Supporting Information

Size Dependent Surface Free Energy and Tolman-Corrected Droplet Nucleation of TIP4P/2005 Water

Mark N. Joswiak,[†] *Nathan Duff*,[†] *Michael F. Doherty*,[†] *Baron Peters*^{†,‡}

†Department of Chemical Engineering, University of California, Santa Barbara, California93106, USA

‡Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, USA

Computational Details

We executed the mitosis method for nanodroplets of 2n = 48, 76, 128, 244, and 560 water molecules, which, assuming bulk density, correspond to radii of 0.70, 0.81, 0.97, 1.2, and 1.6 nm, respectively, when the sub-clusters are conjoined. The smallest system size was chosen such that the sub-clusters were stable (did not evaporate). Additionally, the smallest cluster is large enough that at least one molecule is coordinated to other molecules at all four donor and acceptor hydrogen bond sites.

We performed umbrella sampling^{90,91} along q. Sampling along r was not biased. Within each window, we performed hybrid Monte Carlo molecular dynamics^{S1} (MCMD). A MC move consists of an NVE MD trajectory. Since a TIP4P/2005 water molecule is a rigid body, it has 6 velocity degrees of freedom, 3 translational and 3 angular. To conserve energy, the 6 center of mass (COM) velocities were sampled from the Boltzmann distribution (T = 300 K). Determining the velocity of each atom in a water molecule from the 6 COM velocities is an underdefined problem. This issue was overcome by using the geometry and principle axes of rotation for a TIP4P/2005 water molecule to construct 6 eigenvectors which related the COM velocities to the velocities of each atom. The net velocity of each atom was determined by summing the contributions from each of the 6 COM velocities. This procedure, similar to that of Page and McIver,^{S2} is necessary for energy conservation with rigid molecules. Selecting the velocities of each atom from the Boltzmann distribution will result in a force along the bond between atoms, which cannot be present in a rigid body. The SHAKE algorithm^{S3} will remove this force upon integration which removes energy from the system. More importantly, the equipartition theorem is not obeyed if 3N velocities are drawn when there are only 2N degrees of freedom (here N is the number of atoms).

The MC move from a state 0 to a state 1 is accepted with a probability:

$$P_{0\to 1} = \min\left[1, \ e^{-\beta(E_1 - E_0)}\right]$$
(S1)

where *E* is the sum of the kinetic energy, potential energy, and bias potential. We implemented an infinite square well bias potential in variable *q*. Therefore, nearly all trajectories which finish within the window are accepted. The average *q* bin size was 0.5 and windows contained 3 bins. When the sub-clusters were conjoined, MD trajectories were 50 fs except for 2n = 560, which required 20 fs trajectories to attain an ~40% acceptance rate. Longer trajectories (100-1000 fs) were necessary when the sub-clusters were not in contact to improve sampling in *r*.

NVE dynamics were calculated in LAMMPS^{S4} with a time step of 1 fs. Bond lengths and bond angles were constrained with the SHAKE algorithm.^{S3} We used a 13 Å cutoff with a switching function from 12 to 13 Å for Lennard-Jones interactions. Calculated system properties are altered by the choice of the cutoff distance.^{89,S5} Larger cutoffs yield more accurate results. Tail corrections are often applied in an attempt to account for the truncated interactions. In this work, we do not apply a tail correction for the surface free energy as its implementation for a spherical interface is not straightforward. Moreover, it is unlikely that accounting for the tail correction would undermine our key results. Electrostatic interactions were calculated with the particle-particle particle-mesh solver implemented in LAMMPS with an accuracy of 10⁻⁶. High accuracy electrostatic energy calculations were necessary to diminish energy fluctuations, which can adversely affect detailed balance.

Simulations were performed with periodic boundary conditions. Box lengths were chosen such that each sub-cluster was at least 30 Å away from its nearest periodic image. Periodic box sizes ranged from 52x52x96 Å for 2n = 48 molecules to 77x77x135 Å for 2n = 560 molecules. The z-axis is elongated to allow for separation of the sub-clusters. The sub-clusters are constrained to a

separation axis along the z-direction by restricting the x- and y-components of the center of mass of each sub-cluster to be at most 0.5 Å from the separation axis.

Initial configurations with all of the molecules within a spherical region were generated by Packmol.^{S6} The system was then equilibrated under NVT conditions for 1 ns. Next, we performed preliminary unconstrained sampling of the free energy well using the MCMD algorithm. Based on this unconstrained simulation, we constructed an initial umbrella sampling window by including values of q such that the free energy was within 1.5 $k_{\rm B}T$ of the well minimum. To 'pull' the cluster apart, additional umbrella sampling windows were successively created at lower values of q. Windows were also created to sample clusters with larger q, i.e. to sample clusters which are more compact than typical clusters. For each window, statistics were collected on both the mitosis coordinate q and the distance between the two sub-clusters r.

The bulk density of TIP4P/2005 water at 300 K and 1 bar was determined by equilibrating 903 water molecules in a periodic, cubic box at NPT conditions for 1 ns before performing a collection run of 5 ns. Temperature and pressure was controlled with a Nosé/Hoover thermostat and barostat with damping parameters of 100 and 1000 fs, respectively.^{S7-S9} We calculate the bulk density to be 0.9937 ± 0.0001 g/cm³, in agreement with a study conducted by Vega and de Miguel with a similar force field.⁸⁹ The bulk surface free energy of TIP4P/2005 water at 300 K has been calculated by Vega and de Miguel.⁸⁹ We sought to reproduce their result by using the same number of water molecules as in the aforementioned simulation, but with one box dimension extended (yielding a 30x30x100 Å³ box) and the barostat removed to create a two-phase canonical (NVT) system. The two-phase system was equilibrated for 0.5 ns prior to a collection run of 5 ns. We used the pressure tensor and equation (S2) to calculated the bulk surface free energy,

$$\gamma(\infty) = \frac{L_z}{2} \left(\overline{p}_N - \overline{p}_T \right) \tag{S2}$$

where L_z is the box length in the elongated dimension and \overline{p}_N and \overline{p}_T are the average normal and tangential components of the pressure tensor, respectively. We find that $\gamma(\infty) = 65.7 \pm 0.6$ mJ/m², in agreement with the study of Vega and de Miguel.⁸⁹

Free Energy Landscape

The free energy landscape, $F_L(q,r)$, for 2n = 128 molecules is shown in Figure S1. The free energy is referenced to the landscape minimum, which occurs when the sub-clusters are conjoined. Configurations are included as insets in Figure S1 to illustrate the separation process. Starting from the state with the sub-clusters united, the bridge between the two sub-clusters becomes more constricted as the order parameter decreases. The snapshots visually reinforce a reversible separation of the sub-clusters along q. The free energy landscape near (q = 1, r/Å =22) shows evidence of potential hysteresis effects that could occur if the sub-clusters are separated or combined along r alone.



Figure S1. The free energy landscape, $F_L(q,r)$, for 2n = 128 molecules. The insets are configurations at the indicated points. Oxygen atoms are red and hydrogen atoms muted in off white for clarity.

When *q* is integrated to obtain $F_L(r)$, the lowest free energy state in each *r* bin makes the most significant contribution. In the region of intermediate sub-cluster separation, the lowest free energy bin we sampled corresponds to the highest *q* value. This is apparent from 10 < r/Å < 20 in Figure S1 by moving from left to right at fixed *r*. Therefore, we may not have sampled the lowest free energy state at fixed *r*, which means that the potential of mean force we recover is likely different from the true $F_L(r)$.

We argue that the $F_L(r)$ we obtain produces accurate estimates of ΔF_{mit} because the sub-cluster separation is reversible and we adequately sample states in the free energy well and plateau of $F_L(r)$. A reversible separation ensures that the free energy change between different states is accurate. Since the Helmholtz free energy change is a state function, it is independent of any intermediate states. Thus, sampling the minimum free energy state at fixed r between the well and plateau of $F_L(r)$ is unnecessary so long as the description of the well and plateau is accurate. We use 2n = 128 as an example case. For r < 10 Å, we sample the minimum free energy state with respect to a fixed r (Figure S1). Thus, the shape of $F_L(r)$ near its minimum is accurate. Above r = 25 Å, the sub-clusters are fully separated and at fixed r, the free energy increases rapidly as q increases, indicating that we have sampled the minimum free energy state for that value of r. Therefore, the free energy well and plateau in $F_L(r)$ are accurate and since the subclusters are separated reversibly, the depth of the free energy well is also accurate. The Helmholtz free energy change is given in eq (5). The only significant contributions to the integral come from the minimum free energy states (Figure S2), which are accurately described by the $F_L(r)$ we recover. Therefore, the Helmholtz free energy changes we calculate are accurate.



Figure S2. The percent contribution to the integral in eq (5) by each r bin for 2n = 128.

Significance of a second order term in the Tolman equation

We aim to determine if a second order term^{46,61,82,83} in the Tolman equation (eq (S3)) is needed to explain our results.

$$\frac{\gamma(R)}{\gamma(\infty)} = 1 - \frac{2\delta}{R} + \frac{\kappa}{R^2}$$
(S3)

where κ is a temperature dependent constant. Following the analysis put forth in the letter but with eq (S3) in place of eq (3), we find the Tolman length is -0.66 Å and κ = -0.010 nm². The fit to our data with the Tolman correction is shown in black in Figure S3. The red line is the fit with eq (S3). Within the error of our calculations, we cannot determine if the second order correction is significant.



Figure S3. Comparison of one and two parameter fits to the data. The black line is a fit with the Tolman equation. The red line is a fit utilizing eq (S3).

Broken Bond Model

A negative Tolman length, which indicates that a droplet will have a higher surface free energy than an equally sized bubble, can be explained by the broken bond model.¹⁰² A droplet can hypothetically be created by removing molecules from the bulk. In Figure S4, this would be the removal of the black and interior gray circles. The energy required to accomplish the removal is proportional to the number of bonds that are broken with the surrounding environment. Removing the droplet leaves behind a bubble. The energy required to remove the droplet is the same as that required to form a bubble because the same bonds are broken in each scenario. Note that the bubbles we refer to are actually a vacuum, but we assume that gaseous molecules within this void, due to a much lower density, would have a negligible effect on the number of broken bonds. According to the broken bond model, the surface free energy is proportional to the

number of bonds that are broken.¹⁰² Clearly, the droplet will have a higher surface free energy per surface molecule than the bubble.

Significant fluctuations and small shell volumes make determining the location of a dividing surface extremely difficult for nanoscale droplets by statistical mechanical methods. We suggest that likely choices for the droplet and bubble interfaces tend toward the center of mass of the surface molecules (green and blue lines in Figure S4, respectively). As a result, the droplet has less surface area than the bubble. Therefore, the droplet has a greater surface energy per area than the droplet (i.e. $\delta < 0$). In the limit of infinitely large droplets and bubbles, the broken bond model predicts the surface energies converge to the bulk value.



Figure S4. Diagram illustrating likely interfaces for a droplet (green line) and bubble (blue line) based on removal of the black and interior gray circles. Red circles represent the surface layer of a bubble, black circles represent the surface layer of the droplet, and gray circles represent bulk character.

<u>F* and R* in δ-CNT</u>

Heermann recognized that CNT can be modified by accounting for the Tolman correction to the surface free energy.¹⁰³ Using the unexpanded form of the Tolman equation,

$$\frac{\gamma(R)}{\gamma(\infty)} = \frac{1}{1 + 2\delta/R} \tag{S4}$$

he derived results for the critical cluster size and barrier height without making any simplifications. To more readily realize the importance of the Tolman correction in CNT, we present the Tolman corrected CNT (δ -CNT) predictions of the critical cluster size and barrier height to first-order in δ/R^*_{CNT} . We present a slightly different derivation than Heerman as we employ eq (3) in lieu of eq (S4) for the Tolman equation. As eq (3) is just an expansion of eq (S4), the expressions we present match those of Heermann to first-order in δ/R^*_{CNT} . Utilizing these first-order expressions will yield different results depending on the choice of eqs (3) or (S4) to calculate δ .

CNT describes the nucleation of spherical clusters from a parent phase as follows,

$$F_{\rm CNT}(R) = -\frac{4}{3}\pi R^3 \rho \Delta \mu + 4\pi R^2 \gamma(\infty)$$
(S5)

According to CNT, the critical cluster size is,

$$R^*_{\rm CNT} = \frac{2\gamma(\infty)}{\rho\Delta\mu} \tag{S6}$$

and the height of the free energy barrier is,

$$F^*_{\rm CNT} = \frac{16}{3} \pi \frac{\gamma(\infty)^3}{\left(\rho \Delta \mu\right)^2} \tag{S7}$$

We insert the Tolman equation into CNT and obtain

$$F(R) = -\frac{4}{3}\pi R^{3}\rho\Delta\mu + 4\pi R^{2}\gamma(\infty)\left(1 - \frac{2\delta}{R}\right)$$
(S8)

To calculate the critical size, we set the first derivative (with respect to R) of eq (S8) equal to zero and non-dimensionalize to arrive at

$$0 = -\left(\frac{R^{*}}{R^{*}_{CNT}}\right)^{2} + \frac{R^{*}}{R^{*}_{CNT}} - \frac{\delta}{R^{*}_{CNT}}$$
(S9)

We solve eq (S9) with the quadratic formula and truncate a Taylor expansion to leading order in δ/R^*_{CNT} ,

$$\frac{R^*}{R^*_{\rm CNT}} \approx 1 - \frac{\delta}{R^*_{\rm CNT}} \tag{S10}$$

We solve for the height of the free energy barrier by inserting eq (S10) into eq (S8). We truncate Taylor expansions to first order in δ/R^*_{CNT} and non-dimensionalize by the barrier height predicted by CNT,

$$\frac{F^*}{F^*_{\rm CNT}} \approx 1 - 6 \frac{\delta}{R^*_{\rm CNT}} \tag{S11}$$

We find that, for water, the first order expansion in δ/R^*_{CNT} is in good agreement with exact values obtained from the δ -CNT (Figure S5). By exact values, we are referring to applying eq (3) to δ -CNT without truncating the expressions. As the supersaturation, *S*, increases, eq (S11) begins to deviate from the exact values. However, F^*_{CNT} decreases rapidly as the supersaturation increases. As a result, the first order expression, eq (S11), underestimates the barrier height by only ~2 $k_{\rm B}T$, regardless of supersaturation.



Figure S5. Comparison of the exact solution and first order expression, eq (S11), for the barrier height in δ -CNT for TIP4P/2005 water.

References

- (S1) Mehlig, B.; Heermann, D. W.; Forrest, B. M. Hybrid Monte Carlo Method for Condensed-Matter Systems. *Phys. Rev. E* 1992, 45, 679–685.
- (S2) Page, M.; McIver Jr., J. W. On Evaluating the Reaction Path Hamiltonian. J. Chem. Phys. 1988, 88, 922–935.
- (S3) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* 1977, *23*, 327–341.
- (S4) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. J.
 Comput. Phys. 1995, 117, 1–19.

- (S5) in 't Veld, P. J.; Ismail, A. E.; Grest, G. S. Application of Ewald Summations to Long-Range Dispersion Forces. J. Chem. Phys. 2007, 127, 144711.
- (S6) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. Packmol: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* 2009, *30*, 2157–2164.
- (S7) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. J. Chem. Phys. 1984, 81, 511–519.
- (S8) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* 1985, *31*, 1695–1697.
- (S9) Shinoda, W.; Shiga, M.; Mikami, M. Rapid Estimation of Elastic Constants by Molecular Dynamics Simulation Under Constant Stress. *Phys. Rev. B* 2004, *69*, 134103.