

Supporting Information for:

Dissipative particle dynamics parametrisation using infinite dilution activity coefficients: the impact of bonding

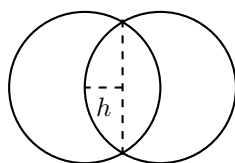
Authors: Rachel Hendrikse,* Carlos Amador and Mark R. Wilson

Department of Chemistry, Durham University, Lower Mountjoy, Stockton Road, Durham, DH1 3LE, United Kingdom.

* Author for correspondence. E-mail: rachel.hendrikse@durham.ac.uk

1 Volume of overlapping spheres

Consider two overlapping spheres, as shown below.



Suppose the distance between the two centres of the spheres is l_0 (i.e. the bond length) and the radius of each sphere is R_C (i.e. the cut-off). The overlapping region can be considered as two spherical caps, each with height h . The volume of one of the caps has a well-known formula, and be calculated by

$$V_{\text{Cap}} = \frac{\pi h^2}{3}(3R_C - h). \quad (1)$$

We can say that

$$l_0 = 2R_C - 2h, \quad (2)$$

and rearranging for h yields $h = R_C - \frac{l_0}{2}$. The total volume of the overlapping region is $V_{\text{OL},2} = 2V_{\text{Cap}}$, and substituting the expression for h finds that

$$V_{\text{OL},2} = \frac{\pi}{12}(4R_C + l_0)(2R_C - l_0)^2. \quad (3)$$

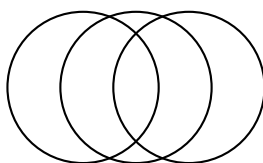
The volume which is outside of the overlapping region can be considered as the total volume of the spheres, minus the part contained in the overlapping region. The total volume of the spheres is

$$V_{\text{T}} = 2 \times \frac{4\pi R_C^3}{3} \quad (4)$$

Since the region inside the overlapping portion is for two spheres, the volume of this part needs to be taken off twice, making the outside region

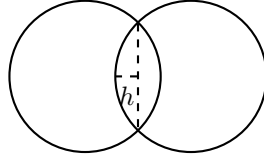
$$V_{\text{OUT}} = V_{\text{T}} - 2V_{\text{OL},2} \quad (5)$$

Now consider three overlapping spheres in a line, as shown below.



The central region where all three spheres overlap, can be calculated in a similar way as described above, only with the separation between the central points being $2l_0$ as opposed to l_0 .

If we remove the central sphere, this can be visualised as below.



In this case, $2l_0 = 2R_C - 2h$, reducing to $h = R_C - l_0$. Using the same formula for V_{Cap} , we find the region of the three overlapping spheres is

$$V_{\text{OL},3} = \frac{2\pi}{3}(2R_C + l_0)(R_C - l_0)^2, \quad (6)$$

which is equivalent to replacing l_0 in Eq. 3 with $2l_0$.

The region concerning two overlapping spheres can, in this case, be calculated as twice the volume calculated using Eq. 3 (to account for the fact that there are two spheres) minus twice the volume $V_{\text{OL},3}$. Finally, the volume outside all of the overlapping regions can be calculated as

$$V_{\text{OUT}} = 3 \times \frac{4\pi R_C^3}{3} - 2V_{\text{OL},2} - 3V_{\text{OL},3}. \quad (7)$$

In the main article, the above equations are generalised to molecules of any length N .

2 Expected relation between excess chemical potential and volume

In our work, we have assumed that the effective excess chemical potential of a bead (or part of a bead) can be related to the number of interactions N_I it has with surrounding beads such that

$$\mu_{\text{ex}} \propto N_I \propto V\rho \quad (8)$$

where V is the volume of the fragment and ρ is the density of interacting particles inside it (i.e. the bulk density $\rho = 3$). Here derive the origin of this assumption. We first do this by considering single, unbonded beads of varying radius R_C , before discussing the implications of this for the situation of bonded beads with constant radius in our article.

2.1 Single beads of varying size

For single beads of varying size, the interaction potential U_{ij} between bead i and particle j is given by:

$$U_{ij} = \frac{1}{2}A_{ij}R_C \left(1 - \frac{r_{ij}}{R_C}\right)^2 \quad \text{for } r_{ij} < R_C \quad (9)$$

where A_{ij} is the interaction parameter and R_C is the cutoff radius. The total interaction potential U_i of i with its surrounding j particles is

$$U_i = \int_0^{R_C} U_{ij} \cdot \rho \cdot 4\pi r^2 dr \quad (10)$$

where ρ is the bulk density (particles per unit volume). Substituting U_{ij}

$$U_i = 2\pi\rho A_{ij}R_C \int_0^{R_C} \left(1 - \frac{r}{R_C}\right)^2 r^2 dr = \frac{2\pi\rho A_{ij}R_C^4}{30} \quad (11)$$

This means that when the cut-off is varying, the potential of an individual bead $U_i \propto R_C^4$. It is important to note that this expression represents the *average* potential a bead has with its surroundings. In reality the potential will fluctuate with different insertion locations due to the fluctuating density and distribution of particles it is surrounded with.

The excess chemical potential μ_{ex} is defined as:

$$\mu_{\text{ex}} = -\ln(\langle \exp(-\Delta U) \rangle). \quad (12)$$

We apply a cumulant expansion and this allows us to express μ_{ex} in terms of the mean and variance of ΔU . The first few terms of the cumulant expansion can be expressed as:

$$\ln\langle \exp(-\Delta U) \rangle \approx -\langle \Delta U \rangle + \frac{1}{2}\text{Var}(\Delta U) + \mathcal{O}(\langle \Delta U^3 \rangle). \quad (13)$$

If we briefly consider the case where $R_C = 1$, this allows us to understand Fig. 1 as presented in the main article and provides a clearer understanding of the non-linear relationship between A_{ij} and μ_{ex} . Although the term $\langle \Delta U \rangle$ scales linearly with the interaction parameter A_{ij} , the second term, $\text{Var}(\Delta U)$, scales with A_{ij}^2 . This leads to an increasing deviation from a linear relationship as A_{ij} becomes larger.

We now return to the case where we vary R_C . Therefore we expect that we can roughly say that we expect the resulting chemical potential $\mu_{\text{ex}} \propto \langle \Delta U \rangle$. and we can choose to estimate the relation between excess chemical potential and cut-off radius as $\mu_{\text{ex}} = CR_C^4$ where C is some constant that varies for different values of the interaction parameter a_{ij} . We test this relation by performing Widom insertion calculations of beads with $a_{ii} = 25, 35, 45$ and of varying size. In each case, the value of C can be determined numerically from when $R_C = 1$. The results of this simulation are plotted in Fig. S1, and also shown is the expected $\mu_{\text{ex}} = CR_C^4$ relationship. We see that there is extremely good matching between the predicted and simulated values. Of course μ_{ex} deviates slightly from the $\mu_{\text{ex}} = CR_C^4$ relation since the variance term is actually $\propto (\Delta U)^2$.

What we have shown is that, in this case, is $\mu_{\text{ex}} \propto R_C^4$ rather than $\propto R_C^3$ (i.e. the volume) as we have assumed in the main article, which is a result of the potential U_{ij} varying as a function of distance. However, we now consider the difference between this and the case of bonded beads of constant size, and the impact that this assumption has.

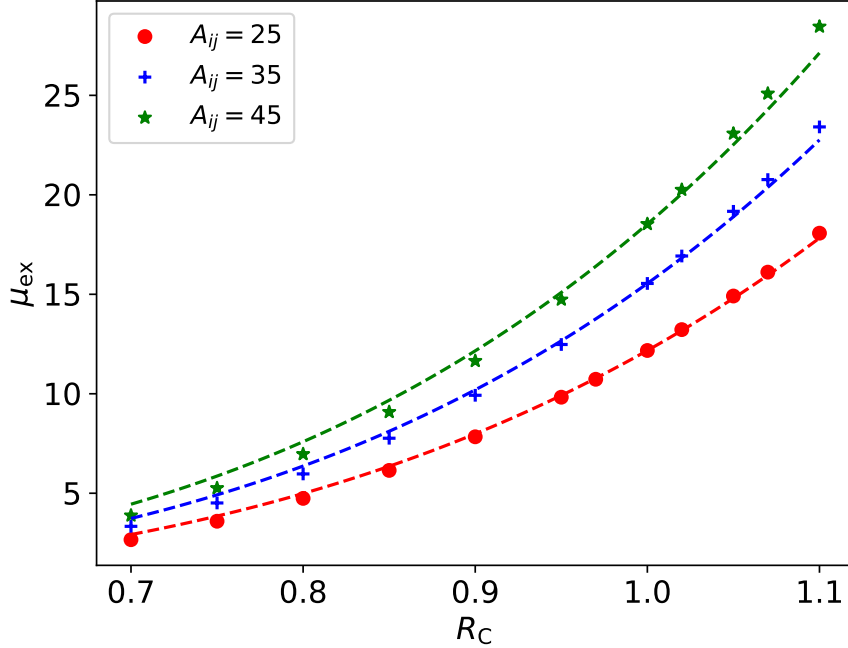


Fig. S1: Relationship between the excess chemical potential μ_{ex} and the bead cut-off R_C for single beads inserted into solvent with self-interaction $a_{ii} = 25$. Scatter points represent the simulated values calculated via Widom insertion, while the dashed line is the expected $\mu_{\text{ex}} = CR_C^4$ relationship (where $C = \mu_{\text{ex}}(R_C = 1)$).

2.2 Bonded beads

We now consider the case relevant to the simulations in the main article, where we have a series of bonded beads which are all of the same size R_C . In this case, the total volume of the inserted molecule is changing due to the insertion of multiple beads at once, rather than altering the size of the beads via the cut-off. If we first consider the situation in which multiple beads are simultaneously inserted but not overlapping, where each bead interacts with the solvent with interaction parameter A_{ij} . In this case, the average potential energy change ΔU_i is (on average) the same for every bead. This implies that μ_{ex} is proportional to the number of beads N_B at fixed radii R_C (i.e. N_B multiples of ΔU in Eq. 11). Now, when the beads are overlapping the situation is slightly more complicated.

If we consider the insertion of two bonded beads which have a region in which they overlap with volume V_{In} and a region where they don't V_{Out} . The number of solvent beads contained within each of those regions is expected to be proportional to their volume. However, the volume is not proportional to the total potential energy of this region because they contain beads which are of varying distances from the centre of the solute beads (and the potential varies with r_{ij} as shown in Eq. 9). This means that, theoretically, one should calculate ΔU for the total molecule for summing up the individual portions of ΔU_{In} and ΔU_{Out} , however, this is a complex calculation. Therefore, in this work, we aim to see if we can simply by scaling each other contributing regions by their volumes instead, neglecting this distance dependence on the potential energy.

Mathematically we now describe this situation. Suppose that, for a given system, the total potential energy change ΔU for inserting molecule i is proportional to the number of solvent beads it interacts with (upon insertion), multiplied by the average potential energy of an interaction. Mathematically, we can write:

$$\Delta U_i \propto N_{\text{beads}} \cdot \langle U_{\text{interaction}} \rangle \quad (14)$$

where N_{beads} is the number of solvent beads interacting with i and $\langle U_{\text{interaction}} \rangle$ is the average potential energy per interaction. Furthermore, we say that the number of beads N_{beads} is proportional to the volume V_i of the molecule, so $N_{\text{beads}} \propto V_i$. Thus, we can approximate ΔU_i as:

$$\Delta U_i \propto V_i \cdot \langle U_{\text{interaction}} \rangle. \quad (15)$$

Where we have here made the simplification to remove the distance dependence of U_{ij} between i and its

interacting particles. When considering a series of bonded beads, we can decompose the total potential energy change ΔU into two components:

- ΔU_{In} , which corresponds to the beads interacting with the molecule within the overlapping region.
- ΔU_{Out} , which pertains to the beads outside the overlap.

If ΔU_{In} and ΔU_{Out} are proportional to their respective volumes V_{In} and V_{Out} , we can express them as:

$$\Delta U_{\text{In}} = V_{\text{In}} u_{\text{In}} \quad \text{and} \quad \Delta U_{\text{Out}} = V_{\text{Out}} u_{\text{Out}} \quad (16)$$

where we now consider the volumes as scaled to the volume of a single bead (i.e the actual volume divided by $4\pi/3$) and u_{In} and u_{Out} are the corresponding energies of single beads. In this representation, we expect that the total excess chemical potential of the molecule could be calculated as

$$-\ln(\langle \exp(- (V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}})) \rangle) \quad (17)$$

Using a cumulant expansion, we get:

$$\begin{aligned} -\ln(\langle \exp(- (V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}})) \rangle) &\approx \langle V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}} \rangle \\ &\quad - \frac{1}{2} (\langle (V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}})^2 \rangle - \langle V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}} \rangle^2) + \dots \end{aligned} \quad (18)$$

Therefore, we have

- **First cumulant (mean):**

$$\langle V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}} \rangle = V_{\text{In}} \langle u_{\text{In}} \rangle + V_{\text{Out}} \langle u_{\text{Out}} \rangle \quad (19)$$

- **Second cumulant (variance):**

$$\langle (V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}})^2 \rangle - \langle V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}} \rangle^2 \quad (20)$$

Expanding the variance term produces

$$V_{\text{In}}^2 (\langle u_{\text{In}}^2 \rangle - \langle u_{\text{In}} \rangle^2) + V_{\text{Out}}^2 (\langle u_{\text{Out}}^2 \rangle - \langle u_{\text{Out}} \rangle^2) + 2V_{\text{In}}V_{\text{Out}} (\langle u_{\text{In}}u_{\text{Out}} \rangle - \langle u_{\text{In}} \rangle \langle u_{\text{Out}} \rangle) \quad (21)$$

Therefore, up to the second cumulant, we have

$$\begin{aligned} -\ln(\langle \exp(- (V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}})) \rangle) &\approx V_{\text{In}} \langle u_{\text{In}} \rangle + V_{\text{Out}} \langle u_{\text{Out}} \rangle \\ &\quad - \frac{1}{2} (V_{\text{In}}^2 \text{Var}(u_{\text{In}}) + V_{\text{Out}}^2 \text{Var}(u_{\text{Out}}) + 2V_{\text{In}}V_{\text{Out}} \text{Cov}(u_{\text{In}}, u_{\text{Out}})) \end{aligned} \quad (22)$$

We will assume that the covariance term is small compared to the variances, allowing us to focus primarily on the variance contributions, such that

$$-\ln(\langle \exp(- (V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}})) \rangle) \approx V_{\text{In}} \langle u_{\text{In}} \rangle + V_{\text{Out}} \langle u_{\text{Out}} \rangle - \frac{1}{2} (V_{\text{In}}^2 \text{Var}(u_{\text{In}}) + V_{\text{Out}}^2 \text{Var}(u_{\text{Out}})) \quad (23)$$

In our main article, we have essentially used the assumption that $\mu_{\text{ex}} \approx V_{\text{In}} \ln(\langle \exp(-u_{\text{In}}) \rangle) - V_{\text{Out}} \ln(\langle \exp(-u_{\text{Out}}) \rangle)$, rather than using $-\ln(\langle \exp(- (V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}})) \rangle)$, which would require the Widom insertion of the entire molecule to calculate. If we take this expression and each term, we have:

$$-V_{\text{In}} \ln(\langle \exp(-u_{\text{In}}) \rangle) - V_{\text{Out}} \ln(\langle \exp(-u_{\text{Out}}) \rangle) \approx V_{\text{In}} \langle u_{\text{In}} \rangle + V_{\text{Out}} \langle u_{\text{Out}} \rangle - \frac{1}{2} (V_{\text{In}} \text{Var}(u_{\text{In}}) + V_{\text{Out}} \text{Var}(u_{\text{Out}})) \quad (24)$$

Which differs from Eq. 23 in the scaling of the variance terms. The original expression can be rewritten as

$$\begin{aligned} -\ln(\langle \exp(- (V_{\text{In}} u_{\text{In}} + V_{\text{Out}} u_{\text{Out}})) \rangle) &\approx -V_{\text{In}} \ln(\langle \exp(-u_{\text{In}}) \rangle) - V_{\text{Out}} \ln(\langle \exp(-u_{\text{Out}}) \rangle) \\ &\quad + \frac{1}{2} (V_{\text{In}}(V_{\text{In}} - 1) \text{Var}(u_{\text{In}}) + V_{\text{Out}}(V_{\text{Out}} - 1) \text{Var}(u_{\text{Out}})) \end{aligned} \quad (25)$$

highlighting the difference between the calculation method used in our article, and the theoretical expression. We essentially assume that the final term is small, relative to the $-V_{\text{In}} \ln(\langle \exp(-u_{\text{In}}) \rangle) - V_{\text{Out}} \ln(\langle \exp(-u_{\text{Out}}) \rangle)$ part. In fact, $V(V - 1)$ is very small and negative when $V < 1$. Recalling that V_{In} and V_{Out} are the volume scaled to the volume of a single bead, neither V_{In} or V_{Out} should exceed a value of 1 until the bond length $0.7 < l_0$. This may explain why our correction method works less well for longer bond lengths (such as $l_0 = 1$), where this unaccounted-for term begins to grow.

3 Relation between μ_{ex} and a_{ij}

We highlighted in the main article, that the relationship presented between μ_{ex} and a_{ij} in Fig. 1 is only valid when $a_{ii} = 25$ and density $\rho = 3$. Here we show the relationship between μ_{ex} and a_{ij} for more solvent cases in Fig. S2.

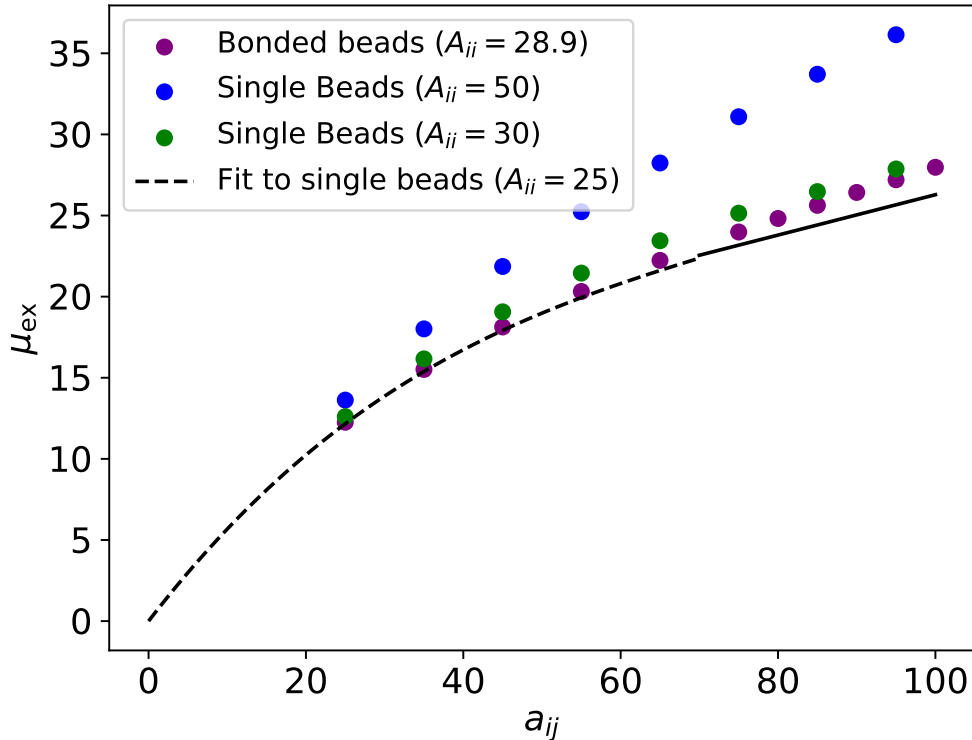


Fig. S2: Calculated relationship between the excess chemical potential μ_{ex} and interaction parameter a_{ij} for the Widom insertion into solvents with different a_{ii} . In the bonded case we use solvent molecules with length $N = 3$ and $l_0 = 0.6$. The fit presented by the black line is for the single bead cases with $a_{ii} = 25$, as shown in the main article.

We see that the divergence between different solvent cases grows with increasing a_{ij} . In particular, the difference between the bonded case at $a_{ii} = 28.9$ and unbonded case at $a_{ii} = 25$ is negligible for $a_{ij} < 45$. In our calculations to determine the partitioning of beads between bonded solvents we only ever use cross interactions of $a_{ij} \leq 45$, meaning we can consider it appropriate to use the fitted expression for single beads.

4 Simulation results for Fig. 5

The simulated values are calculated by fitting a straight line of best fit to the value of $\ln K$ as a function of concentration. We then extrapolate to zero concentration to find our final value for $\ln K$ at infinite dilution. Therefore, the standard deviation listed below is taken as the standard deviation of the y-intercept.

N	a_{AC}	a_{BC}	$\ln(K)$, uncorrected μ	l_0	$\ln(K)$, corrected μ	$\ln(K)$, simulated	Standard deviation
2	25	30	3.41	0.4	2.29	2.418	0.040
				0.6	2.52	2.787	0.007
				1	2.92	3.394	0.029
	30	32.5	1.56	0.4	1.03	1.038	0.026
				0.6	1.13	1.282	0.004
				1	1.32	1.425	0.039
	30	35	3.03	0.4	2.13	1.990	0.035
				0.6	2.31	2.354	0.012
				1	2.63	2.956	0.044
	30	37.5	4.41	0.4	2.97	2.956	0.028
				0.6	3.26	3.373	0.028
				1	3.77	4.248	0.033
32.5	40	4.15	0.4	2.77	2.556	0.038	
			0.6	3.05	3.131	0.022	
			1	3.54	3.946	0.223	
30	40	5.71	0.6	4.18	4.443	0.069	
30	31	0.635	0.6	0.462	0.488	0.012	
3	25	30	5.12	0.4	2.85	2.963	0.008
				0.6	3.31	4.016	0.098
	30	32.5	2.34	0.4	1.32	1.22	0.023
				0.6	1.51	1.631	0.068
				1	1.87	1.968	0.080
	30	35	4.54	0.4	2.69	2.452	0.050
				0.6	3.07	3.166	0.013
	31	35	3.59	0.4	2.15	1.980	0.037
				0.6	2.46	2.435	0.060
				1	2.98	3.113	0.129
	32.5	37	3.87	0.4	2.30	1.966	0.032
				0.6	2.60	2.880	0.077
1				3.16	3.526	0.037	
4	30	35	6.06	0.6	3.84	4.239	0.057
	32.5	36	4.06	0.6	2.62	2.407	0.045

5 Simulation results for Fig. 6

Similarly as described for Fig. 5, we find a standard deviation for each simulated value from its line of best fit. Here we show the standard deviation in brackets for each simulation.

N	a_{AC}	a_{BC}	l_0	Simulated $\ln(K)$ uncorrected solvent	Simulated $\ln(K)$, corrected solvent	Predicted $\ln(K)$
2	30	35	0.6	0.941 (0.004)	1.502 (0.016)	1.71
			1.0	0.929 (0.011)	1.281 (0.017)	
	25	40	0.6	3.905 (0.029)	4.574 (0.060)	4.56
			1.0	4.049 (0.014)	4.440 (0.046)	
	25	35	0.6	2.660 (0.011)	3.239 (0.013)	3.22
			1.0	2.701 (0.008)	3.040 (0.022)	
	30	40	0.6	2.259 (0.010)	2.860 (0.019)	2.86
			1.0	2.285 (0.031)	2.652 (0.041)	
	35	30	0.6	-2.084 (0.007)	-1.586 (0.017)	-1.71
			1.0	-2.190 (0.019)	-1.881 (0.016)	
	40	25	0.6	-5.038 (0.049)	-4.568 (0.031)	-4.56
			1.0	-5.312 (0.087)	-5.275 (0.029)	
4	30	35	0.6	0.615 (0.006)	1.419 (0.008)	1.71
			1.0	0.618 (0.010)	1.151 (0.021)	
	25	40	0.6	3.539 (0.017)	4.554 (0.046)	4.56
			1.0	3.720 (0.037)	4.391 (0.012)	
	25	35	0.6	2.295 (0.004)	3.133 (0.023)	3.22
			1.0	2.421 (0.012)	3.015 (0.023)	
	30	40	0.6	1.856 (0.010)	2.816 (0.026)	2.86
			1.0	1.929 (0.005)	2.471 (0.024)	
	35	30	0.6	-2.366 (0.029)	-1.659 (0.003)	-1.71
			1.0	-2.508 (0.021)	-2.022 (0.023)	
	40	25	0.6	-5.340 (0.048)	-4.781 (0.064)	-4.56
			1.0	-5.727 (0.064)	-5.212 (0.069)	

6 Bonding both the solute and solvent (section 4)

In this section, we expand what is discussed in section 4 of the main article, to study systems with both bonded solute and bonded solvent, combining the bonding effects discussed thus far. We do a number of tests to study two different lengths of the bonded solute ($N = 2$ and $N = 3$, represented by type C beads). In these simulations, one of the solvents is unbonded (type B), and the other (type A) is bonded into molecules of lengths $N = 2, 3$ and 4 beads long. As discussed above, we alter the self-interactions of the bonded beads accordingly, to match the desired pressure. We calculate the expected values of the chemical potential making use of our corrected equations. In this short study, we use a bond length $l_0 = 0.6R_C$ for all calculations.

Fig. S3 shows the value of $\ln(K)$ from simulation results, compared with the predicted value. Generally, we find reasonable agreement between the calculated and simulated values (noting that once

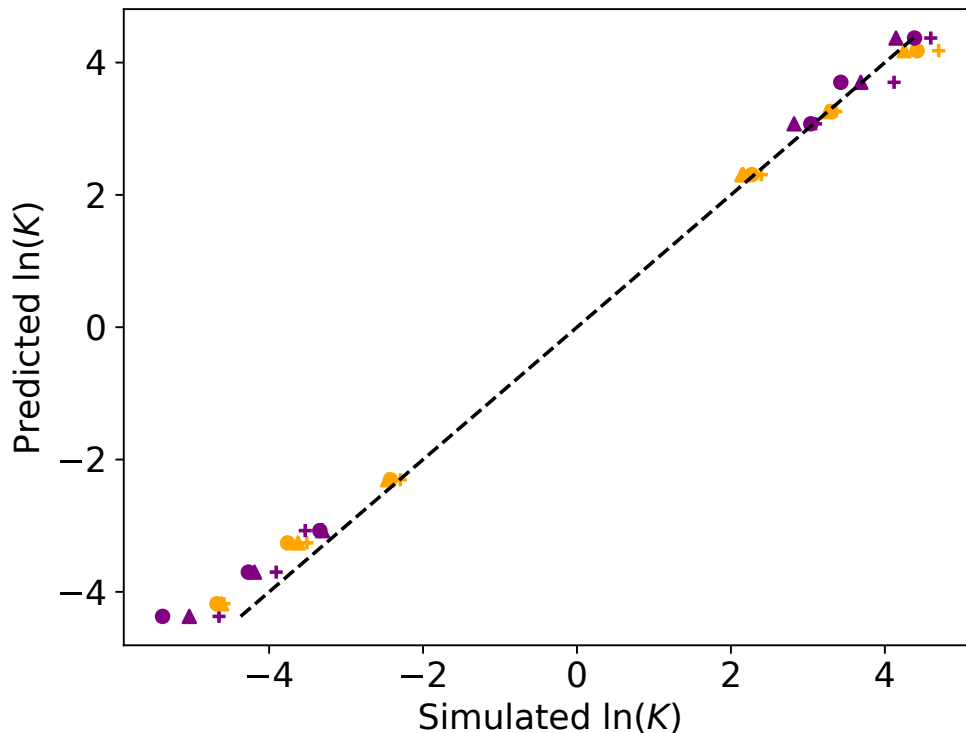


Fig. S3: Partitioning of bonded solutes into two solvents where one solvent is bonded. The length of the solute molecule is represented by colour where $N = 2$ (orange) and $N = 3$ (purple). The length of the solvent molecule is represented by symbols: 2 (+), 3 (O) and 4 (Δ)

again, all values shown in Fig. S3 can be found in the ESI). In all cases, K is defined as the ratio of the concentration in the bonded solvent to the concentration in the unbonded solvent ($K = c^A/c^B$), therefore we can also consider the impact of setting the $a_{AB} < a_{AC}$ compared with $a_{AC} < a_{AB}$. Note that, theoretically, the partitioning should be equivalent for, for example, when $a_{AB} = 25$ and $a_{AC} = 40$ compared with when $a_{AB} = 40$ and $a_{AC} = 25$. The mean squared error for different solute and solvent lengths is shown in Table S.1, highlighting that the error increases with molecular length for the solute.

N (solute)	N (solvent)	MSE
2	2	0.086
	3	0.096
	4	0.057
3	2	0.092
	3	0.25
	4	0.14

Table S.1: Mean squared error (MSE) in $\ln(K)$ for simulations of a bonded solute partitioning into two solvents where one solvent is bonded.

6.1 Simulation results for Fig. S3 (in ESI)

N (solute)	N (solvent)	a_{AC}	a_{AB}	$\ln K$ (simulated)	$\ln K$ (calculated)	
2	2	37.5	30	-3.51	-3.26	
		30	37.5	3.36	3.26	
		40	30	-4.59	-4.18	
		30	40	4.70	4.18	
		35	30	-2.30	-2.31	
		30	35	2.40	2.31	
	3	3	37.5	30	-3.76	-3.26
			30	37.5	3.30	3.26
			40	30	-4.68	-4.18
			30	40	4.42	4.18
			35	30	-2.42	-2.31
			30	35	2.27	2.31
	4	4	35	30	-2.46	-2.31
			30	35	2.14	2.31
			37.5	30	-3.62	-3.26
			30	37.5	3.26	3.26
40			30	-4.62	-4.18	
30			40	4.25	4.18	
3	2	36.2	30	-3.91	-3.70	
		30	36.2	4.12	3.70	
		37.5	30	-4.65	-4.37	
		30	37.5	4.60	4.37	
		35	30	-3.53	-3.07	
		30	35	3.10	3.07	
	3	3	36.2	30	-4.27	-3.70
			30	36.2	3.43	3.70
			37.5	30	-5.38	-4.37
			30	37.5	4.39	4.37
			35	30	-3.34	-3.07
			30	35	3.04	3.07
	4	4	36.2	30	-4.19	-3.70
			30	36.2	3.69	3.70
			37.5	30	-5.04	-4.37
			30	37.5	4.14	4.37
			35	30	-3.31	-3.07
			30	35	2.82	3.07

7 Bond lengths calculations (section 5)

Bonded interactions	N	a_{AC}	a_{BC}	Actual BL
Off	2	30	32.5	0.58218874141325
Off	2	30	35	0.5818081242215943
Off	3	30	32.5	0.5812385388202481
Off	3	31	35	0.5809933109424414
On	2	30	32.5	0.6149366221766263
On	2	30	37.5	0.6149742633567045
On	3	30	32.5	0.6146860837951208
On	3	31	35	0.614537005215961

8 Simulation data for Fig. 12

N	a_{AC}	a_{BC}	Original $\ln K$ (off)	$\ln K$ (on)
3	30	32.5	1.631 (0.068)	1.650 (0.062)
3	31	35	2.435 (0.060)	2.403 (0.070)
3	30	35	3.166 (0.013)	3.052 (0.556)
3	25	30	4.016 (0.098)	3.918 (0.069)
2	30	32.5	1.282 (0.004)	1.231 (0.044)
2	30	37.5	3.373 (0.028)	3.579 (0.023)
2	32.5	40	3.131 (0.022)	3.275 (0.055)
2	30	35	2.354 (0.012)	2.445 (0.016)
2	25	30	2.787 (0.007)	2.853 (0.021)

9 Pressure when bonded interactions turned on (section 5)

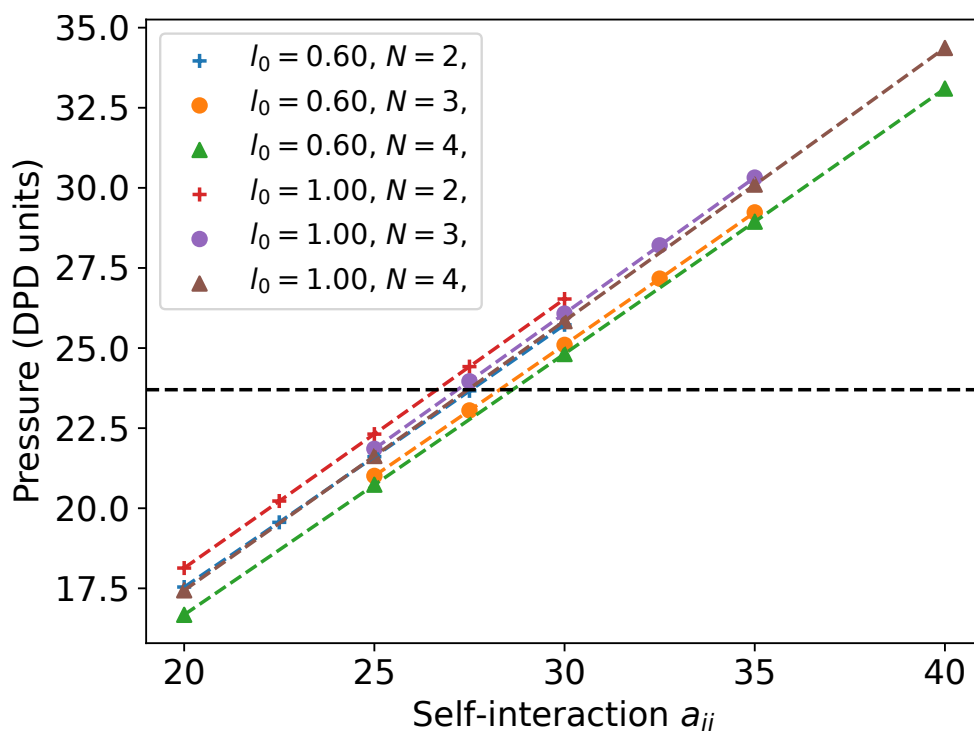


Fig. S4: Bulk pressure when bonded beads are allowed to interact, and where the bond length l_0 and the number of bonded beads N is varied. Also shown (black) is the bulk pressure of unbonded beads at the same number density.

10 Real Molecules

Note that the relation between infinite dilution activity coefficient (IDAC) and $\Delta\mu_{\text{ex}}$ is provided in our previous work,³ and this relationship will be used here to define the cross interaction a_{ij} parameters. The coarse-graining description of all molecules is provided in the main article. In all cases, we need to define the interaction of the solute bead with water and dodecane. The interaction of water and dodecane themselves is simply defined as $a_{ij} = 100$, which is just chosen to correctly reproduce the phase separation of oil and water, and produce a negligible solubility of one within the other. The self-interaction of the solute itself and the self-interaction of dodecane is defined using the ‘pressure matching’ approach described in the main article.

10.0.1 Diethyl carbonate

We use experimental IDAC values to define the interactions of the diethyl carbonate beads with water and dodecane. First, we use the IDAC values to calculate a value for $\Delta\mu_{\text{ex}}$, which is defined as the difference in excess chemical potential of the solute bead at infinite dilution in a solvent, and the excess chemical potential of the solute in a bath of its pure state (see reference³ for more details). The IDAC for diethyl carbonate in water at room temperature is $\gamma = 374$,² meaning we calculate $\Delta\mu_{\text{ex}} = 4.06$. The reported value¹ for the partitioning of diethyl carbonate between water and dodecane is $\ln K = 1.34$, implying that for diethylene carbonate in dodecane $\Delta\mu_{\text{ex}} = 2.72$.

These $\Delta\mu_{\text{ex}}$ values allow us to calculate the cross-interaction parameters of diethyl carbonate with dodecane a_{CD} and diethyl carbonate with water a_{CW} . When no corrections are applied to account for the overlap of bonded bead (what we refer to as ‘uncorrected’ values), we calculate $a_{\text{CW}} = 32.13$ and $a_{\text{CD}} = 34.39$. However, using our equations from the main article we calculate ‘corrected’ a_{ij} parameters (i.e., accounting for the overlap) as $a_{\text{CW}} = 33.9$ and $a_{\text{CD}} = 37.0$.

10.1 Heptanol

Calculating the interaction values for heptanol is slightly more difficult, due to the presence of multiple bead types in the solute. We use a combination of experimental values to find the target excess chemical potentials μ_{ex} for each interaction type. The experimental data is shown in Fig. S5. The target excess

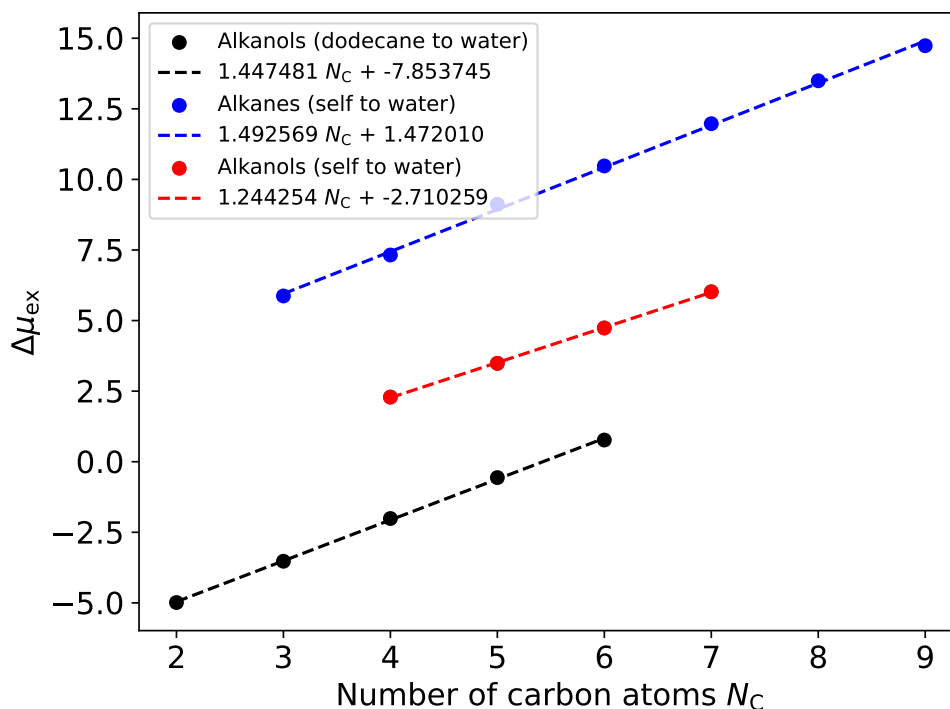


Fig. S5: The experimental data used to calculate the interaction parameters for heptanol in water and dodecane. Data obtained from the infinite dilution activity coefficients of alkanes in water (blue)¹ and alkanols in water (red),⁴ and also the partitioning of alkanols in dodecane/water (black).⁶

chemical potential difference per bead using this data is presented below, where values are presented in units of $k_B T$. Here we have used a combination of extrapolation and gradients from lines of best fit to determine these values.

Bead Name	C3T	C3	COH
Water	5.078448	4.342443	-1.998783
Dodecane	0	0	5.143486

When no corrections are applied to account for the overlap of bonded beads, the parameters are calculated as shown.

Bead Name	C3T	C3	COH	Water	Dodecane
C3T	25	-	-	-	-
C3	25	25	-	-	-
COH	25	25	25	-	-
Water	42.0865	39.1341	19.8074	25	-
Dodecane	25	25	42.3575	100	25

When the overlap of the solute is corrected for, this increases the self-interaction.

Bead Name	C3T	C3	COH	Water	Dodecane
C3T	28.9	-	-	-	-
C3	28.9	28.9	-	-	-
COH	28.9	28.9	28.9	-	-
Water	48.0852	44.7092	23.2402	25	-
Dodecane	28.9	28.9	48.3962	100	29.4

And finally, also correcting for the overlap of the solute as well results in the following final a_{ij} values.

Bead Name	C3T	C3	COH	Water	Dodecane
C3T	28.9	-	-	-	-
C3	28.9	28.9	-	-	-
COH	28.9	28.9	28.9	-	-
Water	60	60	20	25	-
Dodecane	28.9	24.5	60	100	29.4

The exact excess chemical potential for each bead is difficult to match for exactly, due to the large degree of overlap. The estimated excess chemical potential of each bead in both dodecane and water is shown below.

	C3T	C3	COH	Total	Experimental Total
Water	5.53569353	2.37108002	-0.4789928	7.42778	7.42211
Dodecane	-0.22311797	1.07678837	4.2910837	5.14475	5.143486

Our correction tries to capture the variation of excess chemical potential across the molecule, while making sure to match the total excess chemical potential expected for the molecule.

10.2 Benzene

The self-interaction term for ring-like molecules is not presented in the main article. Therefore, we determine the value of this by performing a similar procedure to the other cases, where we find the a_{BB} value required to generate the correct value for the pressure (i.e., the pressure for single beads). Using the bond length and angle potentials described for benzene in the main article, this is determined to be $a_{BB} = 49.7$.

Once again, we use experimental infinite dilution activity coefficient values to define the interactions of the benzene beads with water and dodecane. The IDAC for benzene in water is experimentally reported as $\gamma = 2400$,⁴ and therefore we determine that $\Delta\mu_{\text{ex}} \approx 6.18$. For benzene in dodecane, experimentally $\gamma = 1.3$, and therefore $\Delta\mu_{\text{ex}} = 1.20$.⁵ If we assume that the total excess chemical potential of the ring is simply the sum of the 3 bead excess chemical potentials (no overlap correction), we calculate the a_{ij} values as $a_{BD} = 51.8$ and $a_{BW} = 61.2$

In order to set the a_{ij} values accounting for the overlap, we need to know the volumes V_1 , V_2 and V_3 . Once again, this is not presented in the main article for rings, as the analytical function would be quite complex for their configuration. Therefore, instead we do this by performing a Monte Carlo calculation and determining the volumes as a function of the bond length. In this simulation we simply define the beads in a cuboid simulation domain, and then we generate random points within this space, determining the ones which are generated within the radius beads (or within multiple beads). The volume is calculated using the fraction of points that lie inside the bead’s spheres. The results of these calculations are shown in Fig. S6. Having determined V_1 , V_2 and V_3 , this allows us to determine

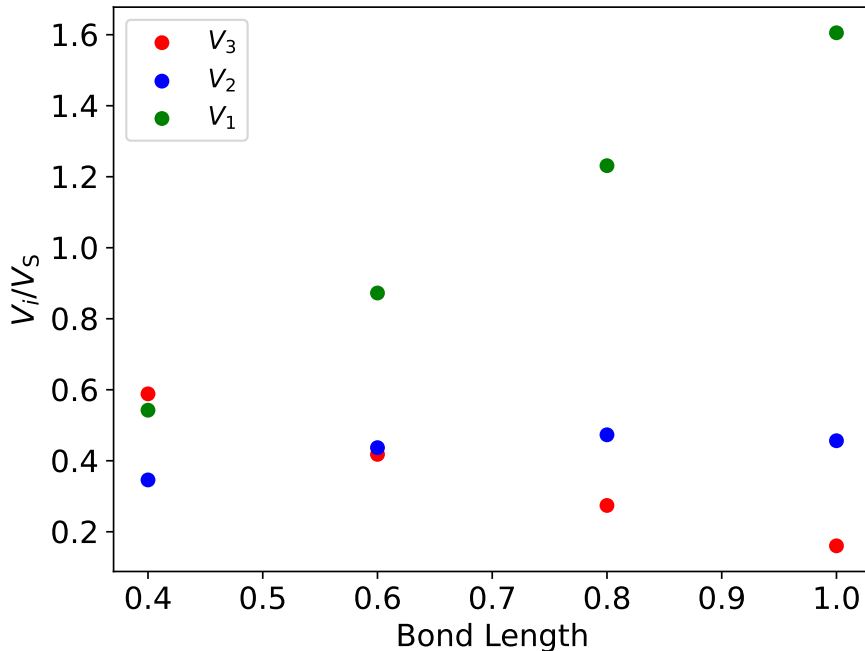


Fig. S6: The relationship between the volumes of overlap V_i (relative to the volume of a single bead) and the bond length in a triangular formation of beads. These volumes are calculated from a simple Monte Carlo simulation.

corrected a_{ij} parameters as $a_{BD} = 52.6$ and $a_{BW} = 65.1$.

However, we note that these a_{ij} values are too high for the μ_{ex} vs. a_{ij} relationship to hold as it is presented in the main article (Eq. 17). Differences between this relationship and bonded cases can emerge, particularly when a_{ij} is larger. Our a_{ij} are relatively large due to the fact that the benzene self-interaction is also large. This is a result of the short distance bonds in the molecule, and large degree of overlap. Since the benzene-water and benzene-dodecane interactions are set relative to this self-interaction, this produces large a_{ij} values for all cross interactions.

However, in order to investigate whether the correct partitioning of rings is produced, we can simply change the reference interaction. The partitioning behaviour is actually only related to the difference in excess chemical potential of the benzene-water and benzene-dodecane interactions, i.e. it depends on $\mu_{\text{ex}}^{\text{BD}} - \mu_{\text{ex}}^{\text{BW}}$ (see Eq. 27 in the main article), rather than the exact values themselves. Therefore, for the study shown here, we choose to define the benzene-dodecane as $a_{BD} = 25$, and adjust the a_{BW} interaction accordingly, to generate the correct $\mu_{\text{ex}}^{\text{BD}} - \mu_{\text{ex}}^{\text{BW}}$. We note once again, that this would produce the wrong behaviour between benzene and dodecane relative to pure benzene, however would still reproduce the correct partitioning behaviour. If one wanted to define this system fully (for studying something other than partitioning behaviour), one should use a μ_{ex} vs. a_{ij} relationship calculated at higher a_{ij} values for a bonded solute. However, as we are interested in studying the partition behaviour (at infinite dilution) only, the choice to set $a_{BD} = 25$ has no impact on the results.

Therefore we recalculate the ‘uncorrected’ cross-interaction a_{ij} parameters to be $a_{BD} = 25$ and $a_{BW} = 29.86$. Correcting for the overlap similarly produces $a_{BD} = 25$ and $a_{BW} = 34.7$. This allows us to study the partitioning behaviour of rings without needing to perform additional simulations to those already presented here and those in the main article.

References

- ¹ M. H. Abraham and W. E. Acree Jr. Correlation and prediction of partition coefficients between the gas phase and water, and the solvents dodecane and undecane. *New J. Chem.*, 28:1538–1543, 2004.
- ² V. Dohnal, P. Vrbka, K. Řehák, A. Böhme, and A. Paschke. Activity coefficients and partial molar excess enthalpies at infinite dilution for four esters in water. *Fluid Phase Equilibria*, 295(2):194–200, 2010.
- ³ R. L. Hendrikse, C. Amador, and M. R. Wilson. Many-body dissipative particle dynamics simulations of micellization of sodium alkyl sulfates. *Soft Matter*, 20:6044–6058, 2024.
- ⁴ K. Kojima, S. Zhang, and T. Hiaki. Measuring methods of infinite dilution activity coefficients and a database for systems including water. *Fluid Phase Equilibria*, 131(1):145–179, 1997.
- ⁵ T. Letcher and W. Moollan. The determination of activity coefficients at infinite dilution using g.l.c. with a moderately volatile solvent (dodecane) at the temperatures 280.15 k and 298.15 k. *The Journal of Chemical Thermodynamics*, 27(9):1025–1032, 1995.
- ⁶ M. Manabe, M. Koda, and K. Shirahama. Partition Coefficients of Alkanols and Polyoxyethylene Alkyl Ethers in the Dodecane–Water System at 25 °C. *Bulletin of the Chemical Society of Japan*, 48(12):3553–3556, 03 2006.