This version is available at https://strathprints.strath.ac.uk/56683/

# Predicting solvation free energies using parameter-free solvent models

Maksim Misin,[†] David S. Palmer,[*,‡] and Maxim V. Fedorov[*,†,¶]

*Department of Physics, SUPA, University of Strathclyde, 107 Rottenrow, Glasgow, G4 0NG, UK, and Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow, G1 1XL, UK*

E-mail: david.palmer@strath.ac.uk; maxim.fedorov@strath.ac.uk

---

[*]To whom correspondence should be addressed
[†]Department of Physics, University of Strathclyde
[‡]Department of Pure and Applied Chemistry, University of Strathclyde
[¶]Skolkovo Institute of Science and Technology, 3 Nobel Street, Moscow 143026, Russian Federation

**Abstract**

We present a new approach for predicting solvation free energies in non-aqueous solvents. Utilizing the corresponding states principle, we estimate solvent Lennard-Jones parameters directly from their critical points. Combined with atomic solutes and pressure corrected three-dimensional reference interaction site model (3D-RISM/PC+), the model gives accurate predictions for a wide range of non-polar solvents, including olive oil. The results, obtained without electrostatic interactions and with a very coarse-grained solvent provide an interesting alternative to widely used and heavily parametrized models.

# Introduction

Solvation free energy is one of the key molecular properties, related to partition coefficients, solubilities, and chemical reactivity[1–3]. Its practical applications in both bioscience and industry have stimulated research in the field for many years[4,5]. Generally, the more sophisticated a representation of a solvent is, the more accurate solvation free energies one can obtain, at the cost of speed[6]. However, for many solvents, one often can get reliable results using relatively simple implicit solvent models[7–10].

Still, the dependence of solvation free energy on temperature, concentration, co-solvent effects and other variables cannot be obtained purely from implicit continuum models, and require at least a coarse-grained description of solvent. After many efforts directed at improving the performance of existing force fields and charge schemes in aqueous systems, we know that molecular dynamics (MD) simulations can predict hydration free energies of various compounds in water quite well[11,12]. At the same time, a recent study suggests that molecular simulations with generic force fields are incapable of predicting solvation thermodynamics in non-aqueous systems with comparable accuracy.[13] A few studies that have shown promising results relied on specifically parametrized force fields, which in general require a lot of effort to develop[14–16].

The three-dimensional reference interaction model (3D-RISM) describes solvation in terms of solute-solvent and solvent-solvent correlation functions. This allows one to rapidly calculate a number of solvation phenomena as irrelevant degrees of freedom are averaged out[1]. Recently, it was shown to predict solvation free energies in water with the same accuracy as MD[17,18]. Despite that, the extension of this method to non-aqueous solvents is difficult due to the convergence issues frequently encountered for solvents with a large number of atoms, as well as the previously mentioned lack of reliable force fields.

In this Article, we propose a method for estimation of solvation free energies in non-polar or weakly polar liquids. Various solvents are approximated as Lennard-Jones spheres or chains, with parameters deduced from their experimental critical points. This makes the presented method applicable to any liquid with known critical parameters and also solves numerical problems frequently encountered in 3D-RISM calculations. We demonstrate the utility of our approach by using it to predict solvation free energies in a variety of pure solvents and also, in olive oil. The model predictions are compared to experimental measurements: for apolar solvents root mean square error is around $1\,\mathrm{kcal/mol}$.

# Models

## Solvents

According to the corresponding states principle, reduced critical temperature

$$T_c^* = \frac{kT_c}{\epsilon} \tag{1}$$

and reduced critical density

$$\rho_c^* = \rho_c \sigma^3 \tag{2}$$

are constants for all classical fluids with orientation independent interaction potentials[19,20]. Here $k$ is Boltzmann constant, $T_c$ and $\rho_c$ are critical temperature and density, $\sigma$ is effective

particle diameter and $\epsilon$ is a constant that determines the strength of intramolecular inter-actions. This principle can be further extended to non-spherical molecules by assuming $T_c^*$ and $\rho_c^*$ are functions of molecular shape and electrostatic properties[19].

It follows that knowing $T_c^*$ and $\rho_c^*$ for a single reference fluid, one can easily obtain intermolecular interaction parameters $\epsilon$ and $\sigma$ for many others from their critical properties. This idea has been used by a number of authors to construct coarse-grained models of real fluids and estimate their properties at a wide range of conditions[21–31]. Most of them came to the conclusion that with the exception of a few simple fluids such as argon, nitrogen, or methane, the majority of the real fluids cannot be adequately described by just two simple parameters and require either additional fittings or more complicated interaction potentials.

However, a precise description of solvent behaviour and phase diagram is frequently not necessary for an accurate estimation of solvation free energy, as can be seen by a number of successful implicit solvation models[7,10]. Statistical mechanics analysis of solvation shows that the free energy of the process does not include contributions due to the solvent reorga-nization and only depends on the solute-solvent interaction energy and its fluctuations[32–34]. Therefore, one would expect that a coarse-grained model of solvent that captured the key solute-solvent interactions would be sufficient to obtain a reliable estimate.

Here we test the simplest solvent models that contain no fitted parameters. Liquids were approximated as spheres interacting via conventional Lennard-Jones potential:

$$U(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \right] \tag{3}$$

To obtain potential parameters, we used equations 1 and 2 as well as critical values for Lennard-Jones fluid obtained by Okumura et al.: $T_{c,LJ}^* = 1.313$, $\rho_{c,LJ}^* = 0.304$[35]. Critical properties of real liquids were taken from Ref. 36. The list of values used is given in table 1. Note that the solvent xylenes is a mixture of isomeric ortho-, meta- and para-forms of xylene.

Table 1: Critical properties and Lennard-Jones parameters.

| Name | $T_c$ [K] | $\rho_c$ [nm$^{-3}$] | $\epsilon$ [kcal/mol] | $\sigma$ [nm] |
|---|---|---|---|---|
| 1,2-dichloroethane | 561.60 | 2.74 | 0.85 | 0.48 |
| acetonitrile | 545.00 | 4.11 | 0.82 | 0.42 |
| benzene | 562.05 | 2.35 | 0.85 | 0.51 |
| bromobenzene | 670.15 | 1.86 | 1.01 | 0.55 |
| carbon disulfide | 552.00 | 3.76 | 0.84 | 0.43 |
| carbon tetrachloride | 556.35 | 2.18 | 0.84 | 0.52 |
| chloroform | 536.40 | 2.52 | 0.81 | 0.49 |
| cyclohexane | 553.80 | 1.96 | 0.84 | 0.54 |
| diethyl ether | 466.70 | 2.15 | 0.71 | 0.52 |
| dimethyl sulfoxide | 729.00 | 2.65 | 1.10 | 0.49 |
| ethyl acetate | 523.30 | 2.11 | 0.79 | 0.52 |
| isooctane | 543.80 | 1.29 | 0.82 | 0.62 |
| isooctane (2-mer) | | | 0.61 | 0.49 |
| n-decane | 617.70 | 1.07 | 0.93 | 0.66 |
| n-decane (4-mer) | | | 0.54 | 0.39 |
| n-heptane | 540.20 | 1.41 | 0.82 | 0.60 |
| n-heptane (3-mer) | | | 0.52 | 0.40 |
| octanol | 652.50 | 1.21 | 0.99 | 0.63 |
| toluene | 591.75 | 1.91 | 0.90 | 0.54 |
| xylenes | 624.57 | 1.59 | 0.95 | 0.58 |
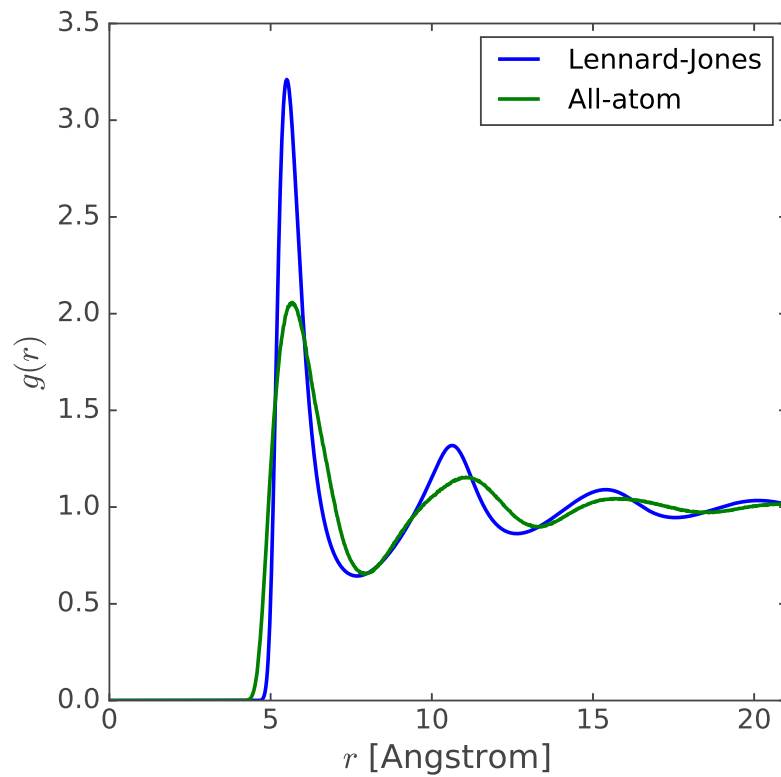
Figure 1: The radial distribution function of $CCl_4$ from HNC 1D-RISM calculation (shown in blue) and all-atom MD simulation (shown in green). MD radial distribution function was computed between centres of mass of molecules.

Besides potential energy between solute and solvent, 3D-RISM calculations also require a density of the solvent and it's radial distribution functions as an input. In this study, the densities were taken from experimental measurements found in Ref. 36. The radial distribution function was calculated using the 1D-RISM method with hypernetted-chain (HNC) closure (note that in the case of a single site solvent, the 1D-RISM model reduces to Ornstein-Zernike equations). For comparison, we also performed molecular dynamics simulations of bulk atomistic solvents. Overall, the structure of radial distribution functions of Lennard-Jones liquids is quite similar to that of their atomistic analogues, however, the height of the peaks is somewhat different. An example is shown in figure 1, more comparisons are given in the supporting information.

Of course, a spherical Lennard-Jones fluid is a poor reference system for most of these liquids. For this reason, isooctane, heptane, and decane were also modelled as chains of Lennard-Jones spheres composed of $m$ segments, each separated by a bond of length $\sigma$. The choice of $m$ was motivated by an equation employed in Statistical Associating Fluid Theory (SAFT) and in some molecular dynamics studies[30,37,38]:

$$m = 1 + \frac{n(C) - 1}{3} \tag{4}$$

where $n(C)$ is the number of carbons in the linear alkane. $m = 3$ for heptane and $m = 4$ for decane follow directly from the equation. We also assumed that $m = 2$ would be a reasonable choice for isooctane. The $\sigma$ and $\epsilon$ parameters for chain beads were similarly obtained using equations 1 and 2, but using critical points for 2-mer ($T^*_{c,LJC2} = 1.78$, $\rho^*_{c,LJC2} = 0.149$, Ref. 39), 3-mer ($T^*_{c,LJC3} = 2.063$, $\rho^*_{c,LJC3} = 0.088$, Ref. 40) and 4-mer ($T^*_{c,LJC4} = 2.26$, $\rho^*_{c,LJC4} = 0.0625$, Ref. 41) Lennard-Jones chain fluids.

We were also interested in applying this approach to solvent mixtures. Unfortunately, we couldn't find a collection of systematic measurements of solvation free energies (or any related quantities) for mixtures of the solvents listed in Table 1. The closest analogue we

could find, for which plenty of experimental data is available, was olive oil.

The major components of olive oil are triglyceride esters of various fatty acids, with exact composition dependent on the cultivar, region and the time of harvest[42,43]. Experimental studies containing olive oil-gas partition coefficients often do not measure the precise composition of employed olive oil[44–46]. Here we assumed that all the experimental measurements were performed in the oil produced from Picual cultivar, one of the most popular varieties in the world[47]. It contains 15% of saturated fatty acids (mostly palmitic acid), 81% of monounsaturated fatty acids (mostly oleic acid) and 4% of polyunsaturated acids (mainly linoleic acid)[48].

**Table 2: Composition of model olive oil.**

| Component | % | $\epsilon$ [kcal/mol] | $\sigma$ [nm] |
|---|---|---|---|
| palmitic acid (6-mer) | 15 | 0.58 | 0.41 |
| oleic acid (6-mer) | 81 | 0.62 | 0.44 |
| linoleic acid (6-mer) | 4 | 0.54 | 0.40 |

We approximated olive oil as a mixture of three 6-mer Lennard-Jones chains, representing fatty acid esters found in an olive oil. The number of beads was again motivated by equation 4. We did not model whole triglycerides (3 fatty acids linked by glycerol) as there are no critical parameters for these type of molecule as well as the convergence of such liquids in 1D-RISM (and obtaining their site-site radial distributions in MD) is rather problematic. The critical parameters for chains were interpolated from the available literature data: $T^*_{c,LJC6} = 2.486$, $\rho^*_{c,LJC6} = 0.03837$ (details are provided in the supporting information). The Lennard-Jones parameters for the beads were obtained using equations 1 and 2. The critical density for the esters was assumed to be the same as the critical density of the corresponding acids. However, one would expect esters found in oil to be less polar than their acid analogues, making fatty acid bulk critical temperatures a poor choice for computing Lennard-Jones $\epsilon$. Using Joback's group contribution method[49], we calculated the change in critical temperature occurring when a carboxylic group is substituted with ester and added

8

the difference to the critical temperatures of pure acids. Critical parameters for palmitic, oleic and linoleic acids were taken from Ref. 36. The final Lennard-Jones parameters and composition of the model olive oil is given in Table 2.

## Solutes and solvation

We approximated solutes as rigid molecules with non-bonded parameters described using the OPLS-2005 force field[50]. We note that the partial charges on solutes were unnecessary as solvents were neutral.

The solvation free energy was evaluated using a recently developed, pressure corrected 3D-RISM free energy functional: 3D-RISM/PC+[18,51], previously referred to as 3D-RISM/ISc[17]. This computation method has been successfully used to predict hydration free energies of a variety of compounds at different temperatures[14,17,18]. The resulting free energies were close to those obtained with thermodynamic integration, at minimal computational cost. Our group maintains a script to simplify solvation free energy computations using this model at https://github.com/MTS-Strathclyde/PC_plus.

The details and derivation of the pressure correction (PC+), and 3D-RISM in general have been described in a number of previous publications and will be only briefly summarized here[1,51,52]. The distribution of solvent density around solute in 3D-RISM is obtained by iteratively solving Ornstein-Zernike-like equation:

$$h_\alpha(\mathbf{r}) = \sum_{\beta=1}^{n_s} (\chi_{\alpha\beta} * c_\beta)(\mathbf{r}), \tag{5}$$

where subscripts $\alpha$ and $\beta$ denote indexes of sites in solvent molecule, $n_s$ is the total number of sites in solvent molecule, $h_\alpha$ is the total correlation function, $c_\alpha$ is the direct correlation function, and the star $*$ denotes convolution[1,18]. Solvent susceptibility functions are given by $\chi_{\alpha\beta}(r) = \omega_{\alpha\beta}(r) + \rho_\beta h_{\alpha\beta}(r)$, where $\omega$ is the bulk solvent intramolecular correlation function and $h_{\alpha\beta}$ is the bulk solvent total correlation function, both obtained from a preliminary

9

1D-RISM computation.

To solve the above equation, it must be coupled with an appropriate approximate closure. One of the better-studied approximations is called hypernetted-chain closure[53], or HNC

$$h_\alpha(\mathbf{r}) + 1 = \exp\left(-\frac{u_\alpha(\mathbf{r})}{kT} + h_\alpha(\mathbf{r}) - c_\alpha(\mathbf{r})\right) \tag{6}$$

where $k$ is the Boltzmann constant, $T$ is temperature, and $u_\alpha$ is the potential energy between solute and solvent site $\alpha$, provided as an input. The exponentiation in HNC closure can sometimes lead to poor convergence. Often, much better convergence can be achieved using its partial series expansion (PSE). Expansion up to the third term (PSE-3) often shows good agreement with HNC[18,54] and we included results obtained with PSE-3 in the supporting information.

Equations 5 and 6 are solved iteratively until both $c$ and $h$ converge to a given threshold. The solvation free energy is then obtained from

$$\Delta G_{3D-RISM} = kT \sum_{\alpha=1}^{n_s} \rho_\alpha \int_V \left[\frac{1}{2}h_\alpha^2(\mathbf{r}) - c_\alpha(\mathbf{r}) - \frac{1}{2}c_\alpha(\mathbf{r})h_\alpha(\mathbf{r})\right] d\mathbf{r} \tag{7}$$

The above expression, while exact in HNC framework, overestimates solvation free energies of non-polar species. It has been speculated[51] that this overestimation is largely due to incorrect solvent pressure in HNC, which is also much larger than experimental.

In pressure corrected model (PC+), the solvation free energy is given by

$$\Delta G_{PC+} = \Delta G_{3D-RISM} - P_{3D-RISM}\Delta V + \rho kT\Delta V \tag{8}$$

where $\rho$ is solvent density and $\Delta V$ is solute partial molar volume, also obtained from 3D-RISM. The pressure, $P_{3D-RISM}$, is computed using the following equation

$$P_{3D-RISM} = \frac{n_s + 1}{2}\rho kT - \frac{kT}{2}\sum_{\alpha=1}^{n_s}\sum_{\beta=1}^{n_s}\rho_\alpha\rho_\beta\hat{c}_{\alpha\beta}(k=0) \tag{9}$$

10

where the $\hat{c}_{\alpha\beta}(k = 0)$ is the integral of the solvent-solvent direct correlation function.

The extra ideal gas expansion work, $\rho kT \Delta V$, which is added alongside $-P_{3D-RISM}\Delta V$ in equation 8, has been somewhat controversial as it cannot be explained purely by theory[51]. However, for water, it was shown to significantly improve results and thus, we kept it for non-aqueous solvents as well[17,18]. We have also computed all our results without this term, but it should be noted that it actually does not make that much of a difference, as liquids studied in this Article have low densities (for comparison, PC results are provided in the Supporting information).

## Calculation details

The optimized geometries of solutes in pure solvents were taken from the Minnesota solvation database[7,55], in which all values are reported for standard temperature and pressure (298.15 K, 1 bar). Note that two molecules: 5-fluorouracil and 5-chlorouracil were excluded from the dataset: for them the average error in solvation free energies across multiple solvents was 7 kcal/mol: much larger than the average errors for other solutes. The solvation free energies of molecules in olive oil were taken from the Ref. 46, in which values are reported for 310 K. For this dataset, the initial solute geometry was generated using Open Babel "gen3d" method[56] and subsequently optimized using Sander[57] and OPLS-2005 force field. We estimated dipole moment of solutes using PM6 method[58] as implemented in Gaussian 09[59].

The majority of solutes in both datasets were organic compounds consisting of 5 – 10 heavy atoms. Besides carbon oxygen and nitrogen, many solutes contained phosphorus, sulfur, and halogens. In total, both datasets had 482 unique compounds and 1247 experimental measurements. The OPLS-2005 force field parameters were assigned automatically to solutes using Maestro[60]. For noble gases, force field parameters were taken from Ref. 61. The Lennard-Jones interaction between solute and solvent was computed using Lorentz-Berthelot

mixing rules[62].

3D-RISM calculations were performed using the rism3d.snglpnt program from Amber-Tools 15 package[63–65]. Similarly to our setup in a previous publication[18], the grid spacing was set to $0.5\,\text{Å}$, the buffer to $25\,\text{Å}$, and tolerance to $1 \times 10^{-5}$. Solvent susceptibility functions (density-density correlations of bulk solvent) required for running 3D-RISM were generated using the rism1d program, which is also included in AmberTools 15 package. Both 1D-RISM and 3D-RISM calculations were performed using HNC closure. Pure solvent densities at $298\,\text{K}$ were taken from Ref. 36. For olive oil, the density at $310\,\text{K}$ was computed using an interpolation relationship[66]. The 1D-RISM equations were solved using HNC closure, with tolerance set to $1 \times 10^{-12}$ and grid spacing to $0.025\,\text{Å}$. As all particles were electrically neutral, setting dielectric constant was not necessary.

## Results and Discussion

A comparison between predicted and experimental solvation free energies is presented in Table 3. This table contains the main results of this Article. Examining accuracies across the solvents, one can see a clear correlation between the validity of assumptions used in the corresponding state principle, and the accuracy of 3D-RISM/PC+ predictions for a given solvent. We split all studied solvents into polar and apolar groups based on their polarity and the ability to specifically interact with the solutes.

Solvation free energy in apolar compounds is predicted with $1.1\,\text{kcal/mol}$ accuracy (omitting single sphere model results for isooctane, heptane, and decane). If one doesn't take into account olive oil, which has an uncertain composition and was approximated rather crudely, the root means squared error drops to $0.98\,\text{kcal/mol}$.

Using a single Lennard-Jones sphere to model solvent molecules, the most accurate predictions were obtained for carbon tetrachloride and carbon disulfide. Both of these compounds have only small orientational correlations[67,68] and can be approximated by the spherical

**Table 3: Accuracies of solvation free energy predictions by 3D-RISM/PC+ for various solvents. RMSE stands for root mean square error, SDE is standard deviation of error. Energies are in kcal/mol.**

| Solvent | N | RMSE | SDE | bias |
|---------|---|------|-----|------|
| Apolar | | | | |
| 1,2-dichloroethane | 39 | 1.16 | 1.07 | 0.47 |
| benzene | 71 | 1.28 | 1.28 | 0.04 |
| bromobenzene | 27 | 1.17 | 1.15 | -0.23 |
| carbon disulfide | 15 | 0.94 | 0.89 | -0.30 |
| carbon tetrachloride | 79 | 0.85 | 0.84 | -0.11 |
| cyclohexane | 103 | 1.01 | 0.75 | -0.67 |
| isooctane | 32 | 0.98 | 0.68 | -0.70 |
| isooctane (2-mer) | 32 | 0.63 | 0.60 | -0.21 |
| n-decane | 39 | 1.70 | 1.23 | -1.17 |
| n-decane (4-mer) | 39 | 0.68 | 0.56 | -0.38 |
| n-heptane | 67 | 0.95 | 0.86 | -0.42 |
| n-heptane (3-mer) | 67 | 0.74 | 0.74 | 0.05 |
| olive oil | 218 | 1.30 | 1.06 | -0.75 |
| toluene | 51 | 1.00 | 0.99 | 0.08 |
| xylenes | 48 | 1.00 | 0.99 | -0.10 |
| Polar | | | | |
| acetonitrile | 7 | 2.23 | 2.13 | 0.67 |
| chloroform | 107 | 1.86 | 1.37 | 1.25 |
| diethyl ether | 70 | 2.21 | 1.65 | 1.47 |
| dimethyl sulfoxide | 7 | 2.65 | 2.65 | 0.01 |
| ethyl acetate | 22 | 3.02 | 2.18 | 2.09 |
| octanol | 245 | 2.24 | 2.22 | 0.31 |

Lennard-Jones potential quite well. At the same time, the assumption that n-decane is orientationally averaged at room temperature is clearly too optimistic. The single sphere approximation for this solvent results in the largest root mean square error among all tested apolar compounds. On the other hand, when n-decane is approximated as a Lennard-Jones chain composed of 4 units, the error in solvation free energy predictions drops by almost a 1 kcal/mol. A similar, but not quite so dramatic effects are observed for n-heptane and isooctane. These findings suggest that using appropriate reference fluids for other solvents might improve results even further.
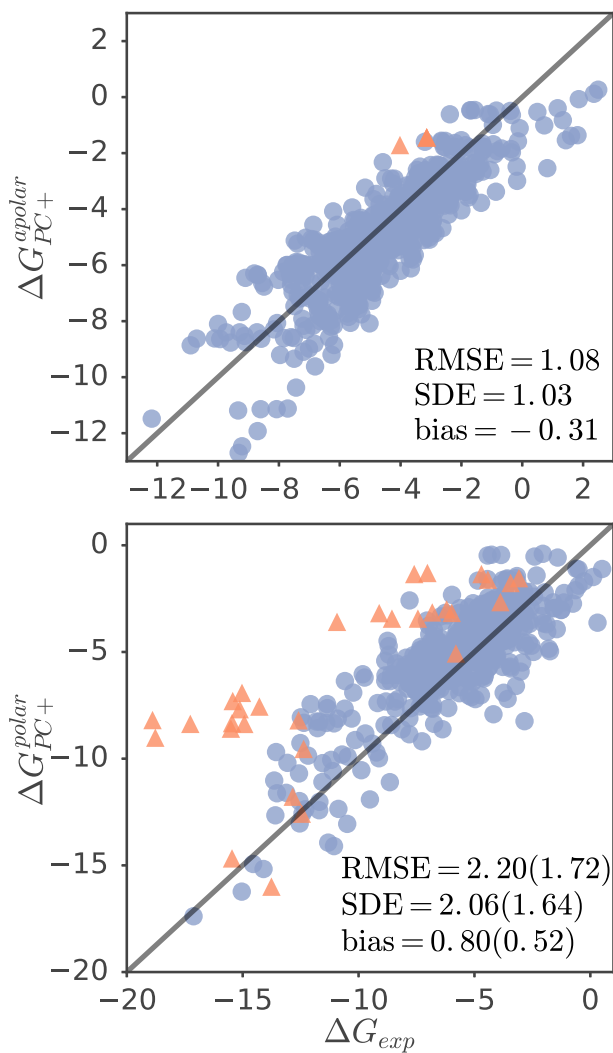


Figure 2: All results for apolar (top) and polar (bottom) solvents. Solutes that can act as donors of two or more hydrogen bonds are marked with triangles. Energies are in kcal/mol.

The fact that solvation free energies are quite inaccurately predicted for benzene solvent is somewhat surprising. Due to its high symmetry, benzene might be expected to be relatively well represented using single sphere approximation. However, it should be noted that benzene among all present aromatic compounds has the highest melting point and a diffraction pattern that resembles that of a crystalline solid[69]. Its comparison with toluene reveals that it is more structured liquid between the two and has more significant $\pi$-$\pi$ interactions[70]. We believe that these results indicate that benzene cannot be successfully treated as a simple apolar liquid.

The accuracy of solvation free energy predictions in olive oil is rather remarkable, considering how simple the model is. The predictions are only slightly less accurate in benzene, mostly due to the large, negative bias. We suspect that the primary source of this bias might be the fact that in our model individual fatty acids are not connected via glycerol, making cavity creation in the oil slightly easier than it actually is. Overall, these results show that PC+ can be used to estimate solvation free energy of molecules in mixtures, and might be used to study, for example, the effect of cosolvents on partition coefficients and activities.

The error of predictions for polar compounds is almost two times larger. This is not surprising, as all of these solvents usually interact with the solute in a very oriented way, making a reference liquid of symmetric Lennard-Jones spheres a poor choice.

Although none of the polar solvents studied here is a good hydrogen bond donor, all of them except chloroform can accept hydrogen bonds. Thus, compounds that can donate hydrogen bonds, particularly compounds donating multiple hydrogen bonds, are the largest outliers among all of the polar solvents studied here. Figure 2 shows the predicted and experimental solvation free energies for all solvents plotted against each other. The top figure contains results for apolar solvents and the bottom one for polar solvents as defined in the Table 3. Compounds capable of donating two or more hydrogen bonds are marked as peach triangles on both figures. It is clear that they are responsible for the majority of the overestimated solvation free energies by the model. A similar figure for each individual

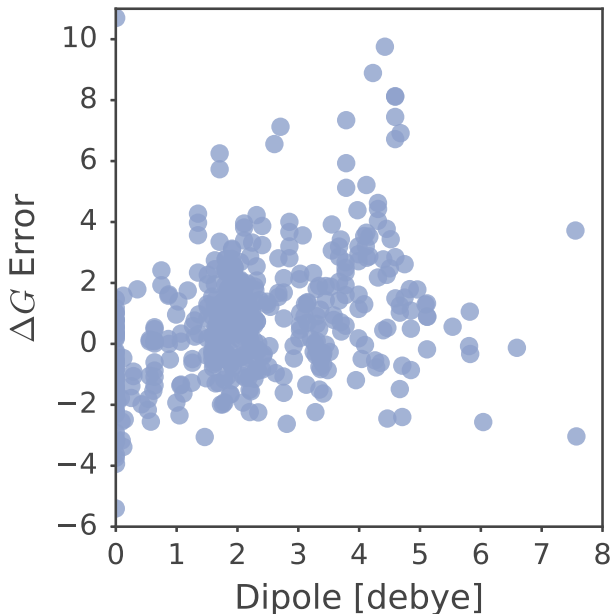solvent is included in the supporting information.



Figure 3: The difference between $\Delta G_{PC+}$ and experimental solvation free energy in all polar solvents, plotted against solutes dipole moment, as estimated using PM6.

Unlike hydrogen bonds, the dipole moment of the solute was only weakly correlated with an error in polar solvents. The correlation between the error in predicted solvation free energy and dipole moment, is shown in figure 3. This suggests that the major problem in our description of solvation in polar solvents was unrelated to the absence of solute polarity, but was rather related to the poor description of the solvents themselves. Thus, potentially, these results might be greatly improved simply by adjusting Lennard-Jones parameters of polar solvents or introducing a small solvent dipole.

We believe that the success of the presented approach is related to the fact that equations 1 and 2 capture the effective average potential of solvents. This indicates that the accuracy with which the liquid model is capable of predicting experimental critical point might be more important to simulations at standard conditions than previously thought. At the same time, correct solvent electrostatic properties and liquid structure might potentially be less relevant for thermodynamics. These conclusions are somewhat reflected by the recent simulations based on the coarse-grained water model from ELBA force-field[71], which can

reproduce water triple point rather well[72]. A few authors have demonstrated that this model estimates solvation free energies quite accurately[73,74], despite its incredible simplicity (single Lennard-Jones sphere and a dipole).

When compared to other models used to predict solvation free energies in non-aqueous solvents such as SMD[7] or COSMO-RS[10], the accuracy of these results might not seem very impressive. Indeed, SMD and SM8 models achieved root mean square error of 0.9 kcal/mol on an even larger dataset, containing 2072 measurements and a huge variety of different nonaqueous solvents. COSMO-RS, had a mean absolute error of 0.5 kcal/mol on the same set. Chamberlin and co-workers also managed to develop a modification of SM8: SM8T-OO[75], for prediction of solvation free energies in olive oil. This model showed 0.8 kcal/mol RMSE on the same dataset as the one used here. The above approaches are better developed and probably should be the first choices for computing solvation free energies in non-aqueous solvents. However, all of them were obtained using heavy parametrization and their accuracy is often reported on solutes similar or identical to those they were trained on. Furthermore, they must be reparameterized for new solvent systems. Thus, whenever one deals with uncommon solvents or conditions, the model presented here might be a sensible choice. Additionally, the present concept is very simplified and can be easily extended using better reference liquids and modifications of equations 1 and 2.

## Conclusion

To summarize, we present a general method for computing solvation free energies in apolar solvents. Approximating solvents as Lennard-Jones spheres and using atomistic solute models, we managed to obtain solvation free energies with 1 kcal/mol accuracy for a broad range of compounds. Parameters for solvents were obtained from their critical points using the corresponding state principle. The method was not only successful for pure solvents but worked relatively accurately for mixtures.

This work warrants a number of future investigations. The 3D-RISM model in which solvent is approximated as a single particle is equivalent to molecular density functional theory in homogeneous reference fluid approximation (MDFT/HRF)[76]. Thus, we believe that this approach can be also applied in classical density functional theories that can readily describe solvents containing dipoles or polarizable charges[77]. Additionally, we expect that the reliability of estimated parameters can be also improved by using polyatomic reference fluids as well as more complicated extensions of the corresponding state principle. Finally, it would be interesting to investigate the properties of the obtained solvent models using molecular dynamics simulations. We plan to address these questions in our future works.

## Acknowledgement

## Supporting Information Available

Supporting information contains pdf file with extra table and figures. We have also included structures (pdb) and topologies (prmtop) of solutes, susceptibility functions for solvents (xvv) and tables (csv) containing results for each individual solute.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

## References

(1) Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. Solvation Thermodynamics of Organic Molecules by the Molecular Integral Equation Theory: Approaching Chemical Accu-

racy. *Chem. Rev.* **2015**, *115*, 6312–6356.

(2) Palmer, D. S.; McDonagh, J. L.; Mitchell, J. B. O.; van Mourik, T.; Fedorov, M. V. First-Principles Calculation of the Intrinsic Aqueous Solubility of Crystalline Druglike Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 3322–3337.

(3) Palmer, D. S.; Llinàs, A.; Morao, I.; Day, G. M.; Goodman, J. M.; Glen, R. C.; Mitchell, J. B. O. Predicting Intrinsic Aqueous Solubility by a Thermodynamic Cycle. *Mol. Pharm.* **2008**, *5*, 266–279.

(4) Sangster, *Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry*; John Wiley & Sons: Chichester ; New York, 1997.

(5) Meunier, M., Ed. *Industrial Applications of Molecular Simulations*; CRC Press: Boca Raton, FL, 2011.

(6) Bardhan, J. P. Biomolecular Electrostatics—I Want your Solvation (Model). *Comput. Sci. Disc.* **2012**, *5*, 013001.

(7) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

(8) Cramer, C. J.; Truhlar, D. G. A Universal Approach to Solvation Modeling. *Acc. Chem. Res.* **2008**, *41*, 760–768.

(9) Mennucci, B., Cammi, R., Eds. *Continuum Solvation Models in Chemical Physics*; John Wiley & Sons, Ltd: Chichester, UK, 2007.

(10) Klamt, A. *COSMO-RS: From Quantum Chemistry to Fluid PhaseThermodynamics and Drug Design*; Elsevier Science: Amsterdam, 2005.

(11) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 350–358.

(12) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 1509–1519.

(13) Zhang, J.; Tuguldur, B.; van der Spoel, D. Force Field Benchmark of Organic Liquids. 2. Gibbs Energy of Solvation. *J. Chem. Inf. Model.* **2015**, *55*, 1192–1201.

(14) Liu, S.; Cao, S.; Hoang, K.; Young, K. L.; Paluch, A. S.; Mobley, D. L. Using MD Simulations to Calculate How Solvents Modulate Solubility. *J. Chem. Theory Comput.* **2016**,

(15) Garrido, N. M.; Economou, I. G.; Queimada, A. J.; Jorge, M.; Macedo, E. A. Prediction of the n-Hexane/Water and 1-Octanol/Water Partition Coefficients for Environmentally Relevant Compounds Using Molecular Simulation. *AIChE J.* **2012**, *58*, 1929–1938.

(16) Garrido, N. M.; Jorge, M.; Queimada, A. J.; Macedo, E. A.; Economou, I. G. Using Molecular Simulation to Predict Solute Solvation and Partition Coefficients in Solvents of Different Polarity. *Phys. Chem. Chem. Phys.* **2011**, *13*, 9155.

(17) Misin, M.; Fedorov, M. V.; Palmer, D. S. Communication: Accurate Hydration Free Energies at a Wide Range of Temperatures from 3D-RISM. *J. Chem. Phys.* **2015**, *142*, 091105.

(18) Misin, M.; Fedorov, M. V.; Palmer, D. S. Hydration Free Energies of Molecular Ions from Theory and Simulation. *J. Phys. Chem. B* **2016**, *120*, 975–983.

(19) Leland, T. W.; Chappelear, P. S. The Corresponding States Principle - A Review of Current Theory and Practice. *Ind. Eng. Chem.* **1968**, *60*, 15–43.

(20) Guggenheim, E. A. The Principle of Corresponding States. *J. Chem. Phys.* **1945**, *13*, 253–261.

(21) Hirschfelder, J. O.; Curtiss, C. F.; Bird, R. B. *Molecular theory of gases and liquids*; Chapman & Hall: New York, 1954.

(22) Flynn, L. W.; Thodos, G. Lennard-Jones Force Constants from Viscosity Data: Their Relationship to Critical Properties. *AIChE J.* **1962**, *8*, 362–365.

(23) Chung, T. H.; Lee, L. L.; Starling, K. E. Applications of Kinetic Gas Theories and Multiparameter Correlation for Prediction of Dilute Gas Viscosity and Thermal Conductivity. *Ind. Eng. Chem. Fund.* **1984**, *23*, 8–13.

(24) van Loef, J. J. Transport Properties of Dense Monatomic and Molecular Fluids and Their Mixtures, and the Corresponding States Principle I. Shear viscosity and Thermal Conductivity. *Physica B+C* **1984**, *124*, 305–326.

(25) Cuadros, F.; Cachadiña, I.; Ahumada, W. Determination of Lennard-Jones Interaction Parameters Using a New Procedure. *Mol Eng* **1996**, *6*, 319–325.

(26) Ben-Amotz, D.; Gift, A. D.; Levine, R. D. Improved Ccorresponding States Scaling of the Equations of State of Simple Fluids. *J. Chem. Phys.* **2002**, *117*, 4632–4634.

(27) Zhu, Y.; Lu, X.; Zhou, J.; Wang, Y.; Shi, J. Prediction of Diffusion Coefficients for Gas, Liquid and Supercritical Fluid: Application to Pure Real Fluids and Infinite Dilute Binary Solutions Based on the Simulation of Lennard–Jones Fluid. *Fluid Phase Equilib.* **2002**, *194–197*, 1141–1159.

(28) Galliéro, G.; Boned, C.; Baylaucq, A. Molecular Dynamics Study of the Lennard-Jones Fluid Viscosity: Application to Real Fluids. *Ind. Eng. Chem. Res.* **2005**, *44*, 6963–6972.

(29) Galliero, G.; Nieto-Draghi, C.; Boned, C.; Avalos, J. B.; Mackie, A. D.; Baylaucq, A.;

Montel, F. Molecular Dynamics Simulation of Acid Gas Mixtures: A Comparison between Several Approximations. *Ind. Eng. Chem. Res.* **2007**, *46*, 5238–5244.

(30) Galliero, G.; Boned, C. Shear Viscosity of the Lennard-Jones Chain Fluid in its Gaseous, Supercritical, and Liquid States. *Phys. Rev. E* **2009**, *79*, 021201.

(31) Mejía, A.; Herdes, C.; Müller, E. A. Force Fields for Coarse-Grained Molecular Simulations from a Corresponding States Correlation. *Ind. Eng. Chem. Res.* **2014**, *53*, 4131–4141.

(32) Ben-Amotz, D.; Underwood, R. Unraveling Water's Entropic Mysteries: A Unified View of Nonpolar, Polar, and Ionic Hydration. *Acc. Chem. Res.* **2008**, *41*, 957–967.

(33) Ben-Amotz, D.; Raineri, F. O.; Stell, G. Solvation Thermodynamics: Theory and Applications†. *J. Phys. Chem. B* **2005**, *109*, 6866–6878.

(34) Qian, H.; Hopfield, J. J. Entropy-Enthalpy Compensation: Perturbation and Relaxation in Thermodynamic Systems. *J. Chem. Phys.* **1996**, *105*, 9292–9298.

(35) Okumura, H.; Yonezawa, F. Liquid–vapor Coexistence Curves of Several Interatomic Model Potentials. *J. Chem. Phys.* **2000**, *113*, 9162–9168.

(36) Yaws, C. L. *Thermophysical Properties of Chemicals and Hydrocarbons*, 2nd ed.; Gulf Professional Publishing: Amsterdam Boston, 2014.

(37) Lafitte, T.; Bessieres, D.; Piñeiro, M. M.; Daridon, J.-L. Simultaneous Estimation of Phase Behavior and Second-Derivative Properties Using the Statistical Associating Fluid Theory with Variable Range Approach. *J. Chem. Phys.* **2006**, *124*, 024509.

(38) Galliero, G. Equilibrium, Interfacial and Transport Properties of n-alkanes: Towards the Simplest Coarse Grained Molecular Model. *Chem. Eng. Res. Des.* **2014**, *92*, 3031–3037.

(39) Dubey, G. S.; O'Shea, S. F.; Monson, P. A. Vapour-liquid equilibria for two centre Lennard-Jones diatomics and dipolar diatomics. *Mol. Phys.* **1993**, *80*, 997–1007.

(40) Blas, F. J.; Vega, L. F. Improved Vapor–Liquid Equilibria Predictions for Lennard-Jones Chains from the Statistical Associating Fluid Dimer Theory: Comparison with Monte Carlo simulations. *J. Chem. Phys.* **2001**, *115*, 4355–4358.

(41) Escobedo, F. A.; Pablo, J. J. D. Simulation and Prediction of Vapour-Liquid Equilibria for Chain Molecules. *Mol. Phys.* **1996**, *87*, 347–366.

(42) Beltrán, G.; del Rio, C.; Sánchez, S.; Martínez, L. Influence of Harvest Date and Crop Yield on the Fatty Acid Composition of Virgin Olive Oils from Cv. Picual. *J. Agric. Food Chem.* **2004**, *52*, 3434–3440.

(43) Ghanbari, R.; Anwar, F.; Alkharfy, K. M.; Gilani, A.-H.; Saari, N. Valuable Nutrients and Functional Bioactives in Different Parts of Olive (Olea europaea L.)—A Review. *Int. J. Mol. Sci.* **2012**, *13*, 3291–3340.

(44) Olivier-Droz, P.; Fernández, J. Solubility of Organic Solvents. I. Gas Chromatographic Determination of Olive Oil-Gas Partition Coefficients. *Helv. Chim. Acta* **1977**, *60*, 454–458.

(45) Abraham, M. H.; Grellier, P. L.; McGill, R. A. Determination of olive oil–gas and hexadecane–gas partition coefficients, and calculation of the corresponding olive oil–water and hexadecane–water partition coefficients. *J. Chem. Soc., Perkin Trans. 2* **1987**, 797–803.

(46) Abraham, M. H.; Ibrahim, A. Gas to Olive Oil Partition Coefficients: A Linear Free Energy Analysis. *J. Chem. Inf. Model.* **2006**, *46*, 1735–1741.

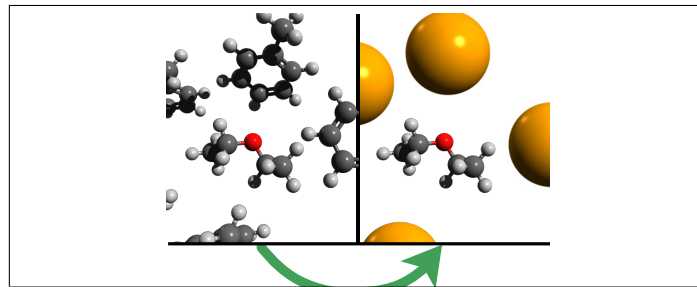(47) García, A.; Brenes, M.; Romero, C.; García, P.; Garrido, A. Study of phenolic com-

pounds in virgin olive oils of the Picual variety. *Eur. Food. Res. Technol.* **2002**, *215*, 407–412.

(48) Quintero-Flórez, A.; Sinausia Nieva, L.; Sánchez-Ortíz, A.; Beltrán, G.; Perona, J. S. The Fatty Acid Composition of Virgin Olive Oil from Different Cultivars Is Determinant for Foam Cell Formation by Macrophages. *J. Agric. Food Chem.* **2015**, *63*, 6731–6738.

(49) Joback, K. G.; Reid, R. C. Estimation of Pure-Component Properties from Group-Contributions. **1987**, *57*, 233–243.

(50) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R. et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005**, *26*, 1752–1780.

(51) Sergiievskyi, V.; Jeanmairet, G.; Levesque, M.; Borgis, D. Solvation Free-Energy Pressure Corrections in the Three Dimensional Reference Interaction Site Model. *J. Chem. Phys.* **2015**, *143*, 184116.

(52) Hirata, F. *Molecular Theory of Solvation*; Kluwer Academic Publishers: New York, 2003.

(53) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids, Fourth Edition: with Applications to Soft Matter*, 4th ed.; Academic Press: Amstersdam, 2013.

(54) Kast, S. M.; Kloss, T. Closed-Form Expressions of the Chemical Potential for Integral Equation Closures with Certain Bridge Functions. *J. Chem. Phys.* **2008**, *129*, 236101.

(55) Marenich, A.; Kelly, C.; Thompson, J.; Hawkins, G.; Chambers, C.; Giesen, D.; Winget, P.; Cramer, C.; Truhlar, D. *Minnesota Solvation Database – version 2012*; University of Minnesota: Minneapolis, 2012.

(56) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J Cheminform* **2011**, *3*, 33.

(57) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheeatham,III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W. et al. AMBER 2015; University of California; San Francisco, 2015.

(58) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J Mol Model* **2007**, *13*, 1173–1213.

(59) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. Gaussian˜09 Revision D.01. 2009.

(60) *Maestro*; Schrödinger, LLC.: New York, NY, 2014.

(61) Straatsma, T. P.; Berendsen, H. J. C.; Postma, J. P. M. Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water. **1986**, *85*, 6720–6727.

(62) Allen, M. P., Tildesley, D. J., Eds. *Computer Simulation of Liquids*; Clarendon Press: Oxford, UK, 1987.

(63) Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J.; Kovalenko, A. Three-Dmensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J. Chem. Theory Comput.* **2010**, *6*, 607–624.

(64) Kovalenko, A.; Hirata, F. Potentials of Mean Force of Simple Ions in Ambient Aqueous Solution. I. Three-Dimensional Reference Interaction Site Model Approach. *J. Chem. Phys* **2000**, *112*, 10391–10402.

(65) Kovalenko, A.; Hirata, F. Self-Consistent Description of a Metal-Water Interface by the Kohn-Sham Density Functional Theory and the Three-Dimensional Reference Interaction Site Model. *J. Chem. Phys.* **1999**, *110*, 10095–10112.

(66) Guignon, B.; Aparicio, C.; Sanz, P. D. Volumetric properties of sunflower and olive oils at temperatures between 15 and 55 C under pressures up to 350 MPa. *High Pressure Res.* **2009**, *29*, 38–45.

(67) Battaglia, M. R.; Cox, T. I.; Madden, P. A. The Orientational Correlation Parameter for Liquid CS2, C6H6 and C6F6. *Mol. Phys.* **1979**, *37*, 1413–1427.

(68) Pusztai, B. L.; McGreevy, R. L. The Structure of Liquid CCl4. *Mol. Phys.* **1997**, *90*, 533–540.

(69) Narten, A. H. Diffraction Pattern and Structure of Liquid Benzene. *J. Chem. Phys.* **1968**, *48*, 1630–1634.

(70) Headen, T. F.; Howard, C. A.; Skipper, N. T.; Wilkinson, M. A.; Bowron, D. T.; Soper, A. K. Structure of π-π Interactions in Aromatic Liquids. *J. Am. Chem. Soc.* **2010**, *132*, 5735–5742.

(71) Orsi, M.; Essex, J. W. The ELBA Force Field for Coarse-Grain Modeling of Lipid Membranes. *PLoS ONE* **2011**, *6*, e28637.

(72) Orsi, M. Comparative Assessment of the ELBA Coarse-Grained Model for Water. *Mol. Phys.* **2014**, *112*, 1566–1576.

(73) Genheden, S. Predicting Partition Coefficients with a Simple All-Atom/Coarse-Grained Hybrid Model. *J. Chem. Theory Comput.* **2016**, *12*, 297–304.

(74) Orsi, M.; Ding, W.; Palaiokostas, M. Direct Mixing of Atomistic Solutes and Coarse-Grained Water. *J. Chem. Theory Comput.* **2014**, *10*, 4684–4693.

(75) Chamberlin, A. C.; Levitt, D. G.; Cramer, C. J.; Truhlar, D. G. Modeling Free Energies of Solvation in Olive Oil. *Mol. Pharmaceutics* **2008**, *5*, 1064–1079.

(76) Sergiievskyi, V. P.; Jeanmairet, G.; Levesque, M.; Borgis, D. Fast Computation of Solvation Free Energies with Molecular Density Functional Theory: Thermodynamic-Ensemble Partial Molar Volume Corrections. *J. Phys. Chem. Lett.* **2014**, *5*, 1935–1942.

(77) Jeanmairet, G.; Levesque, M.; Sergiievskyi, V.; Borgis, D. *Computational Trends in Solvation and Transport in Liquids*; Schriften des Forschungszentrums Jülich: Jülich, 2015; Vol. 28; pp 187–208.

# Graphical TOC Entry



$$\sigma = \sqrt[3]{\frac{\rho_c^*}{\rho_c}} \qquad \epsilon = \frac{kT_c}{T_c^*}$$