

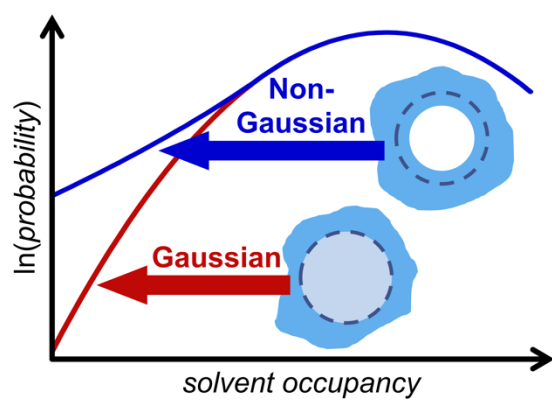
**Gaussian and Non-Gaussian Solvent Density Fluctuations within Solute Cavities in a
Water-like Solvent**

Henry S. Ashbaugh

Tulane University, Chemical and Biomolecular Engineering, New Orleans, LA 70118

Abstract

We report a Monte Carlo simulation study of length-scale dependent density fluctuations in cavities in the coarse-grained mW representation of water at ambient conditions. Specifically, we use a combination of test particle insertion and umbrella sampling techniques to examine the full range of water occupation states in spherical cavities up to 6.3 Å in radius in water. As has previously been observed, water density fluctuations are found to be effectively Gaussian in nature for atomic-scale cavities, but as the cavities get larger they exhibit a non-Gaussian “fat-tail” distribution for lower occupancy states. We introduce a new statistical thermodynamic approach to analyze non-Gaussian fluctuations based on the radial distribution of waters about cavities with varying numbers of waters within its boundaries. It is shown that the onset on these non-Gaussian fluctuations is a result of the formation of a bubble within the cavity as it is emptied that is accompanied by the adsorption of waters onto its interior surface. We revisit a theoretical framework we previously introduced to describe Gaussian fluctuations within cavities to now incorporate bubble formation by including surface tension contributions. This modified theory accurately describes density fluctuations within both atomic and meso-scale cavities. Moreover, the theory predicts the transition from Gaussian to non-Gaussian fluctuations at a specific cavity occupancy in excellent agreement with simulation observations.



TOC Figure.

1. Introduction

Water plays a central role in self-assembly processes in aqueous solution, including detergent assembly into micelles, the folding of proteins, and the formation of larger biomolecular structures. The hydrophobic effect, the limited solubility of oily species in water, provides the impetus for the non-specific aggregation and assembly of the non-polar constituents of these molecules.¹⁻³ The thermodynamic signatures of hydrophobically driven assembly, however, depend on the size and curvature of the aggregating moieties.⁴⁻⁶ On the molecular-scale, the dissolution of non-polar gases and alkanes in water at room temperature is enthalpically favorable but dominated by a large negative entropy that opposes hydration. As the size of the hydrophobic groups increases the roles of entropy and enthalpy are reversed and hydration is opposed by a dominate positive enthalpy and favored by a smaller positive entropy. This switch in the signatures of hydrophobic hydration are illustrated by the work of Li and Walker⁷ who observed from experiments of extending polymers into water using an atomic force microscope that the temperature dependence of the pulling force shifted from entropically towards enthalpically unfavored as the size of the non-polar side chains increased. These observations suggest changes in the hydration mechanism with non-polar solute size. To gain insight into these processes many studies have focused on model solutes to isolate hydrophobic effects from competing interactions, like dispersion and electrostatic forces. Molecular simulations are well suited for this task since individual interactions can be turned off while retaining molecular realism.

The standard approach to describe the hydration of non-polar solutes is to divide the process into two steps.⁸ First, an empty cavity the size and shape of the excluded volume of the solute is created in solution. Second, the attractive interactions between water and the solute placed within the cavity are turned on. For atomic-scale solutes the characteristic thermodynamics of

hydrophobic hydration (e.g., a large negative entropy of hydration) are embedded within the excluded volume contribution to the hydration process, while the attractive contribution can be treated perturbatively. As such, the focus of many studies of hydrophobic hydration have focused solely on the excluded volume, or cavity, contribution to the free energy. The free energy of solvating an empty cavity-like solute is directly related to the probability of finding an empty cavity the size and shape of the solute as

$$\mu_0^{ex} = -k_B T \ln p_0, \quad (1)$$

where $k_B T$ is the product of Boltzmann's constant and the absolute temperature, p_0 is the probability of finding an empty cavity devoid of water within the bulk solvent, and μ_0^{ex} is the free energy of hydrating the empty cavity, i.e., its excess chemical potential of hydration. One approach to evaluate the excess chemical potential, scaled-particle theory (SPT), focuses on the process of growing empty spherical cavities in solution from nothing up to the desired radius.⁹ The original implementation of SPT only utilized water's density and its effective hard sphere diameter to describe the hydration process.^{10, 11} In the early 1970s, however, Stillinger pointed out that this version of SPT does not account for the structure of liquid water and incorrectly predicts the temperature dependence of water's liquid-vapor surface tension.¹² He subsequently provided an empirically corrected SPT utilizing water's experimental radial distribution function determined from X-ray scattering^{13, 14} and the known surface tension of water. Several years later, Ashbaugh and Pratt expanded Stillinger's approach to incorporate many-body correlations into SPT by using multi-body information from molecular simulations.¹⁵⁻¹⁷ These theories helped illuminate the thermodynamic distinction between molecular-scale hydrophobic hydration, which is opposed by a dominant negative hydration entropy, and meso/macro-scale hydrophobic hydration, which is opposed by a dominant positive hydration enthalpy. The distinction between molecular-scale and

meso-scale hydration is described by a crossover length,^{4, 18, 19} which itself is temperature dependent.¹⁷

In the mid-1990s, Hummer et al.²⁰ provided an alternate view of cavity hydration following information theory, where, rather than focusing only on empty cavities, all possible occupancy states are considered, p_n (n is the number of solvents within the cavity). They demonstrated for atomic-scale cavities the p_n distribution is effectively Gaussian in water. While the mean number of solvent molecules in the cavity depends solely on the density of water and the solute's volume, evaluation of the variance in the distribution requires knowledge of water's radial distribution function. The information theory description of hydration was subsequently applied to analyze the thermodynamics of hydrophobic hydration,^{20, 21} the observation of entropy convergence at elevated temperatures,²² and the pressure induced denaturation of proteins.²³ More recently, Ashbaugh, Vats, and Garde²⁴ demonstrated that for state points far from the critical point, the solute-size dependence of the variance of the p_n distribution could be approximated over all size scales using a simple analytical form, referred to as interpolated Gaussian fluctuation theory (IGFT). This theory only requires information on water's density, compressibility, and effective diameter, but not its radial distribution function. IGFT was shown to accurately predict the characteristic thermodynamics of atomic-scale hydrophobic hydration up to 300°C. Moreover, this theory also predicted that in the supercooled regime, the hydration heat capacity could also reverse its sign from positive to negative in agreement with previous simulation observations.²⁵⁻²⁷

Despite its success at addressing atomic-scale hydrophobic hydration, the Gaussian approximation for solvent density fluctuations breaks down as the solute radius increases much beyond that of xenon. This was perhaps first demonstrated by Huang and Chandler²⁸ for the emptying of a Lennard-Jones liquid from cavities significantly larger than the solvent. Notably,

they observed that while the p_n distribution is Gaussian for cavity occupancies near the mean, for cavities twice the solvent's diameter and larger the distribution exhibits a "fat-tail" distribution with lower occupancies being more favorable than anticipated based on Gaussian predictions. They attributed this non-Gaussian tail to the onset of drying as the result of a bubble forming within the cavity as it is emptied. It has subsequently been demonstrated that the onset of non-Gaussian density fluctuations in water between the hydrophobic faces of biomolecules impacts their interactions with each other and interfaces.^{29, 30} Moreover, water density fluctuations within hydrophobic pockets can tilt so that they spontaneously dewet, favoring the binding of non-polar guests to these surfaces.³¹

In a recent paper, Sinha et al.³² examined the process of hydrating large spherical and non-spherical solutes in SPC/E water to determine the point at which water density fluctuations within solute cavities cross over from Gaussian to non-Gaussian behavior. In this analysis, we introduced a theoretical approach utilizing the known interfacial properties of water to describe the onset of bubble formation within the cavity to describe non-Gaussian density fluctuations. That theoretical approach, however, was introduced in a cursory manner and not fully developed. Here we revisit this problem to provide a clearer justification for the proposed changes in the mechanism of cavity emptying in solutions, reporting new molecular simulations and theoretical results for the process of emptying solute volumes in water to gain insights into the role of solvent density fluctuations in the thermodynamics of hydrophobic hydration. Water is modeled here using the coarse-grained mW model developed by Molinero.³³ This model neglects water's hydrogens, capturing the effect of directional hydrogen-bonding using a three-body potential proposed by Stillinger and Weber³⁴ that is shorter-ranged than more traditional representations of water, making it computationally more expedient. More importantly for the work reported here, it has been shown to capture many

of the peculiar thermodynamics of hydrophobic hydration.³⁵ In this work, we use simulations to determine the free energies of removing individual waters from cavities of varying radius to generate a hard-sphere solutes in water, a model solute for understanding hydrophobic hydration. A new statistical mechanical approach is developed to evaluate the work of emptying the cavities based on the structure of water about partially filled and empty cavities, thereby providing new insights into the emptying process. In the final part of the paper, we extend IGFT to incorporate non-Gaussian fluctuations that result from interfacial effects that arise as the solute grows in size.

2. Theory

2.1. Water Partitioning Between a Cavity Interior and the Bulk Solvent. While the empty cavity ($n = 0$) is most relevant for evaluation of the solvation free energy of a hard sphere solute as embodied by eq. (1), as shown below the intermediate n provide significant information on the mechanism by which the volume is emptied. In particular, this chemical potential for creating an empty cavity can be readily extended to consider the process of finding a cavity with n water in it as $\mu_n^{ex} = -k_B T \ln p_n$. The process of removing solvent molecules from a cavity to create a hard-sphere solute can be broken down into a series of steps where the solvent molecules are removed one-by-one from the volume. In this section we consider the emptying process to construct a description of hard sphere solvation based on the packing of solvent molecules both inside and outside the cavity.

To begin, we consider the thermodynamic equilibrium between waters inside and outside a cavity in solution. The canonical partition function of a system of N total water molecules with n waters confined within the boundaries of the cavity, v , and $N - n$ waters outside the cavity is

$$Q_{N,n} = \left[\frac{v^n (V-v)^{N-n} q_{int}^N}{n!(N-n)! \Lambda^{3N}} \right] \left[\frac{\int_v^n \int_{(V-v)^{N-n}} \exp(-\beta U_N) d\mathbf{r}^{N-n} d\mathbf{r}^n}{v^n (V-v)^{N-n}} \right]$$

$$= Q_{N,n}^{id} Q_{N,n}^{ex}, \quad (2)$$

where V is the total system volume, Λ is the thermal de Broglie wavelength, q_{int} is the internal partition function for an individual water molecule associated with vibrational and rotational degrees of freedom, U_N is the system potential energy, $\mathbf{r}$ is the position vector of an individual molecule, and $\beta = 1/k_B T$. Orientational degrees of freedom are neglected in the integral above, but this does not change the results of this analysis. The terms in the first set of square brackets correspond to the ideal gas contribution to the partition function, $Q_{N,n}^{id}$, while the set of square brackets corresponds to the excess contribution to the partition function, $Q_{N,n}^{ex}$, resulting from intermolecular interactions. The ratio of probabilities of observing $n + 1$ and n molecules within the cavity is related to the ratio of their partition functions as

$$\frac{p_{n+1}}{p_n} = \frac{Q_{N,n+1}}{Q_{N,n}} = \frac{Q_{N,n+1}^{id}}{Q_{N,n}^{id}} \frac{Q_{N,n+1}^{ex}}{Q_{N,n}^{ex}}. \quad (3)$$

In the thermodynamic limit ($N \rightarrow \infty$ and $V \rightarrow \infty$) the ratio of the ideal gas partition functions is

$$\frac{Q_{N,n+1}^{id}}{Q_{N,n}^{id}} = \frac{\rho v}{n+1}, \quad (4)$$

where $\rho = N/V$ is the total solvent number density. The ratio of the excess partition functions is

$$\frac{Q_{N,n+1}^{ex}}{Q_{N,n}^{ex}} = \left(\frac{V-v}{v} \right) \frac{\int_v^n \int_{(V-v)^{N-n}} \exp(-\beta U_N) d\mathbf{r}^{N-n-1} d\mathbf{r}^{n+1}}{\int_v^n \int_{(V-v)^{N-n}} \exp(-\beta U_N) d\mathbf{r}^{N-n} d\mathbf{r}^n}. \quad (5)$$

This ratio can be re-expressed as

$$\frac{Q_{N,n+1}^{ex}}{Q_{N,n}^{ex}} = \frac{Q_{N,n+1}^{ex}}{Q_{N-1,n}^{ex}} \frac{Q_{N-1,n}^{ex}}{Q_{N,n}^{ex}} \quad (6a)$$

$$= \left(\frac{V-v}{v} \right) \frac{\int_v^{n+1} \int_{(V-v)^{N-n-1}} \exp(-\beta U_N) d\mathbf{r}^{N-n-1} d\mathbf{r}^{n+1}}{\int_v^n \int_{(V-v)^{N-n-1}} \exp(-\beta U_{N-1}) d\mathbf{r}^{N-n-1} d\mathbf{r}^n} \frac{\int_v^n \int_{(V-v)^{N-n-1}} \exp(-\beta U_{N-1}) d\mathbf{r}^{N-n-1} d\mathbf{r}^n}{\int_v^n \int_{(V-v)^{N-n}} \exp(-\beta U_N) d\mathbf{r}^{N-n} d\mathbf{r}^n} \quad (6b)$$

The energy of the N molecule system can be divided into $N - 1$ and 1 molecules as

$$U_N = U_{N-1} + \Delta U_N. \quad (7)$$

It follows that the first ratio of integrals in eq. (6b) can be expressed as

$$\begin{aligned}
& \frac{\int_v^{n+1} \int_{(V-v)^{N-n-1}} \exp(-\beta U_N) d\mathbf{r}^{N-n-1} d\mathbf{r}^{n+1}}{v \int_v^n \int_{(V-v)^{N-n-1}} \exp(-\beta U_{N-1}) d\mathbf{r}^{N-n-1} d\mathbf{r}^n} \\
&= \frac{\int_v^{n+1} \int_{(V-v)^{N-n-1}} \exp(-\beta \Delta U_N) \exp(-\beta U_{N-1}) d\mathbf{r}^{N-n-1} d\mathbf{r}^{n+1}}{v \int_v^n \int_{(V-v)^{N-n-1}} \exp(-\beta U_{N-1}) d\mathbf{r}^{N-n-1} d\mathbf{r}^n} \\
&= \langle \exp(-\beta \Delta U_N) \rangle_{in},
\end{aligned} \tag{8}$$

which corresponds to the Boltzmann weighting of the mean excess chemical potential of a particle randomly inserted into the cavity. Similarly, the second ratio of integrals in eq. (6c) can be expressed as

$$\begin{aligned}
& \frac{\int_v^n \int_{(V-v)^{N-n}} \exp(-\beta U_N) d\mathbf{r}^{N-n} d\mathbf{r}^n}{(V-v) \int_v^n \int_{(V-v)^{N-n}} \exp(-\beta U_{N-1}) d\mathbf{r}^{N-n-1} d\mathbf{r}^n} \\
&= \frac{\int_v^n \int_{(V-v)^{N-n}} \exp(-\beta \Delta U_N) \exp(-\beta U_{N-1}) d\mathbf{r}^{N-n} d\mathbf{r}^n}{(V-v) \int_v^n \int_{(V-v)^{N-n-1}} \exp(-\beta U_{N-1}) d\mathbf{r}^{N-n-1} d\mathbf{r}^n} \\
&= \langle \exp(-\beta \Delta U_N) \rangle_{out},
\end{aligned} \tag{9}$$

which corresponds to the Boltzmann weighting of the mean excess chemical potential of a particle randomly inserted outside the cavity. Given that the volume of the solvent outside of the cavity is infinitely greater than that inside the cavity, this is simply the excess chemical potential of the bulk solvent. The ratio of cavity occupation probabilities is subsequently

$$\frac{p_{n+1}}{p_n} = \frac{\rho v}{(n+1)} \frac{\langle \exp(-\beta \Delta U_N) \rangle_{in}}{\langle \exp(-\beta \Delta U_N) \rangle_{out}} \tag{10}$$

While the averages in eqs. (8)–(10) are technically for the addition of a water molecule to a system of $N - 1$ waters, in the thermodynamic limit this is indistinguishable for the averages of an N water system.

While eq. (10) could be evaluated from simulation following standard particle insertion techniques, this expression can be more readily evaluated from the occupancy dependent radial

distribution functions (RDFs) between the cavity and the solvent, $g_n(r)$. This collection of RDFs corresponds to the local solvent density as a function of distance from the center of a cavity with n waters residing within it. Associated with each RDF is a function $y_n(r)$, also referred to as a cavity correlation function, that corresponds to the Boltzmann weighting of the potential-of-mean force associated with bringing a solvent particle from a position infinitely far away to a distance r from the center of the cavity with n molecules constrained to reside inside. The cavity correlation function is determined following test particle insertion as

$$y_n(r) = \frac{\int_v^n \int_{(V-v)^{N-n-1}} \exp[-\beta \Delta U_N(r)] \exp(-\beta U_{N-1}) d\mathbf{r}^{N-n-1} d\mathbf{r}^n \frac{Q_{N-1,n}^{ex}}{Q_{N-1,n}^{ex}}}{\int_v^n \int_{(V-v)^{N-n-1}} \exp[-\beta \Delta U_N(\infty)] \exp(-\beta U_N) d\mathbf{r}^{N-n-1} d\mathbf{r}^n \frac{Q_{N-1,n}^{ex}}{Q_{N-1,n}^{ex}}} \\ = \frac{\langle \exp(-\beta \Delta U_N(r)) \rangle}{\langle \exp(-\beta \Delta U_N(\infty)) \rangle}. \quad (11)$$

The numerator corresponds to the Boltzmann weighting of the excess chemical potential for of a water at a distance r from the center of the cavity, while the denominator corresponds to the Boltzmann weighting of the excess chemical potential for a water infinitely far away from the center of the cavity, which is equal to the bulk chemical potential in eq. (9) (i.e., $\langle \exp(-\beta \Delta U_N(\infty)) \rangle = \langle \exp(-\beta \Delta U_N) \rangle_{out}$). For separations outside the cavity, $y_n(r > R)$ corresponds to the RDF between the solvent and cavity with n molecules within it (i.e., $y_n(r > R) = g_n(r > R)$). For distances inside the cavity, $n + 1$ molecules now reside within its boundary. While it is tempting to equate $y_n(r \leq R)$ with $g_{n+1}(r \leq R)$, normalization of the radial distribution function negates this equality. Specifically, within the cavity the radial distribution function obeys the condition

$$\int_0^R \rho g_{n+1}(r) 4\pi r^2 dr = n + 1. \quad (12)$$

This suggests, as we shall see below, that the RDF changes discontinuously across the cavity's boundary, while $y_n(r)$ is expected to be a continuous function. We propose that $y_n(r)$ is related to the radial distribution functions as

$$y_n(r) = \begin{cases} K_n g_{n+1}(r) & r \leq R \\ g_n(r) & r > R \end{cases} \quad (13)$$

where K_n is a constant that ensures $y_n(r)$ is continuous across the cavity's boundary. The question follows, what is K_n ?

Following from eq. (12), the average value of $g_{n+1}(r)$ inside the spherical cavity ($r \leq R$) is

$$\langle g_{n+1}(r) \rangle_{r \leq R} = \frac{n+1}{\rho v}. \quad (14)$$

It follows from eq. (13) that the average value of $y_n(r)$ inside the cavity is

$$\langle y_n(r) \rangle_{r \leq R} = K_n \frac{(n+1)}{\rho v}. \quad (15)$$

Following eq. (11), the average value of $y_n(r)$ inside the cavity is also

$$\langle y_n(r) \rangle_{r \leq R} = \frac{1}{v} \frac{\int_0^R \langle \exp(-\beta \Delta U_N(r)) \rangle 4\pi r^2 dr}{\langle \exp(-\beta \Delta U_N) \rangle_{out}} \quad (16)$$

The mean value of the integral in the numerator over the cavity interior is

$$\frac{\int_0^R \langle \exp(-\beta \Delta U_N(r)) \rangle 4\pi r^2 dr}{v} = \langle \exp(-\beta \Delta U_N) \rangle_{in}. \quad (17)$$

Combining eqs. (10), (16), and (17) it follows that

$$\langle y_n(r) \rangle_{r \leq R} = \frac{\langle \exp(-\beta \Delta U_N) \rangle_{in}}{\langle \exp(-\beta \Delta U_N) \rangle_{out}} = \frac{p_{n+1}}{p_n} \frac{(n+1)}{\rho v}. \quad (18)$$

Comparing eqs. (15) and (18) we find the desired result that

$$K_n = \frac{p_{n+1}}{p_n}. \quad (19)$$

Thus, the discontinuous change in the solvent RDF across the boundary of the cavity can be related to the work associated with adding a solvent molecule into the cavity already containing n molecules.

The derivations above provide a route to determining p_n from molecular simulations, simply from the distribution of solvent molecules about a cavity. Specifically, a cavity may be included within a simulation box to evaluate the $g_n(r)$ distributions. The resulting K_n 's can be determined by enforcing continuity between $g_n(r)$ and $g_{n+1}(r)$ at R to determine $y_n(r)$. In practice, over a small range near the boundary, $-\ln y_n(r)$ can be approximated by a quadratic polynomial. As such we fit the function

$$f(r) = \alpha_2(r - R)^2 + \alpha_1(r - R) + \alpha_0 - \Theta(R - r)\ln K_n, \quad (20)$$

simultaneously to $-\ln g_{n+1}(r)$ for $r < R$ and $-\ln g_n(r)$ for $r \geq R$. Here $\Theta(x)$ is the Heaviside function, and α_0 , α_1 , α_2 , and $\ln K_n$ are fitting constants. This fitting is performed over only a limited range of $R \pm 1$ Å, beyond which the quadratic approximation becomes increasingly inaccurate. Over this limited fitting range, the potential-of-mean force is $-k_B T \ln y_n(r) = \alpha_2(r - R)^2 + \alpha_1(r - R) + \alpha_0$. Once the set of K_n 's across all potential cavity occupation states is determined, the full p_n distribution is given as

$$p_n = \begin{cases} \frac{1}{\Xi} & n = 0 \\ \frac{\prod_{i=0}^{n-1} K_i}{\Xi} & n > 0 \end{cases}, \quad (21a)$$

where

$$\Xi = 1 + \sum_{i=1}^{\infty} (\prod_{j=0}^{i-1} K_j). \quad (21b)$$

While the cavity does exhibit correlations with the solvent that depend on n , on average the cavity itself has no correlations with the solvent. As such the average RDF between the cavity and solvent is expected to be

$$\sum_{n=0}^{\infty} p_n g_n(r) = 1 \quad (22)$$

over all separations from inside to outside the cavity, reflecting the passive nature of the cavity.

To demonstrate the validity of this approach for obtaining the solvent occupancy distribution within a cavity, we consider the ideal gas for which the p_n is known analytically. In the ideal gas where no interactions between molecules are felt, the cavity correlation function between a gas particle and a cavity with n particles within its interior is

$$y_n(r) = 1. \quad (23)$$

As a result of the normalization condition inside the cavity (eq. (14)), however, the n dependent RDFs are given as

$$g_n(r) = \begin{cases} \frac{n}{\rho v} & r \leq R \\ 1 & r > R \end{cases} \quad (24)$$

Comparing eqs. (18), (23), and (24), it follows that

$$K_n = \frac{\rho v}{n+1}. \quad (25)$$

Substituting eq. (25) into eq. (21), we obtain

$$\Xi = 1 + \sum_{i=1}^{\infty} \left(\prod_{j=0}^{i-1} \frac{\rho v}{j+1} \right) = 1 + \sum_{i=1}^{\infty} \frac{(\rho v)^i}{i!} = \exp(\rho v), \quad (26a)$$

and

$$p_n = \left(\prod_{j=0}^{n-1} \frac{\rho v}{j+1} \right) \exp(-\rho v) = \frac{(\rho v)^n}{n!} \exp(-\rho v). \quad (26b)$$

This is the Poisson distribution, the expected probability distribution for a cavity in an ideal gas.

Further, averaging the n dependent RDFs within the cavity ($r \leq R$) yields

$$\begin{aligned} \sum_{n=0}^{\infty} g_n(r \leq R) p_n &= \sum_{n=0}^{\infty} \frac{n}{\rho v} \frac{(\rho v)^n}{n!} \exp(-\rho v) = \exp(-\rho v) \sum_{n=1}^{\infty} \frac{(\rho v)^{n-1}}{(n-1)!} \\ &= \exp(-\rho v) \sum_{n=0}^{\infty} \frac{(\rho v)^n}{n!} = \exp(-\rho v) \exp(\rho v) = 1. \end{aligned} \quad (27)$$

Since the Poisson distribution is normalized, it follows that the average of the radial distribution functions outside the cavity is 1 as well. The average of the cavity radial distribution functions in the ideal gas subsequently obey eq. (22).

2.2. *Interpolated Gaussian Fluctuation Theory.* Here we highlight the key equations underlying the IGFT description of solvent occupation probabilities in atomic-size cavities in water. A more complete development of this theory can be found in ref. ²⁴. Following information theory utilizing only information on the second moment of solvent density fluctuations,²⁰ the solvent occupancy probability distribution within an atomic-sized cavity is expected to be Gaussian in nature. Assuming over this size range n can be approximated as a continuous variable that spans from $-\infty$ to ∞ , the probability of observing n solvent molecules within the cavity can be expressed using the normalized Gaussian distribution^{21, 22}

$$p_n \approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(n-\langle n \rangle)^2}{2\sigma^2}\right] = \frac{1}{\sqrt{2\pi\langle n \rangle\chi}} \exp\left[-\frac{(n-\langle n \rangle)^2}{2\langle n \rangle\chi}\right], \quad (28)$$

where $\langle n \rangle = \rho v$ is the mean number of solvent molecules within the cavity, σ^2 is the variance of the distribution, and $\chi = \sigma^2/\langle n \rangle$ is the normalized variance. The excess chemical potential associated with creating a cavity with n solvent molecules within its boundaries is subsequently

$$\beta\mu_n^{ex} = -\ln p_n = \frac{(n-\langle n \rangle)^2}{2\langle n \rangle\chi} + \frac{1}{2} \ln(2\pi\langle n \rangle\chi). \quad (29a)$$

It follows that the hydration free energy of the empty cavity is

$$\beta\mu_0^{ex} = -\ln p_0 = \frac{\langle n \rangle}{2\chi} + \frac{1}{2} \ln(2\pi\langle n \rangle\chi). \quad (29b)$$

While clearly an approximation, this description of atomic scale solvation has been successfully applied to describe hard sphere solvation in water over a wide range of temperatures.^{22, 36}

While $\langle n \rangle$ is simply determined from the bulk solvent density and the solute volume, χ is determined from a more complex integral over the solvent-solvent RDF and is not analytical. Nevertheless, considering the microscopic and macroscopic limits on the χ integral, Ashbaugh,

Vats, and Garde²⁴ proposed an interpolative analytical approximation for χ in spherical cavities as a function of their radius

$$\chi(R) = \begin{cases} 1 - \eta \left(\frac{2R}{d_{ww}} \right)^3 & R < d_{ww}/2 \\ k_B T \rho \kappa_T + \frac{(3-6\eta-3k_B T \rho \kappa_T)}{2} \left(\frac{d_{ww}}{2R} \right) - \frac{(1-4\eta-k_B T \rho \kappa_T)}{2} \left(\frac{d_{ww}}{2R} \right)^3 & R \geq d_{ww}/2 \end{cases}, \quad (30)$$

where κ_T is water's isothermal compressibility, d_{ww} is water's effective diameter, and $\eta = \pi \rho d_{ww}^3 / 6$ is the solvent packing fraction. The expressions eqs. (28) to (30) are collectively referred to as IGFT. This framework was shown to provide an excellent description of the hydration of hard cavities in water over a wide temperature range and even predicted the unanticipated reversal in the temperature dependence of hydrophobic hydration thermodynamics in the deeply supercooled regime. Interestingly, eq. (30) largely relies only on water's macroscopic equation-of-state while the structure of water is embodied only within d_{ww} (assumed here to be 2.65 Å). This diameter has been found to be effectively independent of temperature.^{24, 25} In previous studies, however, non-Gaussian fluctuations were found to become more significant as the size of the solute grows.^{28, 32, 37} As such, IGFT is expected to become increasingly inaccurate for meso-scale and larger solute volumes.

3. Molecular Simulations

Monte Carlo (MC) simulations³⁸ of mW water³³ with a single cavity in solution were conducted in the isothermal-isobaric ensemble. The temperature and pressure were set to 25°C and 1 atm, respectively. Spherical cavity radii of 2.5 Å to 6.3 Å in 0.2 Å increments were considered. For cavity radii up to 4.1 Å, simulations were conducted with 700 waters, while 1000 waters were considered for the larger solutes. The cavity in principle has no interactions with water, serving to sample solvent density fluctuations within its boundaries. For large volumes, however, the rarity

of large density fluctuations ensure that empty or nearly empty cavities are never observed. To sample large scale density fluctuations we applied a biasing harmonic umbrella potential³⁹ to restrict the range of cavity occupancies observed during the simulation

$$\varphi_{umb}(n) = k_0(n - n_0)^2. \quad (31)$$

Here k_0 is the spring constant and n_0 is the occupancy for which the umbrella potential is a minimum. The value of n_0 was adjusted from 0 to values well above $\langle n \rangle$ in increments of $\Delta n_0 = 2$ to ensure sampling of the full range of relevant occupancy states. The spring constant was adjusted to sample occupancy states ± 3 on either side of n_0 to ensure overlapping fluctuations between simulations with consecutive values of n_0 . As such, k_0 took on values from 2.5 kJ/mol up to 9 kJ/mol depending on n_0 , with larger k_0 values needed as the cavity is emptied. The full p_n distribution was reconstructed using the weighted histogram analysis method.^{40, 41} At least 10^6 MC passes (where 1 MC pass corresponds to 1 attempted move on each water molecule) were performed for equilibration at each value of n_0 , followed by 2.5×10^7 passes for evaluation of thermodynamic averages. Volume moves were performed every 5 MC passes. The water and cavity displacements were modified to ensure acceptance of 30% of moves, while the total simulation volume displacement was similarly modified to ensure acceptance of 30% of attempted changes.

In addition to simulations of an explicit cavity in water, we also performed MC simulations of pure mW water at 25°C and 1 atm. The cubic simulation cell contained 700 water molecules. In addition to gathering information on the pure water properties (e.g., density and compressibility), we evaluated the excess chemical potentials of atomic-sized solutes using Widom test particle insertion.^{42, 43} 10^5 random test particle insertion attempts were made every tenth MC pass. 10^6 MC

passes were conducted for equilibration, followed by 10^8 MC passes for the evaluation of thermodynamic averages.

4. Results and Discussion

4.1 Density Fluctuations within a Spherical Cavity. Figure 1 illustrates the Gaussian nature of solvent density fluctuations within atomic-scale cavities up to 3.7 Å in radius in mW water at 25°C and 1 atm. Specifically, when the probability of observing n water molecules within the cavity, p_n , is plotted on a logarithmic scale as a function of n , we find that the distributions effectively assume a parabolic form consistent with a Gaussian distribution. The simulation results are in excellent agreement with the distributions predicted by IGFT, lending confidence to the fidelity of the theory for describing small-scale solute hydration.

The accuracy of the Gaussian description can be further probed by plotting $\ln K_n (= \ln(p_{n+1}/p_n))$ versus n (Figure 2). The expectation is that if the density fluctuations are Gaussian (i.e., $\ln p_n = an^2 + bn + c$, where a , b , and c are constants) then $\ln K_n$ will be a linear function of n (i.e., $\ln K_n = 2an + a + b$). While IGFT predicts this linear dependence by construction, the simulation results exhibit deviations from linearity with increasing solute size. The simulation results for each cavity considered in Figure 2 exhibit linear behavior for occupation numbers close to the mean ($\langle n \rangle = \rho v = 1.3, 3.4$, and 7.1 for the 2.1 Å, 2.9 Å, and 3.7 Å radius cavities, respectively), indicative of Gaussian-like fluctuations near the mean of the distribution that are well described by IGFT. For more extreme fluctuations away from $\langle n \rangle$, however, this does not necessarily hold. Notably for the 3.7 Å radius sphere, the simulations exhibit a slight positive deviation from linearity for $n = 2$ and 3 , followed by a drop below the linear prediction for $n = 0$ and 1 . Analogous behavior can be observed for the 2.9 Å radius sphere as well, although not as

dramatic as for the larger volume. The 2.1 Å radius cavity, on the other hand, is accurately described by the Gaussian prediction. For these atomic-scale volumes, however, the combination of positive and negative deviations from linearity for $\ln K_n$ are largely compensatory, such that the Gaussian predictions for the probability of observing an empty sphere, i.e., p_0 , are reasonably accurate up to 3.7 Å (Figure 1).

The deviations from Gaussian fluctuations are more significant with increasing cavity size. The occupation probabilities for the 4.3 Å, 5.3 Å, and 6.3 Å radius cavities shown in Figure 3 are parabolic on a logarithmic scale near the maxima in these distributions, and hence are Gaussian close to $\langle n \rangle$. As n decreases, however, the probability distributions for each of these volumes exhibit markedly greater probabilities for observing nearly empty cavities than anticipated based on Gaussian fluctuations. As such, the free energy cost associated with observing an empty mesoscopic-sized cavity would be much lower than predicted following a Gaussian description. This deviation from the Gaussian prediction corresponds to the fat-tail distribution associated with drying phenomena in the context of hydrophobic phenomena.^{28, 29}

The breakdown of the Gaussian description for larger volumes is further scrutinized in the plot of $\ln K_n$ versus n in Figure 4. As above, near the mean occupation number ($\langle n \rangle = 11.1, 20.8$, and 34.9 for the 4.3 Å, 5.3 Å, and 6.3 Å spheres, respectively) the simulation results are linear and well described by IGFT. As n gets smaller, however, each of these cavities exhibit a dramatic break from linearity. The location of this break depends on the size of the sphere. To a first approximation, $\ln K_n$ from $n = 0$ to the break from linearity is (very) roughly constant and equal to ~ 2.5 , although there is an n dependence in this regime. A conclusion that can be drawn is that the crossover from Gaussian to non-Gaussian density fluctuations is not gradual but rather exhibits a marked change as n drops, suggestive of a transition in the cavity emptying mechanism.

The differences in the emptying process of the atomic and meso-scale cavities is readily observed when we consider the hydration free energies of the empty cavities (Figure 5). Up to solutes ~ 3.7 Å in radius, the hydration free energy is well described by IGFT within $\sim 1 k_B T$. This range of solute sizes is coincident with the range of solute sizes for which we reliably observe empty solute cavities in water by test particle insertion. The hydration free energies of solutes larger than 3.7 Å determined using umbrella sampling, however, are markedly lower than what is predicted by IGFT. Indeed, for the 6.3 Å radius solute IGFT predicts a hydration free energy $64 k_B T$ greater than, or nearly twice, the simulation result. Not accounting for non-Gaussian density fluctuations as the solute size increases subsequently leads to increasingly inaccurate predictions of solute hydration free energies.

4.2 Structural Characterization of Cavity Emptying. Insight into the cavity emptying mechanism can be gained by examining the n dependent RDFs between the cavity and water. In Figure 6 we report $g_n(r)$ for $n = 0, 1, 35$, and 45 waters within the 6.3 Å cavity, representative of occupancies from empty to well above $\langle n \rangle$ (RDFs across all cavity occupancy states are reported in Figure S1 in the Supporting Information). The RDF for the empty volume ($g_0(r)$, Figure 6a) corresponds to the RDF between a hard sphere solute and the solvent. Water clearly packs around the 6.3 Å hard solute, with a contact peak approximately 60% greater than the bulk solvent density. This contact density is comparable to that observed for similarly sized hard sphere solutes in multiple different representations of water.^{15, 16, 44} The RDF between water and a cavity with one water molecule within the solute ($g_1(r)$, Figure 6b) is nearly the same as that for the empty volume for separations outside the cavity. As anticipated in section 2.1 above, the water density changes discontinuously across the cavity boundary. Within the cavity we find the lone water tends to

adsorb onto the inner surface of the cavity, indicated by the red arrow pointing to the inner packing peak. This water presumably attaches itself to the inner surface to gain attractive interactions with the waters outside the volume that it no longer finds within the emptied solute. The lower value of the packing density of water against the inner wall compared to that against the outer wall indicates that water entry into the cavity is favorable. Similar results are observed for cavities with $n < \langle n \rangle$, as seen in Figures S1c – S1ai in the Supporting Information.

When the cavity contains approximately a number of waters close to the mean value ($g_{35}(r)$, Figure 6c) the RDF is practically featureless and nearly equal to 1 across all separations, both inside and outside the cavity. This observation reflects that when averaged overall occupancy states the RDF between the cavity and water is 1 (eq. (22)) and that the occupancy states closest to $\langle n \rangle$ comprise the dominant contribution to this average. Since the RDF is essentially uniform, it is anticipated that there is little to no barrier for water entry or exit from the cavity when the contact density inside and outside are nearly equal.

Finally, when the cavity occupancy is significantly greater than that expected by the bulk density the waters within the volume are expected to pack more like they would in a high pressure solid than in a liquid at ambient conditions. When the cavity contains 45 waters the density inside the volume is 29% greater than the bulk. In this case the RDF between the cavity and water ($g_{45}(r)$, Figure 6d) exhibits significant structuring. Inside the volume, the waters pack into approximately two bands as indicated by the two peaks in the RDF for $r < 6.3 \text{ \AA}$. The dominant peak inside the cavity is the one pressed against the inner surface. The waters outside the volume, on the other hand, appear to be repelled from its outer surface. This reflects the fact that the waters inside the volume are so tightly packed against the inner wall they largely present their excluded volumes to the external waters, resulting in the observed depletion. The net result of this is that the

density of water pressed against the inner wall is greater than that against the outer wall, forming a barrier for water entry into the cavity. Similar results are observed for cavities with $n > \langle n \rangle$, as seen in Figures S1ak – S1as in the Supporting Information.

The segregation of the solvent molecule to the inner surface of the cavity as it is emptied can be quantified by considering the radius-of-gyration of the collection of waters inside that volume (Figure 7), evaluated as

$$R_g^2(n) = \frac{\int_0^R r^4 g_n(r) dr}{\int_0^R r^2 g_n(r) dr}. \quad (32)$$

For the case of the 6.3 Å cavity, R_g exhibits a non-monotonic dependence on n (Figure 7a). Beginning near $\langle n \rangle$ (= 34.9 for the 6.3 Å cavity), R_g decreases with decreasing $\langle n \rangle$ before reaching a minimum at $n = 22$. This nearly coincides with the observed deviation from Gaussian behavior of $\ln K_n$ in Figure 4. This indicates from $n = 35$ to 22 the waters within the cavity are segregating towards its middle. For values of n less than 22, however, this trend reverses and R_g increases as water molecules are removed from the cavity (Figure 7a). This effect is such that R_g for the volume with only 1 water molecule in it is greater than that observed near $\langle n \rangle$, let alone for the largest occupancy state considered. This suggests over this range of occupancies the water molecules adsorb onto the inner surface of the cavity as it is emptied, facilitating the formation of a bubble within the cavity.

Given that the deviations from Gaussian-like behavior are more pronounced for larger cavities (e.g., Figures 2 and 4), it stands to reason that the non-monotonic dependence of R_g on n will likewise be more pronounced as well. To facilitate comparison of R_g across different sized cavities, we use the normalized variables $n/\langle n \rangle$, corresponding to the fractional water occupancy, and $R_g/(R\sqrt{3/5})$, where $R\sqrt{3/5}$ is the radius-of-gyration normalized of a sphere with a uniform

density of water inside (i.e., eq. (32) with $g_n(r) = 1$). In Figure 7b we compare solvent R_g 's for spheres 3.3 Å, 4.3 Å, 5.3 Å, and 6.3 Å in radius. For occupancies greater than that for which the minimum in R_g is observed, each of these cavities appear to follow a universal dependence on n using the normalized variables. The values of $n/\langle n \rangle$ for which R_g is minimized, on the other hand, depends on the size of the cavity. For the 3.3 Å volume the minimum occurs when $n = 2$, or $n/\langle n \rangle \approx 0.40$. As the volume gets bigger, the value of $n/\langle n \rangle$ for which R_g is minimized shifts to even larger values, occurring near $n/\langle n \rangle \approx 0.63$ for the 6.3 Å volume. This indicates that on the basis of $n/\langle n \rangle$, the range of fractional occupancy states for which a bubble is stable within the cavity grows as the cavity size increases.

The knitting together of internal and external portions of the RDFs to determine the cavity correlation functions, $y_n(r)$, for the 6.3 Å sphere is illustrated in Figure 8. Shifting $-\ln g_1(r < 6.3 \text{ Å})$ down by 3.5 to meet $-\ln g_0(r > 6.3 \text{ Å})$ at contact yields a smooth result for $-\ln y_0(r)$ across all separations (Figure 8a). The resulting downward shift corresponds to a $-k_B T \ln K_0 = -3.5 k_B T$ drop in the free energy for adding a single water molecule to an empty cavity. This free energy drop reflects the observed lower value of $g_1(R^-)$ at contact with the inner surface of the cavity (Figure 6b) compared to the great value of $g_1(R^+)$ at contact with the outer surface (Figure 6a). Alternately, shifting $-\ln g_{45}(r < 6.3 \text{ Å})$ up by 2.1 to meet $-\ln g_{44}(r > 6.3 \text{ Å})$ at contact yields a smooth result for $-\ln y_{44}(r)$ across all separations (Figure 8b). So, in difference to adding a water molecule to an empty cavity in water, adding a water to an overly packed volume results in an unfavorable $-k_B T \ln K_{44} = 2.1 k_B T$ free energy increase. This, likewise is reflected by differences in the packing of water against the inner surface versus depletion of waters from the outer surface of the cavity for overly packed states (e.g., Figure 6d).

In Figure 9a we compare values of $\ln K_n$ determined by umbrella sampling against that determined from knitting together the cavity RDFs for the 6.3 Å cavity (Similarly good results for the 3.3 Å, 4.3 Å, and 5.3 Å cavities are reported in Figure S2 in the Supporting Information). The agreement between these two approaches is excellent, such that the p_n distributions determined by either technique are indistinguishable (Figure 9a inset). In Figure 9b we show simulation snapshots of slices through the 6.3 Å cavity at n values of 35, 30, 25, 20, 15, 10, 5, and 0. In the Gaussian regime ($n = 35, 30$, and 25), the water molecules inside the cavity appear, to a first approximation, uniformly spread across the cavity. A small bubble appears in the lower lefthand side of the $n = 20$ simulation snapshot, just below the water occupancy level for which $\ln K_n$ breaks from the Gaussian prediction. As n decreases further the bubble within the cavity grows even larger until it is completely empty ($n = 0$). For the $n = 15, 10$, and 5 states, the water molecules within the cavity clearly adsorb to the inner surface of the cavity, largely a result of their attraction to the waters in the bulk solvent and consistent with the observed growth of R_g as the cavity is emptied (Figure 7a).

4.3 Theoretical Accounting of Non-Gaussian Density Fluctuations. Here we develop a theoretical framework to modify IGFT to account for non-Gaussian contributions in the description of solvent density fluctuation within a cavity. Following the observations made above, we expect the onset of non-Gaussian fluctuations to occur over a narrow range of cavity occupancies (e.g., Figure 4), followed by the formation of an empty bubble within the cavity as more waters are removed (e.g., Figure 9). Within the bubble the remaining waters are adsorbed onto the inner surface of the cavity. These observations form the basis for our extension of the thermodynamic description of cavity emptying.

In the Gaussian fluctuation regime, IGFT predicts that the free energy difference associated with removing a water molecule from a cavity is

$$\beta(\mu_n^{ex} - \mu_{n+1}^{ex}) = \ln\left(\frac{p_{n+1}}{p_n}\right) \approx \frac{\partial \ln p_n}{\partial n} = \frac{(\langle n \rangle - n)}{\langle n \rangle \chi}. \quad (33)$$

Note that $\ln(p_{n+1}/p_n) = \ln K_n$ is at best an approximation for the derivative of p_n at $n + 1/2$, since the derivative in this expression assumes n is continuous rather than discrete. The derivation provided here assumes n is a continuous variable as in the development of IGFT above. This assumption, however, does not impact the outcome of the present derivation.

For large fluctuations that sufficiently reduce the density within the cavity we hypothesize that a single bubble is nucleated within its boundary. In this case, the thermodynamics of reducing the solvent occupation number is governed by the work against the bulk pressure and the interfacial tension of the bubble. Assuming the bubble can be treated as a spherical void in the cavity carved out from the solvent, the volume and radius of the empty bubble within cavity are

$$V_{bubble} = \frac{\langle n \rangle - n}{\rho} \quad (34)$$

and

$$R_{bubble} = \left(\frac{3}{4\pi\rho}\right)^{1/3} (\langle n \rangle - n)^{1/3}. \quad (35)$$

The surface upon which the interfacial tension acts is not necessarily determined by R_{bubble} , but rather at an effective radius given as

$$R_{effective} = R_{bubble} + \Delta R = \left(\frac{3}{4\pi\rho}\right)^{1/3} [(\langle n \rangle - n)^{1/3} + \delta], \quad (36)$$

where $\Delta R = (3/4\pi\rho)^{1/3}\delta$ is the difference between the effective and actual bubble radii, and is qualitatively similar to a Tolman length.⁴⁵ The resulting effective surface area of the bubble is

$$A_{effective} = \left(\frac{36\pi}{\rho^2}\right)^{1/3} [(\langle n \rangle - n)^{1/3} + \delta]^2, \quad (37)$$

The free energy associated with creating a cavity with n waters and a bubble within its bounds is subsequently determined by the work against the pressure and surface tension as

$$\beta\mu_n^{ex} = Z(\langle n \rangle - n) + \beta\gamma \left(\frac{36\pi}{\rho^2}\right)^{1/3} [(\langle n \rangle - n)^{1/3} + \delta]^2 + \varepsilon, \quad (38)$$

where $Z = \beta P/\rho$ is the compressibility factor, γ is the surface tension of the bubble interface (assumed here to be the vapor/liquid surface tension), and ε is a constant that ensures continuity of the free energy between the Gaussian and non-Gaussian regimes. As above, the work associated with removing a single water molecule from the cavity is

$$\beta(\mu_n^{ex} - \mu_{n+1}^{ex}) \approx -\beta \frac{\partial \mu_n^{ex}}{\partial n} = Z + \beta\gamma \left(\frac{32\pi}{3\rho^2}\right)^{1/3} \frac{(\langle n \rangle - n)^{1/3} + \delta}{(\langle n \rangle - n)^{2/3}}. \quad (39)$$

We propose that a bubble is nucleated inside the cavity when the free energy of removing a water molecule following eq. (39), the bubble growth path, is equal to that determined following eq. (33), the Gaussian fluctuation path. For states with occupancies less than this transition point, the free energies for removing waters from the cavity following eq. (39) are lower than that determined by eq. (33). The emptying of the cavity subsequently follows the path for removing water that is lowest in free energy. The occupation number at which the bubble is nucleated, n^* , is determined by equating eqs. (33) and (39)

$$\frac{(\langle n \rangle - n^*)}{\langle n \rangle \chi} = Z + \beta\gamma \left(\frac{32\pi}{3\rho^2}\right)^{1/3} \frac{(\langle n \rangle - n^*)^{1/3} + \delta}{(\langle n \rangle - n^*)^{2/3}}. \quad (40)$$

A general solution of this expression requires numerical solution of a quintic equation. Nevertheless, an accurate approximate solution can be derived. Assuming pressure contributions are negligible ($Z \approx 0$), an excellent approximation near atmospheric pressure, and $\delta = 0$ we find

$$n^* = \langle n \rangle - (\beta\gamma\langle n \rangle\chi)^{3/4} \left(\frac{32\pi}{3\rho^2}\right)^{1/4}. \quad (41)$$

Expanding the solution of eq. (40) in Z and δ to second order we find

$$\begin{aligned}
n^* = \langle n \rangle - (\beta\gamma\langle n \rangle\chi)^{\frac{3}{4}} \left(\frac{32\pi}{3\rho^2}\right)^{\frac{1}{4}} - \frac{3(\beta\gamma\langle n \rangle\chi)^{\frac{1}{2}}}{4} \left(\frac{32\pi}{3\rho^2}\right)^{\frac{1}{6}} \delta - \frac{3\langle n \rangle\chi}{4} Z + \frac{9(\beta\gamma\langle n \rangle\chi)^{1/4}}{32} \left(\frac{32\pi}{3\rho^2}\right)^{1/12} \delta^2 \\
- \frac{3\langle n \rangle^2\chi^2}{32(\beta\gamma\langle n \rangle\chi)^{3/4}} \left(\frac{3\rho^2}{32\pi}\right)^{1/4} Z^2 + \frac{3\langle n \rangle\chi}{16(\beta\gamma\langle n \rangle\chi)^{1/4}} \left(\frac{32\pi}{3\rho^2}\right)^{1/12} \delta Z
\end{aligned} \quad (42)$$

For small values of Z and δ , as those found here, use of eq. (42) is simpler than numerical solution of eq. (40). Indeed, use of the terms linear in Z and δ yields results for the free energy (below) that are practically indistinguishable from the numerical solution of n^* . Considering the linear terms in eq. (42), it is apparent that n^* is always less than $\langle n \rangle$. It is possible to find negative values of n^* from eqs. (40) and (42). In this case no bubble is nucleated within the cavity and density fluctuations are Gaussian over the full range of solvent occupancies.

Ensuring that eq. (29a) continuously joins with eq. (38) at n^* to determine ε , the free energy of finding n waters in the cavity is

$$\beta\mu_n^{ex} = \begin{cases} \frac{Z(n^* - n)}{+ \beta\gamma \left(\frac{36\pi}{\rho^2}\right)^{1/3} \left\{ [(\langle n \rangle - n)^{1/3} + \delta]^2 - [(\langle n \rangle - n^*)^{1/3} + \delta]^2 \right\}} & n < n^* \\ \frac{(n^* - \langle n \rangle)^2}{2\langle n \rangle\chi} + \frac{1}{2} \ln(2\pi\langle n \rangle\chi) & \\ \frac{(n - \langle n \rangle)^2}{2\langle n \rangle\chi} + \frac{1}{2} \ln(2\pi\langle n \rangle\chi) & n \geq n^* \end{cases}, \quad (43)$$

Since n^* typically is two or more standard deviations less than $\langle n \rangle$, normalization of the $p_n (= \exp(-\beta\mu_n^{ex}))$ distribution is well approximated by the normalization of the underlying Gaussian distribution. As such, eq. (43) is effectively already normalized. The free energy of creating an empty cavity ($n = 0$) is subsequently

$$\beta\mu_0^{ex} = \begin{cases} \frac{Zn^*}{+ \beta\gamma \left(\frac{36\pi}{\rho^2}\right)^{1/3} \left\{ [\langle n \rangle^{1/3} + \delta]^2 - [(\langle n \rangle - n^*)^{1/3} + \delta]^2 \right\}} & n^* > 0 \\ \frac{(n^* - \langle n \rangle)^2}{2\langle n \rangle\chi} + \frac{1}{2} \ln(2\pi\langle n \rangle\chi) & \\ \frac{\langle n \rangle}{2\chi} + \frac{1}{2} \ln(2\pi\langle n \rangle\chi) & n^* \leq 0 \end{cases}. \quad (44)$$

We refer to this theory as the augmented fluctuation theory (AFT).

We note that AFT draws inspiration from the description of non-Gaussian fluctuations proposed by Huang and Chandler.²⁸ The main differences with their proposal and the present derivation are that we introduce a Tolman-like length to the description of interfacial contributions (eq. (37)), identify the cavity occupancy at which the solvent density fluctuations transition from Gaussian to non-Gaussian (eq. (40)), and neglect contributions from the volume accessible to the bubble within the cavity. The first reason we neglect the volume accessibility contribution is it predicts a divergent hydration free energy for the empty cavity since it assigns a zero probability of observing an empty cavity, i.e., $p_0 = 0$, due to the reduction of the accessible volume to zero when the bubble volume matches the cavity volume. Secondly, we hypothesize that the bubble becomes pinned to the center of the cavity as n approaches zero, as seen in Figure 9. The reason for this pinning is because, for smaller values of n either the bubble of volume $(\langle n \rangle - n)/\rho$ resides entirely within the cavity or a larger bubble that is off center overlaps the cavity so that n solvent molecules on one side of the bubble remain in the cavity. Since the probability of observing bubbles even larger than $(\langle n \rangle - n)/\rho$ becomes negligibly small as the bubble grows, the most likely outcome is that the bubble contained within the cavity is pinned near its center as its occupancy approaches zero.

4.4 Comparison of AFT with Molecular Simulation. We fit AFT to the simulation results for $\beta\mu_0^{ex}$ obtained from simulation using umbrella sampling over the range 2.5 Å to 6.3 Å (Figure 5). This fitting was conducted using the simulation temperature, pressure, and density of pure mW water. Molinero et al. report a value of 66.0 dyne/cm at 300 K. We subsequently estimate a surface tension of 66.3 dyne/cm at 298.15 K for our fitting using the experimental temperature dependence

of the surface tension of water to correct for the slight temperature difference. The fitted value of δ was 0.83, corresponding to a radial increment of $\Delta R = 1.6 \text{ \AA}$. We compare the fitted predictions of AFT for the chemical potential of a hard sphere solute in mW water as a function of the solute size against that determined from simulation using umbrella sampling in Figure 5. AFT provides a significantly improved description of the chemical potential over IGFT with quantitative accuracy over the range of solute sizes examined here, giving us confidence in the physics incorporated into the theory.

In Figure 3 we compare of simulation results for p_n against the predictions of AFT for the 4.3 $\text{\AA}$, 5.3, $\text{\AA}$, and 6.3 $\text{\AA}$ cavities. Overall, the predictions of AFT are improved against that using IGFT. Notably, AFT accurately captures the fat tail of these distributions. We do observe some differences between the simulations and theory in the case of the 6.3 $\text{\AA}$ sphere. These differences are clearer if we consider the $\ln K_n$ distributions for these cavities (Figure 4). Most significantly, AFT accurately captures the break in K_n associated with the transition from the Gaussian to non-Gaussian regime below n^* . For cavity occupancies below n^* , however, AFT captures the n dependence of $\ln K_n$ only semi-quantitatively. In particular, the simulation results observe a shallow minimum in $\ln K_n$ at intermediate values of n between 0 and n^* , while AFT only predicts on weak monotonic increase in $\ln K_n$ with n over this range. We hypothesize that the non-monotonic dependence of $\ln K_n$ over this range is the result of capillary fluctuations at the boundary of the cavity making it harder to remove those final water molecules from the volume, although we have not confirmed this. Nevertheless, AFT accurately threads between the simulation results such that the comparison with simulation in Figure 3 is significantly improved.

In Figure 10 we compare the predictions of n^* by AFT as a function of the cavity radius against that determined from simulation. We estimated n^* from our simulation results two different ways.

The first way, we determined the value of n at which $\ln K_n$ exhibits a peak, shifted up by $1/2$ to better compare the simulation results with the derivative of $\ln p_n$ used to evaluate the transition point in AFT (eq. (40)). The second way we used to evaluate n^* from simulation was to determine the value of n at which R_g is a minimum (Figure 7). As can be seen in Figure 10, AFT accurately captures the size dependence of n^* , especially as the cavity volume grows. This comparison does breakdown for the smallest volumes considered, which is not unexpected given that AFT assumes a continuum treatment of thermodynamics even down to atomic scales. In addition, the agreement between the n^* values determined using either K_n or R_g is quite good, although the R_g estimate tends to be slightly larger. This provides strong evidence that the growth in R_g as the occupation number decreases is directly linked to the onset of non-Gaussian fluctuations in the cavity.

5. Conclusions.

In this paper we presented a molecular simulation study of the emptying of atomic- and meso-scale cavities in water to get a mechanistic understanding of this process and its relationship to the thermodynamics of hydrophobic hydration. As found in previous studies, we demonstrated for sufficiently large enough volumes, the nature of the solvent density fluctuations within a solute cavity transitions from Gaussian to non-Gaussian. The non-Gaussian wing of the distribution as the volume empties is significantly more probable than would be expected if the density fluctuations were normally distributed, which has previously been referred to as a fat-tail distribution. A structural analysis of the waters contained within the cavity finds that the occurrence of a fat-tail in the distribution is accompanied with the formation of a bubble within the cavity. A statistical thermodynamic analysis of solvent packing on either side of the boundary between the cavity's interior and the bulk solvent demonstrated that the formation of the bubble

results in the net adsorption of water onto the cavity’s inner surface, reducing the penalty for removing waters from the cavity thereby fattening the distribution.

The finding that the formation of a bubble within the cavity occurs over a narrow range of occupancy states led us to propose an empirical correction to a theory we previously developed to account for Gaussian solvent density fluctuations within solute cavities, IGFT. This augmented fluctuation theory, or AFT, accounts for the formation of a bubble by smoothly joining IGFT with a macroscopic thermodynamic description of a bubble’s interface at the transition occupancy. AFT successfully describes the free energies of hard sphere solute hydration over a much broader range of solute size scales than IGFT, capturing the fat-tail distribution and predicting the cavity occupancies at which the bubble forms. We note that while we only considered solutes up to 6.3 Å in radius in a coarse-grained representation of water, we previously reported simulations of a more realistic description of water with solutes up to 18 Å in radius where it was demonstrated that a nascent version of AFT accurately describes bubble formation in this system.³² This lends confidence in the accuracy of the theory presented here over a wide range of cavity sizes.

The statistical thermodynamic framework introduced here connecting the p_n distribution to solvent packing both inside and outside a solute cavity provides both a new route for evaluating solvent density fluctuations within the volume and for interpreting their origin. We demonstrated here that the transition from Gaussian to non-Gaussian like behavior in the emptying of a cavity is accompanied with the adsorption of water molecules inside the volume onto the inner surface of the cavity. As a result of this adsorption and accompanying bubble formation, the penalty for removing waters from the cavity is significantly lower than that would be anticipated assuming Gaussian fluctuations. Within the context of this statistical thermodynamic framework, it is worthwhile to consider other reasons large density fluctuations can deviate from Gaussian

behavior. Notably, simulations of alternate solvents that include dense hard spheres,⁴⁶⁻⁴⁸ an isotropic model reproducing water's RDF,⁴⁹ the Jagla model,⁵⁰ and a water-like model with weakened hydrogen-bonds⁵⁰ found that solvent density fluctuations within cavities appear Gaussian near $\langle n \rangle$, but exhibit significantly suppressed occupation probabilities as $n \rightarrow 0$. This would be indicated by large positive deviations from the linear Gaussian response on a $\ln K_n$ plot, in difference to the negative deviations observed here. Compared to water these solvents are more repulsive to one another. For example, the isotropic water model noted above has a pressure of 7500 bar at 1 g/cm³ at 25°C.⁴⁹ Rather than adsorb onto the inner surface of the cavity to gain attractive interactions with the bulk solvent, we anticipate these repulsive solvents would be pushed from the inner surface to positions deeper inside the cavity. As such, we would expect the inner surface adsorption peak observed in Figure 6b to migrate towards the center of the cavity with only a minimal contact density at the inner surface. This would establish a more significant barrier for removing those solvent molecules from the cavity, suppressing larger density fluctuations in accord with the simulation observations.

The present theory does not account for contributions like capillary fluctuations at the bubble interface that can make it harder to fully empty the cavity. In addition, for non-spherical volumes additional contributions associated with Gaussian curvature and higher order corrections may also have to be accounted for.⁵¹ Nevertheless, we expect the present description will still capture the onset of non-Gaussian solvent density fluctuations even in the case of non-spherical cavities since we expect the initial bubble will be spherical, only adopting its final non-spherical shape as the cavity is emptied.

6. Acknowledgements.

We would like to thank the NSF through grant CBET-1805167 for supporting this research. We also like to thank Prof. Shekhar Garde, with whom we have had many insightful conversations about this work and hydrophobic hydration over the years.

Figure Captions

Figure 1. Water occupational probability distributions, p_n , within cavities 2.1 Å, 2.5 Å, 2.9 Å, 3.3 Å, and 3.7 Å in radius at 25°C and 1 atm. Simulation results determined by umbrella sampling are compared against the predictions of IGFT using symbols defined in the figure legend. The simulation errors are smaller than the figure symbols.

Figure 2. Differential change in the water occupational probability, $\ln K_n = \ln(p_{n+1}/p_n)$ for the 2.1 Å, 2.9 Å, and 3.7 Å radius cavities at 25°C and 1 atm. Simulation results determined by umbrella sampling are compared against the predictions of IGFT using symbols defined in the figure legend. The simulation errors are smaller than the figure symbols.

Figure 3. Water occupational probability distributions, p_n , within cavities 4.3 Å, 5.3 Å, and 6.3 Å in radius at 25°C and 1 atm. Simulation results determined by umbrella sampling are compared against the predictions of IGFT and AFT using symbol defined in the figure legend. The simulation errors are smaller than the figure symbols.

Figure 4. Differential change in the water occupational probability, $\ln K_n = \ln(p_{n+1}/p_n)$ for the 4.3 Å, 5.3 Å, and 6.3 Å radius cavities at 25°C and 1 atm. Simulation results determined by umbrella sampling are compared against the predictions of IGFT and AFT using symbol defined in the figure legend. The simulation errors are smaller than the figure symbols.

Figure 5. Excess chemical potentials of hard sphere solutes (i.e., empty cavities) in water as a function of their radius at 25 °C and 1 atm. Simulation results determined by test particle insertion

and umbrella sampling are compared against the predictions of IGFT (eq. (29b)) and AFT (eq. (44)) using the symbols defined in the figure legend. Simulation errors are smaller than the figure symbols.

Figure 6. Cavity/water radial distribution functions, $g_n(r)$, for water occupancies of **a)** $n = 0$, **b)** $n = 1$, **c)** $n = 35$, and **a)** $n = 45$ for the 6.3 Å cavity. The blue and red circles in these figures indicates the outer and inner cavity surface contact values of each radial distribution function. The red arrow in **b** indicates the adsorption peak for one water molecule on the inner surface of the cavity.

Figure 7. Radius-of-gyration of waters within a cavity as a function of n as evaluated by eq. (32) at 25°C and 1 atm determined from simulation. **a)** Results for the 6.3 Å radius spherical cavity. **b)** Results for the 3.3 Å, 4.3 Å, 5.3 Å, and 6.3 Å radius cavities. The radii-of-gyration in **b** are normalized by the radius-of-gyration of a uniform sphere ($R_g/(R\sqrt{3/5})$), while the occupation numbers are normalized by the mean solvent occupation number ($n/\langle n \rangle$).

Figure 8. Construction of the cavity correlation function, $y_n(r)$, from the $g_n(r \geq R)$ and $g_{n+1}(r < R)$ radial distribution functions for the 6.3 Å radius cavity with **a)** $n = 0$, and **b)** $n = 44$. The red arrows indicate the direction which $-\ln g_{n+1}(r < R)$ is shifted to meet $-\ln g_n(r \geq R)$ at R .

Figure 9. a) Differential change in the mW water occupational probability, $\ln K_n = \ln(p_{n+1}/p_n)$, as a function of n for 6.3 Å radius volumes at 25°C and 1 atm. The solid line indicates the results

determined using umbrella sampling, while the points indicate results determined from the cavity radial distribution functions fitted to eq. (20). The red arrows indicate the cavity occupancies shown in the simulation snapshots in **b**. The inset figure shows the p_n distribution function determined using umbrella sampling and from the cavity radial distribution functions following eq. (21). The error bars are smaller than the figure symbols. **b**) Simulation snapshots of the 6.3 Å cavity taken at cavity occupancies of $n = 35, 30, 25, 20, 15, 10, 5$, and 0. These snapshots are taken from a 4 Å thick slice through the center of the solute cavity to more clearly observe the waters inside. The cavity boundary is indicated by the black dashed circle. The waters outside and inside the cavity are colored red and cyan, respectively. The black dot in the middle of the figure indicates the center of the cavity. These images were rendered using VMD.⁵²

Figure 10. The critical bubble nucleation occupancy, n^* , as a function of the cavity radius. Results are reported from the simulations obtained by determination of the peak in $\ln K_n$ (e.g., Figure 4), from the minima in the solvent radii-of-gyration (e.g., Figure 7), and AFT following eq. (42). The difference between the n^* predictions using eqs. (40) and (42) are negligible. The figure symbols are defined in the legend.

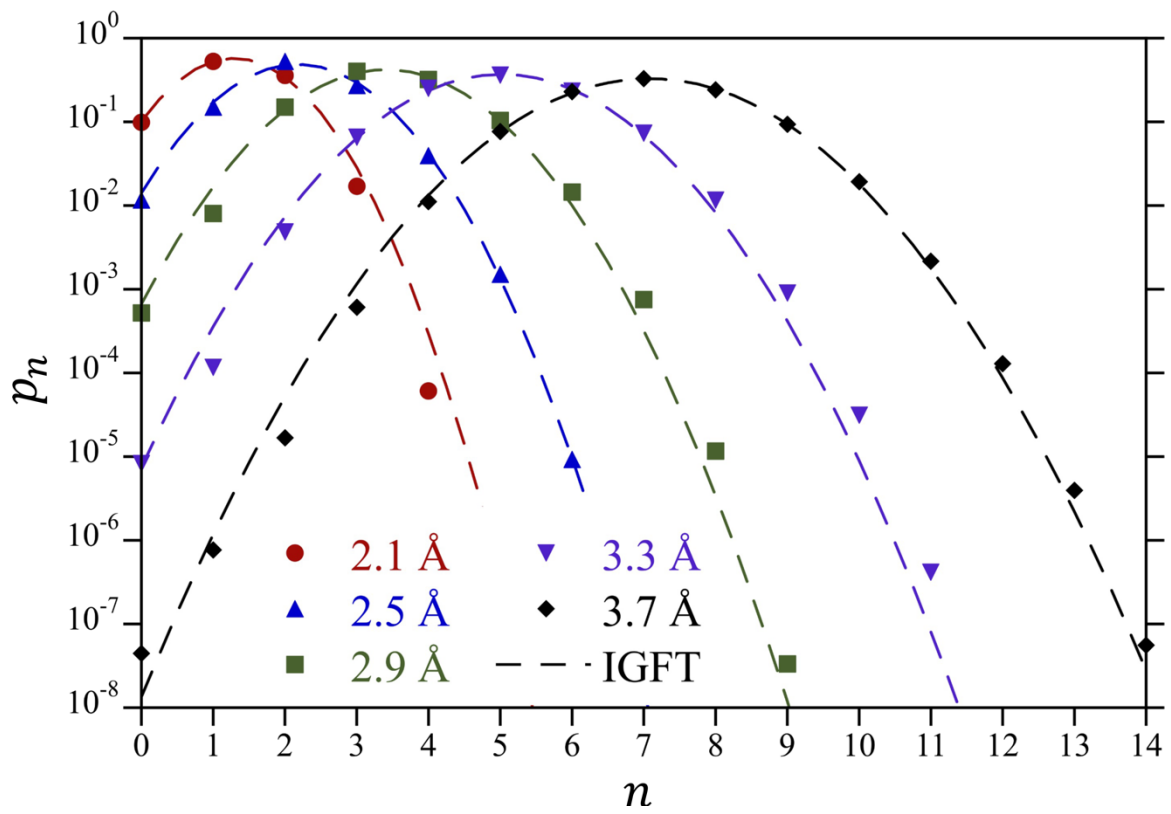


Figure 1.

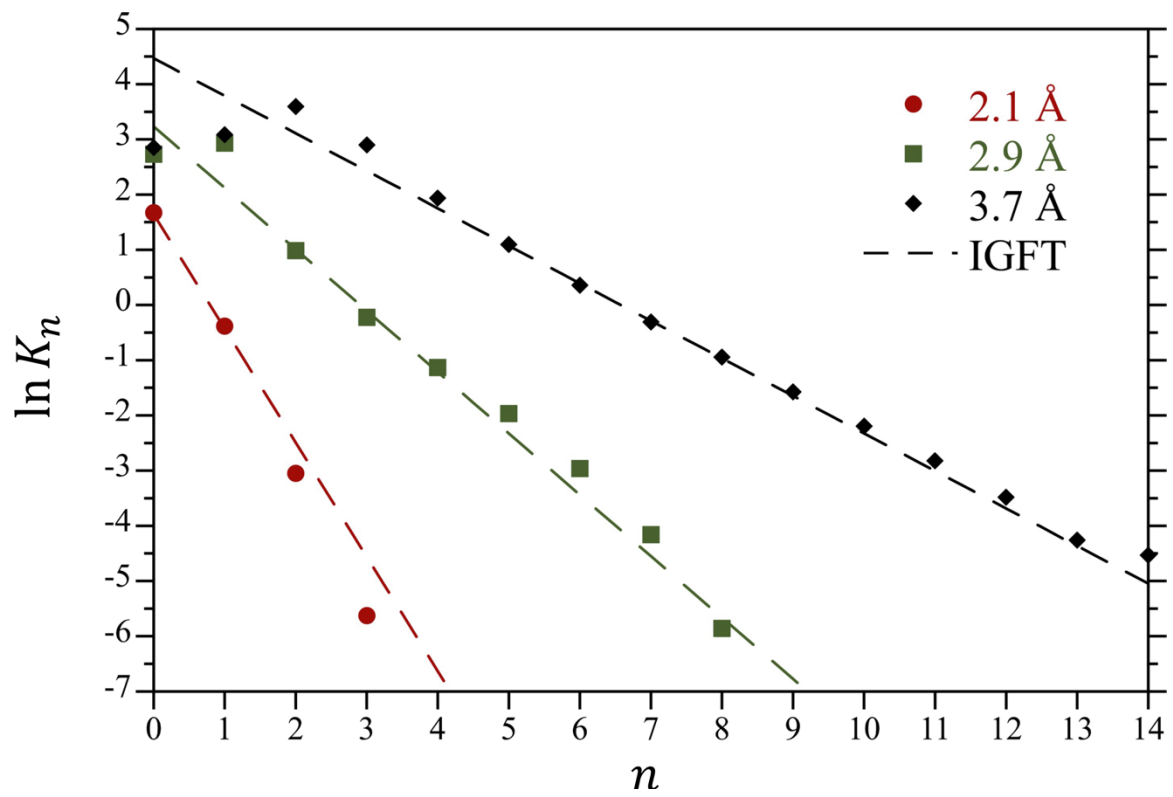


Figure 2.

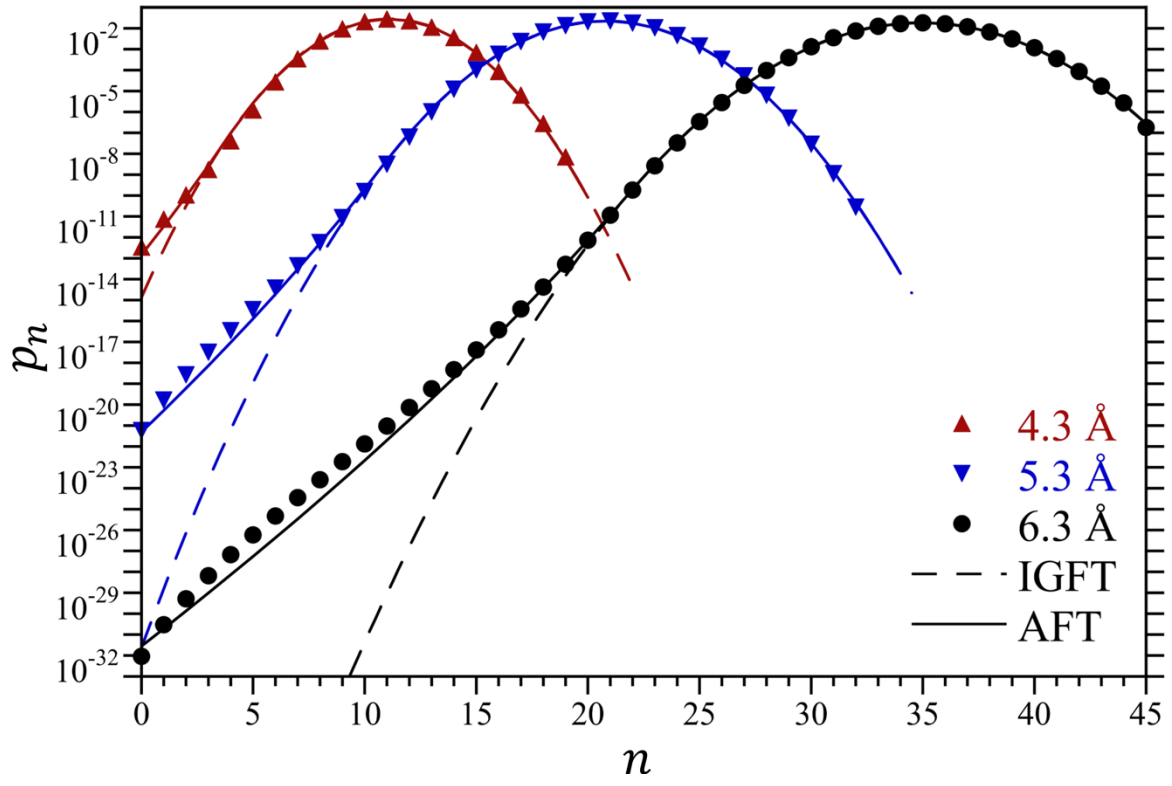


Figure 3.

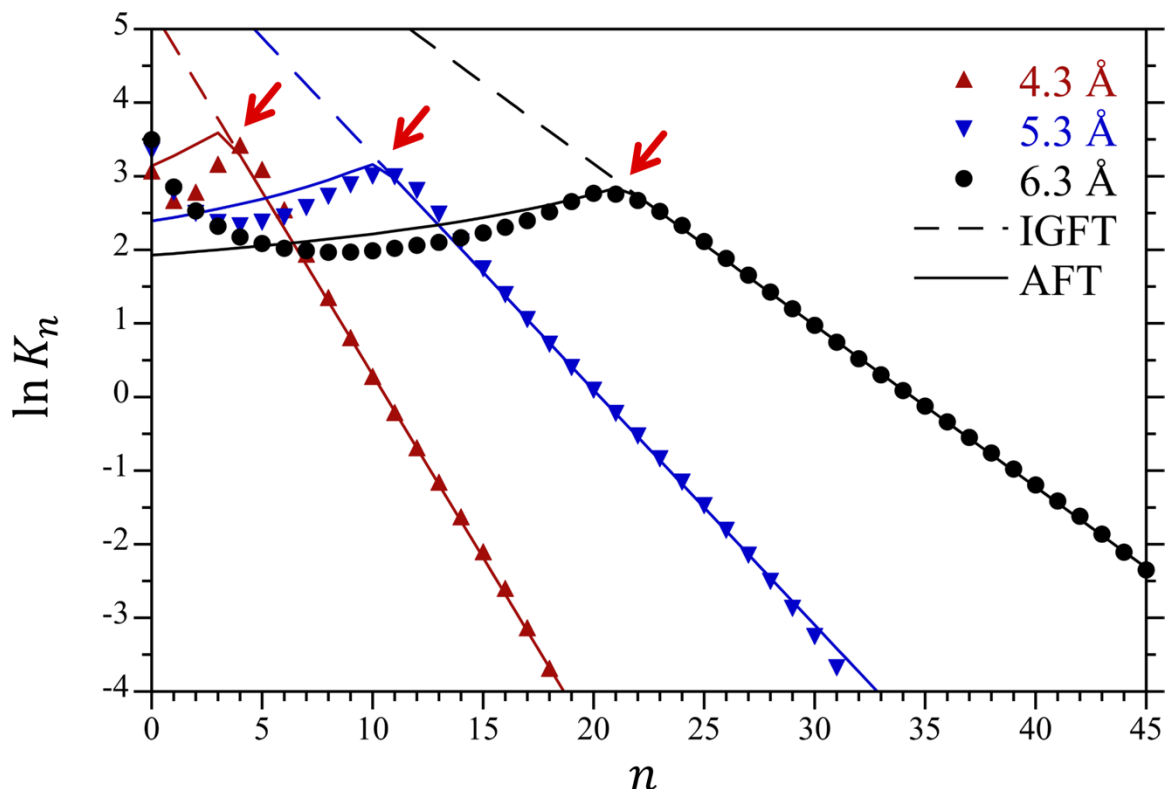


Figure 4.

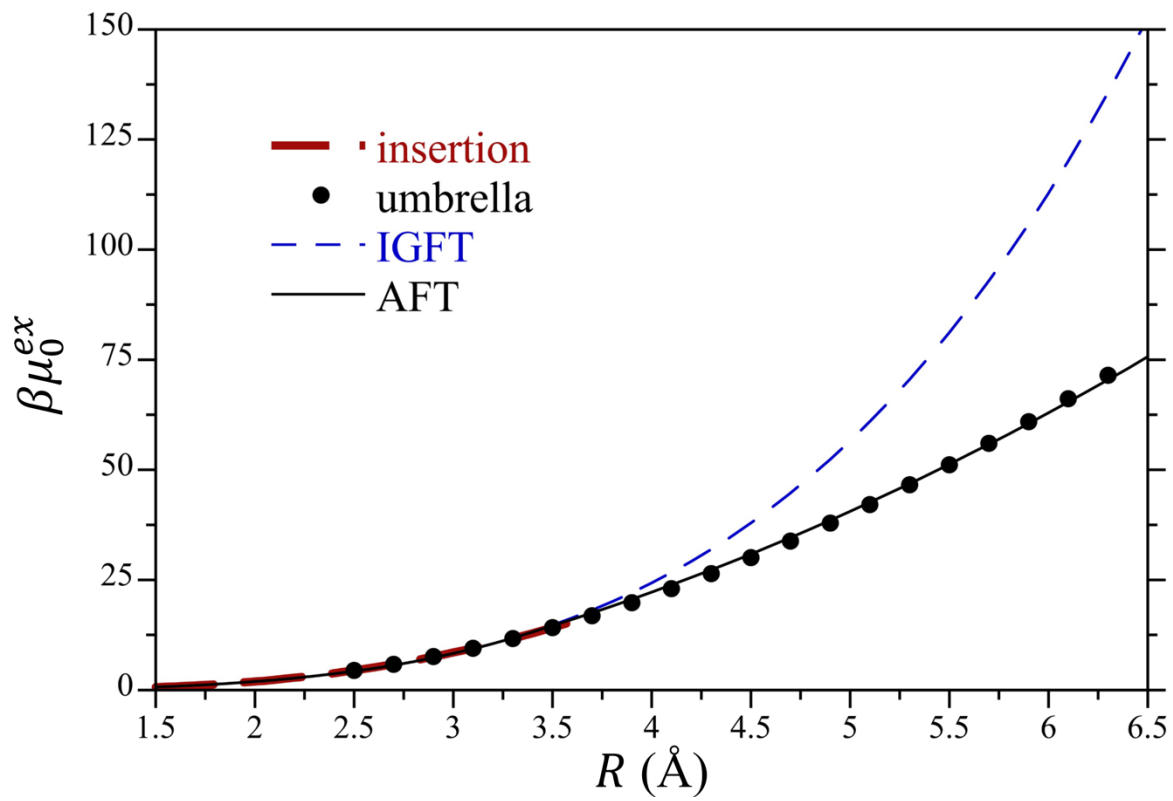


Figure 5.

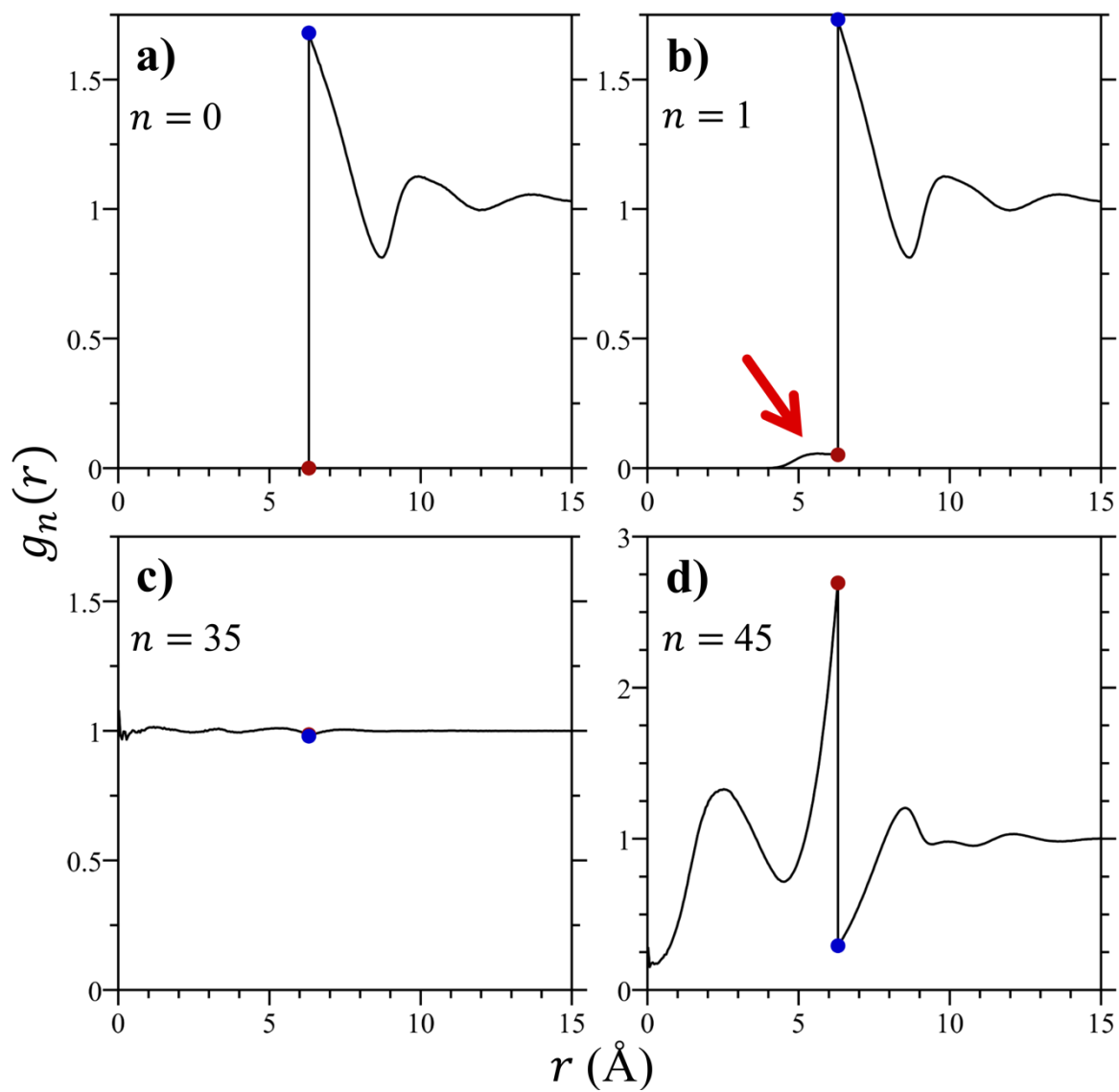


Figure 6.

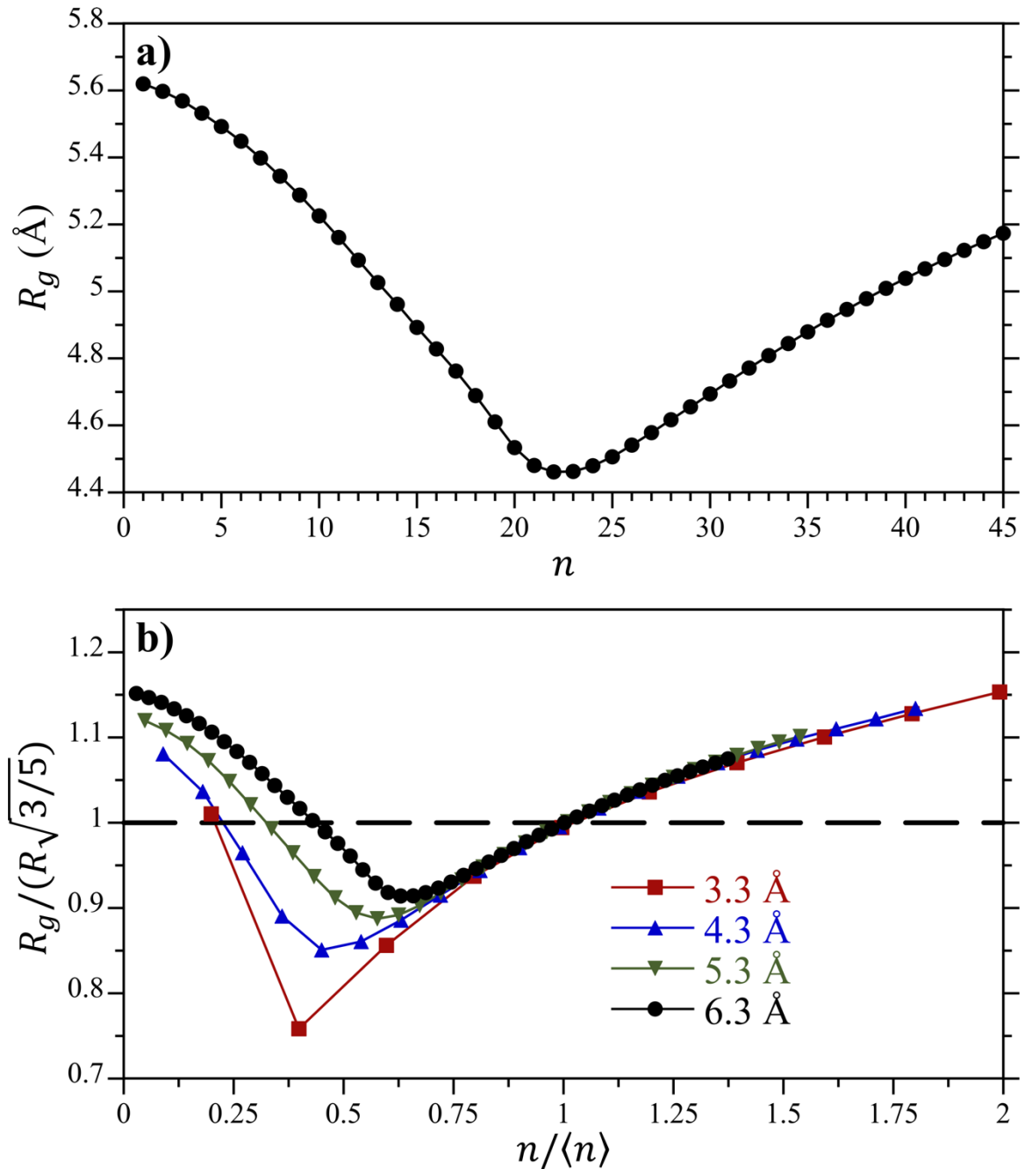


Figure 7.

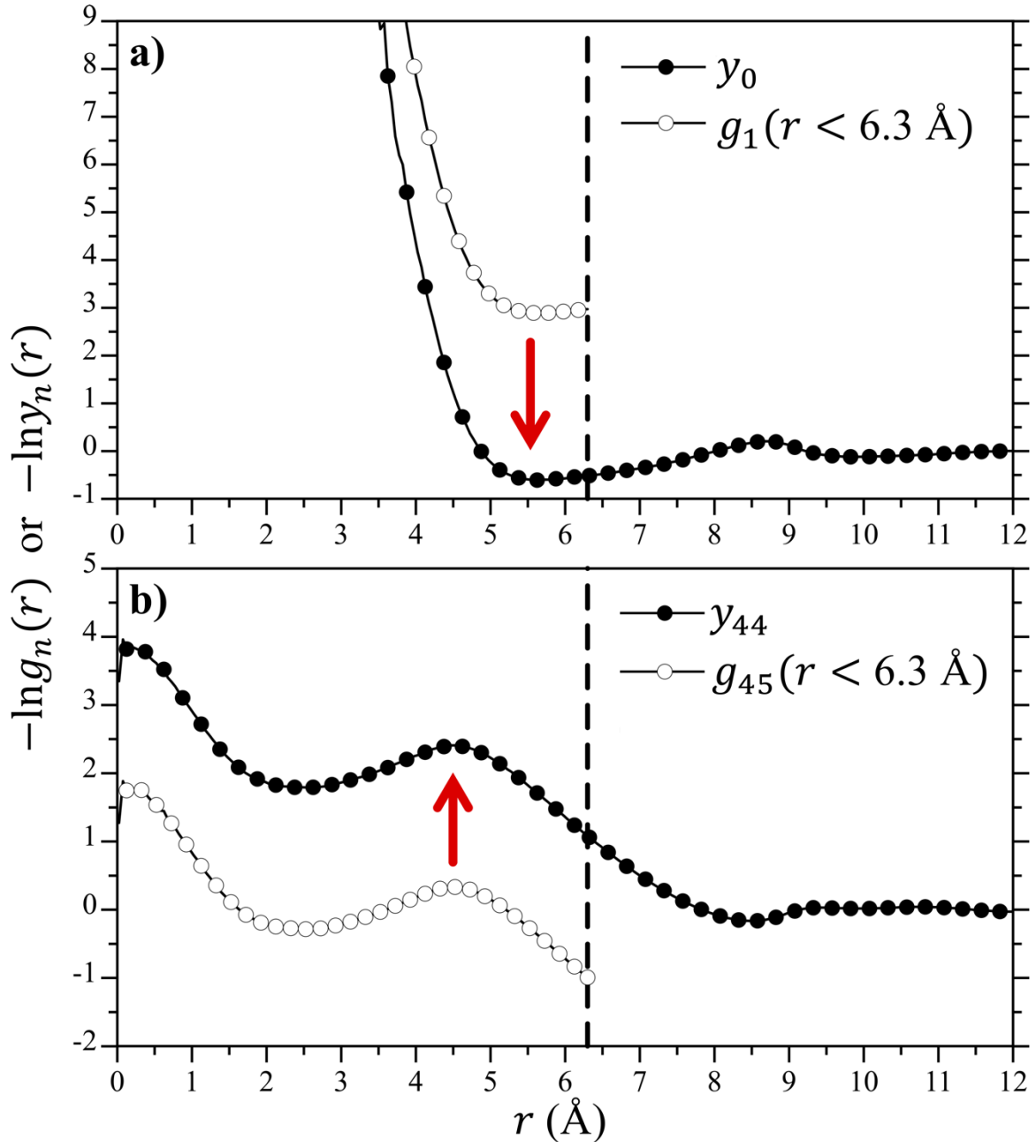


Figure 8.

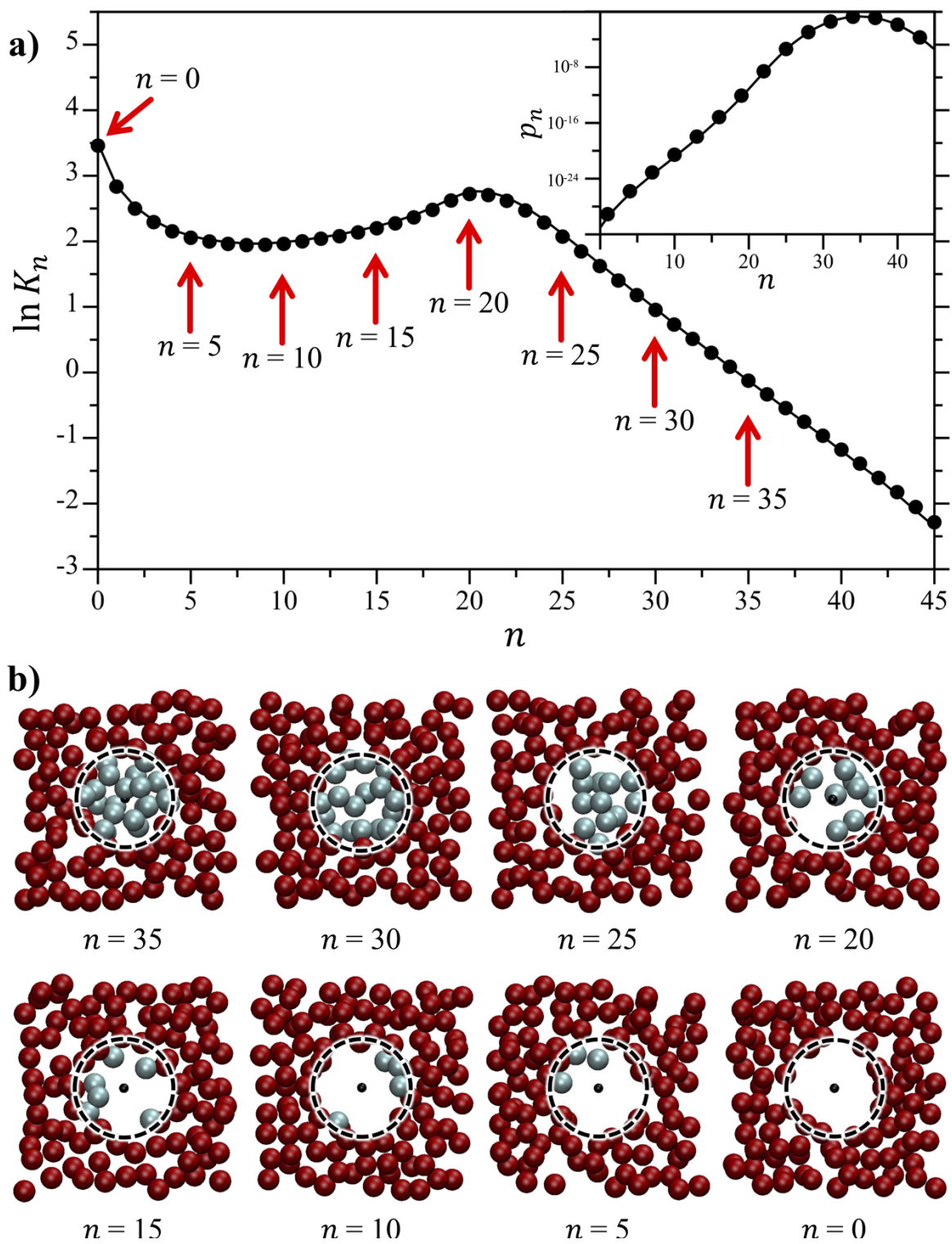


Figure 9.

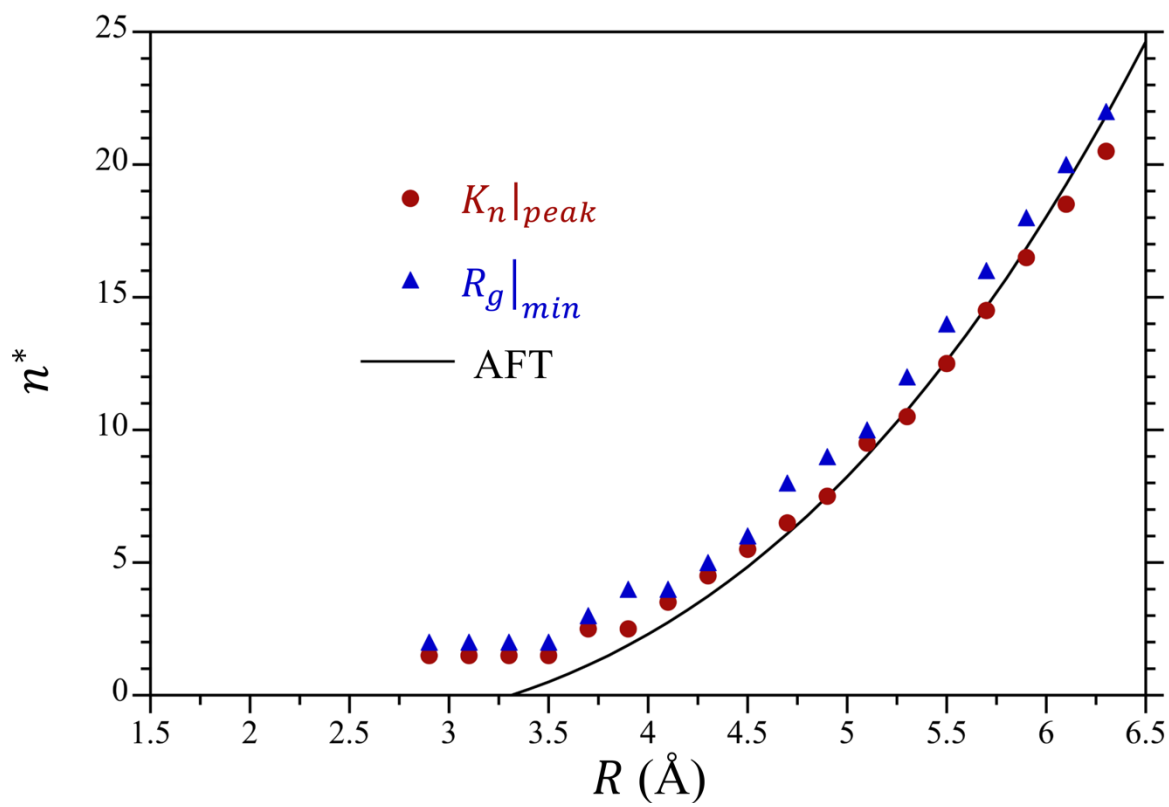


Figure 10.

References.

- (1) Tanford, C. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*; Wiley, 1980.
- (2) Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **1959**, *14*, 1-63. DOI: 10.1016/S0065-3233(08)60608-7.
- (3) Blokzijl, W.; Engberts, J. Hydrophobic effects - Opinions and facts. *Angew. Chem.-Int. Edit.* **1993**, *32* (11), 1545-1579, Review. DOI: 10.1002/anie.199315451.
- (4) Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437* (7059), 640-647. DOI: 10.1038/nature04162.
- (5) Hillyer, M. B.; Gibb, B. C. Molecular Shape and the Hydrophobic Effect. In *Annual Review of Physical Chemistry, Vol 67*, Johnson, M. A., Martinez, T. J. Eds.; Annual Review of Physical Chemistry, Vol. 67; 2016; pp 307-329.
- (6) Biedermann, F.; Nau, W. M.; Schneider, H. J. The Hydrophobic Effect Revisited-Studies with Supramolecular Complexes Imply High-Energy Water as a Noncovalent Driving Force. *Angew. Chem.-Int. Edit.* **2014**, *53* (42), 11158-11171. DOI: 10.1002/anie.201310958.
- (7) Li, I. T. S.; Walker, G. C. Signature of hydrophobic hydration in a single polymer. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108* (40), 16527-16532. DOI: 10.1073/pnas.1105450108.
- (8) Pollack, G. L. Why gases dissolve in liquids. *Science* **1991**, *251* (4999), 1323-1330. DOI: 10.1126/science.251.4999.1323.
- (9) Reiss, H.; Frisch, H. L.; Lebowitz, J. L. Statistical mechanics of rigid spheres. *Journal of Chemical Physics* **1959**, *31*, 369-380. DOI: 10.1063/1.1730361.

- (10) Pierotti, R. A. Scaled particle theory and aqueous and non-aqueous solutions. *Chemical Reviews* **1976**, 76 (6), 717-726. DOI: 10.1021/cr60304a002.
- (11) Graziano, G. On the cavity size distribution in water and n-hexane. *Biophysical Chemistry* **2003**, 104 (2), 393-405. DOI: 10.1016/s0301-4622(03)00027-9.
- (12) Stillinger, F. H. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *Journal of Solution Chemistry* **1973**, 2, 141-158. DOI: 10.1007/Bf00651970
- (13) Narten, A. H.; Levy, H. A. Liquid Water: Molecular Correlation Functions from X-Ray Diffraction. *Journal of Chemical Physics* **1971**, 55, 2263-2269. DOI: 10.1063/1.1676403.
- (14) Narten, A. H.; Levy, H. A. Observed Diffraction Pattern and Proposed Models of Liquid Water. *Science* **1969**, 165, 447-454. DOI: 10.1126/science.165.3892.447.
- (15) Ashbaugh, H. S.; Pratt, L. R. Colloquium: Scaled particle theory and the length scales of hydrophobicity. *Reviews of Modern Physics* **2006**, 78 (1), 159-178. DOI: 10.1103/RevModPhys.78.159.
- (16) Ashbaugh, H. S.; Moura, N. D.; Houser, H.; Wang, Y.; Goodson, A.; Barnett, J. W. Temperature and pressure dependence of the interfacial free energy against a hard surface in contact with water and decane. *Journal of Chemical Physics* **2016**, 145 (12). DOI: 10.1063/1.4963692.
- (17) Ashbaugh, H. S. Entropy crossover from molecular to macroscopic cavity hydration. *Chemical Physics Letters* **2009**, 477 (1-3), 109-111. DOI: 10.1016/j.cplett.2009.06.081.
- (18) Rajamani, S.; Truskett, T. M.; Garde, S. Hydrophobic hydration from small to large lengthscales: Understanding and manipulating the crossover. *Proceedings of the National*

Academy of Sciences of the United States of America **2005**, *102* (27), 9475-9480. DOI: 10.1073/pnas.0504089102.

(19) Lum, K.; Chandler, D.; Weeks, J. D. Hydrophobicity at small and large length scales. *Journal of Physical Chemistry B* **1999**, *103* (22), 4570-4577. DOI: 10.1021/jp984327m.

(20) Hummer, G.; Garde, S.; Garcia, A. E.; Pohorille, A.; Pratt, L. R. An information theory model of hydrophobic interactions. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93* (17), 8951-8955. DOI: 10.1073/pnas.93.17.8951.

(21) Garde, S.; Garcia, A. E.; Pratt, L. R.; Hummer, G. Temperature dependence of the solubility of non-polar gases in water. *Biophysical Chemistry* **1999**, *78* (1-2), 21-32. DOI: 10.1016/s0301-4622(99)00018-6.

(22) Garde, S.; Hummer, G.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. Origin of entropy convergence in hydrophobic hydration and protein folding. *Physical Review Letters* **1996**, *77* (24), 4966-4968. DOI: 10.1103/PhysRevLett.77.4966.

(23) Hummer, G.; Garde, S.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95* (4), 1552-1555. DOI: 10.1073/pnas.95.4.1552.

(24) Ashbaugh, H. S.; Vats, M.; Garde, S. Bridging Gaussian Density Fluctuations from Microscopic to Macroscopic Volumes: Applications to Non-Polar Solute Hydration Thermodynamics. *Journal of Physical Chemistry B* **2021**, *125* (29), 8152-8164. DOI: 10.1021/acs.jpcb.1c04087.

- (25) Ashbaugh, H. S. Reversal of the Temperature Dependence of Hydrophobic Hydration in Supercooled Water. *Journal of Physical Chemistry Letters* **2021**, *12* (34), 8370-8375. DOI: 10.1021/acs.jpcllett.1c02399.
- (26) Galamba, N. Free energy convergence in short- and long-length hydrophobic hydration. *Journal of Molecular Liquids* **2021**, *339*, 116699. DOI: 10.1016/j.molliq.2021.116699.
- (27) Paschek, D. How the liquid-liquid transition affects hydrophobic hydration in deeply supercooled water. *Physical Review Letters* **2005**, *94* (21), 4, Article. DOI: 10.1103/PhysRevLett.94.217802.
- (28) Huang, D. M.; Chandler, D. Cavity formation and the drying transition in the Lennard-Jones fluid. *Physical Review E* **2000**, *61* (2), 1501-1506. DOI: 10.1103/PhysRevE.61.1501.
- (29) Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Hagan, M. F.; Chandler, D.; Garde, S. Sitting at the Edge: How Biomolecules use Hydrophobicity to Tune Their Interactions and Function. *Journal of Physical Chemistry B* **2012**, *116* (8), 2498-2503. DOI: 10.1021/jp2107523.
- (30) Rego, N. B.; Xi, E.; Patel, A. J. Protein Hydration Waters Are Susceptible to Unfavorable Perturbations. *Journal of the American Chemical Society* **2019**, *141* (5), 2080-2086. DOI: 10.1021/jacs.8b11448.
- (31) Barnett, J. W.; Sullivan, M. R.; Long, J. A.; Tang, D.; Nguyen, T.; Ben-Amotz, D.; Gibb, B. C.; Ashbaugh, H. S. Spontaneous drying of non-polar deep-cavity cavitand pockets in aqueous solution. *Nature Chemistry* **2020**, *12* (7), 589-+. DOI: 10.1038/s41557-020-0458-8.
- (32) Sinha, I.; Cramer, S. M.; Ashbaugh, H. S.; Garde, S. Connecting Non-Gaussian Water Density Fluctuations to the Lengthscale Dependent Crossover in Hydrophobic Hydration. *Journal of Physical Chemistry B* **2022**, *126* (39), 7604-7614. DOI: 10.1021/acs.jpcb.2c04990.

- (33) Molinero, V.; Moore, E. B. Water Modeled As an Intermediate Element between Carbon and Silicon. *Journal of Physical Chemistry B* **2009**, *113* (13), 4008-4016. DOI: 10.1021/jp805227c.
- (34) Stillinger, F. H.; Weber, T. A. Computer simulation of local order in condensed phases of silicon. *Physical Review B* **1985**, *31* (8), 5262-5271. DOI: 10.1103/PhysRevB.31.5262.
- (35) Islam, N.; Flint, M.; Rick, S. W. Water hydrogen degrees of freedom and the hydrophobic effect. *Journal of Chemical Physics* **2019**, *150* (1). DOI: 10.1063/1.5053239.
- (36) Hummer, G.; Garde, S.; Garcia, A. E.; Pratt, L. R. New perspectives on hydrophobic effects. *Chemical Physics* **2000**, *258* (2-3), 349-370. DOI: 10.1016/s0301-0104(00)00115-4.
- (37) Patel, A. J.; Varilly, P.; Chandler, D.; Garde, S. Quantifying Density Fluctuations in Volumes of All Shapes and Sizes Using Indirect Umbrella Sampling. *Journal of Statistical Physics* **2011**, *145* (2), 265-275. DOI: 10.1007/s10955-011-0269-9.
- (38) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press, 1987.
- (39) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187-199. DOI: 10.1016/0021-9991(77)90121-8.
- (40) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The weighted histogram analysis method for free-energy calculation on biomolecules. 1. The method. *Journal of Computational Chemistry* **1992**, *13* (8), 1011-1021, Article. DOI: 10.1002/jcc.540130812.
- (41) Roux, B. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* **1995**, *91* (1-3), 275-282. DOI: 10.1016/0010-4655(95)00053-i.
- (42) Widom, B. Some topics in the theory of fluids. *Journal of Chemical Physics* **1963**, *39*, 2808-2812. DOI: 10.1063/1.1734110.

- (43) Shing, K. S.; Chung, S. T. Computer-simulation methods for the calculation of solubility in supercritical extraction systems. *Journal of Physical Chemistry* **1987**, *91*, 1674-1681. DOI: 10.1021/j100290a077.
- (44) Huang, D. M.; Geissler, P. L.; Chandler, D. Scaling of hydrophobic solvation free energies. *Journal of Physical Chemistry B* **2001**, *105* (28), 6704-6709. DOI: 10.1021/jp0104029.
- (45) Tolman, R. C. The Effect of Droplet Size on Surface Tension. *Journal of Chemical Physics* **1948**, *17*, 333-337. DOI: 10.1063/1.1747247.
- (46) Crooks, G. E.; Chandler, D. Gaussian statistics of the hard-sphere fluid. *Physical Review E* **1997**, *56* (4), 4217-4221. DOI: 10.1103/PhysRevE.56.4217.
- (47) Pratt, L. R.; Ashbaugh, H. S. Self-consistent molecular field theory for packing in classical liquids. *Physical Review E* **2003**, *68* (2). DOI: 10.1103/PhysRevE.68.021505.
- (48) Pratt, L. R.; LaViolette, R. A.; Gomez, M. A.; Gentile, M. E. Quasi-chemical theory for the statistical thermodynamics of the hard-sphere fluid. *Journal of Physical Chemistry B* **2001**, *105* (47), 11662-11668. DOI: 10.1021/jp011525w.
- (49) Garde, S.; Ashbaugh, H. S. Temperature dependence of hydrophobic hydration and entropy convergence in an isotropic model of water. *Journal of Chemical Physics* **2001**, *115* (2), 977-982. DOI: 10.1063/1.1379576.
- (50) Head-Gordon, T.; Lynden-Bell, R. M.; Dowdle, J. R.; Rossky, P. J. Predicting cavity formation free energy: how far is the Gaussian approximation valid? *Physical Chemistry Chemical Physics* **2012**, *14* (19), 6996-7004. DOI: 10.1039/c2cp00046f.
- (51) Jin, Z. H.; Kim, J.; Wu, J. Z. Shape Effect on Nanoparticle Solvation: A Comparison of Morphometric Thermodynamics and Microscopic Theories. *Langmuir* **2012**, *28* (17), 6997-7006. DOI: 10.1021/la2051178.

(52) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling* **1996**, *14* (1), 33-38. DOI: 10.1016/0263-7855(96)00018-5.