40, -2.50$ and -2.60 which correspond to red crosses, blue circles and green triangles respectively. Below $\lambda = -2.50$ no cavity is observed. c) shows how the estimated cavity radius varies with λ when growing a cavity in heptane. d) shows the $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ against $U_{(1)}^{cavity}$ and the change in behaviour at $\lambda = \lambda_t$, with a linear fit applied to $\lambda \leq \lambda_t$. Whilst e) shows $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ against $pV + A$ which shows a linear dependence for $\lambda > \lambda_t$ confirming that above $\lambda_t \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ grows as a $pV + A$ term. f) shows the hysteresis present between the growth and annihilation of the cavity.

and shrinking the cavity, λ_t differs and there is a hysteresis in the integrand, see in both table II and figure 3f. For thermodynamic integration to accurately calculate the change in free energy the work undergone must be reversible, *i.e.* $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ must be continuous and well-defined at all values of λ.

Below λ_t, the integrand is proportional to $U_{(1)}^{cavity}$ which can be seen in 3d) that shows $\langle \frac{\partial U}{\partial \lambda} \rangle_\lambda$ against $U_{(1)}^{cavity}$ for growing a cavity in heptane. The constant of proportionality for $\lambda \leq \lambda_t$ is very similar for both growing and shrinking, approximately 65.15 and 64.73 for growing and shrinking the cavity respectively in heptane, whilst for toluene values of 130.29

TABLE II. Estimated values of λ_t for $U_{(1)}^{cavity}$

Solvent	$\lambda_t\,(grow)$	$\lambda_t\,(shrink)$
Heptane	-2.50	-2.56
Toluene	-2.95	-3.05

and 129.25 were obtained. Above λ_t $\langle\frac{\partial U}{\partial\lambda}\rangle_\lambda$ grows much more slowly, although it appears to continue to be proportional to $U_{(1)}^{cavity}$. Figure 3e) shows $\langle\frac{\partial U}{\partial\lambda}\rangle_\lambda$ against $pV + A$ where pV is a pressure-volume term due to expansion work, whilst A is the surface area of the cavity calculated using the cavity radius shown in Figure 3c). From this, it can be seen that below λ_t there is no clear dependence on $pV + A$ due to the lack of a cavity, however with the presence of a cavity above λ_t this term contributes to the integrand which continues to grow at a much slower pace as this term is much larger than that of $U_{(1)}^{cavity}$. Unfortunately, for $U_{(1)}^{cavity}$ is clearly not in this case, thus $U_{(1)}^{cavity}$ it is not an appropriate choice of cavity potential for this system. This demonstrates that although in principle the cavity potential can take any functional form, the choice is highly important and dependent on the system on interest.

B. Fixed WCA cavity results,$U_{(2)}^{cavity}$

To overcome the issues experienced with $U_{(1)}^{cavity}$, a harder potential based on a generalised WCA potential, $U_{(2)}^{cavity}$ shown in equation (13) was employed. As in section III A previously, the cavity/solute is fixed in the centre of the simulation box. Unlike $U_{(1)}^{cavity}$, the cavity radius is defined within the potential (σ_c), therefore the size of the cavity must be specified *a priori* for each case. We have used $\sigma_c = 1.0, 1.3$ and 1.5nm, all of which are large enough to avoid particle-particle overlaps between the inserted pyrene molecule and the solvent molecules. With $U_{(2)}^{cavity}$ the integrand, $\left\langle\frac{\partial U}{\partial\lambda}\right\rangle_\lambda$ is continuous over the entire λ range when either growing or shrinking the cavity (fig 4a). Note that the overall shape of $\langle\frac{\partial U}{\partial\lambda}\rangle_\lambda$ is unimportant in the calculation of ΔG. As with the previous cavity potential, $U_{(1)}^{cavity}$, we calculated the radial solvent density to monitor the cavity radius. ΔG_{insert} was calculated using equation (16) and found to be approximately independent of λ at all cavity sizes (figure 4b). The lack of discontinuity evident in ΔG_{insert} at $\lambda \to 0.0$ indicates that all the cavities investigated are

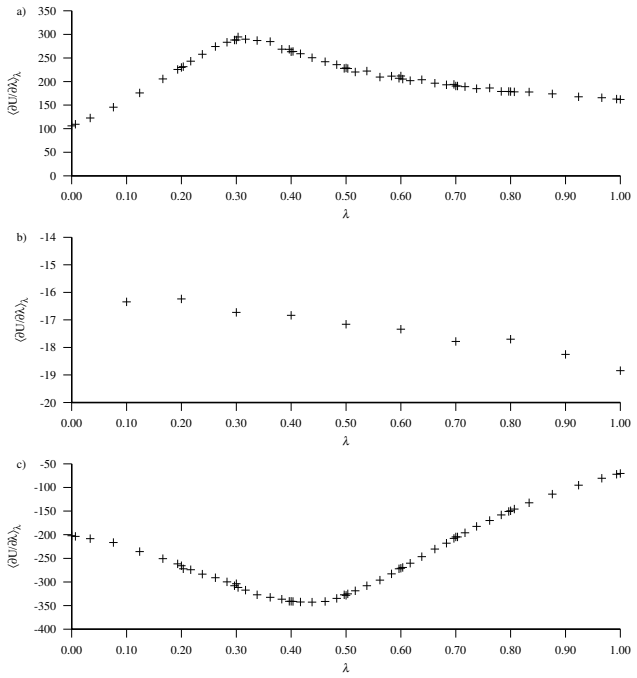

FIG. 4. $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for $U_{(2)}^{cavity}$ for pyrene in heptane with $\sigma_c = 1.0$nm showing a)-c) show stages 1-3 respectively. In all cases the standard error is smaller than the points.

large enough to overcome the end-point singularity due to particle-particle overlaps. The corresponding graphs for $\sigma_c = 1.3, 1.5$nm for heptane and all three values of σ_c for toluene are shown in the Supplementary information (SI Figures 2 and 3).

The μ_{excess} calculated for pyrene in heptane and toluene are given in table III and show a good agreement across the cavity sizes, as well as with experimental results which found $\log P_{TOL/HEP} = 0.79$[30], indicating that pyrene is slightly more soluble in toluene as expected due to their similar aromatic natures. The deviation observed from the experimental value is likely to be due to the coarse-grained nature of the model employed which, as with all coarse-grained potentials, sacrifices atomistic detail for computational tractability. The good agreement of the calculated $\log P$ with the experimental value is particularly impressive given the model did not take into account $\Delta G_{solvation}$ upon parameterization[44,52,53]. Table III includes the values calculated using TI(GQ) for ΔG_{grow} and ΔG_{shrink} and the simple trapezium rule to calculate ΔG_{insert} along with the calculated value of the partition coefficient. The calculated values using TI(Trap) and FEP are within statistical error of those calculated using TI(GQ) and are shown in fig 7.

The correlation time, t_c was estimated using the standard block average method[49,50]

TABLE III. Calculated μ_{excess} for pyrene in heptane and toluene at $T = 298.15$K, $p = 1$bar using a fixed position WCA potential cavity calculated using TI(GQ)

Cavity radius, σ_c /nm	μ_{excess}^{HEP}/kJmol^{-1}	μ_{excess}^{TOL}kJmol^{-1}	$\log P_{TOL/HEP}$
1.0	-45.12 ± 0.17	-48.10 ± 0.21	0.52 ± 0.05
1.3	-44.96 ± 0.26	-47.74 ± 0.36	0.48 ± 0.08
1.5	-44.72 ± 0.33	-50.21 ± 0.52	0.96 ± 0.11

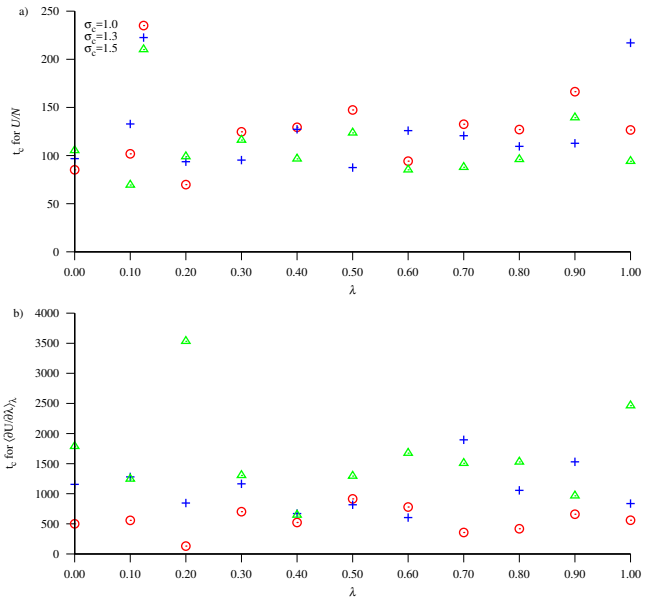

FIG. 5. Estimated t_c for ΔG_{insert} (Stage 2) with heptane as the solvent for a) U/N and b) $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$. The red circles, blue crosses and green triangles correspond to $\sigma_c = 1.0, 1.3$ and 1.5nm respectively.

(detailed in Appendix B) for each stage for $\left\langle \frac{U}{N} \right\rangle$ and $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for $\lambda = 0.0$ to 1.0 in 0.1 intervals. It was found that for all stages t_c for $\left\langle \frac{U}{N} \right\rangle$ is independent of λ, cavity size and solvent identity, for example, t_c for $\left\langle \frac{U}{N} \right\rangle$ for Stage 2 (insertion) in heptane is shown in figure 5a for all σ_c investigated. Calculated values for t_c averaged over all λ values are shown in table IV with a mean value of $t_c \approx 110$. However, it was found that the estimated t_c for $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ shows a dependence on λ for Stages 1 and 3 in which the cavity size varies. For the linear alchemical coupling employed in the thermodynamic integration undertaken in Stage 2 to calculate ΔG_{insert}, t_c is independent of λ, as shown in fig 5b for inserting pyrene into heptane and in the SI for toluene. The overall trends observed for ΔG_{grow} and ΔG_{shrink} mirror one another due to the interchange of λ and $(1 - \lambda)$ in equation (15), with the maximum t_c found for

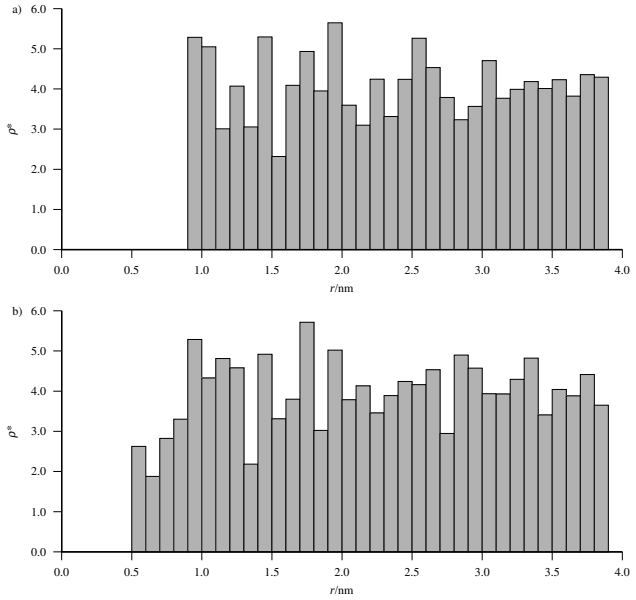

FIG. 6. Histograms showing the instantaneous number density of solvent from the centre of the cavity with $\sigma_c = 1.3$nm and $\lambda = 0.4$. The figures show that the cavity size at intermediate λ fluctuates from approximately 0.4nm to 0.9nm.

intermediate values of λ and a minimum when there is no cavity present ($\lambda = 0$ for ΔG_{grow} and $\lambda = 1$ for ΔG_{shrink} respectively). In all cases there is a maximum in t_c at intermediate λ (although the location of this maxima varies). This is due to a smaller energy difference between different cavity sizes due to the cavity functional form. For example, fig 6 shows the radial solvent density from the centre of the cavity for two configurations for growing a cavity in heptane with $\sigma_c = 1.3$nm, $\lambda = 0.4$, showing a cavity radius of approximately 0.4nm in a, and a much larger cavity radius of approximately 0.9nm in b for the same system. Note that these the histograms are not averaged over multiple configurations as previously shown in figure 3 for $U_{(1)}^{cavity}$. The estimated values of t_c for the different cavity radii investigated qualitatively suggest that t_c increases with increasing σ_c, particularly between $\sigma_c = 1.0$ and 1.3nm. However, it is not possible to quantify the effect of σ_c.

C. Moving WCA cavity results, $U_{(2)}^{cavity}$

The simulations were repeated with $U_{(2)}^{cavity}$ allowing the cavity/solute to undergo centre of mass displacement and rotation around the centre of mass. Note that due to the

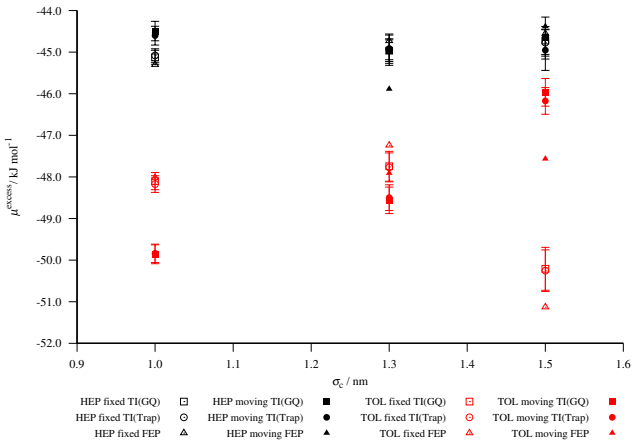

FIG. 7. Calculated values of μ^{excess} for pyrene in heptane (black symbols) and toluene (red symbols). The open symbols donate fixed cavity simulations whilst filled symbols donate moving cavity simulations. TI(GQ), TI(Trap) and FEP are denoted by squars, circles and triangles respectively. The graph shows that μ^{excess} is independent of cavity radius, σ_c.

spherical symmetry of the cavity rotational moves were only performed on the solute. The maximum displacement of the cavity, Δ_{cavity}, is distinct from that of the solvent molecules and constrained so that $0 < \Delta_{cavity} \leq \frac{l}{2}$, where l is the simulation box length.

The calculated values for μ_{excess} using a moving cavity are consistent with those found for the fixed cavity, with only a small deviation occurring for μ_{excess} of pyrene in toluene with $\sigma_c = 1.5$nm, and are shown in fig 7 for TI(Trap), TI(GQ) and FEP for both a moving and fixed WCA cavity. Plots comparing the integrand for $\sigma_c = 1.0nm$ can be found in the SI for both solvents and show excellent agreement, with an overall difference of 0.63kJ/mol (heptane) and 1.76kJ/mol (toluene). Overall, this leads to a calculated value of $\log P_{TOL/HEP} = 0.67$, again in good agreement with the experimental value of $\log P_{TOL/HEP} = 0.79$[30]. As previously discussed in section III B, the deviation from the experimental value is likely to be due to the approximations made in the coarse-grained model. As in section III B for the case of the fixed WCA cavity, t_c was calculated for each stage with respect to $\langle \frac{U}{N} \rangle$ and $\langle \frac{\partial U}{\partial \lambda} \rangle_\lambda$ for $\lambda = 0.0$ to 1.0 in 0.1 intervals. The estimated value of t_c for $\langle \frac{U}{N} \rangle$ was comparable to that of the fixed cavity, as shown in table IV for all three stages.

The estimated t_c values for $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for ΔG_{grow} (Stage 1) are approximately constant in all cases with $\lambda \geq 0.3$, and do not show the maximum at intermediate λ observed with the fixed

TABLE IV. Calculated t_c values for $\left\langle \frac{U}{N} \right\rangle$ for pyrene in heptane or toluene using a fixed or moving WCA cavity averaged over all λ

	t_c			
σ/nm	Heptane		Toluene	
	Fixed Cavity	Moving Cavity	Fixed Cavity	Moving Cavity
ΔG_{grow}				
1.0	130	119	99	96
1.3	110	123	111	102
1.5	115	115	104	101
ΔG_{insert}				
1.0	119	107	111	103
1.3	120	109	111	95
1.5	102	113	97	99
ΔG_{shrink}				
1.0	118	124	111	107
1.3	107	115	98	106
1.5	117	133	100	93

cavity. At small λ the estimated value is much smaller, for example $t_c \approx 2$ when $\lambda = 0.0$ in all cases, in comparison to $t_c \approx 300 - 600$ for the case of the fixed cavity. The significant decrease in t_c at $\lambda \to 0.0$ is due to the large value of Δ_{cavity} possible with small λ that allows the cavity to explore the entire simulation box in a fashion analogous to a random insertion. When $\lambda = 0.0$ there is no energy penalty associated with moving the cavity and so all cavity moves are accepted, leading to $\Delta_{cavity} = \frac{l}{2}$. For $\lambda \geq 0.3$ the estimated values of t_c are on the order of half to a third of that for the fixed cavity, with the largest decrease demonstrated for growing a cavity in heptane with $\sigma_c = 1.5$nm which decreases from $t_c = 1619$ to $t_c = 424$ upon allowing the cavity to move. The mean estimated values of t_c for $\lambda \geq 0.3$ for ΔG_{grow} are shown in table V.

The marked decrease in t_c is down to the removal of the slowest relaxation mode within the simulation, that of the solvent relaxation around the cavity. For a fixed cavity, the periodic boundaries lead to a periodic array of fixed cavities with nanoscale channels between

TABLE V. Mean t_c values for $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for ΔG_{grow} in heptane or toluene using a fixed or moving WCA cavity for $\lambda \geq 0.3$

	t_c			
	Heptane		Toluene	
σ/nm	Fixed Cavity	Moving Cavity	Fixed Cavity	Moving Cavity
1.0	915	233	635	257
1.3	1155	389	1170	469
1.5	1619	424	1330	420

them. It has been shown previously that nanoscale confinement of a fluid can dramatically increase the velocity autocorrelation function and thus the relaxation time[54–57]. A significant reduction in t_c is also observed for ΔG_{insert} in Stage 2. As observed for the fixed cavity, ΔG_{insert}, t_c is independent of λ with the linear alchemical thermodynamic integration scheme (equation (16)) and the mean t_c is shown in table VI below and graphs of t_c vs. λ are shown in the SI. Unlike Stage 1 there is an additional degree of freedom in Stages 2 and 3 with the possibility of the solute undergoing rotational moves around its centre of mass. In order to investigate the importance of this rotation and centre of mass displacements simulations were performed where the solute can either rotate around the centre of mass but remains fixed in space or with the solute orientation fixed and only centre of mass displacement moves (*i.e.* translational motion) is possible. In all cases the cavity was centred on the centre of mass of the solute. In the case of ΔG_{insert} this led to the rotation of the pyrene solute to being the equivalent of a new random orientation being generated as all rotations are accepted as none lead to an extremely high energy configuration with particle-particle overlaps due to the large size of the cavity.

The results indicate that for t_c for $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for ΔG_{insert} increases with cavity size. This is observed in the majority of cases with the largest increase observed between $\sigma_c = 1.0$ and 1.3nm. Both the case of only translational or rotational motion of the solute (and cavity) show a significant decrease from that of a fixed solute, and in all cases the simulations with rotational only moves showed a slightly greater reduction in t_c than those with translation only moves for ΔG_{insert}. The greater decrease in t_c observed for with rotational moves can be explained as due to the large cavity size in comparison to the solute, all solute

TABLE VI. Mean t_c values for $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for ΔG_{insert} in heptane or toluene using a fixed or moving WCA cavity and solute for $\lambda > 0.0$. Also included are the cases where only solute translation or rotation are allowed.

	t_c							
	Heptane				Toluene			
σ/nm	Fixed	Moving	Translation only	Rotation only	Fixed	Moving	Translation only	Rotation only
1.0	482	71	163	81	867	83	303	98
1.3	1071	140	390	211	1299	192	577	358
1.5	1616	275	401	348	1218	288	627	508

rotational moves are accepted, and essentially a new random orientation is generated in each cavity/solute trial move. As $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for ΔG_{insert} depends only on the interaction between the solvent and solute, the new orientation for the solute overcomes the slow relaxation time of the solvent around the fixed cavity. In all cases the calculated ΔG_{insert} was within statistical error of that calculated for the original fixed cavity case.

The correlation time for ΔG_{shrink} (Stage 3) was also estimated in the cases of fixed solute/cavity with no rotational or translational moves, centre of mass displacement moves only, rotation around the centre of mass only and both rotation displacement of the centre of mass moves. Similar trends were again observed for both heptane and toluene as a solvent (see SI). It was found that for $\lambda \to 1.0$ where there is no cavity t_c tends to the same value in all cases (*i.e.* fixed or moving solute), with $t_c = 255$ and $t_c = 260$ for pyrene in heptane and toluene at $\lambda = 1.0$ respectively. In the cases where the solute/cavity is able to undergo translational motion, with or without rotational motion, the estimated t_c is approximately constant for all values of σ_c and λ investigated see fig 8 for a representative case. Results for the other systems are shown in the SI.

In the case of rotational trial moves only there still a strong dependence on λ and σ_c, indicating that it is the centre of mass displacement that significantly increases the efficiency of the simulation. Indeed, it is the centre of mass translational moves not the rotational trial moves that remove the slow relaxation between nanoscale channels that is responsible for the large relaxation times. At $\lambda \to 0.0$ (*i.e.* a cavity much larger than the solute), for rotation about the centre of mass only, the estimated t_c is slightly below that of the fixed

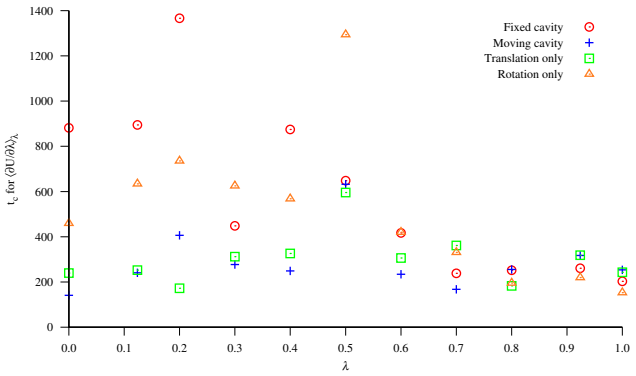

FIG. 8. Estimated t_c for ΔG_{shrink} for pyrene in heptane with $\sigma_c = 1.0$nm with a fixed (red circle) or moving cavity/solute (blue plus). Also shown are the case with cavity/solute translational (green cross) or rotational (orange triangle) moves only. Note in both these cases the solvent is free to undergo both translational and rotational trial moves.

solute. This slight decrease is due to the solute the large size of the cavity that means all rotational moves are accepted and in effect a new orientation is generated each time, reducing t_c as observed in the case of ΔG_{insert}. As λ increases, the size of the cavity shrinks the estimated value of t_c increases, becoming comparable to that of the fixed solute/cavity simulations. This is because the large rotations that increase the efficiency at low λ lead to particle-particle overlaps with the smaller cavity and are thus rejected. These results for ΔG_{shrink} clearly show that, whilst rotational moves around the centre of mass help to decrease t_c at low λ when there is a cavity large enough to accept all rotational trial moves, it is the centre of mass translational moves that drastically increase the efficiency of the simulations.

IV. CONCLUSION

The partition coefficient, $\log P_{TOL/HEP}$, of pyrene between toluene and heptane has been calculated using a novel cavity-based method previously employed by Li *et al*[25], thus avoiding particle-particle overlaps that can reduce the accuracy of free energy calculations. Good agreement was found with experimental results. The cavity-based technique provides an elegant methodology for free energy and excess chemical potential calculations that have applications in a wide range of scientific problems, notably in the area of solubility calcula-

tions. Here we have demonstrated that the cavity-based method is independent of model and underlying molecular simulation technique, having previously been employed using atomistic molecular dynamics simulations[25]. In addition, it can easily be integrated into both pre-existing software and bespoke programs. Here we have investigated the influence of the cavity attributes on the accuracy and efficiency of the calculation. We have shown that, although in principle, the cavity attributes play little role in the calculation of $\Delta G_{solvation}$ as the cavity is both created and destroyed, providing the cavity is large enough for the solute to fit inside and thus avoid any overlaps with the solvent molecules, in practice that is not the case with the accuracy and efficiency of the calculation affected by the cavity attributes. The choice of cavity potential is important on the accuracy of the calculation. We have shown that a poor choice of cavity functional form can lead to a hysteresis and discontinuity in the $\langle \frac{\partial U}{\partial \lambda} \rangle_\lambda$ due to a nucleation event. Previously utilised exponential cavity potentials employed in atomistic simulations have been found to be unsuccessful with the coarse-grained model employed here. The hysteresis is overcome by the use of a harder cavity potential based on the WCA potential. We found that by allowing the cavity/solute molecule to move can greatly increase the efficiency of the simulations by removing the slow relaxation mode of the solvent confined in nanoscale channels bounded by an infinite fixed array of cavities. The reduction in t_c was most striking for ΔG_{grow}, which demonstrated up to an order of magnitude decrease in t_c.

V. SUPPLEMENTARY MATERIAL

See supplementary material for a complete set of results for all cavity sizes and potentials investigated in both solvents.

ACKNOWLEDGMENTS

We would like to thank Dr Guadalupe Jiménez-Serratos, Professor Erich Müller and Professor George Jackson from Imperial College, London for the development of the forcefield employed in these simulations along with helpful discussions throughout. The authors would also like to thank and acknowledge funding and technical support from BP, including BP's High Performance Computing Facility through the BP International Centre for Advanced

Materials (BP-ICAM) which made this research possible. All the simulations carried out within this work were conducted using the HPC resources from BP and the Department of Chemistry, Cambridge.

Appendix A: Excess chemical potential and partition coefficient

In the thermodynamic the chemical potential of a solute in solution is given by the change in free energy upon the addition of a solute molecule[38,58], that is,

$$\mu_{solution} = G_{solution}\left(T, p, N_{solute}+1, N_{solvent}\right) - G_{solution}\left(T, p, N_{solute}, N_{solvent}\right) \tag{A1}$$

Whilst the free energy, G is given by

$$G\left(T, p, N_{solute}, N_{solvent}\right) = -k_B T \ln\left(\Delta_0\right) \tag{A2}$$

where Δ_0 is the isobaric-isothermal NpT partition function for state 0 where $T, p, N_{solute}, N_{solvent}$

$$\begin{aligned}\Delta_0 = &\frac{q_{solute}^{N_{solute}} q_{solvent}^{N_{solvent}}}{\Lambda_{solute}^{3N_{solute}} \Lambda_{solvent}^{3N_{solvent}} N_{solvent}!\, N_{solute}!} \int \exp\left(-\beta p V\right) dV \\ &\times \int \exp\left[-\beta U\left(N_{solute}, N_{solvent}\right)\right] d\mathbf{R}\left(N_{solute}, N_{solvent}\right)\end{aligned} \tag{A3}$$

where q is the intramolecular (vibrational and electronic) partition function and Λ the de Broglie wavelength for the solute or solvent as indicated by the subscript. The free energy for state 1, $T, p, N_{solute}+1, N_{solvent}$, is given by an analogous expression to that shown in ((A3)) with

$$U\left(N_{solute}+1, N_{solvent}\right) = U\left(N_{solute}, N_{solvent}\right) + U_{solute-solvent}\left(\mathbf{R}_{solute}\right) \tag{A4}$$

where $\mathbf{R}_{solute}$ is the fixed position of the extra solute and $U_{solute-solvent}\left(\mathbf{R}_{solute}\right)$ is the interaction energy of the extra solute with the rest of the solution. Following equation (A1) and (A2),we can express $\mu_{solution}$ as

$$\begin{aligned}\mu_{solution} &= -k_B T \ln\left(\Delta_1/\Delta_0\right) = -k_B T \ln\left\{q_{solute}/\left[\Lambda_{solute}^3\left(N_{solute}+1\right)\right]\right\} \\ &- k_B T \ln\left\{\frac{\int \exp\left(-\beta p V\right) dV \int \exp\left[-\beta U\left(N_{solvent}, N_{solute}+1\right)\right] d\mathbf{R}\left(N_{solvent}, N_{solute}+1\right)}{\int \exp\left(-\beta p V\right) dV \int \exp\left[-\beta U\left(N_{solvent}, N_{solute}\right)\right] d\mathbf{R}\left(N_{solvent}, N solute\right)}\right\} \\ &= -k_B T \ln\left\{q_{solute}/\left[\Lambda_{solute}^3\left(N_{solute}+1\right)\right]\right\} - k_B T \ln\left\langle V \exp\left[-\beta U_{solute-solution}\left(\mathbf{R}_{solute}\right)\right]\right\rangle_0\end{aligned} \tag{A5}$$

The first term in equation (A5) is the ideal gas contribution to the chemical potential, μ_{ideal}. The second term is a free energy perturbation (FEP) expression to compute the free energy difference between the solution with and without the extra solute with $\langle \ldots \rangle_0$ indicating an ensemble average at state 0. Away from the critical points the correlation between insertion energy and volume is weak,[59,60] thus

$$\mu_{solution} = \mu_{ideal} + k_B T \ln\left(\rho_{solute}\right) - k_B T \ln \left\langle \exp\left[-\beta U\left(\mathbf{R}_{solute}\right)\right]\right\rangle_0 \tag{A6}$$

where $\rho_{solute} = \left(N_{solute}+1\right)/V \approx N_{solute}/V$ and is the number density concentration and equation (A6) is equivalent to equation (3). The derivation of the partition coefficient, $\ln P_{I/II}$ follows as detailed in section II in the main paper by noting the equivalence of (total) chemical potentials of the solute in the two phases at equilibrium.

Appendix B: Correlation time

When computing ensemble averages from simulations the statistical accuracy of the value obtained from the is important. Here we employ a method to calculate the correlation time, t_c, based on block averages[49,50,61]. There are, however, many alternative ways to estimate the statistical error, including the explicit calculation of correlation functions[49,50,62]. We have chosen to employ the block average approach as it can be applied directly to the results obtained from Monte Carlo simulations without the need to calculate additional quantities.

Assuming that $A\left(t\right)$ is a Gaussian process, the ensemble average of the run $\left\langle A\right\rangle_{run}$ and variance in the mean of the run, $\sigma^2\left(A\right)$, are given by

$$\left\langle A\right\rangle_{run} = \frac{1}{t_{run}} \int_0^{t_{run}} A\left(t\right) \mathrm{d}t \tag{B1}$$

$$\begin{aligned} \sigma^2\left(\left\langle A\right\rangle_{run}\right) &= \frac{1}{t_{run}} \left(\left\langle A^2\right\rangle_{run} - \left\langle A\right\rangle_{run}^2\right) \\ &= \frac{1}{t_{run}^2} \iint_0^{t_{run}} \left\langle \left[A\left(t\right) - \left\langle A\right\rangle\right]\left[A\left(t'\right) - \left\langle A\right\rangle\right]\right\rangle \mathrm{d}t\mathrm{d}t' \end{aligned} \tag{B2}$$

where the integrand in (B2) is simply the time correlation function, $C_A\left(t-t'\right)$. Note that the Gaussian assumption holds for both the potential energy, U, and $\frac{\partial U}{\partial \lambda}$ considered here. By normalizing $C_A\left(t\right)$ and noting that t_c is the integral from $-\infty$ to $+\infty$ (otherwise known as the characteristic decay time of C_A or, in our case, the correlation time) in (B2), in the

case where t_{run} is much larger than t_c we can rewrite (B2) as

$$\sigma^2\left(\langle A\rangle_{run}\right) \approx \frac{t_c}{t_{run}}\left(\left\langle A^2\right\rangle_{run} - \langle A\rangle_{run}^2\right) \tag{B3}$$

where $\sigma^2(A)$ is the estimated error and t_c is currently unknown. To calculate t_c we split the simulation into n_b blocks of size b such that $n_b t_b = t_{run}$. For each block we calculate the mean, $\langle A\rangle_b$ and estimate the variance, $\sigma^2\left(\langle A\rangle_b\right)$ by

$$\langle A\rangle_b = \frac{1}{t_b}\sum_{t=1}^{t_b} A(t) \tag{B4}$$

$$\sigma^2\left(\langle A\rangle_b\right) \approx \frac{1}{n_b}\sum_{b=1}^{n_b}\left(\langle A\rangle_b - \langle A\rangle_{run}\right)^2 \tag{B5}$$

By comparing (B3) and (B5) we can estimate t_c by computing the function

$$P(t_b) = \frac{t_b\sigma^2\left(\langle A\rangle_b\right)}{\left\langle A^2\right\rangle_{run} - \langle A\rangle_{run}^2} \tag{B6}$$

which approaches t_c in the limit $t_b \gg t_c$.

REFERENCES

[1]M. Lísal, W. R. Smith, and J. Kolafa, "Molecular simulations of aqueous electrolyte solubility: 1. the expanded-ensemble osmotic molecular dynamics method for the solution phase," The Journal of Physical Chemistry B **109**, 12956–12965 (2005).

[2]J. C. Dearden, "In silico prediction of aqueous solubility," Expert opinion on drug discovery **1**, 31–52 (2006).

[3]D. S. Palmer, A. Llinàs, I. Morao, G. M. Day, J. M. Goodman, R. C. Glen, and J. B. Mitchell, "Predicting intrinsic aqueous solubility by a thermodynamic cycle," Molecular pharmaceutics **5**, 266–279 (2008).

[4]F. Moucka, M. Lísal, J. Skvor, J. Jirsák, I. Nezbeda, and 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**, 7849–7861 (2011).

[5]J. Aragones, E. Sanz, and C. Vega, "Solubility of nacl in water by molecular simulation revisited," The Journal of chemical physics **136**, 244508 (2012).

[6]H. M. Manzanilla-Granados, H. Saint-Martín, R. Fuentes-Azcatl, and 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**, 8389–8396 (2015).

[7]Z. Mester and 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**, 044505 (2015).

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

[9]J. Espinosa, J. Young, H. Jiang, D. Gupta, C. Vega, E. Sanz, P. Debenedetti, and A. 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**, 154111 (2016).

[10]G. L. Amidon, H. Lennernäs, V. P. Shah, and 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**, 413–420 (1995).

[11]M. H. Abraham and J. Le, "The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship," Journal of pharmaceutical sciences **88**, 868–880 (1999).

[12]W. L. Jorgensen and E. M. Duffy, "Prediction of drug solubility from monte carlo simulations," Bioorganic & medicinal chemistry letters **10**, 1155–1158 (2000).

[13]W. L. Jorgensen and E. M. Duffy, "Prediction of drug solubility from structure," Advanced drug delivery reviews **54**, 355–366 (2002).

[14]M. Esmaili, S. M. Ghaffari, Z. Moosavi-Movahedi, M. S. Atri, A. Sharifizadeh, M. Farhadi, R. Yousefi, J.-M. Chobert, T. Haertlé, and 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**, 2166–2172 (2011).

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

[16]O. C. Mullins, A. E. Pomerantz, J. Y. Zuo, and C. Dong, "Downhole fluid analysis and asphaltene science for petroleum reservoir evaluation," Annual review of chemical and biomolecular engineering **5**, 325–345 (2014).

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

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

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

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

[21]T. Headen, E. Boek, G. Jackson, T. Totton, and E. Muller, "Simulation of asphaltene aggregation through molecular dynamics: Insights and limitations," Energy and Fuels **31**, 1108–1125 (2017).

[22]S. Senkan and 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**, 188–190 (2003).

[23]A. Benavides, J. Aragones, and C. Vega, "Consensus on the solubility of nacl in water from computer simulations using the chemical potential route," The Journal of chemical physics **144**, 124504 (2016).

[24]M. Ferrario, G. Ciccotti, E. Spohr, T. Cartailler, and P. Turq, "Solubility of kf in water by molecular dynamics using the kirkwood integration method," The Journal of chemical physics **117**, 4947–4953 (2002).

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

[26]F. Moučka, J. Kolafa, M. Lísal, and W. R. Smith, "Chemical potentials of alkaline earth metal halide aqueous electrolytes and solubility of their hydrates by molecular simulation: Application to cacl2, antarcticite, and sinjarite," The Journal of Chemical Physics **148**, 222832 (2018).

[27]A. Ghoufi and 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**, 6496–6502

(2010).

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

[29]L. J. Dunne and G. Manos, "Statistical mechanics of binary mixture adsorption in metal-organic frameworks in the osmotic ensemble." Philosophical Transactions A: Mathematical, Physical and Engineering Sciences (2017).

[30]H. K. Hansen, C. Riverol, and W. E. Acree, "Solubilities of anthracene, fluoranthene and pyrene in organic solvents: comparison of calculated values using unifac and modified unifac (dortmund) models with experimental data and values using the mobile order theory," The Canadian Journal of Chemical Engineering **78**, 1168–1174 (2000).

[31]A. Jouyban, A. Shayanfar, and W. E. Acree, "Solubility prediction of polycyclic aromatic hydrocarbons in non-aqueous solvent mixtures," Fluid Phase Equilibria **293**, 47–58 (2010).

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

[33]A. Leo, C. Hansch, and D. Elkins, "Partition coefficients and their uses," Chemical reviews **71**, 525–616 (1971).

[34]E. Baka, J. E. Comer, and 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**, 335–341 (2008).

[35]J. Alsenz and M. Kansy, "High throughput solubility measurement in drug discovery and development," Advanced drug delivery reviews **59**, 546–567 (2007).

[36]A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," The journal of physical chemistry B **102**, 3586–3616 (1998).

[37]G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides," The Journal of Physical Chemistry B **105**, 6474–6487 (2001).

[38]B. Widom, "Some topics in the theory of fluids," The Journal of Chemical Physics **39**,

2808–2812 (1963).

[39]T. Steinbrecher, D. L. Mobley, and D. A. Case, "Nonlinear scaling schemes for lennard-jones interactions in free energy calculations," The Journal of chemical physics **127**, 214108 (2007).

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

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

[42]M. Zacharias, T. Straatsma, and J. McCammon, "Separation-shifted scaling, a new scaling method for lennard-jones interactions in thermodynamic integration," The Journal of chemical physics **100**, 9025–9031 (1994).

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

[44]E. Mller and A. Meja, "Extension of the saft-vr mie eos to model homonuclear rings and its parametrization based on the principle of corresponding states." Langmuir (2017), 10.1021/acs.langmuir.7b00976.

[45]C. L. Kong, "Combining rules for intermolecular potential parameters. ii. rules for the lennard-jones (12–6) potential and the morse potential," The Journal of chemical physics **59**, 2464–2467 (1973).

[46]P. J. Linstrom and W. G. Mallard, eds., *NIST Chemistry WebBook, NIST Standard Reference Database Number 69* (National Institute of Standards and Technology, Gaithersburg MD, 20899, 2005).

[47]J. P. Postma, H. J. Berendsen, and J. R. Haak, "Thermodynamics of cavity formation in water. a molecular dynamics study," in *Faraday Symposia of the Chemical Society*, Vol. 17 (Royal Society of Chemistry, 1982) pp. 55–67.

[48]J. D. Weeks, D. Chandler, and H. C. Andersen, "Role of repulsive forces in determining equilibrium structure of simple liquids," Journal of Chemical Physics **54**, 5237–+ (1971).

[49]D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Vol. 1 (Academic press, 2001).

[50]M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (Oxford university press, 2017).

[51]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, "Interaction models for water in relation to protein hydration," in *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*, edited by B. Pullman (Springer Netherlands, Dordrecht, 1981) pp. 331–342.

[52]E. A. Müller and K. E. Gubbins, "Molecular-based equations of state for associating fluids: A review of saft and related approaches," Industrial & engineering chemistry research **40**, 2193–2211 (2001).

[53]O. Lobanova, C. Avendano, T. Lafitte, E. A. Muller, and G. Jackson, "Saft-gamma force field for the simulation of molecular fluids: 4. a single-site coarse-grained model of water applicable over a wide temperature range," Molecular Physics **113**, 1228–1249 (2015).

[54]I. Bitsanis, S. A. Somers, H. T. Davis, and M. Tirrell, "Microscopic dynamics of flow in molecularly narrow pores," The Journal of Chemical Physics **93**, 3427–3431 (1990).

[55]A. L. Demirel and S. Granick, "Glasslike transition of a confined simple fluid," Physical review letters **77**, 2261 (1996).

[56]L. Bocquet and J.-L. Barrat, "Hydrodynamic properties of confined fluids," Journal of Physics: Condensed Matter **8**, 9297 (1996).

[57]M. Frank and D. Drikakis, "Solid-like heat transfer in confined liquids," Microfluidics and Nanofluidics **21**, 148 (2017).

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

[59]J. G. Kirkwood, "Statistical mechanics of fluid mixtures," The Journal of Chemical Physics **3**, 300–313 (1935).

[60]K. Shing and S. Chung, "Computer simulation methods for the calculation of solubility in supercritical extraction systems," Journal of Physical Chemistry **91**, 1674–1681 (1987).

[61]G. Jacucci and A. Rahman, "Comparing the efficiency of metropolis monte carlo and molecular-dynamics methods for configuration space sampling," Il Nuovo Cimento D **4**, 341–356 (1984).

[62]R. Zwanzig and N. K. Ailawadi, "Statistical error due to finite time averaging in computer experiments," Physical Review **182**, 280 (1969).

Addressing hysteresis and slow equilibration issues in cavity-based calculation of chemical potentials

C. R. Wand,[1] T. S. Totton,[2] and D. Frenkel[1]

[1)] *Department of Chemistry, University of Cambridge, Cambridgeshire, CB2 1EW, United Kingdom* [a)]

[2)] *BP Exploration Operating Co. Ltd., Sunbury-on-Thames, TW16 7LN, United Kingdom*

(Dated: 5 June 2018)

SUPPLEMENTAL MATERIAL

[a)] Author to whom correspondence should be addressed: df246@cam.ac.uk

I. EXPONENTIAL CAVITY RESULTS FOR TOLUENE, $U_{(1)}^{cavity}$

This section shows the results obtained when using toluene as a solvent for $U_{(1)}^{cavity}$. As discussed in the main manuscript, the trend observed is consistent with that shown in Figure 3 for heptane.

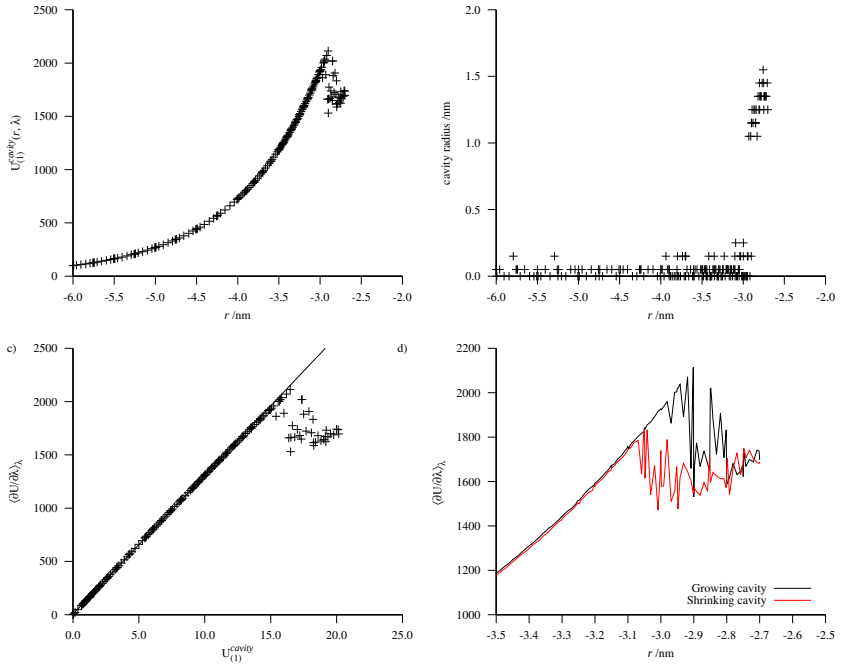

FIG. 1. Results for growing a cavity in toluene for $U_{(1)}^{cavity}$. a) $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for showing the discontinuity at $\lambda_t \approx -2.95$, b) shows how the estimated cavity radius varies with λ. Below $\lambda = -2.95$ no cavity is observed. c) shows the $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ against $U_{(1)}^{cavity}$ and the change in behaviour at $\lambda = \lambda_t$, with a linear fit applied to $\lambda \leq \lambda_t$. d) shows the hysteresis present between the growth and annihilation of the cavity.

II. WCA FIXED CAVITY RESULTS FOR HEPTANE, $U_{(2)}^{cavity}$

This section shows the results obtained when using toluene as a solvent for $U_{(1)}^{cavity}$. As discussed in the main manuscript, the trend observed is consistent with that shown in Figure 4 for $\sigma_c = 1.0$nm.

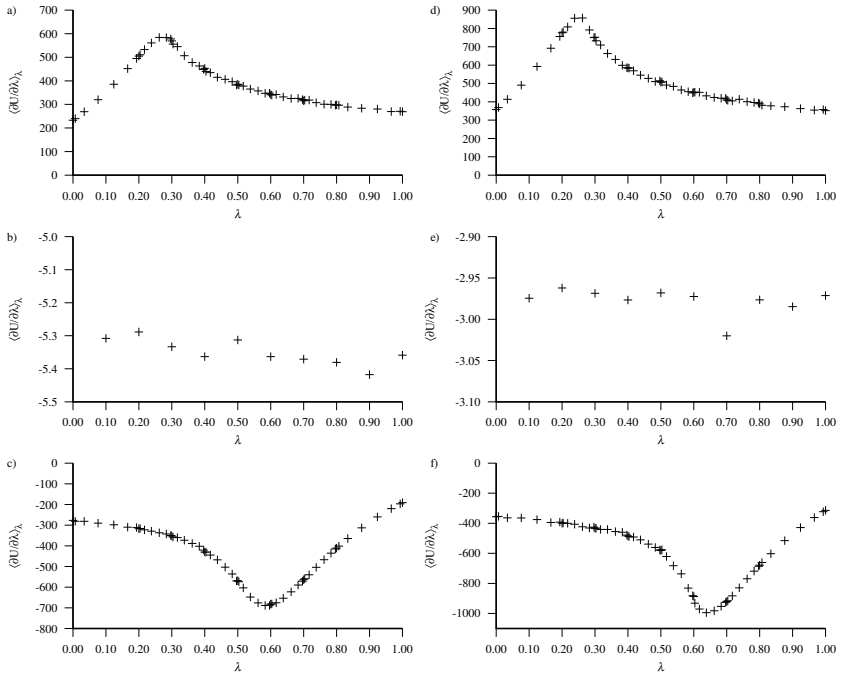

FIG. 2. $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for $U_{(2)}^{cavity}$ growing a cavity in heptane (top row), inserting pyrene into heptane (middle row) and shrinking the cavity (bottom row). For a)-c) $\sigma_c = 1.3$nm whilst $\sigma_c = 1.5$nm for d)-f).In all cases the standard error is smaller than the points.

III. WCA FIXED CAVITY RESULTS FOR TOLUENE, $U_{(2)}^{cavity}$

This section shows the results obtained when using toluene as a solvent for $U_{(1)}^{cavity}$. As discussed in the main manuscript, the trend observed is consistent with that shown in Figure 4 for heptane $\sigma_c = 1.0$nm and in the previous supporting information section for heptane at other σ_c values.

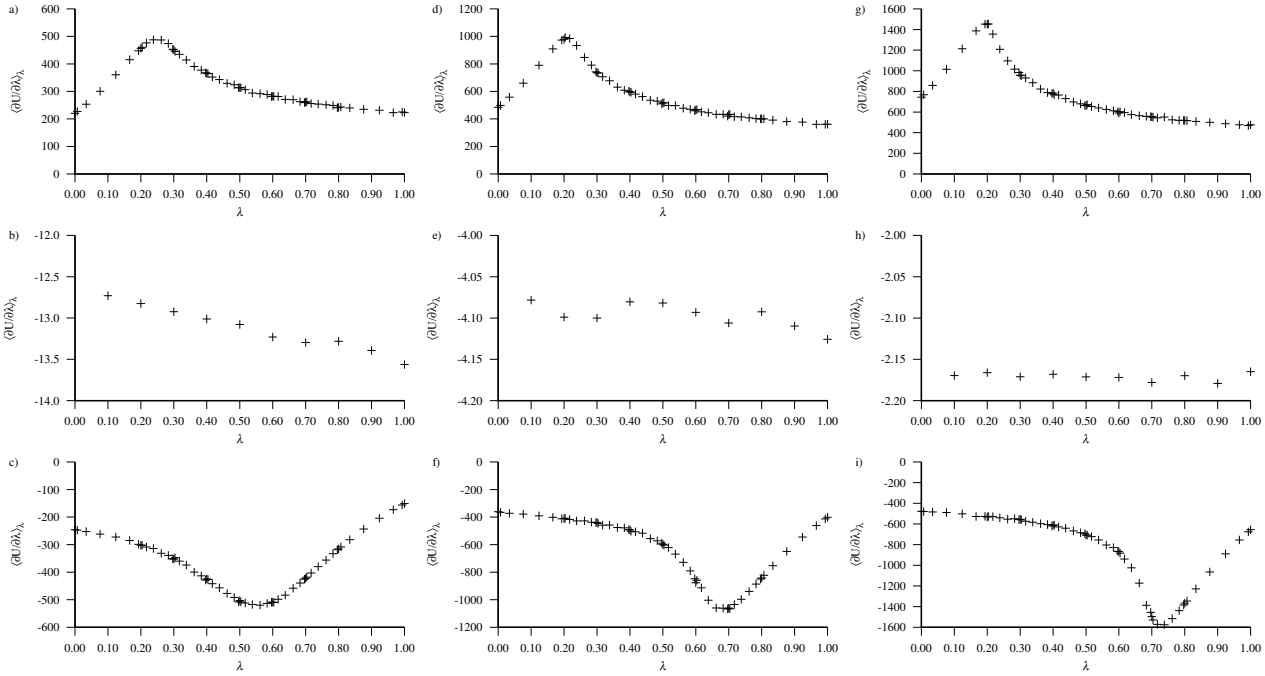

FIG. 3. $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for $U_{(2)}^{cavity}$ growing a cavity in toluene (top row), inserting pyrene into heptane (middle row) and shrinking the cavity (bottom row). For a)-c) $\sigma_c = 1.0$nm whilst $\sigma_c = 1.3$nm for d)-f) and $\sigma_c = 1.5$nm for g)-i). In all cases the standard error is smaller than the points.

IV. COMPARISON OF FIXED AND MOVING CAVITY RESULTS FOR WCA CAVITY POTENTIAL, $U_{(2)}^{cavity}$

This section shows a comparison between results obtained using a fixed or moving cavity for $\sigma_c = 1.0$nm for both heptane and toluene.

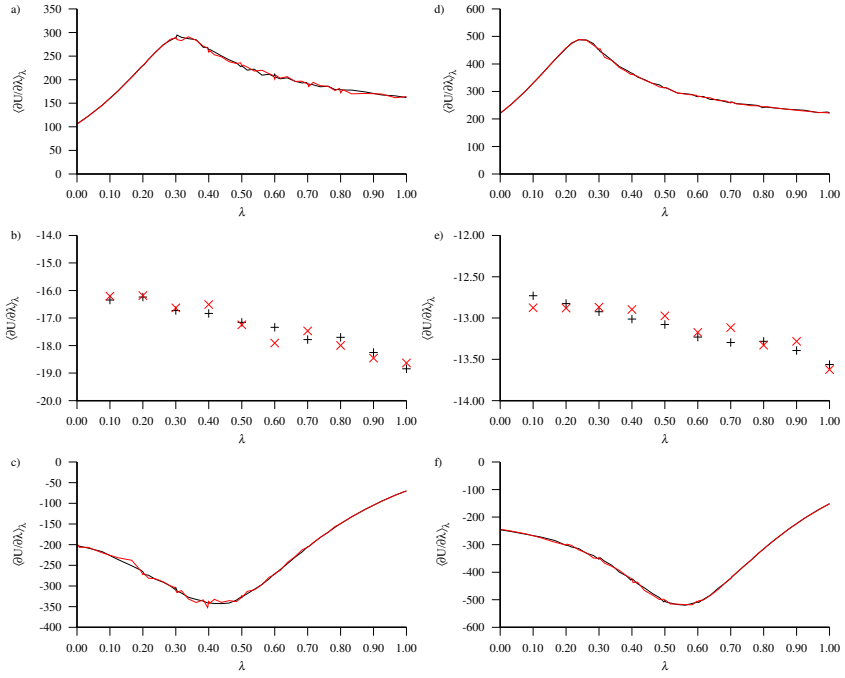

FIG. 4. $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for $U_{(2)}^{cavity}$ growing a cavity (top row), inserting pyrene into heptane (middle row) and shrinking the cavity (bottom row) for $\sigma_c = 1.0$nm. In each case the results obtained from a fixed cavity are shown in black, and in red for a moving cavity. a)-c) are for heptane as a solvent, whilst d)-f) are for toluene as a solvent.In all cases the standard error is smaller than the points.

V. CORRELATION TIME, t_c FOR FIXED AND MOVING WCA CAVITY POTENTIAL,$U_{(2)}^{cavity}$

This section shows the estimated correlation times, t_c for fixed and moving cavities. Also shown for stages 2 and 3 are solute translational moves only and rotational moves only.

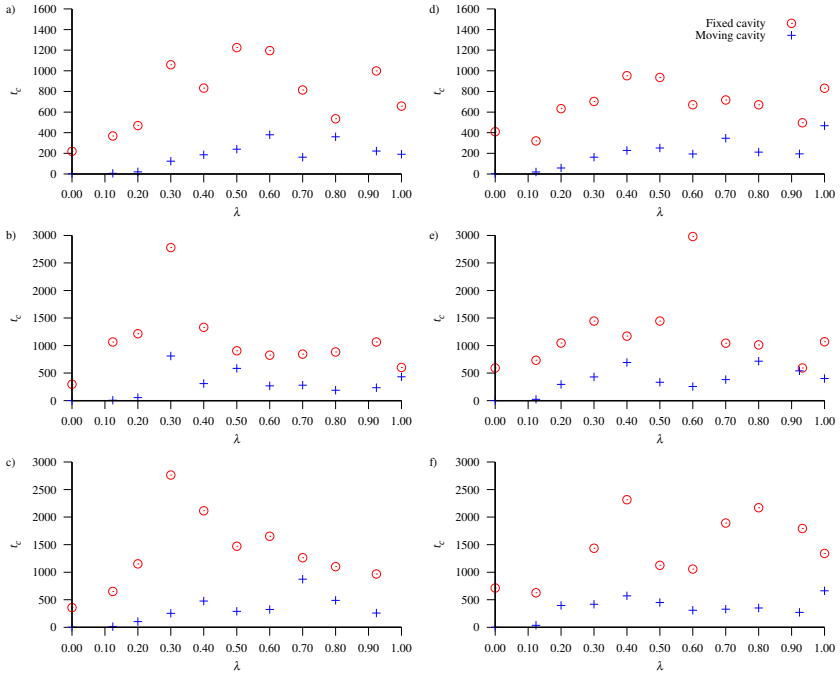

FIG. 5. The correlation time, t_c, for $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for $U_{(2)}^{cavity}$ for growing a cavity (Stage 1)in heptane (a-c) and toluene (d-f). The top row (a,d) have $\sigma_c = 1.0$nm, middle (b,e) have $\sigma_c = 1.3$nm and bottom (c,f) have $\sigma_c = 1.5$nm respectively. The results for a fixed cavity are shown by red circles, whilst those for a moving cavity are shown by blue pluses.

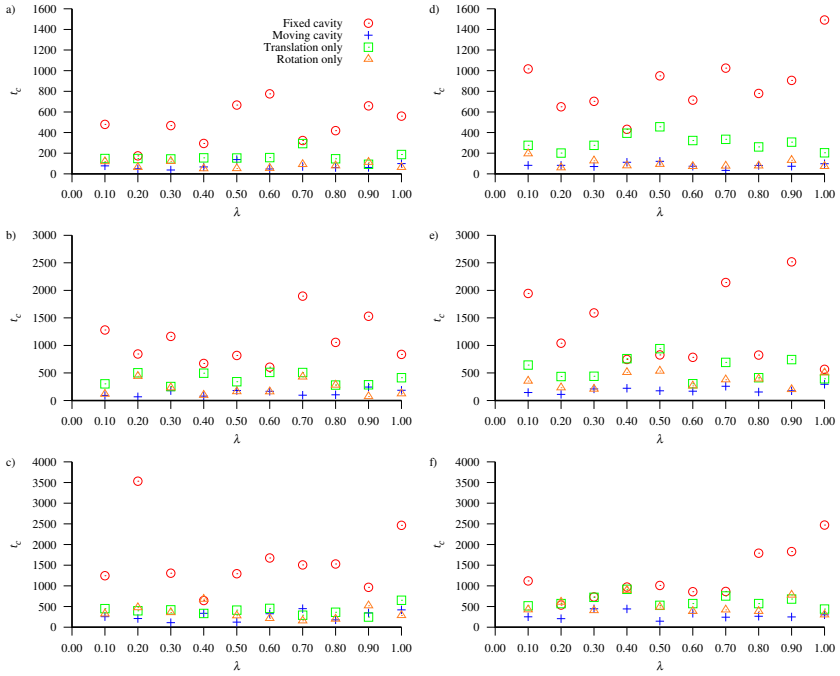

FIG. 6. The correlation time, t_c, for $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for $U_{(2)}^{cavity}$ for inserting pyrene (Stage 2) in heptane (a-c) and toluene (d-f). The top row (a,d) have $\sigma_c = 1.0$nm, middle (b,e) have $\sigma_c = 1.3$nm and bottom (c,f) have $\sigma_c = 1.5$nm respectively. The results for a fixed cavity are shown by red circles, whilst those for a moving cavity are shown by blue pluses. Also shown are the results for translational moves only (green squares) and rotational moves only (orange triangles).

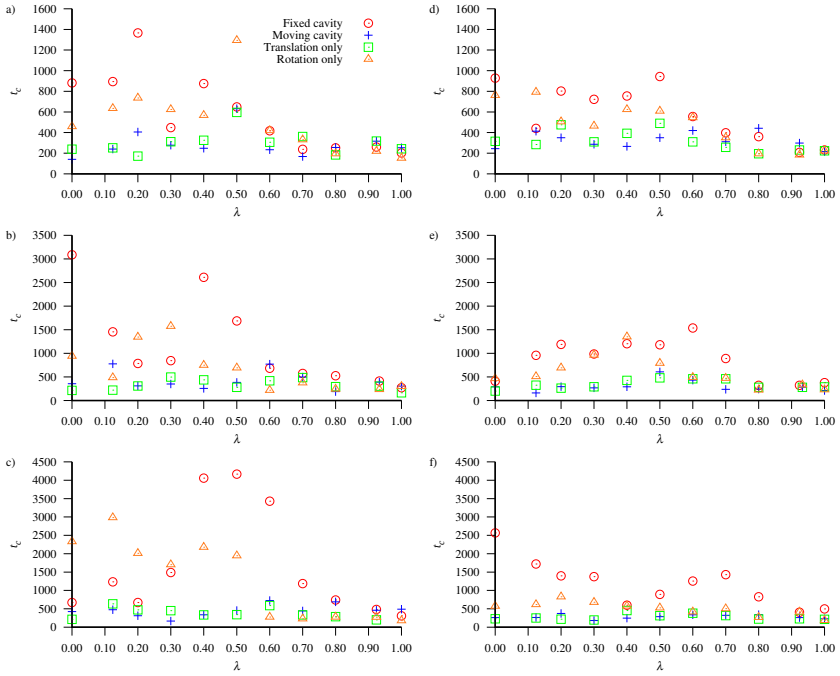

FIG. 7. The correlation time, t_c, for $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda$ for $U_{(2)}^{cavity}$ for shrinking a cavity (Stage 3) for pyrene in heptane (a-c) and toluene (d-f). The top row (a,d) have $\sigma_c = 1.0$nm, middle (b,e) have $\sigma_c = 1.3$nm and bottom (c,f) have $\sigma_c = 1.5$nm respectively. The results for a fixed cavity are shown by red circles, whilst those for a moving cavity are shown by blue pluses. Also shown are the results for translational moves only (green squares) and rotational moves only (orange triangles).