Order and correlation contributions to the entropy of hydrophobic solvation

ARTICLE *in* THE JOURNAL OF CHEMICAL PHYSICS · MARCH 2015

Impact Factor: 2.95 · DOI: 10.1063/1.4908532 · Source: PubMed

4 AUTHORS, INCLUDING:

Quinn Alexander Besford

University of Melbourne

3 PUBLICATIONS **13** CITATIONS

AIP | The Journal of Chemical Physics

Order and correlation contributions to the entropy of hydrophobic solvation

Maoyuan Liu, Quinn Alexander Besford, Thomas Mulvaney, and Angus Gray-Weale

Citation: The Journal of Chemical Physics **142**, 114117 (2015); doi: 10.1063/1.4908532
View online: http://dx.doi.org/10.1063/1.4908532
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/142/11?ver=pdfcov
Published by the AIP Publishing

Order and correlation contributions to the entropy of hydrophobic solvation

Maoyuan Liu, Quinn Alexander Besford, Thomas Mulvaney, and Angus Gray-Weale[a)]
School of Chemistry, The University of Melbourne, Victoria 3010, Australia

(Received 24 October 2014; accepted 5 February 2015; published online 20 March 2015)

The entropy of hydrophobic solvation has been explained as the result of ordered solvation structures, of hydrogen bonds, of the small size of the water molecule, of dispersion forces, and of solvent density fluctuations. We report a new approach to the calculation of the entropy of hydrophobic solvation, along with tests of and comparisons to several other methods. The methods are assessed in the light of the available thermodynamic and spectroscopic information on the effects of temperature on hydrophobic solvation. Five model hydrophobes in SPC/E water give benchmark solvation entropies via Widom's test-particle insertion method, and other methods and models are tested against these particle-insertion results. Entropies associated with distributions of tetrahedral order, of electric field, and of solvent dipole orientations are examined. We find these contributions are small compared to the benchmark particle-insertion entropy. Competitive with or better than other theories in accuracy, but with no free parameters, is the new estimate of the entropy contributed by correlations between dipole moments. Dipole correlations account for most of the hydrophobic solvation entropy for all models studied and capture the distinctive temperature dependence seen in thermodynamic and spectroscopic experiments. Entropies based on pair and many-body correlations in number density approach the correct magnitudes but fail to describe temperature and size dependences, respectively. Hydrogen-bond definitions and free energies that best reproduce entropies from simulations are reported, but it is difficult to choose one hydrogen bond model that fits a variety of experiments. The use of information theory, scaled-particle theory, and related methods is discussed briefly. Our results provide a test of the Frank-Evans hypothesis that the negative solvation entropy is due to structured water near the solute, complement the spectroscopic detection of that solvation structure by identifying the structural feature responsible for the entropy change, and point to a possible explanation for the observed dependence on length scale. Our key results are that the hydrophobic effect, i.e. the signature, temperature-dependent, solvation entropy of nonpolar molecules in water, is largely due to a dispersion force arising from correlations between rotating permanent dipole moments, that the strength of this force depends on the Kirkwood g-factor, and that the strength of this force may be obtained exactly without simulation. [http://dx.doi.org/10.1063/1.4908532]

I. INTRODUCTION

At around room temperature, entropy limits the solubility in water of a large class of atoms and nonpolar molecules, for example, xenon and methane.[1–6] On heating, this entropy of solvation approaches zero from below, as enthalpy increases and dominates the free-energy barrier for dissolution. The crossover from entropy to enthalpy occurs roughly between room temperature and the boiling point of water.[3,7] This hydrophobic effect also inhibits other processes that move water molecules from the bulk, where their neighbours are virtually always other waters, to environments where they have at least one nonpolar neighbour.[4] Examples include the unfolding of a protein to expose hydrophobic groups,[8,9] the formation of an interface between long-chain alcohols and water,[10] and the breakup of a bilayer cell membrane that would expose its inner hydrophobic layers to the surrounding water.[4,11] In addition to the temperature dependence, the size dependence of hydrophobic solvation free energy is distinctive.[12,13] It is proportional to the volume of the solute up to a radius of approximately 1 nm and to surface area for larger solutes.[11] A smooth crossover from microscopic solvation free energy to the macroscopic interfacial tension is seen. Several explanations and analyses of this phenomenon have been published, many along with methods of calculation of the hydrophobic solvation entropy. We report here a comparison of the performance of several of these methods, including a new approach that calculates the solvation entropy from correlations between dipoles. We discuss the results in the light of recent experimental detection of changes in water structure near a hydrophobe.[14]

Computer simulations using relatively simple models for the forces have been shown to describe hydrophobic solvation, including the temperature[6,7] and size dependences of the solvation free energy.[11–13] Paschek studied models for neon, argon, krypton, xenon, and methane and found all agree reasonably well with experiment. The entropies closest to experiment were obtained for the SPC/E model.[6,15] Chandler and coworkers

[a)]Electronic mail: gusgw@gusgw.net

used the SPC/E model to study several aspects of hydrophobic solvation and interface structure,[11] and argued for its merit. Taken together, these studies show that SPC/E water exhibits the temperature and size dependences of solvation entropy and free energy similar to those seen in experiment. We choose the SPC/E model for solvating waters and follow Paschek in the choice of five model hydrophobes, deferring the study of larger solutes and interfaces.

The first explanation of the hydrophobic effect of which we are aware is the "iceberg" model described by Frank and Evans,[16] who inferred from the negative solvation entropy of nonpolar molecules an ordering of the nearby waters, even perhaps the formation of ice-like structures. Davis *et al.* detected enhanced water structure in hydrophobes' solvation shells at low temperature using Raman scattering with multivariate curve resolution.[14] They saw also the disappearance of this structure on heating, mirroring the observed approach to zero from below of the solvation entropy. An increase in tetrahedral order around a simulated hydrophobe, similar to that predicted by the iceberg model, and seen by Galamba, has been reported.[17] A further recent innovation in the calculation of solvation entropies from simulation is the use of the electric field as an order parameter.[18] Both the tetrahedral order parameter and the electric field experienced by a molecule in the presence of a solute are functions of the configurations of the other water molecules. Distributions of these order parameters therefore define macrostates for which entropies may be calculated. We investigate the magnitudes of these entropies and compare them to other approaches. With other entropy estimators, these provide a clear test of the Frank and Evans hypothesis and allow us to determine which, if any, structural changes are reflected in thermodynamic[3,7] and spectroscopic evidence.[14]

The tendency of water molecules' dipole moments to point towards or away from a nonpolar solute is one possible source of solvation entropy. Otten *et al.* have shown using simulations that this particular entropy is insignificant for an ion near a planar water surface.[19] This same entropy near a small hydrophobic solute is examined here, but so are other correlations involving water dipoles. Consideration of higher order dipole correlations first involves correlations between dipoles and the solute. Second, we consider correlations between dipoles on pairs of water molecules, which affect solvation thermodynamics because waters near the solute have fewer neighbours with which they may correlate. We show here that these two entropy contributions are together much larger than the term considered by Otten *et al.*[19] and consider the possibility that the hydrophobic solvation may be understood in terms of polarisation fluctuations. This mechanism is largely a dispersion force between rotating permanent dipole moments. The force resembles the Keesom force in that it results from correlated permanent moments[20] but differs in that its strength and temperature dependence are strongly influenced by additional neighbour correlations. Keesom considered only the correlation between two isolated dipole moments.[20] We report an exact result for the effect of strong correlations, which result in a renormalization of the dipole moment that appears in the dispersion interaction.

The entropy of a fluid may be written as an infinite series in which each term is an integral that depends on only one n-particle correlation function.[21,22] The part of this series that depends on pair density correlations, that is on the radial distribution function (RDF), has been considered as an approximation to the full entropy.[23] This part is usually denoted S_2. We assess the ability of this pair entropy to describe the hydrophobic solvation entropy. The particular advantage of this method is that it reveals the importance of many-body correlations. This is possible because the difference between the pair and many-body entropies is the sum of the three-body and higher terms in the series. The estimation of the entropies of certain characteristic structures is another approach to solvation entropy calculation that includes "many-body" effects. We assess here this method using the entropies published by Godec and Merzel.[18]

Changes in hydrogen bonding are a possible contribution to the free energy cost of solvation.[24] There is some freedom in the choice of the definition of the hydrogen bond for the analysis of simulations, with both geometric and energetic definitions available, and no guarantee that a given definition corresponds to that implicit in spectroscopic experiments.[14] The enthalpy and entropy of formation of the hydrogen bond are usually parameters chosen to match experiment or simulation. We select representative models and assess their ability to reproduce the solvation entropy.

The Lum-Chandler-Weeks model recognised the importance of density fluctuations and was based on the assumption of different behaviours at small and large length scales.[25] Later approaches, including information theory and scaled particle theory,[26,27] are based at small scales on the probability distribution of waters found within a spherical volume. This approach is closely related to Widom's test-particle insertion method applied to hard spheres. The approach uses simulation data to characterise small scale behaviour, with experimental data to characterise large scale behaviour, and using statistical mechanical principles to deduce the form of the solvation free energy. The strength of this approach is its accuracy for any solute whose "hard-sphere" diameter is known. The crossover from small to large length scales prominent in this family of theories has been found also for a hard-sphere solute in a Lennard-Jones (LJ) solvent.[13]

We discuss methods of calculation of the entropy of solvation in Sec. II, describe methods of extraction of essentially exact entropies from simulation in Sec. III. Results are set out in Sec. IV, and conclusions in Sec. V.

II. METHODS OF SOLVATION ENTROPY CALCULATION

We describe here each method of entropy estimation that we consider. Previously published methods are only briefly summarised. A new method is described at greater length in Secs. II C 2–II C 4.

A. Tetrahedral order

A tetrahedral order parameter, $q(i)$, for molecule i, has been defined[17,22,23,28]

$$q(i) = 1 - \frac{3}{8}\sum_{j}^{3}\sum_{k=j+1}^{4}\left[\cos\eta_{jik} + \frac{1}{3}\right]^2, \qquad (1)$$

where η_{jik} is the angle formed by the nearest neighbours j and k to the central molecule i, calculated along the oxygen-oxygen displacement vectors. The double sum in Eq. (1) is over the four nearest neighbours closest to molecule i. Kumar *et al.* proposed the tetrahedral-order entropy[28]

$$\frac{s_{\text{tet}}^{\text{bulk}}}{k_{\text{B}}} = \frac{s_{\text{tet}}^{\text{id}}}{k_{\text{B}}} + \frac{3}{2}\int_{q_{\min}}^{q_{\max}} \log(1-q)\, p(q)\, \mathrm{d}q, \tag{2}$$

where $s_{\text{tet}}^{\text{id}}/k_B = \log(\Omega_0) + 3/2\log(8/3)$ is a constant, k_B is the Boltzmann's constant, and $p(q)$ is the distribution of q under given conditions. The smallest and largest possible values of the order parameter q are, respectively, $q_{\min} = -3$ and $q_{\max} = 1$. To calculate the change in S_{tet} upon solvation, we define a local analog of the Kumar tetrahedral entropy

$$\frac{s_{\text{tet}}^{\text{loc}}(r)}{k_{\text{B}}} = \frac{s_{\text{tet}}^{\text{id}}}{k_{\text{B}}} + \frac{3}{2}\int_{q_{\min}}^{q_{\max}} \log(1-q)\, p(r;q)\, \mathrm{d}q, \tag{3}$$

where $s_{\text{tet}}^{\text{loc}}(r)$ is the per solvent molecule entropy at a distance r from the solute, and $p(r;q)$ is the distribution of q at r. The total change in S_{tet} upon solvation is then

$$\frac{\Delta S_{\text{tet}}}{k_B} = 4\pi\rho \int_0^\infty \frac{s_{\text{tet}}^{\text{local}}(r) - s_{\text{tet}}^{\text{bulk}}}{k_{\text{B}}} g_{\text{sw}}(r) r^2\, \mathrm{d}r, \tag{4}$$

where ρ is the number density of solvent molecules and g_{sw} is the solute to water oxygen RDF.

B. Electric field distribution

Godec and Merzel showed the distribution of magnitudes of the electric fields on water molecules near a simulated hydrophobic solute is perturbed from the bulk distribution.[18] The local entropy using this approach is defined as

$$s_{\text{E}}^{\text{loc}}(r) = -k_B \int_{|\boldsymbol{E}|_{\min}}^{|\boldsymbol{E}|_{\max}} p(r;|\boldsymbol{E}|) \log(p(r;|\boldsymbol{E}|))\, \mathrm{d}|\boldsymbol{E}|, \tag{5}$$

where $|\boldsymbol{E}|$ is the magnitude of the electric field experienced by a molecule, and $p(r;|\boldsymbol{E}|)$ is the distribution of electric field strengths at a distance r from the solute. We find that the observed values of the electric field magnitude in our simulations are within the range of 0.0–0.1 atomic units, and so choose these values as the limits in the integral. The total change in entropy is then

$$\frac{\Delta S_{\text{E}}}{k_{\text{B}}} = 4\pi\rho \int_0^\infty \frac{s_{\text{E}}^{\text{local}}(r) - s_{\text{E}}^{\text{bulk}}}{k_B} g_{\text{sw}}(r) r^2\, dr, \tag{6}$$

where $s_{\text{E}}^{\text{bulk}}$ is the limit of $s_{\text{E}}^{\text{local}}(r)$ at large r.

C. Dipole correlations and dispersion forces

1. Leading order dipole-solute correlations

The orientation of dipole moments around a dissolved solute, or at an interface, can be described by the distribution of $\cos(\theta)$, where θ is the angle between the dipole moment and the displacement $\boldsymbol{r}$ from the dissolved solute to the water molecule carrying that moment. We follow the treatment of Otten *et al.*,[19] who studied correlations near an ion at an interface, and define a local entropy

$$s_\theta^{\text{loc}}(r) = -k_{\text{B}} \int_{-1}^{1} p(r;\cos\theta) \log(p(r;\cos\theta))\, \mathrm{d}\cos\theta, \tag{7}$$

where $p(r;\cos\theta)$ is the distribution of water dipole orientations at a distance r from the solute. The total change in entropy is then

$$\frac{\Delta S_\theta}{k_{\text{B}}} = 4\pi\rho \int_0^\infty \frac{s_\theta^{\text{loc}}(r) - s_\theta^{\text{bulk}}}{k_{\text{B}}} g_{\text{sw}}(r) r^2\, \mathrm{d}r, \tag{8}$$

where $s_\theta^{\text{bulk}} = -k_{\text{B}} \int_{-1}^{1} p_{\text{bulk}}(\cos\theta) \log\left(p_{\text{bulk}}(\cos\theta)\right)\, \mathrm{d}\cos\theta$, with $p_{\text{bulk}}(\cos\theta) = \frac{1}{2}$ the distribution of dipole orientations in the bulk.

2. Expansion of dipole correlations over a basis set

Consider a RDF that depends on distance as well as on dipole orientations, $g^{ij}(\boldsymbol{r},\boldsymbol{\mu},\boldsymbol{\mu}')$, where the correlation is between particles of species i and j, joined by a vector $\boldsymbol{r}$, and carrying dipole moments $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$. Contributions to the dipole correlation entropy may be evaluated by expanding this full, dipole-dependent RDF. For a fluid of polar molecules, the expansion may be written

$$g^{ij}(\boldsymbol{r},\boldsymbol{\mu},\boldsymbol{\mu}') = g_{\text{S}}^{ij}(r) + \sum_{l_1}\sum_{l_2}\sum_{l\geq|l_1-l_2|}^{l_1+l_2} h_{l_1 l_2 l}^{ij}(r) \times \Phi_{l_1 l_2 l}(\hat{\boldsymbol{r}},\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\mu}}'). \tag{9}$$

Here, g_{S}^{ij} is the usual RDF, independent of dipole orientations, and $h_{l_1 l_2 l}^{ij}$ is an expansion coefficient describing preferential orientations of the moments. The rotational invariants $\Phi_{l_1 l_2 l}$ are defined here in terms of dipole orientations only, not the full set of Euler angles.[29] They are explicitly

$$\Phi_{l_1 l_2 l}(\hat{\boldsymbol{r}},\hat{\boldsymbol{\mu}}_1,\hat{\boldsymbol{\mu}}_2) = \sum_{m=-\min(l_1,l_2)}^{\min(l_1,l_2)} 4\pi\langle l_1,m,l_2,-m|l,0\rangle \times Y_{l_1,m}(\theta_1,\phi_1) Y_{l_2,-m}(\theta_2,\phi_2), \tag{10}$$

where the orientation of a unit vector parallel to the dipole moment, $\hat{\boldsymbol{\mu}}_i$, is given by the angles θ_i and ϕ_i, and $\langle l_1,m,l_2,-m|l,0\rangle$ is the Clebsch-Gordan coefficient. Further details on this expansion are provided by Gray and Gubbins in Appendix A of their book.[29]

The set of rotational invariants is a basis set which spans the rotational group of SO(3). Rotational invariants as given in Eq. (10) have a normalisation of 1, given by $\langle \Phi^2_{l_1 l_2 l}\rangle_{\hat{\boldsymbol{r}},\hat{\boldsymbol{\mu}}_1,\hat{\boldsymbol{\mu}}_2} = 1$. Its orthogonality relationship is thus $\langle \Phi_{l_1 l_2 l}\Phi_{l_1' l_2' l'}\rangle_{\hat{\boldsymbol{r}},\hat{\boldsymbol{\mu}}_1,\hat{\boldsymbol{\mu}}_2} = \delta(l_1,l_1')\delta(l_2,l_2')\delta(l,l')$, where δ is the Kronecker delta function. Rotational invariants have been used in molecular Ornstein-Zernike theory with other normalisations,[30–34] however we choose a normalisation of 1 to simplify higher order expansions in $\Phi_{l_1 l_2 l}$.

We are interested in dipole correlations, which include the longest range correlations between moments, and leave shorter ranged correlations between the higher moments for future study. Note that higher moments are not neglected, they and all other components of the intermolecular interactions influence the dipoles' correlations. Richardi *et al.* find the quadrupolar correlation function $g_{220}(r)$ in water and methanol

to be an order of magnitude smaller than dipolar correlation functions.[33]

3. Dipole correlations

The angular-dependent RDF may be written in terms of an angular-dependent potential of mean force (PMF) as

$$g^{ij}(\boldsymbol{r},\boldsymbol{\mu}_1,\boldsymbol{\mu}_2) = \exp(-\beta\mathcal{W}^{ij}(\boldsymbol{r},\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)), \tag{11}$$

where $\beta = (k_BT)^{-1}$, T is the absolute temperature, and $\mathcal{W}_{ij}$ is the PMF. Functions such as $g^{ij}(\boldsymbol{r},\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)$ and $\mathcal{W}^{ij}(\boldsymbol{r},\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)$ can be expanded in the set of rotational invariants.[29] The expansion of the natural logarithm of Eq. (11) in rotational invariants gives[35]

$$-\beta\mathcal{W}_S^{ij}(r) - \sum_{l_1}\sum_{l_2}\sum_{l\geq|l_1-l_2|}^{l_1+l_2}\beta\mathcal{W}_{l_1l_2l}^{ij}(r)\Phi_{l_1l_2l} = \log\left(g_S^{ij}(r) + \sum_{l_1}\sum_{l_2}\sum_{l\geq|l_1-l_2|}^{l_1+l_2}h_{l_1l_2l}^{ij}(r)\Phi_{l_1l_2l}\right). \tag{12}$$

We seek the contribution of dipole correlations $h_{l_1l_2l}^{ij}(r)$ to $\omega^{ij}(r) = -k_BT\log(g_S^{ij}(r))$, the PMF averaged over correlated dipole configurations. Note that $\mathcal{W}_S^{ij}(r) \neq \omega^{ij}(r)$. Recall from above (Eq. (9)) that the subscript "S" denotes the term independent of dipole orientation. This angular contribution can be expressed in terms of dipole correlation functions $h_{l_1l_2l}^{ij}(r)$,

$$\omega^{ij}(r) = \mathcal{W}_S^{ij}(r) - k_BT \times\left\langle\log\left(1+\sum_{l_1}\sum_{l_2}\sum_{l\geq|l_1-l_2|}^{l_1+l_2}\frac{h_{l_1l_2l}^{ij}(r)}{g_S^{ij}(r)}\Phi_{l_1l_2l}\right)\right\rangle_{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2}. \tag{13}$$

The first non-zero term in the expansion of the logarithm term is the second order perturbation in $h_{l_1l_2l}^{ij}(r)$.

The expansion of Eq. (13) in the case of correlations between pairs of water molecules yields

$$\omega^{\mathrm{ww}}(r) = \mathcal{W}_S^{\mathrm{ww}}(r) + \phi(r) + O\left((h_{l_1l_2l}^{\mathrm{ww}}(r))^3/(g_S^{\mathrm{ww}}(r))^3\right), \tag{14}$$

$$\phi(r) = -k_BT\sum_{l_1}\sum_{l_2}\sum_{l\geq|l_1-l_2|}^{l_1+l_2}(h_{l_1l_2l}^{\mathrm{ww}}(r))^2/2(g_S^{\mathrm{ww}}(r))^2, \tag{15}$$

where $\phi(r)$ arises from correlations between dipoles. $\phi(r)$ is always negative, so it produces an attractive force between pairs of polar molecules. It is implicit here and below that the sum includes only real basis functions. $\phi(r)$ contributes as does any other long-ranged attraction to the Ornstein-Zernike direct correlation functions $c_S^{\mathrm{ww}}(r)$, which is the second functional derivative of the free energy. Changes in free energy due to ϕ may therefore be estimated as a sum over pairs (see Sec. II C 5).[34]

To describe the correlation between a solute and dipole moments in the solvent, we need only one dipole moment. The above treatment applies, though many of the correlation functions $h_{l_1l_2l}^{ij}$ now vanish. For solute-water interactions, we replace ϕ in Eq. (14) with ψ,

$$\omega^{\mathrm{sw}}(r) = \mathcal{W}_S^{\mathrm{sw}}(r) + \psi(r) + O\left((h_{l_1l_2l}^{\mathrm{sw}}(r))^3/(g_S^{\mathrm{sw}}(r))^3\right), \tag{16}$$

$$\psi(r) = -k_BT\frac{1}{2(g_S^{sw}(r))^2}\sum_{n=1}^{\infty}(h_{n0n}^{sw}(r))^2, \tag{17}$$

where $g_{\mathrm{sw}}(r)$ is the solute-water RDF, and $h_{n0n}^{sw}(r)$ are the coefficients of expansion obtained from molecular dynamics (MD) configurations.

These expansions are the same as those in the Quadratic Hypernetted Chain (QHNC) method, where terms in higher powers of the correlation functions are neglected. We do not neglect the higher power contributions in calculating the fluid properties, rather we evaluate the contribution of the terms quadratic in the dipole correlations to the free energy for simulated structures (Eqs. (15) and (17)).

4. Solvent-solvent correlations

In a fluid of neutral, rigid, polar molecules, the slowest decaying term in the potential energy of interaction of two molecules corresponds to the interaction of two dipoles. Using the expansion described in Sec. II C 2, this term is $u_{112}(r) \propto (\mu^2/r^3)$. The slowest decaying term in the expansion of the RDF, Eq. (9), is[35]

$$\lim_{r\to\infty}h_{112}(r) = \sqrt{\frac{2}{3}}\frac{(\epsilon-1)^2}{4\pi y\,\rho\epsilon}\frac{1}{r^3}, \tag{18}$$

where $y = 4\pi\rho\mu^2/9k_BT$, and ϵ is the relative static permittivity. It is convenient here to introduce the Kirkwood g-factor, defined as

$$g_K = \frac{\langle|\boldsymbol{M}|^2\rangle}{N\mu^2}, \tag{19}$$

where $\langle|\boldsymbol{M}|^2\rangle$ is the mean-squared total dipole moment of N molecules. The g-factor describes the average alignment of dipoles. For an uncorrelated dipole fluid, the mean squared total moment is the sum of the squares of the molecules' individual moments, so $g_K = 1$. For $g_K > 1$, the dipoles tend to align in parallel; for $g_K < 1$, the dipoles tend to align anti-parallel. g_K has a value of 2.5 for SPC/E water, which should be compared to the value for real water of 2.6.[36,37] For rigid dipoles in an infinite system, the g-factor is related to the permittivity by the Kirkwood equation

$$\frac{(\epsilon-1)(2\epsilon+1)}{9\epsilon} = yg_K. \tag{20}$$

In Sec. II C 3, it is shown that the leading contribution to the free energy from dipole correlation functions is proportional to the sum of the squares of the correlation functions. The water-water part of this contribution is denoted $\phi(r)$. The asymptotic form of $\phi(r)$ can therefore be expressed in terms of g_K as

$$\phi_\infty(r) = \lim_{r\to\infty}\phi(r) = -\frac{k_BT}{2}\left(\lim_{r\to\infty}h_{112}(r)\right)^2 \tag{21}$$

$$= -k_BT\frac{27g_K^2(\epsilon-1)^2}{16\pi^2(1+2\epsilon)^2\rho^2}\frac{1}{r^6} \tag{22}$$

$$= -k_BT\frac{27g_K^2}{64\pi^2}\left[1+O\left(\frac{1}{\epsilon}\right)\right]\frac{1}{\rho^2r^6}. \tag{23}$$

Equation (23) is an exact result for a fluid of rigid dipoles, and the strength of attraction is proportional to g_K^2. $\phi_\infty(r)$ becomes stronger if dipoles have a tendency to align more in parallel.

Note that g_K does not depend on the magnitude of the dipole μ, but the correlation of dipole orientations. It is possible for a fluid to have a large dipole moment but $g_K \approx 1$, or a small dipole moment but a large g_K.[37] The factor of g_K^2 therefore represents an amplification of the attractive $\phi(r)$ compared to an uncorrelated polar fluid of the same dipole moment. In SPC/E water, this amplification is a factor of $g_K^2 = 6.25$.

A simple physical picture of this effect is that each dipole moment is renormalized because it tends to carry with it parallel components of its neighbours' dipoles.

5. The contributions of ψ and ϕ to the free energy of solvation

Following the treatment of free energy changes given by Hansen and McDonald (see especially Secs. 3.4, 3.5, and 4.3 in Ref. 34), we can write the contribution of ϕ_∞ to the free energy of the fluid as

$$G_\phi = 4\pi\rho \int_0^\infty \mathrm{d}r\, r^2 \phi_\infty(r)\, g_{\mathrm{S}}^{\mathrm{ww}}(r). \tag{24}$$

A form more convenient to the analysis of simulations is the equation

$$G_\phi = \left\langle \sum_{i>j} \phi_\infty(r_{ij}) \right\rangle, \tag{25}$$

where the sum is over pairs of water molecules has been made explicit, and the average is over an NpT ensemble. We performed two separate calculations using Eq. (25) with $\phi_\infty(r)$ and $\phi(r)$ and find the results within k_B, similar to the errors in ΔS from particle insertion. A quantitative comparison of ϕ and ϕ_∞ appears below.

The Gibbs free energy contributed by solute-dipole correlation to solvation, denoted as ΔG_ψ, may be obtained similarly from the equation

$$\Delta G_\psi = G_\psi = 4\pi\rho \int_0^\infty dr\, \psi(r)\, g_{\mathrm{S}}^{\mathrm{sw}}(r) r^2, \tag{26}$$

where ρ is the number density of water.

The changes in G_ϕ and G_ψ upon addition of the solute, $\Delta G_\phi(T)$ and $\Delta G_\psi(T)$, are fit to the form[7]

$$\Delta G = \Delta C_{\mathrm{p}} \left(T - T_{\mathrm{h}} - T \log \frac{T}{T_{\mathrm{s}}} \right) \tag{27}$$

to extract the solvation entropy $\Delta S_\phi(T)$ by differentiation. This form has been shown to match experimental solvation free energies closely, and as shown below, also closely follows our calculated free energies.

D. Pair entropy

The excess entropy may be obtained from an infinite series of integrals, each of which depends on a single n-body correlation function,[23]

$$S^{\mathrm{ex}} = S - S^{\mathrm{id}} = S_2 + S_3 + \cdots, \tag{28}$$

where S is the total entropy of the fluid, $S^{\mathrm{id}} = 5/2 - \log(\rho\Lambda^3)$ is the ideal gas entropy, Λ is the thermal de Broglie wavelength, S^{ex} is the total excess entropy, and S_2 and S_3 are the two-body and three-body entropies. The pair correlation entropy has been calculated for simulation models of water[22,23] and has been shown to approximate well with the total excess entropy S^{ex}. We consider here only the two-body S_2 entropy, given by

$$\frac{S_2}{Nk_B} = -2\pi\rho \sum_{\alpha,\beta} x_\alpha x_\beta \int_0^\infty (g_{\mathrm{S}}^{\alpha\beta}(r) \log[g_{\mathrm{S}}^{\alpha\beta}(r)] - [g_{\mathrm{S}}^{\alpha\beta}(r) - 1]) r^2\, dr, \tag{29}$$

where $g_{\mathrm{S}}^{\alpha\beta}(r)$ is the RDF between species α and β, and ρ is the total number density. The change in S_2 upon solvation is given by

$$\frac{\Delta S_2}{k_B} = \left[\frac{S_2^{\mathrm{ww}}(\boldsymbol{r}^{N+1})}{k_B} + \frac{S_2^{\mathrm{sw}}(\boldsymbol{r}^{N+1})}{k_B} \right] - \frac{S_2^{\mathrm{ww}}(\boldsymbol{r}^N)}{k_B} \tag{30}$$

$$= \frac{\Delta S_2^{\mathrm{ww}}}{k_B} + \frac{S_2^{\mathrm{sw}}}{k_B}, \tag{31}$$

where $\boldsymbol{r}^N$ denotes the configuration of N water molecules, and $\boldsymbol{r}^{N+1}$ denotes the configuration of those molecules with the solute added, S_2^{ww} is calculated from only the water-water RDFs (O–O, O–H, H–H), and S_2^{sw} is calculated from the solute-water RDFs (solute–O, solute–H).

E. Configurational entropy from clusters

Godec and Merzel also recently discussed the entropy of many-body correlations in hydrogen bonded clusters.[18] They report the correlation entropy of different types of clusters in the solvation shell of hard-sphere solutes in SPC/E water. We note that our model solutes for Ne and Xe (see Sec. III) resemble Godec and Merzel's "s1" and "s2" solutes in that the size of the first and second solvation shells is the same. We compute from our Ne and Xe simulations the average number of each type of hydrogen bond (HB)-clusters in the first two solvation shells. The total change in entropy due to many-body correlations in HB-clusters can be written down as

$$\Delta S_{\mathrm{c}} = \sum_i \sum_{s=1}^{2} -\alpha_i \Delta S_i^{\mathrm{C}}(s) n_i(s), \tag{32}$$

where i is the type of HB-clusters as defined by Godec and Merzel, s is the first or second solvation shell, ΔS_i^C is the correlation entropy of clusters in the solvation shell relative to clusters in the bulk, and α_i is the degeneracy of each type of HB-cluster.

F. Hydrogen-bond models

The number of HBs per water molecule can be obtained from MD simulations as a function of distance r from the solute. Djikaev and Ruckenstein recently proposed a probabilistic hydrogen bond model.[38] They consider the difference between the numbers of HB per water molecule near a solute and in the bulk. The change in free energy due to this change in the number of hydrogen bonds can be written as

$$\Delta G_{\mathrm{HB}} = 4\pi\rho \int_0^\infty \frac{1}{2}\left[G_{\mathrm{HB}}(r) n_{\mathrm{HB}}(r) - G_{\mathrm{HB}}^{\mathrm{bulk}} n_{\mathrm{HB}}^{\mathrm{bulk}}\right] \times g_{\mathrm{S}}^{\mathrm{sw}}(r) r^2\, dr, \quad (33)$$

where $n_{\mathrm{HB}}(r)$ is the number of hydrogen bonds per water molecule as a function of distance r from the solute, and $G_{\mathrm{HB}}(r)$ is the Gibbs free energy per hydrogen bond. For Muller-type HB models,[39,40] $G_{\mathrm{HB}}(r) = \Delta H(r) - T\Delta S(r)$ is parametrized to be different in the first solvation shell and outside the first solvation shell, and so the model has four parameters. For other parameterisation of hydrogen bonds, no distinction is made between solvation shells and the bulk, so $G_{\mathrm{HB}}(r) = G_{\mathrm{HB}}^{\mathrm{bulk}} = \Delta H - T\Delta S$. Many parameterisation of ΔH and ΔS have been published.[40–42] In Djikaev and Ruckenstein's treatment, the temperature dependence of $G_{\mathrm{HB}}(r)$ is ignored, and so $G_{\mathrm{HB}}(r) = U_{\mathrm{HB}}$. In this treatment, the ratio $\Delta G_{\mathrm{HB}}/U_{\mathrm{HB}}$ is the number of HBs broken upon solvation, given by

$$\Delta n_{\mathrm{HB}} = \frac{\Delta G_{\mathrm{HB}}}{U_{\mathrm{HB}}} = 4\pi\rho \int_0^\infty \frac{1}{2}\left[n_{\mathrm{HB}}(r) - n_{\mathrm{HB}}^{\mathrm{bulk}}\right] g_{\mathrm{S}}^{\mathrm{sw}}(r) r^2\, \mathrm{d}r. \quad (34)$$

There are also several different definitions of hydrogen bonds.[18,43] We take the definition of Luzar and Chandler,[44] where two water molecules are said to be hydrogen-bonded if their O–O distance is less than 3.5 Å and the H-O-O angle is less than 30°. We also perform calculations using a different and more recent definition by Godec and Merzel,[18] where two water molecules are said to be hydrogen-bonded if their O-O distance is less than 3 Å and the O-H-O angle is greater than 150°. Other definitions include criteria based on H–O distances, other angular constraints, functional combinations of angular and distance constraints, and interaction energies. These definitions are discussed by Kumar *et al.*[43]

G. Information theory

We compare our simulation results also to scaled particle theory[27] (SPT) and information theory (IT).[26] SPT predictions of hydrophobic solvation free energy, change in entropy, and change in heat capacity are given in Ashbaugh and Pratt[27] as gradients with respect to solvent accessible surface area (SASA), i.e., $\partial\Delta G/\partial A$, $\partial\Delta S/\partial A$, and $\partial\Delta C_p/\partial A$. We convert these to $\Delta G/A$, $\Delta S/A$, and $\Delta C_p/A$ by numerical integration. IT predictions of hydrophobic solvation free energy are given in Hummer *et al.*[26] for noble gases as $\mu(T)$, modelled as hard sphere solutes in the SPC model of water. Changes in entropy and heat capacity are obtained by numerical differentiation.

III. SIMULATION METHODS

A. Molecular dynamics

Classical MD simulations were performed using the SPC/E[15] and TIP4P[45] models of water in the isobaric-isothermal (NpT) ensemble. MD simulations were performed using a Nosé-Hoover chain thermostat and barostat[46,47] with relaxation times of 1.5 and 2.4 ps, respectively. The electrostatic potential is calculated using the P3M method on a grid of $16 \times 16 \times 16$ with an Ewald convergence parameter $\eta = 0.18$.[48] Bond constraints are handled using the SHAKE algorithm,[49] and the integration is performed using a timestep of 1 fs. Systems of 512 water molecules are simulated at temperatures from 260 K to 360 K at 5 K intervals in the NpT ensemble at 1 atm pressure. Each system were equilibrated for 8 ns, and statistics were collected over 24 ns trajectories. Statistics are sampled at 1 ps (1000 MD steps), an interval selected using the block averaging method.[50] The structure of bulk water was also investigated at a higher level of theory using density functional theory MD (DFT-MD). The Carr-Parrinello MD (CPMD) program[51] was used to perform Born-Oppenheimer MD simulations using the BLYP functionals.[52] The simulation contained 128 water molecules at experimental density, simulated at 298 K in the NVT ensemble using a Nosé-Hoover chain thermostat[46,47] with a frequency of 1200 cm^{-1}. A plane-wave cutoff of 70 Ry was used. To increase the sampling efficiency of MD steps, the mass of the hydrogen was increased to 10 amu to reduce high-frequency oscillations, thus allowing a timestep of 3 fs. This optimisation has been shown to give the correct thermodynamic behaviour of water.[53] A total simulation length of 70 ps was performed, of which the first 50 ps was discarded as equilibration, and Wannier centres were calculated for the latter 20 ps of trajectory to give accurate dipole moments of water.[54,55]

We follow Paschek in the choice of models for hydrophobes: the solvation thermodynamics of Ne, Ar, Kr, Xe, and CH_4 are calculated using methods described above. The models are of the LJ type and are optimised for their pure component properties.[6,56] Their interaction potentials with water molecules (ϵ_{sw}, σ_{sw}) are obtained from the Lorentz-Berthelot mixing rules, $\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$ and $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$. Tabulated interaction potentials have been provided by Paschek,[6] and we include them as the supplementary material.[57]

B. Thermodynamic integration

We perform thermodynamic integrations (TIs) by "growing" a soft-sphere solute in SPC/E water, and then a second integration during which we turn on the electronic dispersion interactions between solute and surrounding waters. We define two thermodynamic paths, turning on the soft-sphere repulsion and turning on the van der Waals (vdW) attraction. The water-solute pair potentials are, respectively,

$$U_{\mathrm{ws}}(\lambda_1, 0; r) = 4\lambda_1 \epsilon_{\mathrm{ws}} \frac{\sigma_{\mathrm{ws}}^{12}}{r^{12}}, \quad (35)$$

$$U_{\mathrm{ws}}(1, \lambda_2; r) = U_{\mathrm{ws}}(1, 0; r) - 4\lambda_2 \epsilon_{\mathrm{ws}} \frac{\sigma_{\mathrm{ws}}^{6}}{r^{6}}, \quad (36)$$

where ϵ_{ws} and σ_{ws} are the usual Lennard-Jones parameters, and $0 \le \lambda_1, \lambda_2 \le 1$ are the switching parameters. The free energies of the two paths are the free energies of creating a soft-sphere cavity (ΔG_{ss}, repulsive) and inserting a van der Waals solute into the cavity then allowing the system to relax (ΔG_{vdw}, attractive)

$$\Delta G_{\mathrm{ss}} = \int_0^1 \mathrm{d}\lambda_1 \left\langle \frac{\partial U_{\mathrm{ws}}(\lambda_1, 0; r)}{\partial \lambda_1} \right\rangle_{\lambda_1, 0}, \quad (37)$$

$$\Delta G_{\mathrm{vdW}} = \int_0^1 \mathrm{d}\lambda_2 \left\langle \frac{\partial U_{\mathrm{ws}}(1,\lambda_2;r)}{\partial \lambda_2} \right\rangle_{1,\lambda_2}, \quad (38)$$

$$\Delta G_{\mathrm{sol}} = \Delta G_{\mathrm{ss}} + \Delta G_{\mathrm{vdW}}. \quad (39)$$

Thermodynamic integrations were performed at the temperatures 260 K, 275 K, 290 K, 298 K, 300 K, 305 K, 310 K, 325 K, and 350 K. We evaluate Eqs. (37) and (38) using the midpoint rule with 8 points for each of the integrals. Simulations were performed at a constant pressure of 1 atm and each consisted of 511 SPC/E water molecules and one solute. Statistics were collected for 1 ns after an initial 0.3 ns of equilibration.

C. Particle insertion

The solvation free energy of a solute can be obtained from MD simulations using the Widom's test-particle insertion (PI) method.[6,58] The solute of interest is inserted at position $\boldsymbol{x}_s$ into an equilibrated configuration of the pure solvent. Without allowing the system to relax, the change in potential energy ΔU on insertion is calculated. For MD simulations in the NpT ensemble, the solvation free energy ΔG_{sol}, equivalent to the excess chemical potential μ_{ex}, is given by[6]

$$\Delta G_{\mathrm{sol}} = -k_{\mathrm{B}}T \log\left(\frac{1}{\langle V\rangle}\left\langle \int_V d\boldsymbol{x}_s \exp(-\Delta U/k_BT)\right\rangle\right), \quad (40)$$

where $\langle\cdots\rangle$ represents the NpT ensemble average, V the volume of the simulation cell, and ΔU the change in potential energy on addition of a particle at $\boldsymbol{x}_s$. The simplest numerical scheme for approximating the integral in Eq. (40) is by integrating over an evenly spaced grid of insertion points. Since the simulation box fluctuates in size, care must be taken when considering the magnitude of $dV = |\mathrm{d}\boldsymbol{x}_s|$.

We choose a regular grid for $\boldsymbol{x}_s$ with mesh spacing of approximately 1 bohr. Decreasing the mesh spacing further does not improve the convergence of $\exp(-\Delta U/k_BT)$. At this mesh spacing, we find the rate of successful insertion per configuration ranges from 3 per solvent molecule for large solutes, such as Xe, to 20 per solvent molecule for small solutes, such as Ne, even though there are more than 200 insertions per solvent molecule. A successful insertion is defined as an insertion that gives a non-zero $\exp(-\Delta U/k_BT)$. To improve the computational efficiency of this calculation, we use the excluded volume map technique[59,60] by taking $\exp(-\Delta U/k_BT)$ to be zero for grid points that are within $0.7\sigma_{\mathrm{sw}}$ of oxygen sites. This eliminates the need to compute ΔU for 90% to 95% of grid points, thus increasing computational speed by a factor of 10 to 20. Solvation entropies and changes to heat capacity are calculated by fitting $\Delta G_{\mathrm{sol}}(T)$ to Eq. (27).

D. Accumulation of averages and distributions

The Boltzmann factors obtained from the particle insertion method, $\exp(-\Delta U/k_BT)$, can be used to obtain other quantities via the usual free energy perturbation (FEP) method. For any quantity q, its ensemble average in the solution $(N+1)$ of one solute in N solvent molecules is given by

$$\langle q\rangle_{\boldsymbol{x}_{N+1}} = \frac{\left\langle \int_V d\boldsymbol{x}_s\, q(\boldsymbol{x}_s;\boldsymbol{x}_N)\exp(-\Delta U/k_BT)\right\rangle_{\boldsymbol{x}_N}}{\left\langle \int_V d\boldsymbol{x}_s \exp(-\Delta U/k_BT)\right\rangle_{\boldsymbol{x}_N}}, \quad (41)$$

where $q(\boldsymbol{x}_s;\boldsymbol{x}_N)$ is some order parameter of the solute position $\boldsymbol{x}_s$ and the solvent positions $\boldsymbol{x}_N$. The same grid mesh and excluded volume map technique can be used here. We note that quantities calculated from Eq. (41) have significantly better convergence compared to those obtained from thermodynamic integration and related methods. This is since Eq. (41) samples 10^3–10^4 solute positions per MD frame, whereas explicit solute simulations sample only one. In addition to the calculation of averages, functions and probability distributions may also be calculated using the weights $\exp(-\Delta U/k_BT)$. Distributions of order parameters described previously in Secs. II A–II C 1 are calculated this way. This approach gives the same distribution as sampling along a molecular dynamics trajectory but is much more efficient. Note that a consequence of this method is that averages for smaller solutes, such as neon, will converge more quickly than those for larger solutes, such as xenon. This is because there are more accessible solvation sites for the smaller solute in the bulk structure.

IV. RESULTS AND DISCUSSION

A. Calculation of ΔG_{sol}

Before examining the performance of the entropy estimators (see Sec. IV B, methods in Sec. II), we describe the calculations used to obtain benchmark entropies.

The Widom test-PI method gives the change in excess Gibbs free energy for the solvation of a solute. The PI method may converge slowly for dense liquids,[61] as few insertions lead to acceptable solvation structure. We therefore compare $\Delta G_{\mathrm{sol}}(T)$ calculated by the PI method and by TI, where the solute is gradually "grown" to its full size. The PI method agrees with TI (see the supplementary material[57]).

The PI and TI methods converge to the correct free energy change for the chosen simulation model and, in principle, apply just as well to any laws of force between the particles. For this reason they do not alone provide much information about the physical mechanism of solvation. These methods are ideal as tests of other approaches. The relative performance of methods of calculation of the solvation entropy in predicting the PI result provides a test of the hypotheses on which the methods are based. As the PI method is relatively inexpensive computationally, we were able to obtain excellent statistics at many temperatures, and ΔS_{sol} is calculated to within about $1k_{\mathrm{B}}$.

Thermodynamic integration reveals the magnitude of free energy changes from turning on the repulsive and attractive terms in the Lennard-Jones potential. Figures 1(a) and 1(b) show these contributions to the free energy, calculated as described in Sec. III B. Individually, the repulsive and attractive components of ΔG are 2–5 times larger than ΔG_{sol}. This large cancellation takes place as the electronic dispersion forces draw the solvent closer to the solute.

The correlation between the size of the solute and the free energy of hydrophobic solvation is the basis of many analyses of the phenomena.[4] For example, the information theory treatment analyses the solvation free energy of a hard sphere, and so its application to real molecules or more realistic models requires the determination of the appropriate radius of the hard sphere. The cancellation of free energies in Figures 1(a)

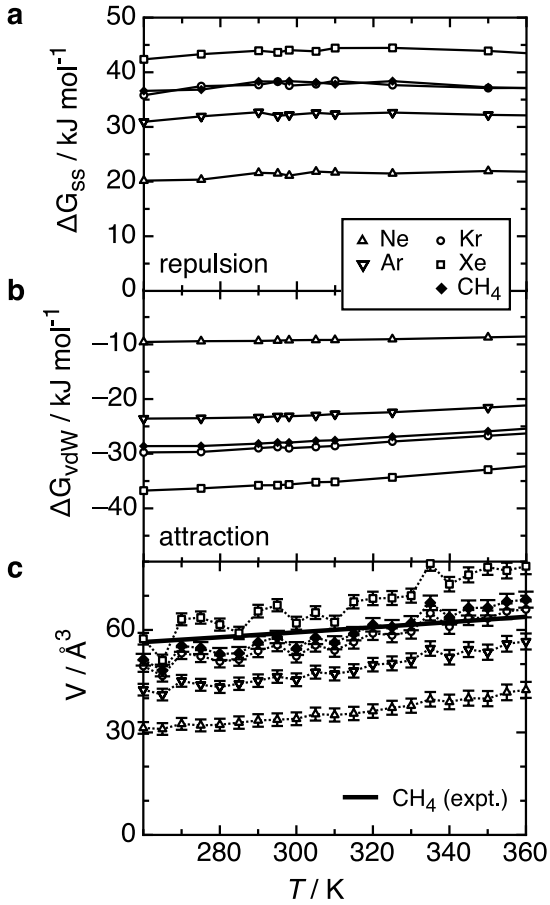

FIG. 1. (a) ΔG_{ss} from turning on the repulsive LJ potential, Eq. (37). (b) ΔG_{vdW} from turning on the attractive LJ potential, Eq. (38). ΔG_{ss} and ΔG_{vdw} calculated via thermodynamic integration, Sec. III B. (c) Partial molar volumes calculated via the particle insertion method, experimental methane data from Ref. 62.

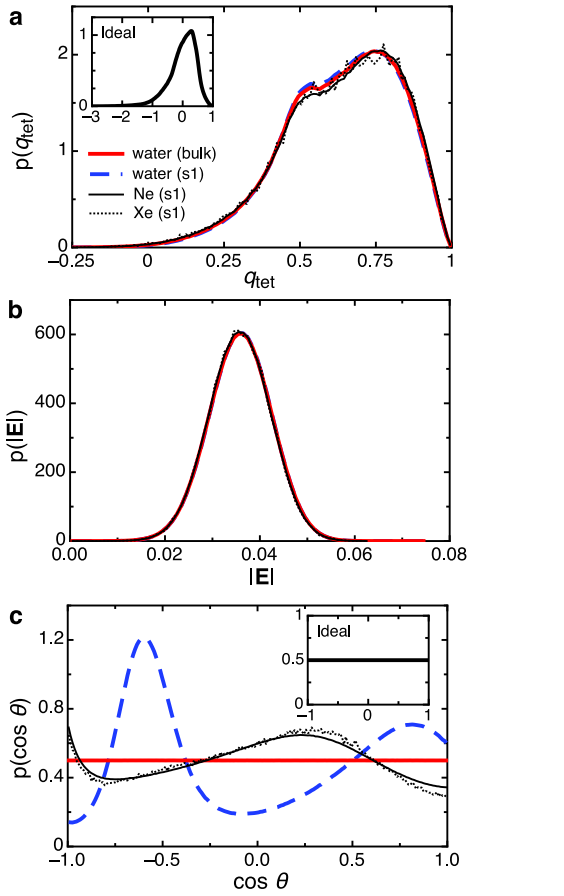

FIG. 2. (a) Distribution of the tetrahedral order parameter of water molecules, Eq. (1). (b) Distribution of the magnitude of the local electric field experienced by water molecules. (c) Distribution of the orientation of waters' dipole moment with respect to the solute. Distributions are given for water molecules in the first shell (s1) of a water molecule, a Ne atom, and a Xe atom. The bulk water distribution is also given. The ideal gas distributions are calculated using random coordinates.

and 1(b) suggests that the effective size depends not only on repulsive interactions but also on the strength of electronic dispersion attractions between the solute and the solvent. This behaviour has been examined in detail for a different simulation model by Gallicchio *et al.*[5] We return to these free energies in discussing information and scaled-particle theories in Sec. IV F.

The convergence of the PI method can be demonstrated by calculating the partial molar volumes of the solutes. We find that the simulation model of CH_4 agrees with the experimental partial molar volumes,[62] and all calculated partial molar volumes increase with temperature as expected (see Figure 1(c)).

B. Distributions of order parameters

1. Tetrahedral order

The distributions of the tetrahedral order parameter, q, calculated using Eq. (1), are shown in Figure 2(a). The peak in the distributions at $q \sim 0.75$ represents water with four neighbours in a tetrahedral arrangement. The shoulder in the distributions at $q \sim 0.5$ represents non-tetrahedrally arranged waters, which includes pentameric geometries that are not well characterised by Eq. (1). The inset shows the distribution of the tetrahedral order parameter for an ideal gas, which peaks at $q \sim 0.2$ and extends far into the negative region.

The change in tetrahedral ordering in the hydration shell of a hydrophobe is small. The statistical noise in "Xe (s1)" is comparable to its difference to bulk water, while the better converged "Ne (s1)" differs from the bulk water distribution by a few percent at most. The larger statistical noise in "Xe (s1)" is a consequence of the PI method, where the sampling efficiency decreases with solute size (see Sec. III D). Discernible differences between water molecules near a hydrophobe and in the bulk include a slight decrease of the shoulder at $q \sim 0.5$, accompanied by a slight broadening of the $q > 0.75$ tail. This is an overall enhancement of tetrahedrality near the solute, in agreement with previous reports[17] and experimental observations.[14]

The enhancement of tetrahedrality has been argued to contribute towards the entropy cost of hydrophobic solvation,[17,63,64] though the structural enhancement is not of an ice-like rigidity as initially hypothesised by Frank and Evans. We discuss in Sec. IV D the contribution of this enhancement of tetrahedrality towards the entropy and change in heat capacities of hydrophobic solvation, and its relative magnitude compared to other contributions.

2. The distribution of electric fields

Godec and Merzel discussed the distribution of electric field strengths experienced by a water molecule. Figure 2(b) shows this distribution for bulk waters and for those close to two types of hydrophobic solute. Any changes in this distribution are hardly perceptible. We therefore expect little change in entropy from this source, but return to this in Sec. IV D.

3. Leading order dipole ordering

Distributions of $\cos(\theta)$, where θ is the angle between a water dipole and the vector joining that water to a solute or another water molecule, are shown in Figure 2(c). The central particle is either the solute or another water molecule. There is a distinct difference between the distribution for waters near the solute and that for waters in the bulk. This change, far more pronounced than those described above for tetrahedral order or electric field, points to correlations in dipoles playing a significant role in the hydrophobic effect. However, Otten *et al.* examined this entropy for the case of waters solvating an ion at a planar surface and found it insignificant. We report and discuss the contribution of this distribution to the entropy for the case of waters near small hydrophobic solutes in Sec. IV D. We also examine higher order dipole correlation contributions to the entropy (see Secs. IV C 1 and IV D).

C. Pair functions

Figure 3 shows the functions that contribute to the free energy as sums over solute-water pairs or over water-water pairs.

1. Dipole contributions to the free energy

The function $\psi(r)$, shown in Figure 3(a), is above ~0.1 k_BT for a range of only ~0.5 Å outside the solute, and then decays roughly exponentially with a range of ~4 Å for each of the hydrophobes studied. The decay is faster than a power law. Note that this function contains contributions to the free energy of solvation, and to the entropy, which are not captured by the analysis of $\cos\theta$ discussed in Sec. IV B 3. For example, the contribution from h^{sw}_{202} included the tendency of the dipole moment to align with the vector joining solute to solvent, whether parallel or anti-parallel.

The function $\phi(r)$ is calculated from simulated coordinates via Eq. (15) and shown in Figure 3 (centre) for three simulation models of water. These are SPC/E, used elsewhere in this work, TIP4P, and a DFT-MD simulation (see methods Sec. III A). The asymptotic form, $\phi_\infty(r)$, from Eq. (23), was also calculated from SPC/E and real water parameters and is shown in Figure 3 (centre). The experimental ϕ_∞ agrees well with the DFT-MD ϕ, as does the SPC/E ϕ_∞ with SPC/E ϕ. We note that the SPC/E model underestimates $\phi_\infty(r)$ relative to that calculated from real water properties. Crucially, $\phi_\infty(r)$ is close to $\phi(r)$ across the entire range of interaction. The most slowly decaying part of $\phi(r)$ is known exactly, $\phi_\infty(r)$, and so this function may be used to estimate the water-water dipole correlation free energy with any distribution of waters.

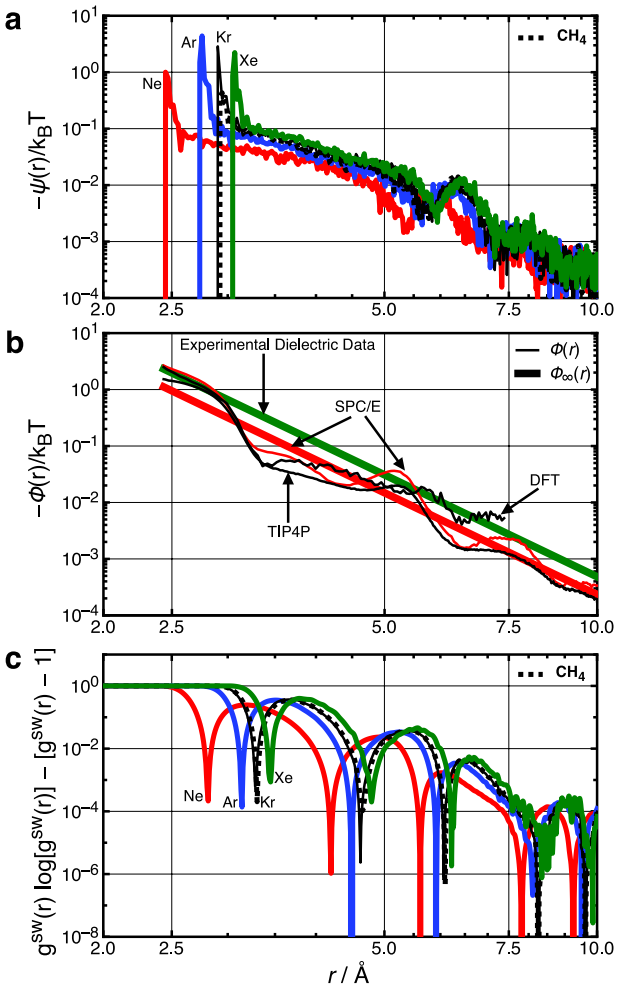

FIG. 3. Three types of functions of the distance between pairs of molecules from which we calculate the free energy or entropy. (a) The function $\psi(r)$, where r is the distance from solute to solvent that gives the contribution of solute-dipole correlations to the solvation free energy. (b) The function $\phi(r)$ as a function of distance between two water molecules, calculated using the asymptotic approximation ($\phi_\infty(r)$, Eq. (23)) and from MD simulations using a large basis set of rotational invariants ($\phi(r)$, Eq. (15)). Classical MD simulations were performed at 298 K and 1 atm pressure, DFT simulations were performed using BLYP functions and Born-Oppenheimer dynamics. Convergence of $\phi(r)$ is achieved when the addition of further basis functions does not change the result. (c) The integrand in the expression for the solute-water contribution to the pair entropy, S_2^{sw} for each solute studied.

The full function $\phi(r)$, in contrast, requires the calculation of correlation functions and, therefore, is probably only useful in the analysis of simulations. We therefore report below results calculated from $\phi_\infty(r)$ (see Sec. II C 5). The difference in entropies found from ϕ and ϕ_∞ is smaller than the uncertainty in the particle insertion result.

2. Pair entropy

The calculation of ΔS_2 involves integration over $g_S^{\alpha\beta}(r) \log [g_S^{\alpha\beta}(r)] - [g_S^{\alpha\beta}(r) - 1]$. The absolute value of this function, for solute-water correlations, is shown in Figure 3, for comparison to the functions ψ and ϕ. Note that S_2 depends on pair correlations in the number density, where ψ and ϕ depend on dipole correlations.

3. Temperature dependence of dipole correlation contributions

Figure 4 (top) shows the free energy of solute-dipole correlation $\Delta G_\psi(T)$ calculated via Eq. (26). This free energy arises from the tendency for dipoles to align near the influence of the solute, described by the coefficients of Φ_{n0n} in the expansion of Eq. (9). The Φ_{n0n} are functions of $\cos\theta$, where θ is the angle between water's dipole moment and the displacement between water and the solute. Of the known changes in structure near a hydrophobe that captured by the correlation functions in $\psi(r)$ is the most thermodynamically significant contribution examined so far. Note though that this contribution to the free energy is negative. Its contribution to the entropy is also negative and is discussed in Sec. IV D.

Figure 4 (bottom) shows also the contribution of ϕ to the solvation free energy, $\Delta G_{\rm sol}$, via Eq. (25), given in the order of their size. The curvature of ΔG_ϕ increases with size, consistent with the observation that the experimental $\Delta C_{\rm p}$ of solvation increases also with size.[8] ΔG_ϕ includes only the effect of water-water dipole correlations, that is, of the dispersion attraction of correlated permanent dipoles. Other important changes, such as vdW attraction from solute-solvent interaction, are not included. ΔG_ϕ is not expected to account for the full solvation free energy $\Delta G_{\rm sol}$ but captures its temperature dependence. This dependence is examined in further detail below. Figure 4 shows also fits of Eq. (27) the calculated $\Delta G_\phi(T)$ and $\Delta G_\psi(T)$.

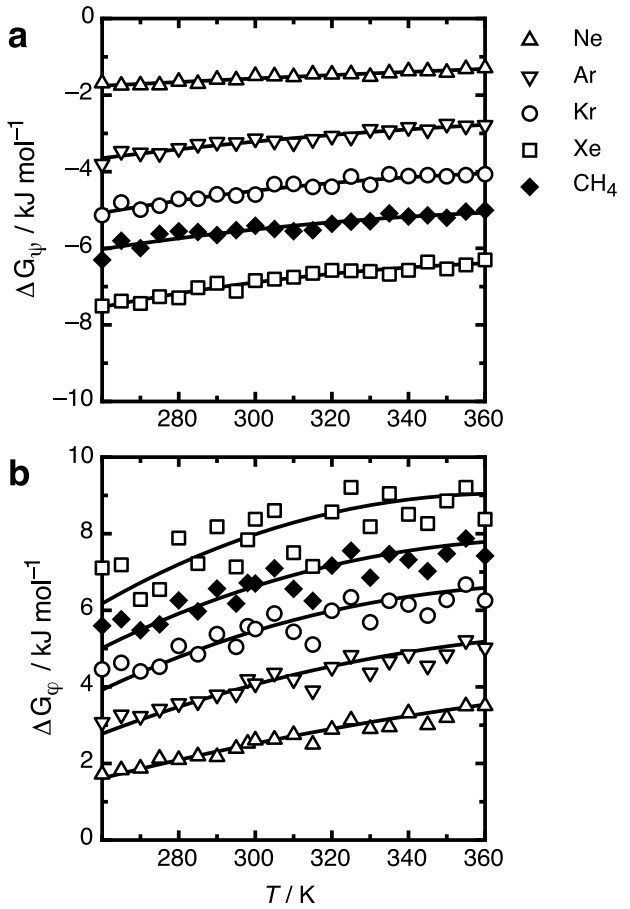

FIG. 4. The contributions of $\phi(r)$ and $\psi(r)$ to the free energy of solvation, denoted ΔG_ϕ and ΔG_ψ, calculated for model solutes in SPC/E water at 1 atm. Solid curves are fits to Eq. (27).

D. Comparison of parameter-free models

Different entropic contributions for the solvation of Ne and Xe are given in Figure 5, where entropies based on order parameters (symbols) are compared to the benchmark PI solvation entropy (dashed black lines). The comparison for the remaining solutes is provided in the supplementary material.[57] The entropy of the different contributions at 300 K is given in Table I. The change in partial molar heat capacity on solvation at 300 K ($\Delta C_{\rm p}$ from Eq. (27), related to the gradient of entropies with temperature), is given in Table II. We find that the tetrahedral ordering, the orientation of dipole moments, and the magnitude of electric fields all give entropies far smaller than the solvation entropy.

The hydrophobic effect has been explained in terms of structural changes in water, specifically the entropy from tetrahedral ordering in water.[4,7] Recent experimental evidence shows the structure of water molecules near a hydrophobe is indeed different to bulk water.[14] Galamba reports the tetrahedral order parameters of simulated water molecules near to and far from a solvated hydrophobe and also shows that the structure of water near hydrophobes is different.[17] We too find differences in the distribution of the tetrahedral order parameter near and far from the hydrophobe and confirm these reports that the structure of water near a hydrophobe is different to bulk water. However, we have directly calculated here the entropy cost of this enhancement of tetrahedral ordering near hydrophobes. The tetrahedral entropy $\Delta S_{\rm tet}$ does not show a negative entropy decreasing with temperature, and its magnitude is comparable to the magnitude of the uncertainty (see Table I). Contribution of $\Delta S_{\rm tet}$ to the hydrophobic solvation entropy is at least an order of magnitude smaller than other contributions, despite the change in tetrahedral order for solvation waters (Sec. IV B 1 and Figure 2, see also Ref. 17). On the other hand, the $\Delta C_{\rm p}$ of tetrahedral order to appears to be significant (see Table II). This suggests a possible interpretation that the enhancement of tetrahedral ordering around a solute does not significantly contribute to the hydrophobic solvation entropy but may contribute to the hydrophobic change in heat capacity.

Otten *et al.* have previously calculated the change in entropy due to dipole orientation (ΔS_θ) for the adsorption of an ion to an interface and found that this entropy is negligibly small compared to the magnitude of the effect.[19] Our results show that in the solvation of small hydrophobic particles, this leading order orientational correlation has an entropy that is small compared to the size of the effect but significantly larger than the contribution of tetrahedral order. Its contribution to the heat capacity is also small but significant. Crucially, both ΔS_θ and $\Delta C_{{\rm p},\theta}$ show a dependence with the size of the solute. This hints at dipole correlations being an important factor in hydrophobic solvation, for which ΔS_θ and $\Delta C_{{\rm p},\theta}$ are the leading order contributions. We examine higher order dipole correlations below.

ΔS_ψ may be interpreted as the entropy arising from the alignment of water molecules under the influence of solute, since it is calculated from distributions of functions of $\cos\theta$. The first term in $\psi(r)$, $-k_B T h_{101}(r)^2/(2g(r)^2)$, may be considered to capture orientational order similar to that captured by ΔS_θ, since the rotational invariant $\Phi_{101} = \cos\theta$ is the same as

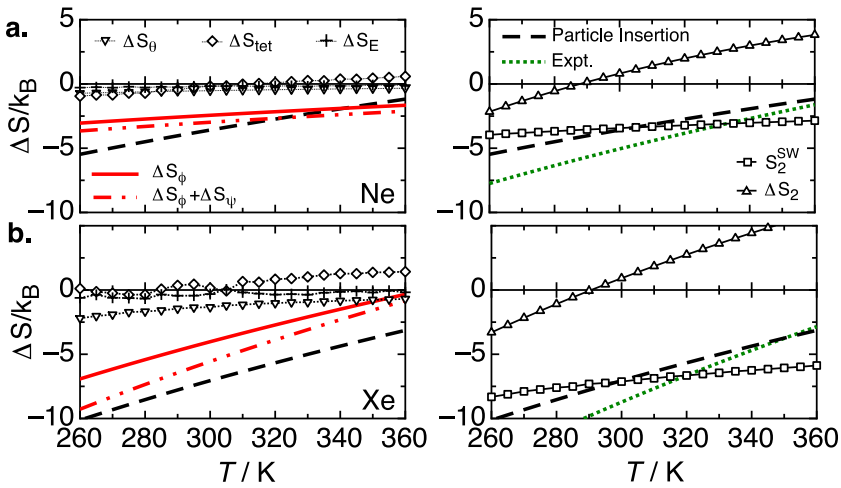

FIG. 5. Entropic contributions (Sec. II) compared to the "benchmark" solvation entropies from the particle insertion method (dashed). Left: contributions from dipole orientation (ΔS_θ, ▽), electric field strength (ΔS_E, +), tetrahedral order (ΔS_{tet}, ◇), and higher-order dipole correlations (ΔS_ϕ, solid and $\Delta S_\phi + \Delta S_\psi$, dotted-dashed). Right: pair correlation entropies from only the solute-water RDF (S_2^{sw}, □) and all RDFs (ΔS_2, △). Experimental entropies (dotted) from Ref. 70. Plots for solutes not shown here are provided in supplementary material.[57]

TABLE I. Contributions to the solvation entropies of small gas molecules from MD simulations. Experimental entropies are obtained from gas phase solubilities.[6,70] Particle insertion entropies are the benchmark values for simulations. Symbols for the different entropic contributions are defined in Sec. II. Estimated uncertainties for dipole correlations are obtained from standard errors of fit parameters (see Eq. (27)), estimated uncertainties for other entropies obtained by integrating the magnitude of statistical fluctuations over the RDF. All entropies are given at 300 K and 1 bar in units of k_B.

	Ne	Ar	Kr	Xe	CH_4
			Benchmarks		
Expt.	−5.04	−7.36	−8.14	−8.72	−7.55
Particle insertion	−3.59	−5.54	−6.29	−7.05	−6.17
			Dipole correlations		
$\Delta S_\phi + \Delta S_\psi$	−2.9 ± 0.1	−4.3 ± 0.1	−5.0 ± 0.2	−5.5 ± 0.3	−5.0 ± 0.2
ΔS_ϕ	−2.4 ± 0.1	−3.2 ± 0.1	−3.6 ± 0.2	−4.0 ± 0.3	−3.8 ± 0.2
ΔS_ψ	−0.54 ± 0.02	−1.12 ± 0.04	−1.38 ± 0.04	−1.54 ± 0.04	−1.26 ± 0.06
			Other contributions		
ΔS_2	+0.81	+0.52	+0.63	+0.98	+0.92
ΔS_θ	−0.53	−0.81	−0.98	−1.34	−1.00
ΔS_E	−0.17	−0.18	−0.18	−0.32	−0.19
ΔS_{tet}	−0.17	+0.03	+0.11	+0.16	+0.13
ΔS_c	−3.4			−3.3	
Estimated uncertainty	±0.1	±0.2	±0.2	±0.4	±0.3

TABLE II. Contributions to ΔC_p of solvation of small gas molecules from MD simulations. Experimental change in heat capacities is obtained from gas phase solubilities.[6,70] Particle insertion entropies are the benchmark values for simulations. Symbols for the different contributions are defined in Sec. II. Estimated uncertainties are obtained from standard errors of fit parameters (see Eq. (27)). All ΔC_p are given at 300 K and 1 bar in units of k_B.

	Ne	Ar	Kr	Xe	CH_4
			Benchmarks		
Expt.	18.8	26.8	27.8	21.9	32.2
Particle insertion	13.2 ± 0.2	17.3 ± 0.6	19.2 ± 1.0	21.4 ± 1.6	19.6 ± 1.1
			Dipole correlations		
$\Delta C_{p,\phi} + \Delta C_{p,\psi}$	4.8 ± 1.4	12.6 ± 4.4	19.7 ± 3.6	25.9 ± 6.5	20.2 ± 4.1
$\Delta C_{p,\phi}$	4.3 ± 1.3	10.0 ± 4.3	14.6 ± 3.5	20.2 ± 6.4	15.5 ± 3.8
$\Delta C_{p,\psi}$	0.5 ± 0.5	2.6 ± 0.9	5.1 ± 1.0	5.7 ± 1.0	4.7 ± 1.5
			Other structural contributions		
$\Delta C_{p,2}$	18.3 ± 0.6	24.0 ± 1.5	27.6 ± 2.0	32.7 ± 3.1	28.5 ± 2.1
$\Delta C_{p,\theta}$	1.2 ± 0.1	2.0 ± 0.3	2.6 ± 0.7	4.9 ± 1.5	2.7 ± 0.7
$\Delta C_{p,E}$	0.5 ± 0.1	0.7 ± 0.1	0.9 ± 0.2	1.5 ± 0.3	0.9 ± 0.2
$\Delta C_{p,tet}$	4.9 ± 0.1	5.7 ± 0.2	5.9 ± 0.3	5.8 ± 0.5	6.0 ± 0.3

the order parameter in ΔS_θ. We find the entropy ΔS_ψ is negative, consistent with the negative entropy of hydrophobic solvation, though the magnitude of ΔS_ψ is about one quarter of the simulated benchmark entropy, in the range of $-0.5 \sim -2k_B$. Though this contribution to the entropy is relatively small, it is not negligible and its temperature dependence does approach zero from below as temperature increases, following the changes in structure seen by Davis *et al.*[14] The presence of this entropy confirms in part the hypothesis of Frank and Evans, in that a tendency to order of waters near the hydrophobic solute is partly responsible for the entropy of solvation.

Figure 5 (left column) shows the temperature dependence of the contribution to the entropy of water-water dipole correlations, $\Delta S_\phi(T)$. The solvent-solvent component ΔS_ϕ approaches zero from below with heating and vanishes near the boiling point of water, both features of the hydrophobic solvation entropy. The left-hand column of Figure 5 also compares the sum of the two dipole correlation entropies to the simulated solvation entropies. The two dipole correlation terms together, $\Delta S_\phi + \Delta S_\psi$, account for the majority of the solvation entropy across all solutes studied also see Table I.

Dipole correlations, ϕ, make a large contribution to the ΔC_p of solvation, given as $\Delta C_{p,\phi}$ in Table II. Along with the solute-dipole correlation ψ, $\Delta C_{p,\phi} + \Delta C_{p,\psi}$ makes up the majority of ΔC_p of solvation. This is also evident in the similar gradients between $\Delta S_\phi + \Delta S_\psi$ and the benchmark particle insertion $\Delta S(T)$ in Figure 5. We note that the large uncertainty estimates for $\Delta C_{p,\phi}$ are due to the difficulty in converging a second derivative. With the exception of Ne and Ar, $\Delta C_{p,\phi} + \Delta C_{p,\psi}$ agrees very well with the benchmark particle insertion ΔC_p, though in the case of Ar, the error margins are just overlapping. It is possible that another source of ΔC_p makes a significant contribution for the smallest solutes, such as $\Delta C_{p,tet}$, or other contributions not considered in this study.

Nevertheless, the dipole correlation contribution $\Delta S_\phi + \Delta S_\psi$ is the only contribution that gives reasonable agreement with both the entropy and change in heat capacity of hydrophobic solvation across a wide range of temperature and for a variety of solutes. Calculations of entropy based on ψ and ϕ are not based on any freely adjustable parameters. The temperature dependence emerges without additional assumptions, and matches thermodynamic[3,7] and spectroscopic experiments.[14]

In the right-hand column of Figure 5, the change in two-body correlation entropy upon solvation, ΔS_2 (symbols), is compared with the benchmark PI solvation entropy (black dashed lines) and the experimental entropy (dotted green lines). Near room temperature, the magnitude of S_2^{sw} is very close to the benchmark PI ΔS, but S_2^{sw} underestimates the temperature dependence. The temperature gradient of ΔS_2 has the correct sign, but the gradient is ~50% steeper than the temperature gradient of the benchmark ΔS. Cancellation between the solute-water and water-water contributions to ΔS_2 yields an overall entropy that is in the range from 0 to 1 k_B at 300 K (see Table I).

Chakravarty and coworkers have shown in their simulations that S_2 is a good estimator for the total excess entropy s^{ex} in pure water (see Eq. (28)).[22,23] We note that this result is not in contradiction with the findings of Chakravarty and coworkers. While S_2 agrees well with the total s^{ex} of the liquid across a much larger section of the phase diagram, hydrophobic solvation entropies arise from the cancellation of large s^{ex} of a bulk liquid and a dilute solution to yield a small solvation entropy.

Borgis and coworkers recently published a molecular density functional theory (MDFT) for modelling point-charge models of water.[65,66] The MDFT approach considers the excess free energy of a liquid in terms of the number and polarisation density up to the second order but required a three-body correction to reproduce the tetrahedral structure of water.[66] Our treatment of ϕ and ψ also begins by considering the polarisation density, but we capture many-body interactions by expanding correlation functions in the basis set of rotational invariants. We find correlations in polarisation density, described by ϕ, accounts for a majority of the ΔS and ΔC_p of solvation. Contributions from cross terms in density and polarisation density,[65] described by ψ, are also significant, and the combination of ϕ and ψ gives an excellent agreement for the signature temperature dependence of hydrophobic solvation. Crucially, our treatment differs from the MDFT approach as we consider a molecular mechanism for the hydrophobic solvation, that of pairwise attraction between rotating dipole moments on water molecules.

E. Hydrogen bonding models

The recently published probabilistic hydrogen bond model by Djikaev and Ruckenstein is based on classical density functional theory, which adds the free energy of HBs to the grand thermodynamic potential.[38] Djikaev and Ruckenstein's model is able to reproduce the size dependence of the hydrophobic effect, including the volume-to-surface-area cross-over. Djikaev and Ruckenstein did not include any temperature dependence of the energy of HB, and so the temperature dependence of their solvation energy is proportional to temperature dependence of the number of HBs broken upon solvation, $\Delta n_{HB}(T)$ (see Eq. (34)). The left-hand panels of Figure 6 show $\Delta n_{HB}(T)$ calculated from our calculations, using two different definitions of HBs. We find that $\Delta n_{HB}(T)$ is a decreasing function with temperature, and so the free energy $\Delta G_{HB}(T) = U_{HB}\Delta n_{HB}(T)$ is an increasing function with temperature for an energetically favourable HB. This implies that $\Delta G_{HB}(T) = \Delta H - T\Delta S$ must have a negative ΔS, in agreement with the observation that hydrophobic solvation has a negative entropy.

There is little consensus in the literature on the free energy of hydrogen bond formation, or its temperature dependence. Silverstein *et al.* tabulated 10 different parameterisations of the HB free energy, $G_{HB} = H_{HB} - TS_{HB}$ from spectroscopic data.[40] Suresh and coworkers recently re-parameterised the HB free energy using dielectric data.[41,42] Additionally, the two-state models, where the HB free energy in the solvation shell is parameterised separately from the bulk, have been parameterised by Muller and Silverstein.[39,40] Kumar *et al.* reviewed a large range of geometry definitions based on distance, angular and energetic constraints,[43] and showed that these different definitions can give widely varying results.

We select some of the popular choices of HB free energy: the Muller two-state model,[39] the recent re-parameterisation by Suresh and coworkers,[42] and the parameterisation by Hare and

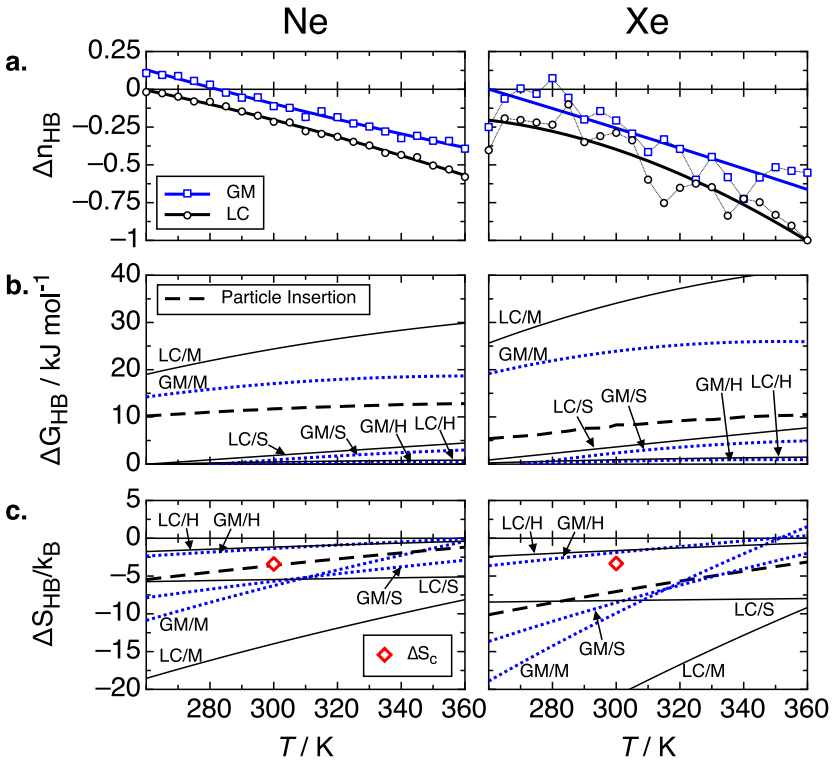

FIG. 6. (a) Change in the number of hydrogen bonds upon solvation (Eq. (34)), using geometric definitions of hydrogen bonds from Luzar and Chandler[44] and the Godec and Merzel.[18] (b) Free energy change from broken hydrogen bonds. The hydrogen bond free energy is parameterised by Muller,[39] Suresh *et al.*[41,42] and Hare and Sorensen.[67] (c) Entropy change from broken hydrogen bonds. Many-body correlation entropies from hydrogen bond clusters, ΔS_c, are calculated according to Godec and Merzel (Eq. (32)).[18]

Sorensen from Raman spectroscopy[67] that was noted to be the only parametrization to be consistent with the heat capacity of water.[40] For these choices of HB free energies, we count the hydrogen bonds in our simulations using the Luzar and Chandler[44] and the Godec and Merzel[18] definition of hydrogen bonds and calculate the free energy. We compare these results to the benchmark particle insertion solvation free energies and entropies (see centre and right-hand panels in Figure 6). We find that ΔG_{HB} and ΔS_{HB} are highly sensitive to both the choice of the HB free energy, as well as the choice of the HB geometry definition. The few calculations attempted here do not represent the total spread in ΔG_{HB} and ΔS_{HB} if all combinations of HB free energy parameterisations and HB geometry definitions were compiled. The parametrization of Hare and Sorensen,[67] though consistent with the heat capacity of water, significantly underestimates the hydrophobic solvation entropy. The choice of HB models corresponds to a sufficiently wide range of free energies that it amounts to a fit, though no single HB model that we are aware of is consistent with both bulk and solvation properties.

Godec and Merzel reported that the many-body correlation entropy of H-bonded clusters near a hydrophobic solute is significantly enhanced compared to the same clusters in the bulk.[18] We obtain a total entropy change ΔS_c by calculating the number of each type of H-bonded clusters in the solvation shells of hydrophobic solutes from MD simulations. We find that ΔS_c has the correct sign and is close to the benchmark particle insertion entropy for neon but does not grow with the solute size. The predictions for neon and xenon are similar.

F. Information theory

IT predictions of hydrophobic solvation thermodynamics are given in Hummer *et al.*[26] for noble gases and methane in SPC water. Due to the similarity of solvation thermodynamics for many hydrophobes of different sizes and compositions, the hydrophobic part of solvation is thought to be due to the work required to open up a cavity in water.[5,26,68,69] Solutes in SPT and IT are hard-spheres with radius R, where R is the distance of closest approach between the solute center and the oxygen atom on a water molecule. They capture well the hydrophobic cavity formation.

This cavity free energy is similar to the free energy of turning on the repulsive force of the LJ solute, ΔG_{SS} (see Figure 1(a)). Figure 7(a) shows that ΔG_{SS} is closest to the IT/SPT estimate of the solvation free energy of the hard sphere. Our model solutes, following Paschek,[6] use a LJ forcefield. For these solutes R is taken to be the location of the first maximum in the solute-oxygen RDF. The solvent-accessible surface area is taken to be $A = 4\pi R^2$.

The total solvation free energy, $\Delta G_{SS} + \Delta G_{vdW}$, for the Lennard-Jones solute is the result of cancellation between a positive cavity free energy and a large negative vdW attractive part.[5] SPT and IT do not include ΔG_{vdW}, so their predictions need not coincide with the experimental and PI solvation free energies. It is for this reason that the IT/SPT predictions are much higher than experiment (Figure 7(a)), and this should not be taken as a failing of the IT/SPT approach. Similarly, the dipole correlation contribution does not include other influences and so is lower than the experimental and particle insertion solvation free energies.

SPT and IT have been shown to predict well the hydrophobic thermodynamics of small solutes,[7] as is confirmed in Figures 7(b) and 7(c). The IT/SPT predictions for the solvation entropy are similar to the experimental entropy, but roughly twice the particle insertion SPC/E value. Note that real solutes experience forces that attract them to the solvents, but this attraction is neglected in the SPT/IT result. The agreement

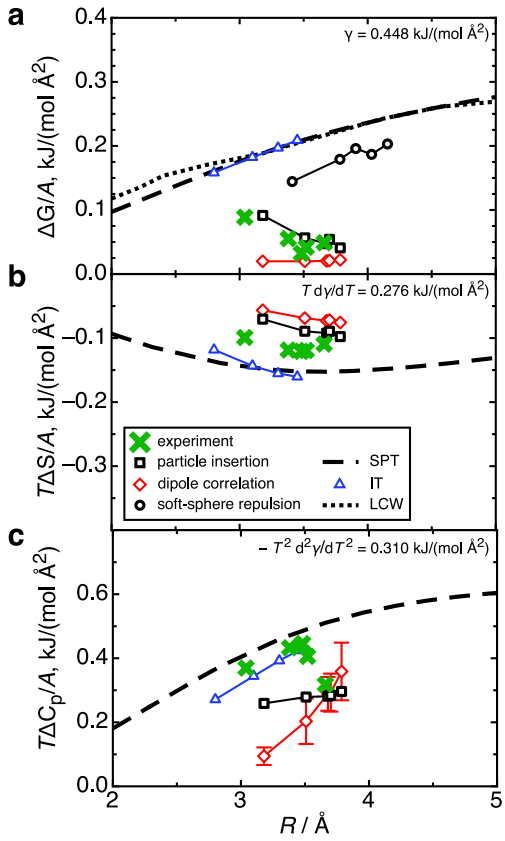

FIG. 7. The size-dependence of scaled particle theory[27] (dashed), information theory[39] (△), Lum-Chandler-Weeks theory[25] (dotted), experimental solubilities[70] (×), and results from this work via the particle insertion method (□), thermodynamic integration of a soft-sphere solute (○), and dipole correlations $\phi(r)$ (◇). (a) Solvation free energy per unit surface area. (b) Solvation entropy per unit surface area. (c) Solvation change in heat capacity per unit surface area. The macroscopic surface tension and other relevant quantities are given in the top right corner of each panel.

with experiment is therefore due in part to a cancellation of errors. The best agreement is between the dipole correlation estimate and particle insertion estimates of solvation entropy (Figure 7(b)).

SPT and IT also give the long-range interfacial behaviour of water. The asymptotic value of SPT is the macroscopic surface tension by definition. Their predictions at the intermediate length-scales are obtained by fitting higher order corrections to the pressure in a cavity, such as the Tolman length, which are adjustable parameters. In order to capture the temperature dependence, separate fits of adjustable parameters at each temperature are necessary.[27] The ability of IT and SPT to describe hydrophobic solvation across a wide range of length scales is useful.

The hydrophobic solvation free energy exhibits a volume proportionality at small scales, and a surface area proportionality at large scales.[11,13,25] The crossover in this scaling behaviour occurs at ~1 nm in water. This crossover length is well reproduced by SPT and IT.[26,27] Huang *et al.* showed using simulations of hard sphere solutes in Lennard-Jones fluids that the volume to surface area crossover can emerge from dispersion forces.[13] As shown above, the dipole correlation contribution to the entropy reproduces well the simulated temperature dependent solvation entropy for five model solutes. It is dominated by a dispersion interaction, ϕ_∞, and the contribution of a dispersion interaction to the free energy of solvation crosses from volume to area dependence with solute radius (see supplementary material[57] or Ref. 13 for more details). The crossover is at around three or four molecular diameters, or in the case of water around 1 nm. We defer detailed examination of larger scales to a later paper, but note that the dipole correlation entropy provides a possible explanation for the crossover short to long length scale behaviour, as it captures both the temperature and size dependences.

V. CONCLUSIONS

The Frank and Evans hypothesis states that the negative entropy of hydrophobic solvation near room temperature is due to structuring around the hydrophobic solute. We confirm the simulation result that tetrahedral order, quantified as q (see Sec. II A), is increased near the hydrophobe,[17] but find the associated entropy relatively small (see Figures 2 and 5). The only changes to water structure and alignment around the hydrophobe that we find contribute substantially to the entropy are those captured by the correlation functions that contribute to $\psi(r)$ (see Figure 3). These changes in the orientations of water molecules are detected by solute-correlated Raman experiments[14] because they affect the interactions of water molecules with the solute and with other solute neighbours. This contribution, with higher order terms, is about one quarter of the simulated benchmark entropy of solvation. Though the Frank-Evans hypothesis explains the hydrophobic effect in part, the relevant structure is not at all as expected. We note that the entropy associated with order is only around one quarter of the total.

A further contribution is obtained by considering correlations between water molecules' dipoles, included in the function $\phi_\infty(r)$. Analysis of the asymptotic behaviour of these correlations with a standard result identifies it as a dispersion force, modified by the correlations captured in the Kirkwood g-factor. This allows the estimation of this contribution for any distribution of water molecules. The sum of this entropy, ΔS_ϕ, and ΔS_ψ, is within about $1.2k_B$ of the benchmark entropy at 300 K for all solutes studied, close to "chemical" accuracy. It remains in agreement with the benchmark across conditions studied and exhibits the distinctive temperature dependence of measured hydrophobic entropies. No free parameters are required for the calculation of the dipole correlation entropy, and it is obtained by systematic expansion of the free energy in correlation functions. A crucial distinction between the two contributions to dipole correlation entropy is their range. The bulk of the contribution of ψ is within 1 Å of the hydrophobic solute, while the contribution of ϕ is a sum over all water pairs, with the relatively slow r^{-6} decay characteristic of dispersion interactions.

The wide variety of possible models based on hydrogen bonding makes it difficult to assess the approach. Of the models examined, we find that the Godec-Merzel definition, with the Suresh hydrogen bond free energy most closely reproduces the solvation entropy (Figure 6). It has been argued elsewhere that the only hydrogen bond model consistent with the heat capacity of water is that of Hare and Sorenson.[67] We have been

unable to find a hydrogen bond model that captures both hydrophobic solvation entropy and bulk water thermodynamics.

The information and scaled particle theories are based on simulated water properties at small scales and experimental properties at large scales. They are connected to the application of the particle insertion method to a hard sphere solute and so are capable of accurate prediction of all the thermodynamics of solvation provided that the appropriate hard sphere radius, or other information on solute shape can be included. An important feature of information and scaled particle theories is the crossover from small to large length scales. We have examined most closely the temperature dependence of the solvation entropy for relatively small solutes, but the dipole correlation entropy is dominated by a dispersion interaction, and this leads naturally to a length-scale crossover. We defer a more detailed investigation of larger length scales, but note that the dipole correlation mechanism is consistent with information theory.

Our key results are that the hydrophobic effect, i.e., the signature, temperature-dependent, solvation entropy of nonpolar molecules in water, is largely due to a dispersion force arising from correlations between rotating permanent dipole moments, that the strength of this force depends on the Kirkwood g-factor, and that the strength of this force may be obtained exactly without simulation.

ACKNOWLEDGMENTS

This research was supported by a grant from the Australian Research Council (Grant No. DP110103388), and by grants of computer time from the Victorian Life Sciences Computation Initiative (Grant No. VR0252), and from the National Computational Infrastructure-National Facility.

[1]J. A. V. Butler, Trans. Faraday Soc. **33**, 229 (1937).
[2]D. D. Eley, Trans. Faraday Soc. **35**, 1281 (1939).
[3]S. J. Gill, N. F. Nichols, and I. Wadsö, J. Chem. Thermodyn. **8**, 445 (1976).
[4]C. Tanford, in *The Hydrophobic Effect: Formation of Micelles and Biological Membranes* (Wiley Interscience Publications, Wiley, 1980).
[5]E. Gallicchio, M. M. Kubo, and R. M. Levy, J. Phys. Chem. B **104**, 6271 (2000).
[6]D. Paschek, J. Chem. Phys. **120**, 6674 (2004).
[7]N. T. Southall, K. A. Dill, and A. D. J. Haymet, J. Phys. Chem. B **106**, 521 (2002).
[8]R. S. Spolar, J. H. Ha, and M. T. Record, Proc. Natl. Acad. Sci. U. S. A. **86**, 8382 (1989).
[9]K. P. Murphy, P. L. Privalov, and S. J. Gill, Science **247**, 559 (1990).
[10]D. Villers and J. K. Platten, J. Phys. Chem. **92**, 4023 (1988).
[11]D. Chandler, Nature **437**, 640 (2005).
[12]N. T. Southall and K. A. Dill, J. Phys. Chem. B **104**, 1326 (2000).
[13]D. M. Huang, P. L. Geissler, and D. Chandler, J. Phys. Chem. B **105**, 6704 (2001).
[14]J. G. Davis, K. P. Gierszal, P. Wang, and D. Ben-Amotz, Nature **491**, 582 (2012).
[15]H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[16]H. S. Frank and M. W. Evans, J. Chem. Phys. **13**, 507 (1945).
[17]N. Galamba, J. Phys. Chem. B **117**, 2153 (2013).
[18]A. Godec and F. Merzel, J. Am. Chem. Soc. **134**, 17574 (2012).
[19]D. E. Otten, P. R. Shaffer, P. L. Geissler, and R. J. Saykally, Proc. Natl. Acad. Sci. U. S. A. **109**, 701 (2012).
[20]W. H. Keesom, Phys. Z. **22**, 129 (1921).
[21]R. E. Nettleton and M. S. Green, J. Chem. Phys. **29**, 1365 (1958).
[22]M. Agarwal, M. P. Alam, and C. Chakravarty, J. Phys. Chem. B **115**, 6935 (2011).
[23]R. Sharma, M. Agarwal, and C. Chakravarty, Mol. Phys. **106**, 1925 (2008).
[24]Y. S. Djikaev and E. Ruckenstein, J. Phys. Chem. B **116**, 2820 (2012).
[25]K. Lum, D. Chandler, and J. D. Weeks, J. Phys. Chem. B **103**, 4570 (1999).
[26]G. Hummer, A. Garde, A. E. Garcia, M. E. Paulaitis, and L. R. Pratt, J. Phys. Chem. B **102**, 10469 (1998).
[27]H. Ashbaugh and L. Pratt, Rev. Mod. Phys. **78**, 159 (2006).
[28]P. Kumar, S. V. Buldyrev, and H. E. Stanley, Proc. Natl. Acad. Sci. U. S. A. **106**, 22130 (2009).
[29]C. G. Gray and K. E. Gubbins, *Theory of Molecular Fluids: Fundamentals* (Oxford University Press, 1984).
[30]L. Blum and A. J. Torruella, J. Chem. Phys. **56**, 303 (1972).
[31]G. Patey, Mol. Phys. **35**, 1413 (1978).
[32]P. H. Fries and G. N. Patey, J. Chem. Phys. **82**, 429 (1985).
[33]J. Richardi, C. Millot, and P. H. Fries, J. Chem. Phys. **110**, 1138 (1999).
[34]J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic Press, London, 2006), p. 416.
[35]J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic Press, London, 1986).
[36]N. J. English, Mol. Phys. **103**, 1945 (2005).
[37]Y. Marcus, J. Solution Chem. **21**, 1217 (1992).
[38]Y. S. Djikaev and E. Ruckenstein, J. Phys. Chem. B **117**, 7015 (2013).
[39]N. Muller, Acc. Chem. Res. **23**, 23 (1990).
[40]K. Silverstein, A. D. J. Haymet, and K. A. Dill, J. Am. Chem. Soc. **122**, 8037 (2000).
[41]S. J. Suresh and V. M. Naik, J. Chem. Phys. **113**, 9727 (2000).
[42]A. Rastogi, S. Yadav, and S. J. Suresh, J. Chem. Phys. **135**, 086101 (2011).
[43]R. Kumar, J. R. Schmidt, and J. L. Skinner, J. Chem. Phys. **126**, 204107 (2007).
[44]A. Luzar and D. Chandler, Nature **379**, 55 (1996).
[45]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[46]S. Nose, J. Chem. Phys. **81**, 511 (1984).
[47]W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[48]U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[49]J. Ryckaert, G. Ciccotti, and H. Berendsen, J. Comput. Phys. **23**, 341 (1977).
[50]M. Allen and D. Tildesley, *Computer Simulation of Liquids*, 1st ed. (Oxford University Press, 1987).
[51]"CPMD, Copyright IBM Corp. 1990-2008, Copyright MPI für Festkörperforschung Stuttgart 1997-2001," 2012.
[52]C. T. Lee, W. T. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[53]H. A. N. Zheng, S. Wang, Y. Zhang, and N. York, J. Comput. Chem. **30**, 2706 (2009).
[54]N. Marzari and D. Vanderbilt, Phys. Rev. B: Condens. Matter Mater. Phys. **56**, 12847 (1997).
[55]P. Silvestrelli and M. Parrinello, Phys. Rev. Lett. **82**, 3308 (1999).
[56]B. Guillot and Y. Guissani, J. Chem. Phys. **99**, 8075 (1993).
[57]See supplementary material at http://dx.doi.org/10.1063/1.4908532 for $\Delta S(T)$ for solutes not shown in Fig. 5, the derivation of the length-scale crossover of ΔG_ϕ, fit parameters to Eq. (27), validation of the particle insertion method, and a glossary of symbols.
[58]B. Widom, J. Chem. Phys. **39**, 2808 (1963).
[59]G. L. Deitrick, L. E. Scriven, and H. T. Davis, J. Chem. Phys. **90**, 2370 (1989).
[60]G. Deitrick, L. Scriven, and H. Davis, Mol. Simul. **8**, 239 (1992).
[61]D. Frenkel and B. Smit, *Understanding Molecular Simulations* (Academic Press, 2001).
[62]T. Rettich, Y. Handa, R. Battino, and E. Wilhelm, J. Phys. Chem. **85**, 3230 (1981).
[63]G. Graziano, J. Phys. Chem. B **118**, 2598 (2014).
[64]N. Galamba, J. Phys. Chem. B **118**, 2600 (2014).
[65]G. Jeanmairet, M. Levesque, and D. Borgis, J. Chem. Phys. **139**, 154101 (2013).
[66]G. Jeanmairet, M. Levesque, R. Vuilleumier, and D. Borgis, J. Phys. Chem. Lett. **4**, 619 (2013).
[67]D. E. Hare and C. M. Sorensen, J. Chem. Phys. **93**, 6954 (1990).
[68]L. Pratt and D. Chandler, J. Chem. Phys. **67**, 3683 (1977).
[69]B. Lee, Biopolymers **31**, 993 (1991).
[70]R. F. Prini and R. Crovetto, J. Phys. Chem. Ref. Data **18**, 1231 (1989).