Insight into the molecular mechanism of water evaporation via the finite temperature string method

Citation: Musolino, Nicholas, and Bernhardt L. Trout. "Insight into the Molecular Mechanism of Water Evaporation via the Finite Temperature String Method." The Journal of Chemical Physics 138, no. 13 (2013): 134707. © 2013 American Institute of Physics

As Published: http://dx.doi.org/10.1063/1.4798458

Publisher: American Institute of Physics (AIP)

Persistent URL: http://hdl.handle.net/1721.1/92390

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Insight into the molecular mechanism of water evaporation via the finite temperature string method

Nicholas Musolino and Bernhardt L. Trout

Citation: The Journal of Chemical Physics **138**, 134707 (2013); doi: 10.1063/1.4798458
View online: http://dx.doi.org/10.1063/1.4798458
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/138/13?ver=pdfcov
Published by the AIP Publishing

Insight into the molecular mechanism of water evaporation via the finite temperature string method

Nicholas Musolino and Bernhardt L. Trout[a)]
Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Rm. E19-502, Cambridge, Massachusetts 02144, USA

(Received 5 November 2012; accepted 11 March 2013; published online 4 April 2013)

The process of water's evaporation at its liquid/air interface has proven challenging to study experimentally and, because it constitutes a rare event on molecular time scales, presents a challenge for computer simulations as well. In this work, we simulated water's evaporation using the classical extended simple point charge model water model, and identified a minimum free energy path for this process in terms of 10 descriptive order parameters. The measured free energy change was 7.4 kcal/mol at 298 K, in reasonable agreement with the experimental value of 6.3 kcal/mol, and the mean first-passage time was 1375 ns for a single molecule, corresponding to an evaporation coefficient of 0.25. In the observed minimum free energy process, the water molecule diffuses to the surface, and tends to rotate so that its dipole and one O–H bond are oriented outward as it crosses the Gibbs dividing surface. As the water molecule moves further outward through the interfacial region, its local density is higher than the time-averaged density, indicating a local solvation shell that protrudes from the interface. The water molecule loses donor and acceptor hydrogen bonds, and then, with its dipole nearly normal to the interface, stops donating its remaining hydrogen bond. At that point, when the final, accepted hydrogen bond is broken, the water molecule is free. We also analyzed which order parameters are most important in the process and in reactive trajectories, and found that the relative orientation of water molecules near the evaporating molecule, and the number of accepted hydrogen bonds, were important variables in reactive trajectories and in kinetic descriptions of the process. © *2013 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4798458]

I. INTRODUCTION

The evaporation of water at its interface with air has been studied because it plays an important role in atmospheric processes, as well as in technological and analytical applications. In the field of microfluidic technology, for example, excess evaporation leads to crystallization of ink components in inkjet print heads, resulting in efforts to develop additives to address the so-called "inkjet decap problem."[1] Controlling the rate of evaporation from aqueous interfaces is also advantageous in drying operations, to diminish the risk of surface cracking.[2] Likewise, when evaporation is used to create supersaturated solutions in protein crystallization, diminished evaporation rates can favor nucleation of crystallites over the precipitation of aggregates.[3–5]

Despite the importance and ubiquity of the aqueous evaporation process, little is known about its molecular-level mechanism(s).[75] In fact, there is currently no consensus as to the actual *rate* of evaporation of water into dry air or vacuum. Rates of evaporation are often expressed in terms of the evaporation coefficient γ_E (or E), which is equal to the mass accommodation coefficient α.

This study is an effort to understand, at a molecular level, the process of evaporation, which can be thought of as the inverse of accommodation of water itself. Our motivation for understanding evaporation in this way is to aid in designing soluble solution additives to diminish the rate of evaporation into air.

The evaporation coefficient is the ratio of the actual evaporation rate to the theoretical maximum rate calculated from the Hertz-Knudsen[6,7] equation:

$$G_{max} = \frac{1}{4}\frac{P_v}{k_B T}\left(\frac{8k_B T}{\pi M}\right)^{1/2} = \frac{P_v}{(2\pi M k_B T)^{1/2}}.$$

In this equation, G_{max} is molar flux, M the molecular weight, T the temperature of the surface, and P_v the corresponding vapor pressure. This maximum rate is derived by considering dynamic equilibrium under liquid-vapor coexistence, and neglecting any vapor- or liquid-phase resistance: the evaporation rate is set equal to the rate at which gas molecules condense, assuming that every molecule that strikes the liquid surface enters the liquid. At a surface temperature of 298 K, the vapor pressure is 0.00317 MPa, and this value is, on a mass basis, 0.108 g/(cm^2 s).[8]

Early measurements of the evaporation rate, performed in the decade 1925–1935, obtained values of about 0.4[9] and 0.04 for this coefficient.[10,11] Since that time, practitioners have observed values between about 0.001 and 1.0,[12,13] with most measurements falling between 0.04 and 1.0. The challenging in measuring G and $\gamma_E = G/G_{max}$ is that this rate should be measured under conditions where heat and mass transfer are negligible, and this often requires compensating for non-zero heat or mass transfer resistance in either bulk phase. Evaporation rate measurements have been carried out

a)Electronic mail: trout@mit.edu

by monitoring an evaporating droplet's size,[10,11] isotope exchange in droplet flow train reactors[14] or in jetted streams,[15] Raman thermometry,[16] and monitoring droplet expansion by Mie scattering.[17] Typically, studies with dynamically renewing surfaces result in values of γ_E close to unity, while those with quasi-static surfaces display values of about 0.1 or less.[12,13,18] Summarizing the state of affairs in their 2011 update[13] to their 2006 review of mass transfer at interfaces,[18] Davidovits and co-authors write, using the notation α for the evaporation/condensation coefficient, "The question still remains, why do some studies yield α_{H_2O} significantly smaller than 1 while others point to a value of $\alpha = 1$"?

Because of the difficult nature of these experiments, elucidating the molecular-level details of the evaporation process is, in the main, a future goal of experimental work. To date, experimental studies have proposed that evaporation (and condensation) is mediated by the formation of "small clusters or aggregates" of non-bulk liquid water at the interface,[14] and (in separate work) that molecules in such a cluster become "weakly-bound surface species," then finally evaporate to become gas molecules.[16] This stepwise process leads to a free energetic barrier to evaporation, and in light of previous experimental evidence that water leaves the interface with Boltzmann-distributed kinetic energy,[19] the authors identified the barrier as possibly entropic in nature, "due to possible geometric requirements for the evaporation of a water molecule."[16]

In terms of intermolecular interactions, citing the dependence of the empirical evaporation coefficient on isotopic composition, Cappa and co-workers highlighted the "importance of the first solvation shell in controlling evaporation."[15] "Specifically," they wrote, "the nature of acceptor and donor hydrogen bonds, and their influence on librational and hindered translational motions, will determine evaporation rates." The evaporating molecule's accepted hydrogen bonds, they write, would exhibit a strong dependence on isotopic composition, while donated hydrogen bonds would be only indirectly sensitive to composition,[15] which could suggest that accepted hydrogen bonds are more important than donated ones in "determin[ing] evaporation rates."

Molecular simulations are often applied to experimentally challenging physical problems, because they provide molecular-scale spatial and temporal resolution, and allow precise control over physical conditions like temperature and pressure. Indeed, over the past two and a half decades, computer simulations have been used to study water's interfaces with air, where evaporation takes place, because they allow researchers to observe, at a molecular level of detail, the behavior of this ubiquitous substance outside its well-studied bulk state.

Early molecular dynamics (MD) studies focused on the structural properties of the interface, such as the length scale of density variation, the distribution of surface molecules' orientation, and surface tension.[20–25] Later, first-principles MD simulations examined surface molecules' polarization at interface, in addition to structural properties.[26–29]

Simulations have also been used to obtain a picture of the dual processes of evaporation and condensation or mass accommodation; the latter term denotes the transfer of a water or solute molecule from the gas phase into the solution phase. Mass accommodation of atmospherically relevant solutes, including water vapor itself, has been examined using molecular dynamics simulations.[30–35] These studies examined the potential of mean force (PMF) of such a system as a function of the height of the solute above or below the interface; when a single variable is restrained in this way, the PMF is equal to the free energy. These studies found that (1) as the water molecule in the vapor approaches the interface, the system loses free energy, with only a minority of the total free energy (FE) change occurring inside the Gibbs dividing surface (the plane where time-averaged density is equal to $\frac{1}{2}(\rho_{liq} + \rho_{vapor})$); and (2) no activated state was observed between the liquid and vapor states, and no significant minimum in the free energy profile was present on the liquid side of the interface, as for other solutes.

Employing a single coordinate, such as the distance above or below the interface, however, does not provide physical insight into the molecular-level picture of evaporation, as restraining this variable averages over all other physical quantities. In particular, such restraints do not identify the physical conformations (relative to the interface and its neighbors) the water molecule typically passes through during the evaporation process, or what forces (or entropic considerations) are most influential during this process.

Another method of studying evaporation and accommodation has been directly observed such events in long MD simulations. Because evaporation is a somewhat rare event on the time scales accessible to molecular simulations, obtaining a representative ensemble of such trajectories can be challenging. For example, at the maximum rate discussed above, an evaporation event would take place, on average, once every 2.8 ns from an interface with area 1000 Å^2, which is typical of the systems in MD simulation studies. Gathering a representative ensemble of "reactive" trajectories would therefore require long simulation times and, with frequently-saved coordinate data, concomitantly large trajectory data files.[36]

Accordingly, several MD studies have focused on mass accommodation instead of evaporation, by repeatedly placing water molecules in the vapor region above a liquid slab, and "firing" them at the water surface with Boltzmann-distributed linear and angular momenta. In most cases, very few water molecules are scattered or deflected, so that most remain on the surface or enter the bulk within the time interval of observation (typically 10–20 ps), leading to values of the accommodation coefficient near unity.[33,37–39] As practitioners have pointed out, however, the appropriate length of time to monitor the simulated systems for accommodation or desorption back into the vapor phase is not known *a priori*. Other simulation studies monitored the evaporation flux and obtained values of 0.99[40] for TIP3P water at 300 K and 0.3[41] for TIP4P water. (A study by Matsumoto, with few methodological details, reported a value of 0.3.[42])

More recently, Caleman and van der Spoel focused on the structural and energetic results of evaporation,[43,44] and found that evaporated molecules had a surfeit of kinetic energy, compared to the entire system's temperature. Mason observed 74 evaporation events from a 4890-molecule spherical droplet in a non-periodic simulation of TIP3P water.[36]

These events occurred after unusually close oxygen–oxygen or hydrogen–hydrogen contacts, suggesting a transfer of van der Waals or electrostatic potential energy into the kinetic energy needed to overcome the energetic barrier to evaporation. In most cases, the molecules in question had a coordination number of 1 or 2 at the start of the evaporation process.

To understand how evaporation proceeds, we have employed the string method in collective variables[45] (SMCV), using ten descriptive variables designed for water molecule at an interface. This method, coupled with milestoning, has the ability to sample rare events, measure free energy profiles, and calculate expected first-passage times (a kinetic measure). This approach is described in Sec. II, and the results are given in Secs. III A and III B below. A detailed analysis of the trajectories which contribute to the reactive flux is presented in Sec. IV. Finally, in Sec. V we summarize our conclusions, and discuss their implication for the design of additives to retard evaporation.

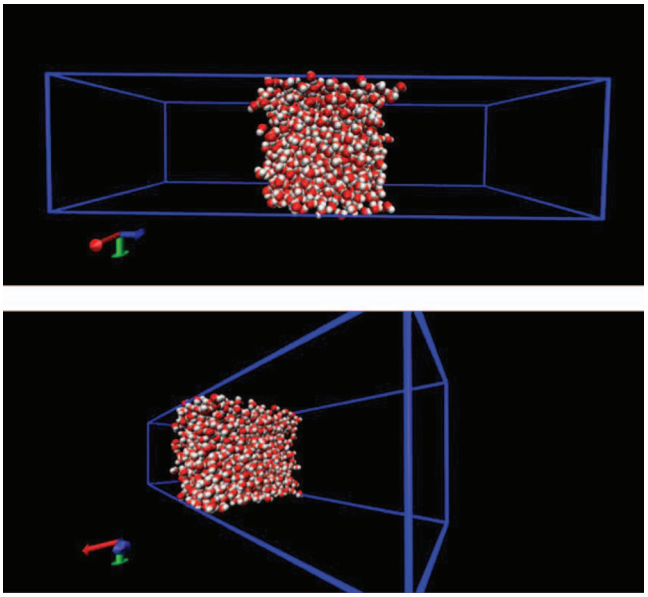

FIG. 1. Rendering of 1025 water molecules in the 31 × 31 × (4 × 31 Å) unit cell. The z-axis is in the horizontal direction.

II. ORDER PARAMETERS AND SIMULATION SETUP

The three-site extended simple point charge model (SPC/E)[46] of water was used throughout the simulations in this study. This particular model was chosen because it exhibits an enthalpy of vaporization, self-diffusion coefficient, and dielectric constant in close agreement with water's actual values,[35] and for its previous success in reproducing an interfacial thermodynamic property, namely the surface tension of water.[23,47] A recent comparison of six water potentials[47] found that the SPC/E and TIP6P potentials "provide the best agreement [of surface tension] with experiment at all temperatures." As Alejandre *et al.* write,[23] such simulations truly challenge potentials, as they are extensions of a potential beyond the bulk-liquid conditions to which it was parameterized.

The general procedure to simulate an interface was to begin with a bulk-liquid simulation, and then extend the simulation cell size in the z direction, thereby creating a "slab" of water in the primary cell (see Figure 1 of Ref. 35). When periodic boundary conditions are applied, a lamella with thickness ≈30 Å was formed, with about 95 Å of vacuum separating it in either direction from the next periodic image of the lamella. The primary unit cell, along with its boundaries, is shown in Figure 1, and its time-averaged density profile is shown in Figure 2.

In order to validate our simulation procedure, the surface tension exhibited by the system was measured. To do so, the components of the pressure tensor were calculated at each recorded time step: at each position z (sampled at 1 Å intervals), intermolecular forces from a pair of atoms contributed to the pressure tensor when the line segment joining those atoms crossed through the z-plane in question. Then, the surface tension was calculated from the components of the pressure tensor at each point in space:[25,48]

$$\gamma = \frac{1}{2}\int_{-\infty}^{\infty} (P_N(z) - P_T(z))\, dz, \qquad (1)$$

where $P_N(z)$ and $P_T(z)$ are the normal (in this case, zz) and transverse (xx and yy) components of the pressure tensor, respectively, at z.

These simulations and those described below were carried out in the canonical ensemble, using a time step of 1.0 fs and rigid bond lengths. During equilibration and production, temperature was controlled using Langevin dynamics (298 K, damping coefficient 4 ps^{-1}) in NAMD.[49] Electrostatics were treated with the particle mesh Ewald (PME) procedure,[50,51] with grid size 32 × 32 × 128. The use of PME has been shown to be important for obtaining accurate values of the surface tension in such systems.[23,47] Simulations of bulk liquid water were carried out in the *NPT* ensemble at 1.0 atm, with pressure controlled with by the Langevin piston approach implemented in NAMD, with oscillation period 200 fs and decay time 100 fs.

The surface tension measured in this way from a 1025-molecule, 2-ns simulation was $\gamma = 61 \pm 2$ dyn/cm; the statistical error was taken to be one standard deviation of the value

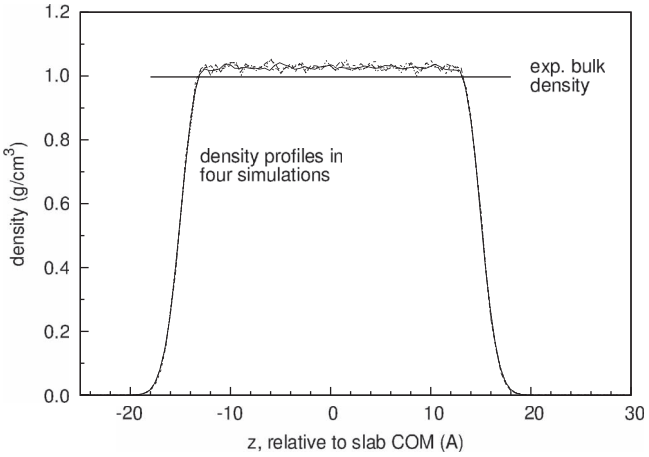

FIG. 2. Time-averaged density profiles from four 2.0-ns simulations, with frames recorded every 5 ps.

TABLE I. Description of order parameters used to describe state of water molecule near interface. Order parameters 8 and 9 have definitions that do not permit the imposition of forces, but listed force constants were used to calculate distances in order parameter-space.

OP	Quantity measured for evaporating molecule	Force constant range (kcal/mol/[OP]^2)
q_0^z	z-position of COM relative to slab COM	2.5–5.0
q_1^{lden}	Local density	2.5–20
q_2^{ravg}	Average of relative orientation to neighbors	2.0–10
q_3^{rstd}	Standard deviation of relative orientation to neighbors	6.0–10
q_4^{thet}	Orientation of dipole relative to interface normal	10–40
q_5^{omeg}	Orientation of molecular normal relative to interface normal	15–75
q_6^{don}	Number of H-bonds donated	10–20
q_7^{acc}	Number of H-bonds accepted	10–20
q_8^{tdist}	Homogeneity of distance of four nearest neighbors	20–100
q_9^{tang}	Angular tetrahedrality of four nearest neighbors	20–100

of γ over eight block averages. This was in good agreement with Chen and Smith's "final value" of 61.3 dyn/cm for the same potential,[47] and the values of 61–62 dyn/cm obtained in other studies.[52–54]

To study the evaporation process, we employed the string method in collective variables,[45] which is based on the finite-temperature string method.[55–58] Our goal was to identify the minimum free energy path (MFEP) from a single molecule's bulk liquid state to its evaporated state; this path is the most likely path for transitions from the former state to the latter. This path was identified as a series of order parameter values. *Order parameters*, also called collective variables, are functions of the simulated system's atomic coordinates, and aim to quantitatively characterize the state of system, in this case as either liquid, vapor, or some intermediate state.

The order parameters (OPs) used for this study are listed in Table I, and detailed definitions can be found in Appendix. All order parameters are measured for a specific, pre-selected molecule, which is denoted with superscript "a" in Appendix and which is called "the evaporating molecule" below. Order parameter zero described the z-position of the evaporating molecule relative to the center of mass of the "slab" of other molecules; for reference, the Gibbs dividing surface (GDS) was located at $z = 15.1$ Å, using the same datum. The GDS was identified by finding the point at which the time-averaged density of the slab first reached $\rho_{\mathrm{GDS}} = \frac{1}{2}(\rho_{\mathrm{liq}} + \rho_{\mathrm{vapor}}) \approx \frac{1}{2}\rho_{\mathrm{liq}}$, since $\rho_{\mathrm{vapor}} \approx 0.03\%\rho_{\mathrm{liq}}$ at standard temperature and pressure. This location and the density profile itself were reproducible from simulation to simulation.

Order parameter 1 described the local density in the vicinity of the selected water molecule, using a smooth weighting function to count molecules within about 3.5 Å. In essence, this order parameter includes contributions from the oxygen–oxygen and oxygen–hydrogen radial distribution functions, both measured at each frame from the selected molecule's oxygen atom. Order parameters 2 and 3 summarize the distribution of the angles between the evaporating molecules dipole vector and the dipole vectors of nearby molecules, as depicted in Figure 3(a). Figure 3(b) shows that the decay length for dipole correlation is between 3 and 3.5 Å, which informed the weighting function used in the definition of these variables.

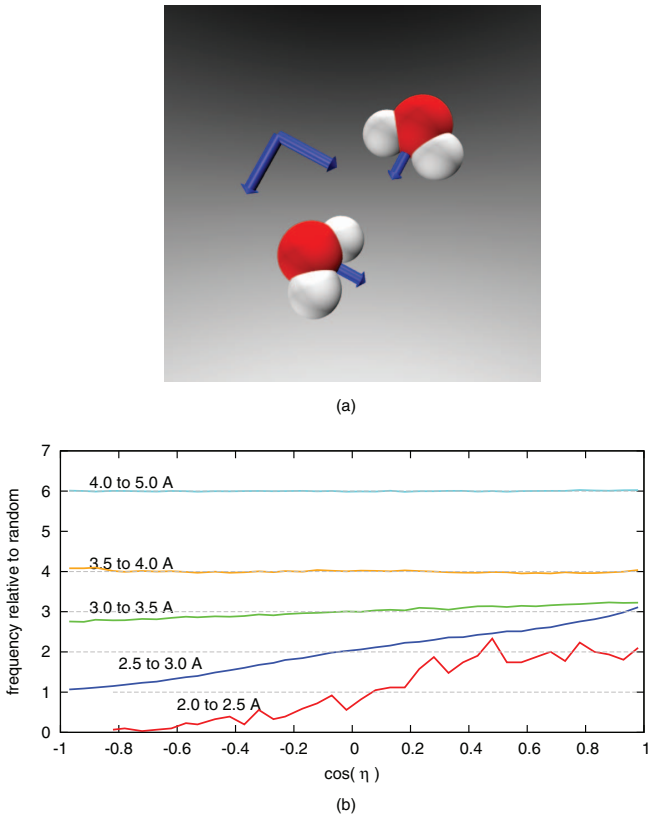

FIG. 3. The dipole-dipole angle η and its distribution in bulk water. (a) The angle η between dipole vectors (blue cylinders) of two water molecules; for clarity, the dipole vectors are translated and reproduced. (b) Distribution of cosine of dipole-dipole angle η for water molecules with indicated O–O separation distances in a 1.0-ns bulk SPC/E water simulation.

Order parameters 4 and 5 summarize the "absolute" orientation of the evaporating molecule, that is, its orientation relative to the interfacial normal $\hat{z}$. OP 4 measures the direction of the molecule's dipole vector as outward or inward facing, while OP 5 measures how outward-facing is the vector ν, perpendicular to the H-O-H plane. Graphical definitions of these angles are shown in Figure 4(a), and the dipole vector and molecular normal ν are shown in Figure 4(b).

Order parameters 6 and 7 count the number of hydrogen bonds the evaporating molecule is donating and accepting, respectively. Contributions are counted in a continuous manner using a weighting function, based on O $\cdots$ H distance. Finally, order parameters 8 and 9 were the distance- and angle-based tetrahedrality measures of Chau and Hardwick.[59] These order parameters can take values between zero, representing a perfectly tetrahedral arrangement of water's oxygen atoms, to a maximum value of 1 in a disordered state.

These order parameters, along with their derivatives with respect to atomic coordinates, were implemented in non-parallel C++ code in NAMD version 2.6.[49] Their derivatives with respect to atomic coordinates were also implemented,

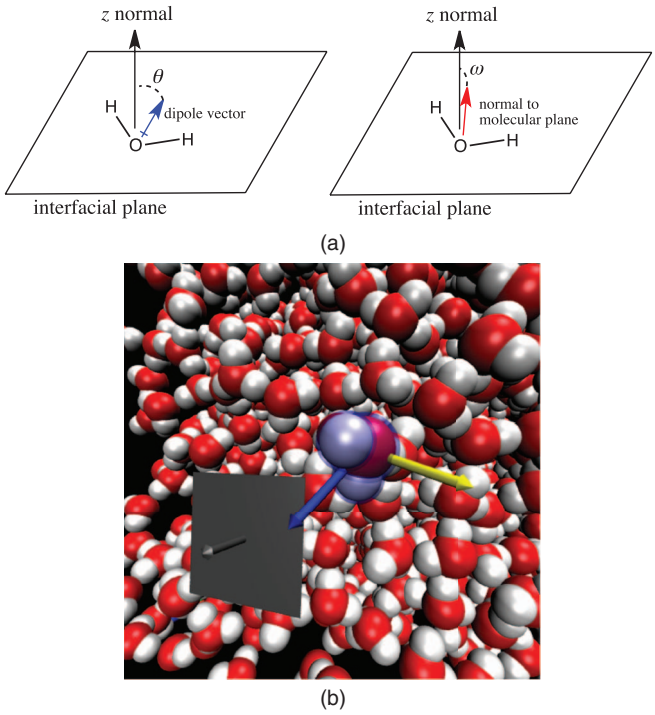

FIG. 4. The two "absolute" orientation variables θ and ω, used to define $q_4 = \cos(\theta)$ and $q_5 = \cos^2\omega$. The two angles are defined in relation to the interfacial normal vector and the evaporating molecule's dipole vector and the molecular normal, respectively. (a) Schematic illustration of the two "absolute orientation" angles. (b) Rendering of the dipole vector μ (blue cylinder), the molecular normal vector ν (yellow cylinder), and the interfacial normal $\hat{z}$ (gray cylinder).

and allowed us to restrain the values of the order parameters. This source code can be made available upon request.

For a physical process, the minimum free energy path can be determined by performing restrained dynamics simulations with several copies of the system (called "replicas" or "images") along a pathway constituting a transition, using the string method in collective variables. Throughout the rest of this paper, the word "image" refers to these copies of the system, which are subjected to independent MD simulations, and does not refer to a spatially translated set of coordinates under periodic boundary conditions.

In these simulations, the order parameters are restrained to target values using a restraint potential of the form $U_{rest} = \sum_i \frac{1}{2}k_i(q_i - q_i^*)^2$, where q_i^* is each OP's target value, and where gradients $\nabla_{\mathbf{x}}(q_i)$ of the order parameters with respect to atomic coordinates used to calculate forces on individual atoms. The restraint forces in all images along the string are then used to calculate the next iteration of the string, which should be closer than its predecessor to the MFEP. In this study, the initial string was created using OP measurements from a series of previous simulations in which the single order parameter q_0 was restrained.

Once the evolving string converged to a final MFEP, the free energy profile and the mean first-passage time for the evaporation process was computed using milestoning,[60–63] since the boundaries of Voronoi cells supported by the MFEP points in order parameter-space are optimal milestones.[61] In these simulations, the system was unrestrained when it was inside the Voronoi cells, and only those unrestrained segments of the trajectory were used to calculate the mean first-passage times, in accordance with the milestoning procedure.[61]

Further details about the MFEP identification and Voronoi milestoning are provided in Secs. III A and III B below.

III. RESULTS

A. Minimum free energy path

To identify the minimum free energy path, a string comprising $N_{img} = 16$ images was constructed. The initial target values for the order parameters in each image were chosen based on previous simulations in which only the relative z-coordinate q_0^z was restrained. Restrained molecular dynamics was performed, with production times of 125–500 ps in each iteration.

In accordance with the SMCV procedure,[45] values of the order parameters, restraint forces, and metric tensor were recorded every 100 fs. In general, restraint forces reached a steady value after 10–20 ps; examples are shown in Figure 5. After each iteration, recorded data were used to compute the potential of mean force and the target OP values for the next iteration of the string, placing images at equal arc-length intervals along the string. This process was repeated until the new string was not far from its predecessor, as measured by Frechét distance.[64]

Two changes were implemented over the course of the string evolution procedure. First, after string 17, the definition of $q_2^{ror-avg}$ was modified to return values in the range $[0, 2\pi)$, rather than in the range $(-\pi, \pi]$, as it originally did. Target

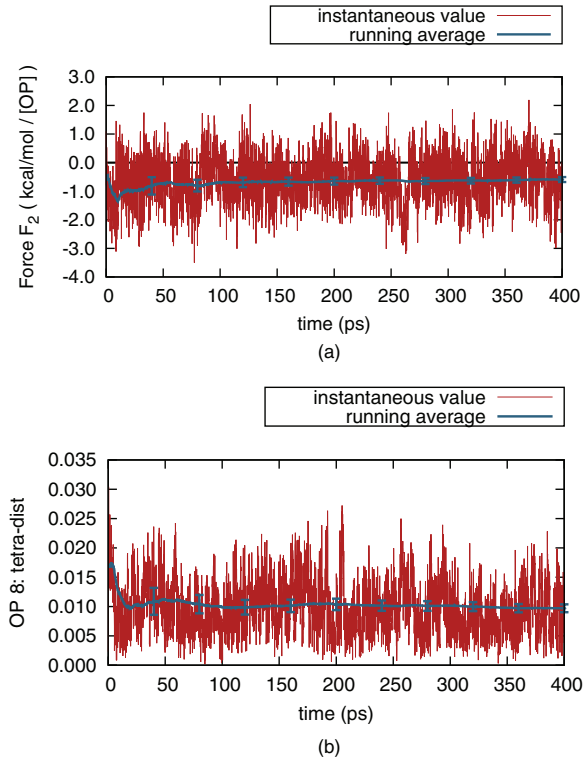

FIG. 5. Examples of restraint force and order parameter convergence from image 9 of string 4. (a) Restraint force for order parameter $q_2^{ror-avg}$. (b) Value of order parameter $q_8^{tetra-ang}$.

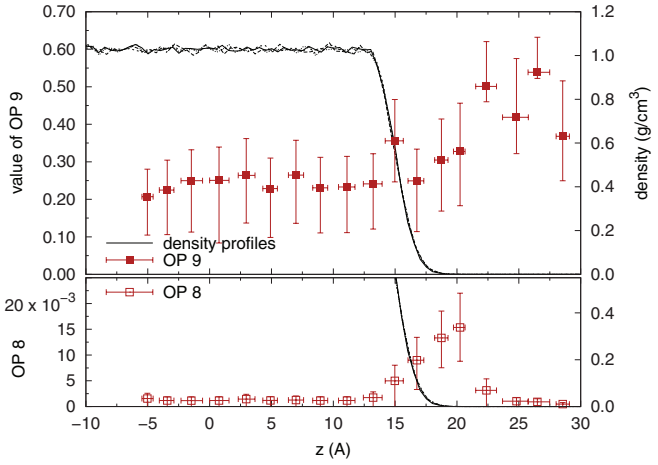

FIG. 6. Values of tetrahedrality order parameters in each Voronoi dynamics image. Bars indicate one semi-standard deviation over simulation trajectory.

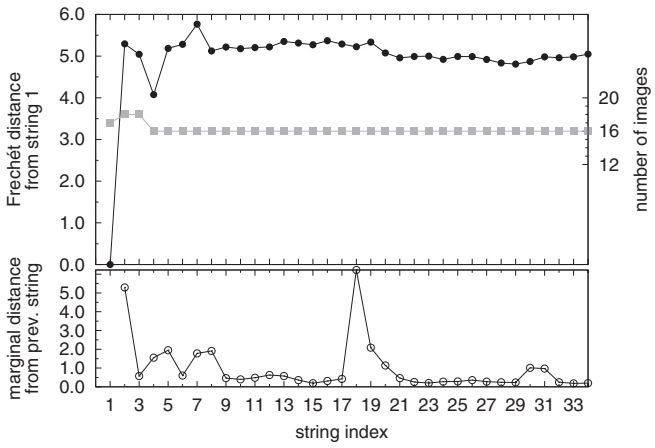

FIG. 7. Frechét distance from initial string, and Frechét distance from each string's predecessor. See text for notes about methodological adjustments at string 18 and string 29.

values in string 18 were shifted to match the new definition. Second, after string 29, the recorded values of $q_8^{tetra-dist}$ and $q_9^{tetra-ang}$ in the string were set to their average values in the last simulation; before that, their values were simply the result of the movement and parameterization of the string in order parameter space. The values of $q_8^{tetra-dist}$ and $q_9^{tetra-ang}$ along the final string, measured in non-restrained simulations (described in Sec. III B), are shown in Figure 6.

The distance of the evolving string from the initial string is shown in the upper panel of Figure 7. Because the string

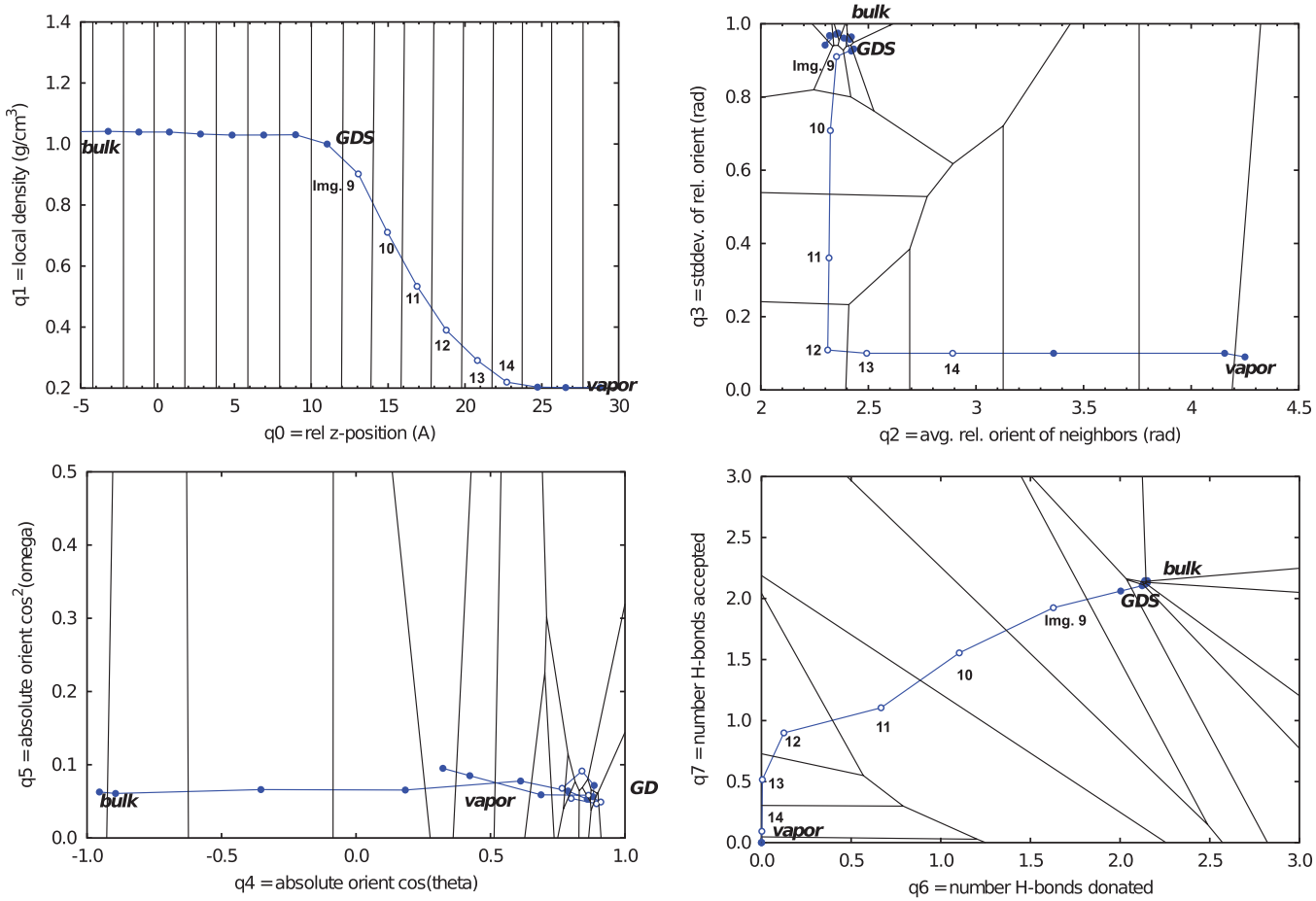

FIG. 8. Minimum free energy path for evaporation, along with Voronoi cell boundaries between images, projected onto two order parameter dimensions at a time. The point labeled "GDS" is image 9, which contains the plane $q_0 = z_{GDS}$, the Gibbs dividing surface. The free energy changed most dramatically over the images (numbers 9–14), highlighted with white centers. Note that Voronoi cell boundaries do not necessarily appear normal to the string because they respect the scaling of order parameters (see text), and because of the plots' axis scaling.

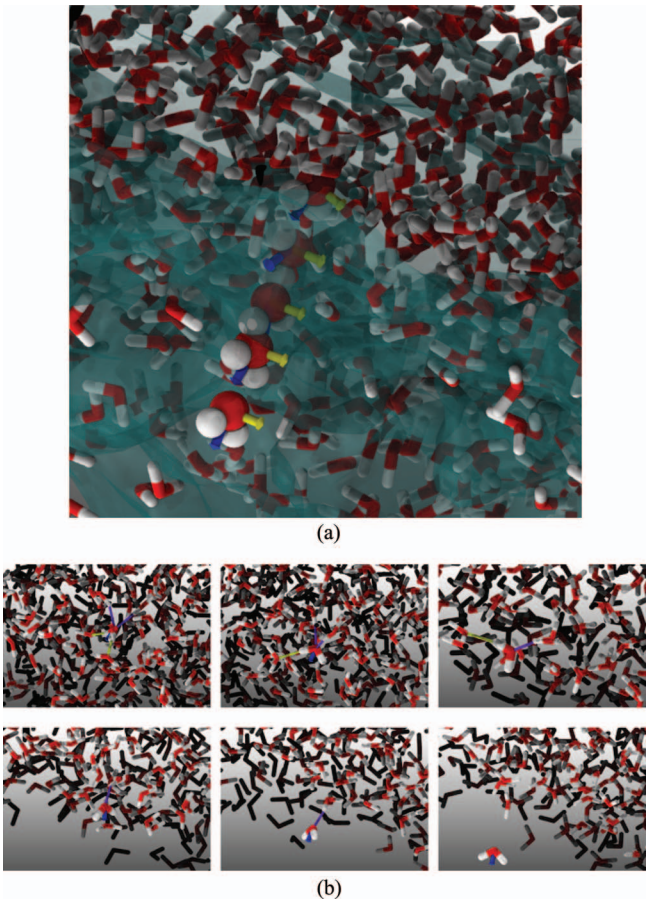

FIG. 9. Snapshots from the frames in images 8–13 in which the system was closest to its OP target values, as measured by minimal restraint energy. In these images, $q_0 = z$ varied from 13 to 21 Å. (a) Water molecule's orientation during evaporation; the molecule's position and orientation (subject to translation) from images 9–13 is shown in a single figure. The other molecules' configuration is from image 9. The blue and yellow vectors are the dipole and molecular normal vectors, respectively. The transparent surface is the water surface. (b) Hydrogen bonds that an evaporating water molecule donates (yellow) and accepts (purple) in images 8–13. The blue arrow is the dipole vector of the evaporating molecule.

took a large step from its initial state in the first SMCV iteration, we also measured the distance from the second or third string, and confirmed that these distances were not evolving when iteration was stopped. The two changes in string evolution mentioned above explain the positive deviations in the lower panel of Figure 7.

The MFEP obtained from SMCV is depicted in Figure 8, and the transition from a bulk liquid to a vapor state can be described as follows:

1. From the bulk, the water molecule diffuses toward the interface, with increasing q_0 values. During this time, the values of q_2, q_3 describing the orientation of nearby molecules are approximately constant, as are the hydrogen bond counts and the tetrahedrality OPs. The absolute orientation OPs change, although their values are not physically important in the bulk phase. This diffusion is represented by a gradually decreasing mean first-passage time, although it is difficult to see in Figure 14 below.
2. The water molecule enters "inner interface" region, (inside the Gibbs dividing plane), and its dipole vector gradually shifts from being somewhat in-plane to become outward-facing. This is *unlike* the typical interfacial molecule, which has its dipole in-plane.[24] The local density is not significantly lower than the bulk, suggesting first solvation shell still surrounds the water molecule; accordingly, its hydrogen bond values are bulk-like as well, with $\mathrm{D} + \mathrm{A} \approx 2 + 2$.
3. Next, the water molecule loses one of the hydrogen bonds it is donating, and it rotates around its normal vector ν to becomes more outward-directed, with a dipole directed about 40° from the normal $\hat{z}$. At this point, $\cos\omega \approx 0.3$, about half of its maximum possible value given the value of θ, indicating one O–H vector is more nearly in plane than the other, outward-facing hydrogen.[65] In this position, the molecule necessarily stops donating a second H-bond with its outward-facing O–H, and on average accepts one H-bond.
4. As the water molecule moves to the outer fringes of the interface, it rotates (again, about its molecular normal axis) so that its dipole is more outward-facing, about 20° from $\hat{z}$, and no longer makes the H-bond it had been donating, leaving only one accepted H-bond, At this point, the time-averaged density is about 0.05 g/cm^3; and there are few atoms within the 3.25- Å density averaging radius, except for the donor hydrogen.
5. With both O–H bonds facing outward, the single hydrogen bond from a neighbor, which had been holding the molecule in place, can break, and at this point, the molecule is free.

The recorded frames in each image in which the system was closest to its target OP values, as measured by minimum restraint energy, were used to generate the snapshots shown in Figures 9(a) and 9(b).

B. Evaporation thermodynamics and kinetics

Once the MFEP was obtained, the image points were used as the "support points" for Voronoi dynamics, since the boundaries between such Voronoi cells are expected to be, in general, optimal milestones.[60] Two additional images were added, one at each end of the string, to ensure that the final milestone, which separates image N from image $N - 1$, was outside the region of free energy change. In these Voronoi dynamics simulations, all entries and departures to and from Voronoi cells were recorded, along with the number of steps during which the simulated system was within its home cell. Instead of reversing the system's velocities at cell boundaries, half-pseudoharmonic soft-wall restraints were used, as described in Ref. 63, with force constant $k_w = 14.0$ kcal/mol.

In dividing any space into Voronoi cells, it is necessary to establish a distance metric, because each cell is defined as the set of points in the space closer to one central point than to any other points. The ten different order parameters used in this study had different natural ranges of variation—for example, the number of Hydrogen bonds donated by a molecule might vary from 0.0 to 2.5, while the distance-based tetrahedrality

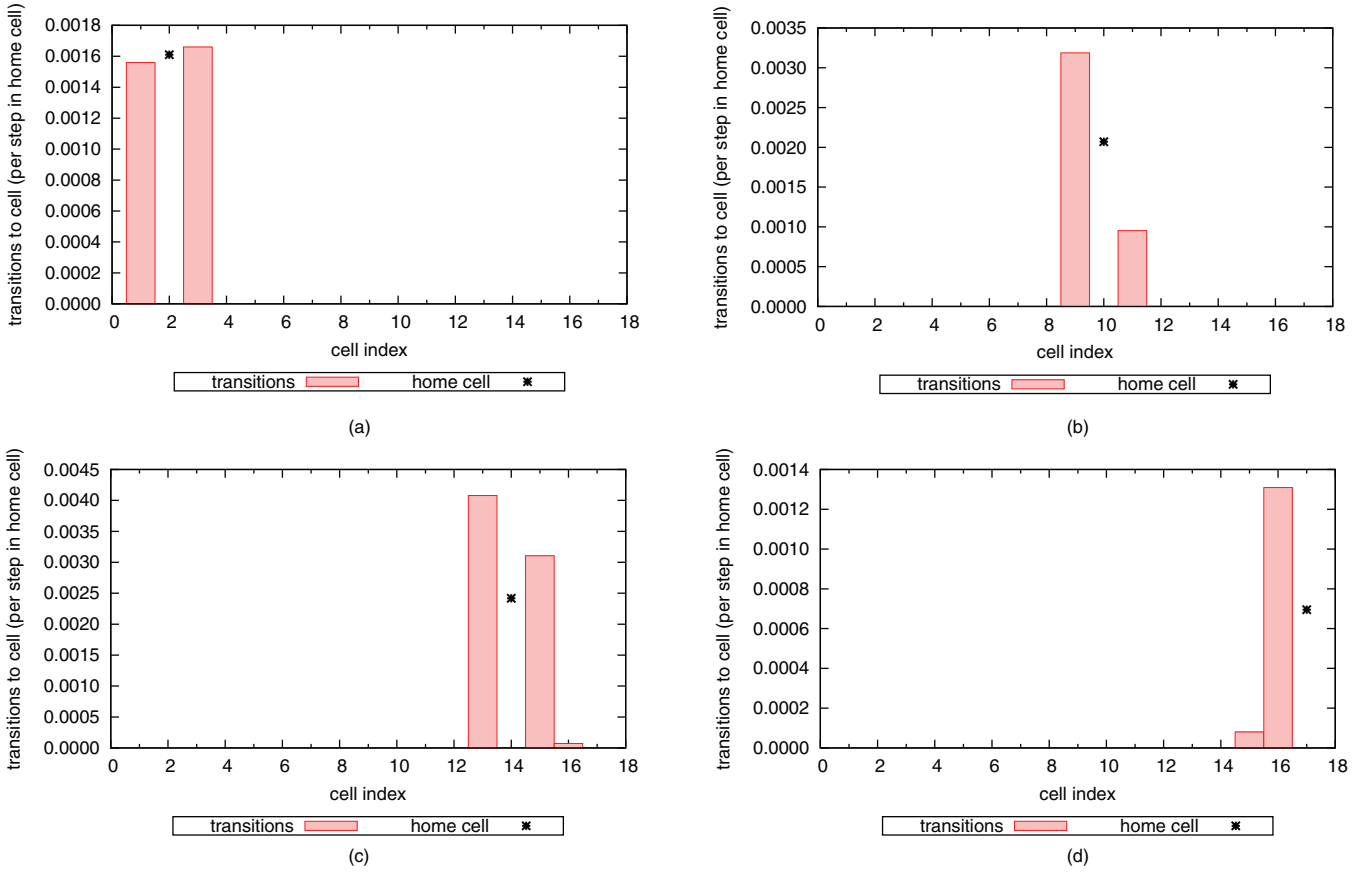

FIG. 10. Transition frequencies from home cell to other cells for four selected images during Voronoi dynamics simulations. Simulations for other images exhibited transitions to sequential cells, as in panels (a) and (b). (a) Image 2. (b) Image 10. (c) Image 14. (d) Image 17.

measure varies from 0.000 to 0.002. Because of this, a scaling factor was used in each dimension in order parameter-space. This scaling factor was taken to be the inverse of the force constant:

$$d(\mathbf{q}, \mathbf{r}) = \left(\sum_{\text{OPs } i} \frac{1}{k_i^2} (q_i - r_i)^2 \right)^{1/2}.$$

The values of each force constant are listed in Table I.

Production MD was carried out for 2.0 ns. Transition events to neighboring cells were counted, and transition rates from four simulations are shown in Figure 10. In most images, almost all transitions took place to sequential cells, i.e., from cell j to cells $j - 1$ and $j + 1$.

Using milestoning analysis, the free energy (FE) of the system at each milestone was determined, as shown in Figure 11, along with mean first-passage time to the final milestone, shown in Figure 14. The free energy profile is

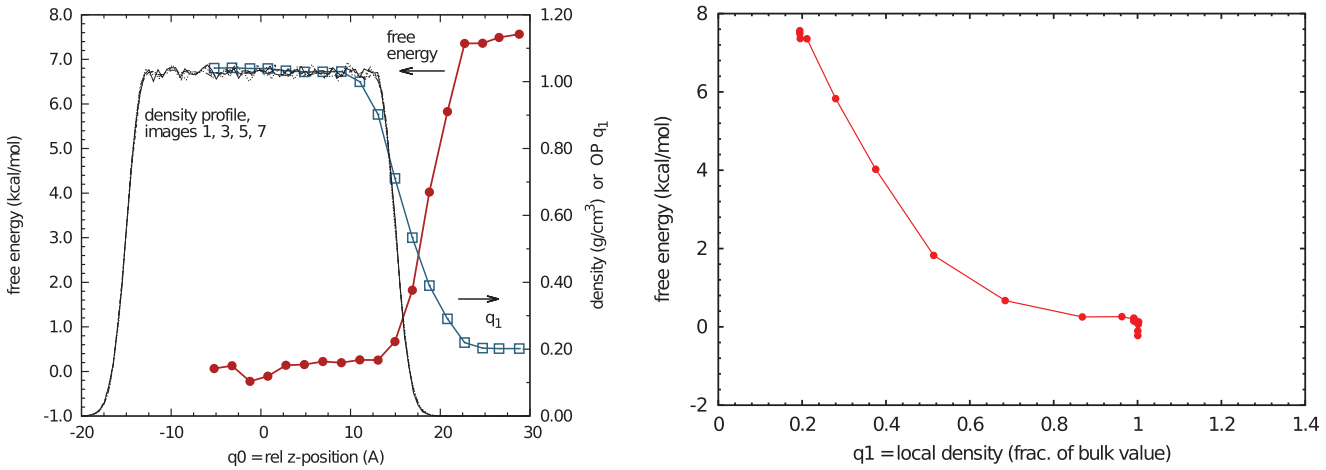

FIG. 11. Free energy measured through Voronoi milestoning, as a function of order parameter q_0 = relative z-position (left) and order parameter q_1 = local density (right). The local density achieves a minimum value in the vapor phase when only the evaporating molecule itself is contributing to the local density.

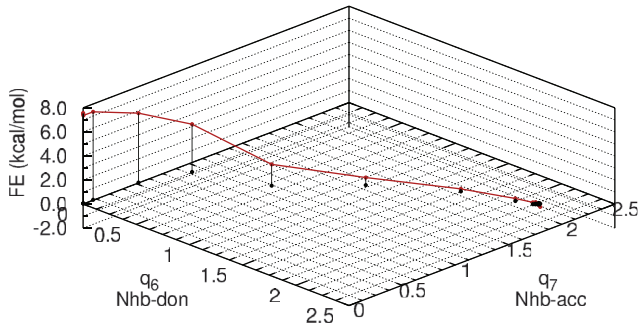

FIG. 12. Free energy as a function of order parameters 6 and 7.

shown in Figure 11, and contains a flat region in the bulk phase, a change in the interfacial region of changing density, and then levels off once the molecule has broken free into the vapor phase, with no FE maximum. The character of the profile is similar to others computed using similar simulations, with only one parameter (z-position) restrained.[31,32,34] The total free energy change is 7.4 kcal/mol, which is in good agreement with the same value measured by Taylor and Garrett,[31] and slightly larger in magnitude than the value of 6.8 kcal/mol observed by Vacha *et al.*,[32] both for SPC/E water.

The free energy profile is reproduced in Figures 12 and 13. The majority of the free energy change takes place after the evaporating molecule has reached the surface, where the number of donor and acceptor hydrogen bonds is (D, A) = (1, 1). In fact, about 2 kcal/mol of FE change occurs as the molecule transitions from (1, 1) to (0, 1), and about 1.5 kcal/mol between (0, 1) and (0, 0).

Figure 13 shows that the average energy, which was measured only when each simulation was inside its respective home cell and free of restraint energy, increases by about 11.5 kcal/molas the water molecule evaporates. This energy penalty, then, must be offset by a corresponding increase in entropy upon evaporation: in the liquid phase, the water molecule is part of a tetrahedral network which extends throughout the bulk, and therefore is severely restricted in its rotational degrees of freedom. As the water molecule leaves the bulk, these restrictions are loosened, although even with two hydrogen bonds, the molecule may have only one or zero unrestricted rotational degrees of freedom.

There also appears to be a peak in the energy profile at around z = 20–24 Å, and lower energy values at 27 and 29 Å. It appears that these higher energy values occur because as the evaporating restrained water is restrained a few angstroms above the interface, other water molecules continue to solvate the evaporating molecule, making a total of 3, 2, or 1 hydrogen bonds. The other molecules in this shell extend beyond the GDS, and have *their own* local hydrogen bond networks disrupted. Once the evaporating molecule loses all its hydrogen bonds, the "protrusion" of solvating shell can reform into the flat interface, thereby minimizing the number of molecules with fewer than a full complement of hydrogen bonds. This is analogous to the role of surface tension effects in the separation of a macroscopic droplet from a bulk liquid phase.

The mean first-passage time (MFPT) is plotted in Figure 14. The overall MFPT from the first milestone, in the bulk region, is 1375 ns. While the evaporating molecule is in the bulk liquid portion of the slab, the MFPT slowly decreases, although this behavior is difficult to see with the scale of Figure 14. This portion of the MFPT profile corresponds to diffusion in the z-direction. Then, beginning at the milestone

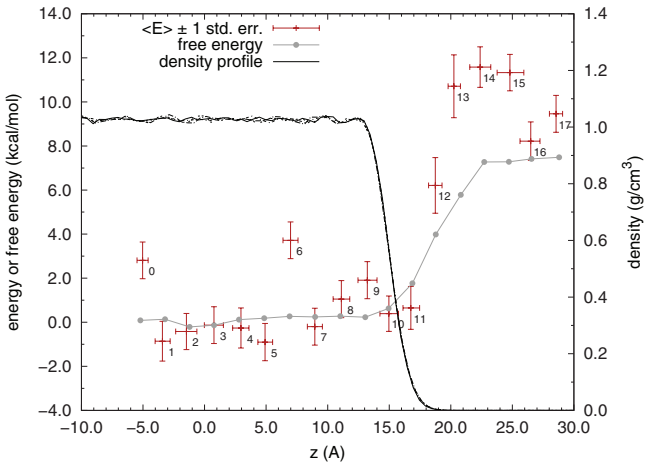

FIG. 13. Free energy profile, along with average system energy values. Error bars are 1.5 standard errors.

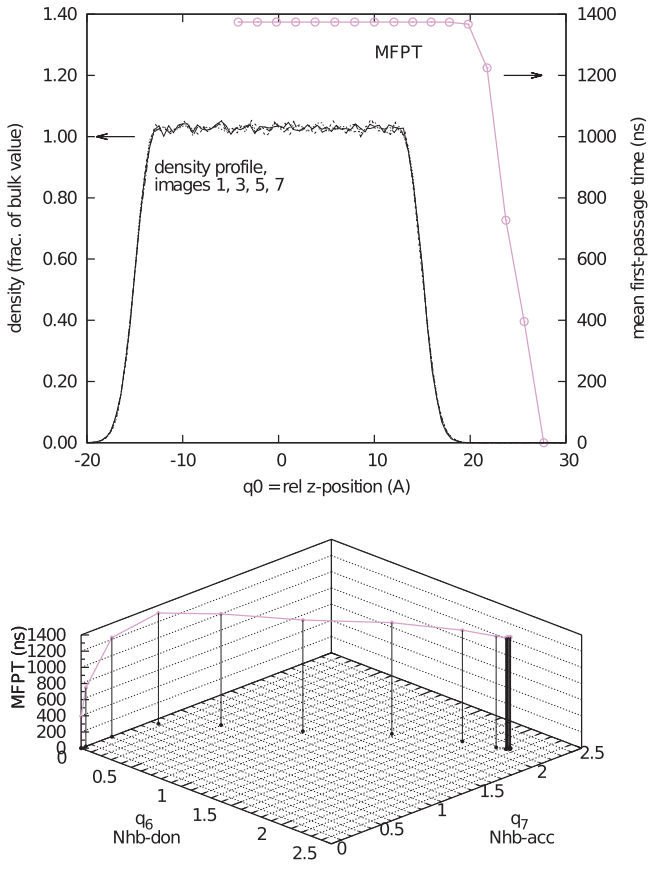

FIG. 14. Mean first passage time to the final milestone as a function of order parameter q_0 = relative z-position (top) and as a function of q_6 and q_7, the number of hydrogen bonds accepted and donated (bottom).

where $(D, A) = (1.4, 1.8)$, the MFPT starts to decrease more dramatically; the greatest change in the MFPT, indicating the *slowest* part of the evaporation process, occurs when the water molecule loses its *final*, accepted hydrogen bond, *i.e.* the transition from (0, 1) to (0, 0). This corresponds to one of the larger (not the largest) changes in free energy discussed above.

1. Comparison to experimental results

The free energy difference corresponding to the transfer of a solute molecule from a vapor into solution has been termed the "free energy of solvation" by Ben-Naim and Marcus,[66] who showed that values can be obtained from vapor-liquid equilibrium and other data, interpreted through thermodynamic arguments. The evaporation process is the reverse of the self-solvation process for water, with $\Delta G_{evap} = -\Delta G_{solv}$. The free energy change measured in this study, and its enthalpic and entropic components, are compared with experimental values in Table II.

In these simulations, the formal system size was fixed, which would suggest the simulations were carried out in the canonical ensemble, leading to measured FE values that are Helmholtz free energy differences. However, because the volume physically occupied by the system of molecules could fluctuate, practitioners have argued that the systems exhibit behavior as if in the *NPT* ensemble, so that the free energies measured should be directly compared to experimental Gibbs free energy values.[31,39]

The error bars reported in Table II come from examining the free energy profile in the bulk-liquid region of the system, where it is expected to be constant. Overall, the results obtained show good agreement with the actual values for water, considering the simplicity of the water model used, and in particular its lack of polarizability.

The evaporation flux implied by these simulation measurements can also be calculated, using the mean first-passage time. We chose a particular water molecule to evaporate, so the mass flux corresponding to our MFPT is $G = M\frac{1}{\tau}\frac{1}{a}$, where a is the specific area occupied by a water molecule. Counting the water molecules in the bulk liquid phase intersected by the plane $z = 0$ (within the SPC/E molecule's van der Waals radius of 1.76 Å) at each frame in 2-ns slab simulations, a was 8.28 Å^2. This leads to a mass flux of $G = 0.026\ \mathrm{g/(cm^2\ s)}$, and an evaporation coefficient $\gamma_E = 0.24$. The MFPT measured in this series of simulations therefore corresponds to an evaporation rate within the (broad) range of measured values.

TABLE II. Comparison of simulation measurements to experimental values for the evaporation or "desolvation" process at 298 K. All values are in kcal/mol.

	ΔG_{evap}	ΔH_{evap}	$-\Delta S_{evap}T$
SPC/E water	7.4 ± 0.4	11.5 ± 1.0	-4.2 ± 1.4
Actual[66]	6.23	9.97	−3.64

IV. IDENTIFICATION OF IMPORTANT ORDER PARAMETERS

A. Principal component analysis

The objective of this analysis was to determine what order parameter(s) varied the most over the critical part of the MFEP. Because evaporation is *not* an activated process with a transition state, we examined images 10–13 (where the entire string comprised images 0–16). This region of the string accounted for half the free energy change and about half the change in MFPT values.

The trajectories analyzed were "contributing trajectories," as shown in Figure 15. These contributing trajectories are defined as those that contribute to the forward or backward reaction rate in the milestoning scheme, by, for example, starting at one milestone in a Voronoi cell, and reaching the opposite milestone *before* intersecting the original milestone again. The label "forward" indicates the direction from reactant to product along the string or within an image, i.e., in the direction of evaporation, from the liquid to vapor state, while the label "backward" indicates the reverse direction.

For example, in a 2.0-ns simulation of image 10, there were 47 forward contributing and 47 backward contributing trajectory segments observed, with average length 1.6 ps for both. During these simulations, order parameter values were recorded every 5 fs, to provide greater resolution in time; the simulation code also printed OP values whenever a system entered or left its home cell. The union of all these contributing

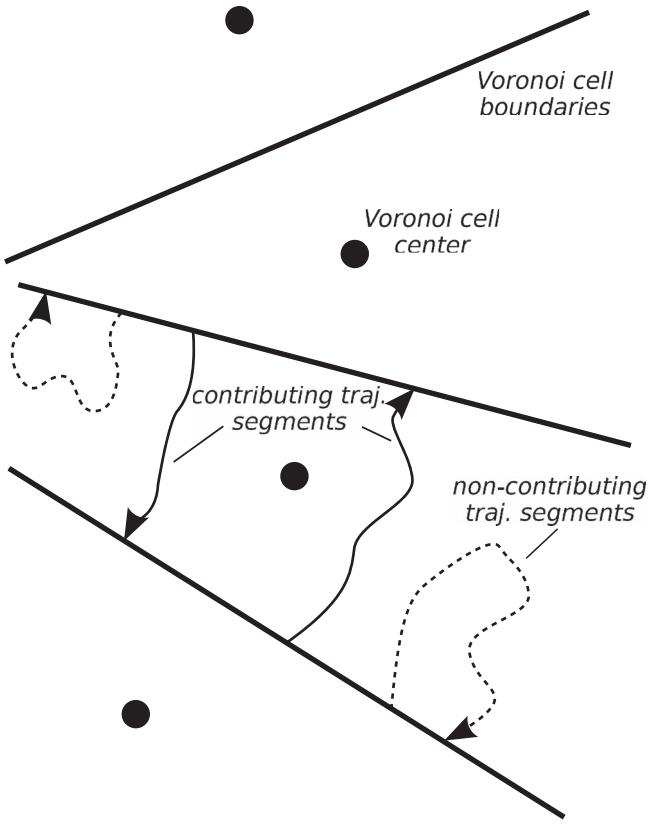

FIG. 15. Schematic showing contributing trajectories in both reactant-to-product and product-to-reactant directions (solid curves) and non-contributing trajectories (dashed curves).

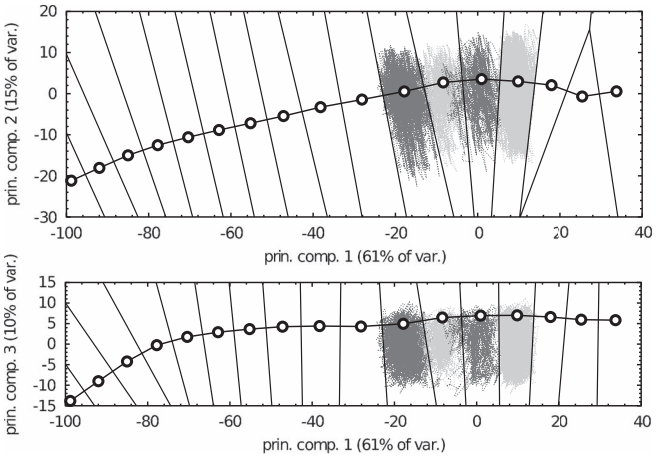

FIG. 16. Projection of contributing trajectory segments in the forward (evaporating) direction onto principle components. Image centers (Voronoi support points) are the black points, while the trajectories from images 10–13 are shown in alternating shades of gray. The rightmost point represents the final, vapor-phase image.

trajectories from the four simulation cells was analyzed with principal component analysis (PCA).

To normalize differently-scaled order parameters, each OP was scaled by $\frac{1}{\sqrt{k_i}}$, as in the simulations themselves, after subtracting OPs' mean values from all recorded points. This scaling approach was used to reflect the original dynamics used to create the trajectory points.

The points from all forward contributing trajectories, after being projected onto the first three principle components, are shown in Figure 16. This shows that the first principle component (PC) is aligned along the length of the string. In addition, the string (represented by the image centers, which serve as Voronoi support points) lies in the middle of the "tube" of reactive trajectories, as would be expected under the SMCV methodology. The eigenvalues from PCA, which (after normalization so that their sum is unity) represent the amount of variance captured by each principal component are shown in Figure 17(a).

Figure 17(b) shows projections of the first two principal components onto the original order parameters. The first principal component is aligned most closely with OPs q_0^z, q_6^{don}, and q_7^{acc}. Order parameter q_2^{ravg} is directly nearly parallel with PC2, although PC2 explains only about one third as much variation in the trajectory points' OP values. All order parameters' projections along PC1 and PC2 are listed in Table III, which shows that similar results were obtained by examining backward trajectories.

This analysis suggests that the order parameters can be divided into a "first tier" of importance, containing the z-position and the hydrogen bond counts. However, the case of OP q_2^{ravg} is less clear: PC2 is, by construction, orthogonal to PC1. Because all reactive trajectories were aggregated together, it is not clear whether the presence of q_2^{ravg} as the main component of PC2 reflects its importance. That is, q_2^{ravg} could appear in PC2 because (1) its value changes over the course of the reactive trajectories, which themselves cover a significant region of FE/MFPT changes, or (2) the order parameter does not change much along the trajectories, and the reaction tube is a collection of many parallel trajectories with many different values of q_2^{ravg}, unchanging along each trajectory. Possibility (1) would suggest that q_2^{ravg} is important, while possibility (2) would suggest that it is not.

To address this question, we applied two other analyses, which are described in Subsections IV B and IV C.

B. Directional analysis

To disaggregate the collections of OP values in many trajectory segments, we focused on the trajectories one at a time. Initially, we attempted to apply PCA to each individual trajectory, but because of their typically nonlinear behavior in the ten-dimensional space, projections of the trajectory

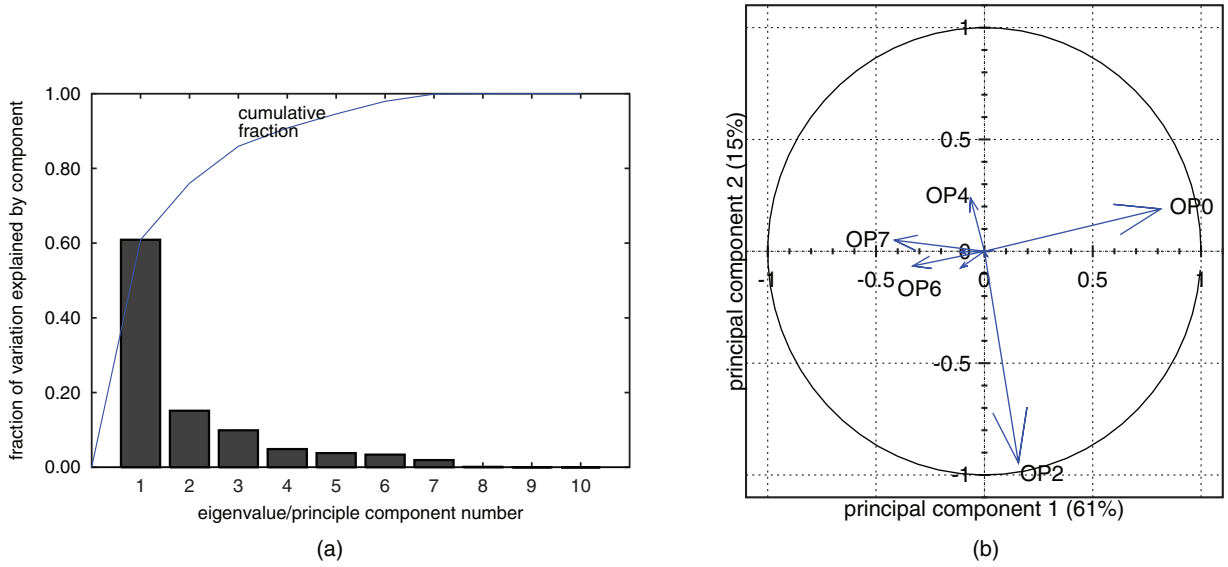

FIG. 17. Summary of PCA results. (a) Amount of variance explained by principal components. (b) Contributions of each order parameter to principal components 1 and 2.

TABLE III. Order parameter components of first and second principle components in analysis of contributing trajectories in Images 10–13. The three largest components in PC1 and the two largest in PC2 are highlighted with boldface type.

	Forward (evaporating) direction		Backward (reverse) direction	
OP	PC1	PC2	PC1	PC2
0	**0.81**	**0.20**	**0.82**	**0.19**
1	−0.11	−0.0084	−0.11	−0.0098
2	0.18	**−0.96**	0.18	**−0.97**
3	−0.11	−0.068	−0.11	−0.045
4	−0.065	0.16	−0.045	0.12
5	0.015	−0.014	−0.018	−0.0058
6	**−0.33**	−0.068	**−0.33**	−0.04
7	**−0.41**	0.017	**−0.40**	−0.027
8	−0.0015	0.0004	−0.0015	0.0001
9	0.038	0.0047	0.037	0.0024
$(\lambda_i/\sum_j \lambda_j)$[a]	0.61	0.15	0.61	0.15

[a]Percent of data variance explained by this component.

points onto their principal components often appeared unsatisfactory.

Instead, to understand the nature of reactive trajectory segments, and what order parameters were changing in this most interesting region of the string, we looked at the vector $\Delta\mathbf{q}'$ for each of the forward- or backward-contributing segments. This vector is the overall direction from the point in OP space where the trajectory *enters* the Voronoi cell, to the point where the trajectory *leaves* the cell:

$$\Delta\mathbf{q}' = \mathbf{q}'_{f,segment} - \mathbf{q}'_{0,segment}.$$

The prime mark ($'$) indicates that OP scaling was applied, in the same manner discussed above.

These directions were then normalized, and the mean direction in each cell was calculated, using techniques for directional variables.[67] These directions are listed in Table IV, which also shows that these vectors were closely grouped around the mean in each image, as the "circular variance" listed in the last column of Table IV was typically ∼0.2.

This shows that the OPs which changed most during the forward trajectories were different in each image: initially, the molecule loses its first hydrogen bond (images 8 and 9), and the trajectories are directed along the two H-bond OPs; next, the molecule continues moving outward, and drops it remaining donated bond (images 10 and 11); next, the alignment of nearby molecules undergoes a shift, as the value of q_2^{r-avg} increases, indicating *decreasing* alignment with neighbors. Examining the OP values, the average dipole-dipole angle η increases from about 65°–100° (images 11–13). Once again, by the time this point along the string is reached, less than one H-bond is being accepted, and the molecule is ready to evaporate.

While the underlying data examined in this trajectory direction analysis and the PCA approach described above, they appear to paint a consistent picture, in which the z-position, average relative orientation, and the hydrogen bond numbers are the most important order parameters.

C. Examining MFPT as a function of order parameters

A common goal in characterizing reactive systems with collective variables is to identify how the reaction committor probability, p_B (sometimes written p_{fold} in the protein simulation literature) can be related to those collective variables. The Voronoi boundaries between images points along the

TABLE IV. Results of local direction analysis for forward-directed transitions in Images 8–14. The two largest components of the mean vector in each image are underlined.

					Components of normalized mean direction $\hat{\mu}$ along OPs											
Image	N_{trans}	q_6^{don}	q_7^{acc}	F[a]	0	1	2	3	4	5	6	7	8	9	$\delta(\hat{\mu})$[b]	$(1-R)$[c]
8	74	2.0	2.1	−0.1	0.50	−0.18	−0.16	−0.13	−0.10	−0.01	−0.60	−0.55	0.00	0.00	(0.99)	0.216
9	76	1.6	1.9	0.3	0.42	−0.18	−0.05	−0.32	0.03	−0.04	−0.61	−0.55	0.00	−0.00	(0.99)	0.202
10	47	1.1	1.6	1.5	0.49	−0.16	−0.08	−0.31	0.15	−0.12	−0.63	−0.44	0.00	−0.01	(0.98)	0.209
11	34	0.7	1.1	3.7	0.64	−0.13	0.37	0.04	0.12	−0.18	−0.48	−0.39	0.00	0.01	(0.96)	0.374
12	32	0.1	0.9	5.7	0.57	−0.05	0.77	0.09	−0.08	0.12	−0.09	−0.22	−0.00	−0.00	(0.97)	0.312
13	213	0.0	0.5	7.0	0.47	−0.00	0.84	0.02	−0.28	0.02	0.00	−0.01	−0.00	−0.00	(0.997)	0.254
14	438	0.0	0.1	7.0	0.11	−0.00	0.94	0.04	−0.26	0.01	0.00	−0.00	−0.00	0.00	(0.999)	0.126

[a] Free energy, in kcal/mol.
[b] 95% confidence "cap" for mean; in image 8, for example, the 95% confidence interval is given by $(\mu_{sample} \cdot \hat{\mu}) > 0.99$.
[c]Measure of variance around mean direction, which takes values between 0 (no variance) and 1 (random distribution).

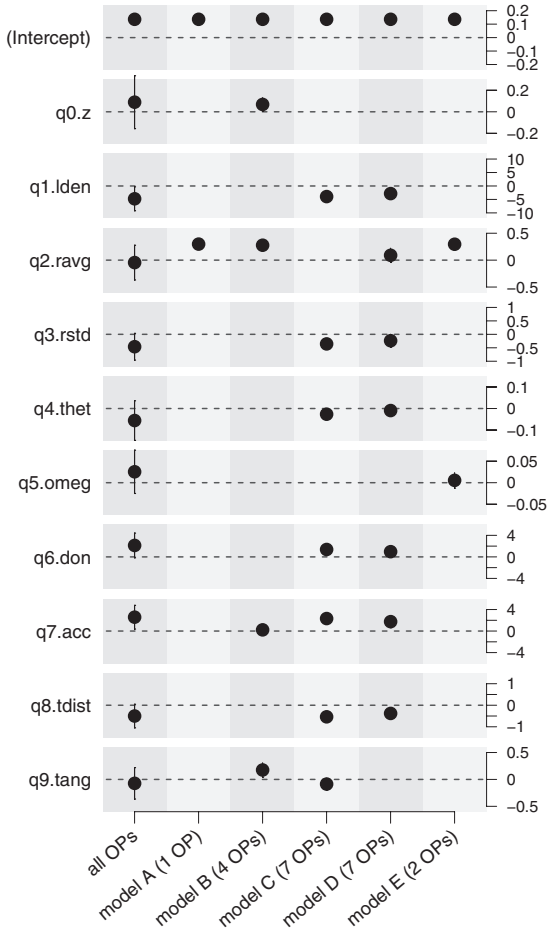

FIG. 18. Coefficients of order parameters for the five models *A*–*E* with best BIC values, and the linear model containing all order parameters. The coefficients are for the normalized order parameters, and the error bars are 95% confidence intervals.

reaction path (string) identified through SMCV, and which serve as ideal milestones for measuring kinetic properties, are isocommittor surfaces in the volume of space local to the string.[60,61,68]

After performing milestoning calculations, the mean first-passage time to the final milestone—in this case, the evaporated state—can be examined as a function of collective variables, as the MFPT is monotonically related to the committor probability p_B. While the analyses above described how the evaporating water molecule's state changed during evaporation events, it did not include information from the MFPT, which reflects a water molecule's likelihood of evaporation at its different CV states along the reaction path.

To understand the relationship between MFPT and the order parameters, the MFPT values at each milestone were used for linear regression. The order parameters at each milestone were calculated as the midpoint between the image centers (support points) on either side of the milestone. (While an ideal approach might be to use the point of maximal hitting point density[60] on the milestone for the $\{q_i\}$ values at each milestone, Figure 16 shows that the string and its constituent image points are within the main reaction channel identified.)

The order parameter values were then centered and scaled by their standard deviations, and the MFPT was transformed by taking $\tau' = 1 - \tau/\tau_{bulk}$, where τ_{bulk} is the MFPT value in the milestone farthest from the evaporated state. Subtracting this ratio from one simply allowed τ' to increase from 0.0 (bulk state) to 1.0 (evaporated state), so that evaporation is in a "positive" direction, consistent with the rest of this paper. All milestone points were equally weighted, although the final four points, where the MFPT changed the most, typically had a relatively large influence on regressions, with values of Cook's distance[69,70] of approximately 1.

As in any multivariate regression, identifying a model requires a compromise between model simplicity (parsimony) and goodness of fit. We identified the two best-fitting combinations of 1, 2, …, 10 order parameters using an exhaustive search; these are listed in Table V, where they are sorted by the values of the Bayes information criterion (BIC). The coefficients on each OP for the top five models are shown in Figure 18.

In Table V, the two most frequently appearing variables are the average relative orientation q_2^{ravg} and the number of hydrogen bonds accepted q_7^{acc}, and model *B*, the second-best model, contains these two order parameters, along with q_0^z and q_9^{t-ang}. This is consistent with the results obtained above, while it should be noted that this analysis, instead of looking

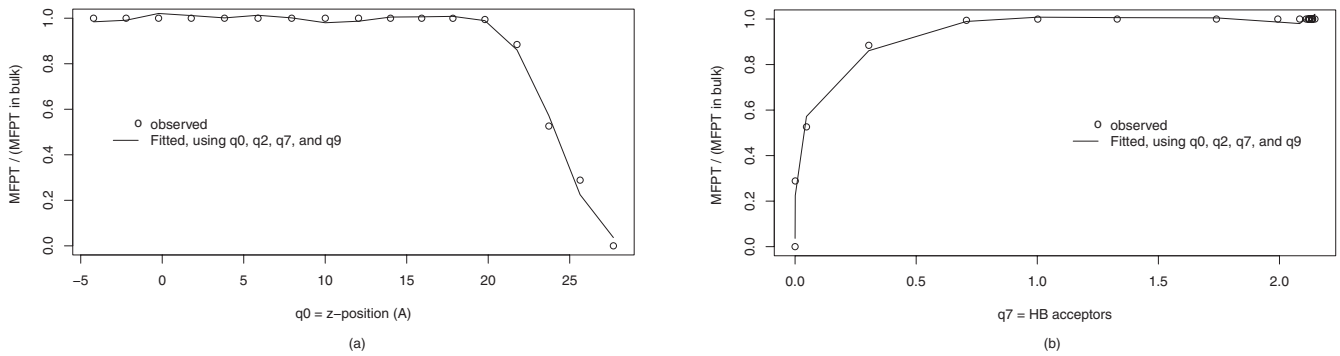

FIG. 19. Observed and fitted values of the MFPT values at milestones, plotted against (a) local density and (b) number hydrogen bonds accepted. The fitted data were from model *B* of Table V and Figure 18. (a) MFPT and *z*-position. (b) MFPT and number hydrogen bonds accepted.

TABLE V. Best models of $\tau' = (1 - \tau/\tau_{bulk})$ with different numbers of order parameters used. The combinations are sorted by BIC value.

OPs	Order parameters in linear model	BIC	Designation[a]
1	q_2^{ravg}	−71.5	Model A
4	$q_0^z + q_2^{ravg} + q_7^{acc} + q_9^{tang}$	−70.5	Model B
7	$q_1^{lden} + q_3^{rstd} + q_4^{thet} + q_6^{don} + q_7^{acc} + q_8^{tdist} + q_9^{tang}$	−69.8	Model C
6	$q_1^{lden} + q_2^{ravg} + q_3^{rstd} + q_6^{don} + q_7^{acc} + q_8^{tdist}$	−69.5	Model D
2	$q_2^{ravg} + q_9^{tang}$	−69.2	Model E
2	$q_2^{ravg} + q_5^{omeg}$	−69.1	
7	$q_1^{lden} + q_2^{ravg} + q_3^{rstd} + q_4^{thet} + q_6^{don} + q_7^{acc} + q_8^{tdist}$	−68.9	
8	$q_0^z + q_1^{lden} + q_3^{rstd} + q_4^{thet} + q_5^{omeg} + q_6^{don} + q_7^{acc} + q_8^{tdist}$	−68.9	
4	$q_2^{ravg} + q_4^{thet} + q_7^{acc} + q_9^{tang}$	−68.8	
3	$q_2^{ravg} + q_5^{omeg} + q_9^{tang}$	−68.7	
6	$q_1^{lden} + q_3^{rstd} + q_4^{thet} + q_6^{don} + q_7^{acc} + q_8^{tdist}$	−68.7	
8	$q_1^{lden} + q_3^{rstd} + q_4^{thet} + q_5^{omeg} + q_6^{don} + q_7^{acc} + q_8^{tdist} + q_9^{tang}$	−68.2	
5	$q_1^{lden} + q_2^{ravg} + q_3^{rstd} + q_7^{acc} + q_8^{tdist}$	−68.0	
3	$q_2^{ravg} + q_7^{acc} + q_9^{tang}$	−68.0	
5	$q_0^z + q_2^{ravg} + q_4^{thet} + q_7^{acc} + q_9^{tang}$	−67.9	
9	$q_0^z + q_1^{lden} + q_3^{rstd} + q_4^{thet} + q_5^{omeg} + q_6^{don} + q_7^{acc} + q_8^{tdist} + q_9^{tang}$	−66.5	
9	$q_0^z + q_1^{lden} + q_2^{ravg} + q_3^{rstd} + q_4^{thet} + q_5^{omeg} + q_6^{don} + q_7^{acc} + q_8^{tdist}$	−66.1	
10	$q_0^z + q_1^{lden} + q_2^{ravg} + q_3^{rstd} + q_4^{thet} + q_5^{omeg} + q_6^{don} + q_7^{acc} + q_8^{tdist} + q_9^{tang}$	−63.9	All OPs

[a]These names are used in a comparison of the models' coefficients in Figure 18.

at local, trajectory-based data, identifies these order parameters using the MFEP points themselves, along with the quantitative values of the MFPT.

The purpose of this regression analysis is not necessarily to construct a quantitative model for the MFPT profile, but rather to identify order parameters that may be important in determining the value of the mean first-passage time, which is related to p_B. Nonetheless, the coefficients for the top five models listed in Table V, as well as the model containing all 10 OP terms, are given in Figure 18, and Figure 19 shows that even a simple linear model in four terms can reasonably reproduce the shape of the MFPT profile.[76]

V. CONCLUSIONS AND OUTLOOK

In this study, the minimum free energy path for evaporation (under the eight restrained interfacial order parameters) was determined. During evaporation, the evaporating molecule sheds its second donated and second accepted hydrogen bonds, and rotates (relative to the interfacial normal) into a position unlike most water molecules in the interfacial region. It then loses its remaining donated hydrogen bond, and then loses its final accepted hydrogen bond at a time that its two hydrogen atoms are pointing outward. For details, see Figure 8 in this article.

During this evaporation process, the orientation of nearby molecules relative to the evaporating molecules becomes *less aligned*, as measured by dipole-dipole angle η. In particular, the mean dipole-dipole angles shifts values of partial alignment, at ∼60°, to an partially anti-aligned state (average value ∼60°).

Using Voronoi milestoning, the evaporation process was found to take place on a plateau-like free energy landscape, with $\Delta F = 7.4$ kcal/mol for the SPC/E water model. The mean first-passage time for evaporation was found to be 1375 ns for an individual molecule. This corresponds to an evaporation coefficient of $\gamma_E = 0.24$.

The directions in which order parameters most varied were analyzed, by examining contributing trajectories collectively in the portion of the string where FE and MFPT values changed, and by examining those trajectories directionality on an individual basis. These analyses suggested that the relative z-position, the orientation of nearby water molecules, and the number of hydrogen bonds accepted and donated (q_0^z, q_2^{ravg}, q_7^{acc}, and q_6^{don}, respectively) were the order parameters that changed most in this important region of the string.

When the MFPT values were regressed against the OP values at each milestone, the orientation of nearby molecules (q_2^{ravg}) and the number of accepted hydrogen bonds (q_7^{acc}) appeared most frequently in the combinations of OPs as explanatory variables with best BIC values.

Together, these results suggest that the loss of accepted hydrogen bonds, and the reorganization of the first solvation shell, play a critical role in the evaporation process. This conclusion would be consistent with the those of Cappa *et al.*,[15] cited above, which were based on strong isotopic dependence of the evaporation coefficient.

Based on the understanding of the evaporation mechanism suggested by the above conclusions, specific

features of an additive to impede evaporation are put forward below.

Finally, we note that future work could include refinements of the order parameters used in this study. For example, one could measure the z-coordinate of an evaporating water molecule with respect to a calculated instantaneous or time-averaged interfacial surface,[71] rather than using a fixed position. Recently-published general order parameters for molecular crystals[72] could be used to measure the location and orientation of a water molecule's individual neighbors,[73] rather than measuring the mean and variance of nearby molecules' orientation, as was done here. And the unusual oxygen–oxygen or hydrogen–hydrogen distances which played a role in evaporation in a previous study[36] could be examined further as order parameters using a technique other than the SMCV, because the string method and milestoning would not be appropriate for a case in which kinetic energy propels a system along a reaction path.

A. Implications for additive design

Based on the mechanistic understanding of the evaporation process described above, we believe that a successful additive to impede evaporation would possess the following features. Note that the following features are based on the reasoning that the inhibiting additive would target the existing, natural kinetic bottleneck in the evaporation process, in order to induce an energetic barrier there.

i. The additive would exhibit a strong propensity to form a "second" hydrogen bond with water, i.e., to *donate* a hydrogen bond to a water molecule which has only one "natural" (accepted) H-bond from the liquid phase.
ii. The feature above may require that the additive's donor group exhibit a certain orientation relative to the interface, such as facing generally outward.
iii. The hydrogen-bond donating feature would also ideally be placed well into the outer half of the interfacial region, beyond the Gibbs dividing surface, since most of the evaporating molecule's free energy change (and passage time) takes place there.
iv. The additive should adsorb at the liquid-vapor interface, i.e. be surface active, in order to impede evaporation there. In designing a soluble additive, the hope is that it will be possible to "tune" surface activity separately from the particular evaporation-inhibiting features suggested above, by adjusting the number or degree of hydrophobicity of those functional groups which do not participate in the interactions with interfacial water, but instead lead to the surface active or amphiphilic nature the additive.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Erik E. Santiso for guidance in modifying NAMD, for sharing a software utility to effect SMCV iterations, and for many helpful technical discussions. Molecular graphics were produced using VMD.[74]

The authors would like to kindly acknowledge support from the DuPont-MIT Alliance, and from the NIH/NIGMS Biotechnology Training Program at MIT.

APPENDIX: INTERFACIAL ORDER PARAMETERS

Let "a" denote the molecule whose evaporation is being simulated, and the z-axis be normal to the interfacial plane. The first order parameter is the distance between the center-of-mass of the evaporating molecule and the center-of-mass of all the other molecules, which collectively are designated the "slab":

$$q_0 = r^a_{\mathrm{COM},z} - r^{\mathrm{slab}}_{\mathrm{COM},z} \qquad \text{z pos.}$$

The local density order parameters involve a sum of the masses of all atoms, weighted by their distance to the oxygen atom the evaporating molecule:

$$q_1 = \frac{1}{V_w} \sum_{\text{atoms } i} m_i w_{lcl}\left(r_i - r^a_{\mathrm{O}}\right) \qquad \text{local density,}$$

where the normalization factor V_w is the bulk density integrated using the weighting factor describing the local vicinity:

$$V_w = \int_0^\infty \rho_{bulk} w_{lcl}(r) 4\pi r^2 dr.$$

The w_{den} function is one of two smoothing functions used to define what neighbors are local, and what pairs of atoms are hydrogen bonding (see below):

$$w_{lcl}(\mathbf{u}) = \frac{1}{1 + \exp\left(\kappa(|\mathbf{u}| - R_{lcl})\right)},$$

$$w_{hb}(\mathbf{u}) = \frac{1}{1 + \exp\left(\kappa(|\mathbf{u}| - R_{hb})\right)},$$

where $\kappa^{-1} = 0.2\,\text{Å}$; $R_{lcl} = 3.25\,\text{Å}$; and $R_{hb} = 2.3\,\text{Å}$. These smoothing functions, rather than sharp distance cutoffs, are used to make the order parameters differentiable functions of atomic coordinates, which is required to apply conservative restraint forces during the SMCV procedure. The functions are graphed in Figure 20 below. The local density order parameter includes the mass of the evaporating molecule itself, and thus has a minimum value of M_a/V_w.

The next order parameters involve the relative orientation of the evaporating molecule a to its neighbors. For any two water molecules, this relative orientation is described by the angle between their dipole moments:

$$\eta_{a,i} = \arccos \frac{\mu^a \cdot \mu^i}{|\mu^a||\mu^i|},$$

where μ^j is the geometric dipole given by $\mu^j = r^j_{\mathrm{H1}} - r^j_{\mathrm{O}} + r^j_{\mathrm{H2}} - r^j_{\mathrm{O}}$. Once a set $\{\eta_{a,j}\}_{j=1}^{N-1}$ of relative orientations has been generated, their average and variance are calculated

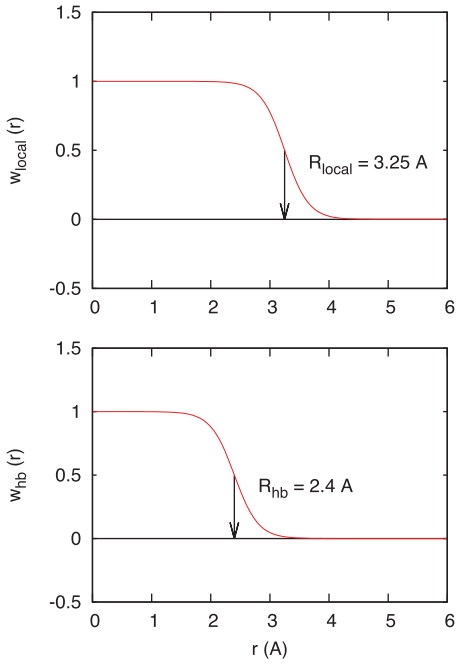

FIG. 20. Smooth weighting functions used for calculating local density and local relative orientational order (top), and number of hydrogen bonds donated and accepted (bottom).

using special approaches for angular random variables:[67]

$$\bar{x} = \frac{1}{W_{tot}} \sum_{\text{mol. } j \neq a} w_{lcl}]big(r_{\text{O}}^{j} - r_{\text{O}}^{a}\big) \cos 2\eta_j,$$

$$\bar{y} = \frac{1}{W_{tot}} \sum_{\text{mol. } j \neq a} w_{lcl}\big(r_{\text{O}}^{j} - r_{\text{O}}^{a}\big) \sin 2\eta_j,$$

$$W_{tot} = \sum_{\text{mol. } j \neq a} w_{lcl}\big(r_{\text{O}}^{j} - r_{\text{O}}^{a}\big) \qquad \text{normalization factor,}$$

$$q_2 = \arctan \frac{\bar{y}}{\bar{x}} \qquad \text{average rel. orient.,}$$

$$q_3 = (-2 \ln (\bar{x}^2 + \bar{y}^2)^{1/2})^{1/2} \qquad \text{std. dev. of rel. orient,}$$

where the arctan function is calculated in the interval from $[0, 2\pi)$ based on the signs of both $\bar{x}$ and $\bar{y}$, using the `atan2` function of the C++ standard library.

The orientation of the water molecule relative to the interfacial normal $\hat{z}$ can be completely specified using two angular variables: the angle θ between the evaporating molecule's dipole and the normal, and the angle ϕ between the interfacial normal and the normal to the plane formed by the molecule's three atoms,

$$q_4 = \cos\theta = \frac{\mu \cdot \hat{z}}{|\mu|} \qquad \text{angular orient,}$$

$$\mu = (r_{\text{H1}} - r_{\text{O}}) + (r_{\text{H2}} - r_{\text{O}}),$$

$$q_5 = (\cos\omega)^2 = \left(\frac{\nu \cdot \hat{z}}{|\nu|}\right)^2 \qquad \text{angular orient,}$$

$$\nu = (r_{\text{H1}} - r_{\text{O}}) \times (r_{\text{H2}} - r_{\text{O}}).$$

These definitions are shown in Figure 4.

The reason that the order parameter q_5 is defined with the square of the cosine is to account for symmetry. Because of this, the proper approach is to take the angle between the *directed* z-normal and the *directionless* normal to the molecular plane. This OP gives the same value, whether the computationally used normal is facing inward or outward. Using the square of the cosine is equivalent (in information content) to using the cosine of twice the measured angle.

Order parameters q_6 and q_7 use a smoothing function to count the number of hydrogen bonds. The H-bond weighting function w_{hb} (defined above) smoothly changes from a value of 1 to 0 at a cutoff of $r_{hb} = 2.3$ Å. The value of this cutoff was chosen based on the O–H RDF of water

$$q_6 = \sum_{\text{H} \notin a} w_{hb}\left(r_{\text{H}} - r_{\text{O}}^{a}\right) \qquad \text{H-bonds accepted,}$$

$$q_7 = \sum_{\text{H1, H2} \in a} \sum_{\text{O} \notin a} w_{hb}\left(r_{\text{H}}^{a} - r_{\text{O}}\right) \qquad \text{H-bonds donated.}$$

Finally, two order parameters measure the tetrahedrality of the evaporating molecule's local environment. For purposes of these OP measurements, the local environment is defined as the four nearest neighbors, as measured by oxygen–oxygen distance.

The first tetrahedrality OP, designated q_8, measures the variance of the oxygen–oxygen distances from their mean value. For a perfect tetrahedral arrangement, this order parameter would be zero. The second tetrahedrality OP, q_9, measures the deviation of the neighbor-central molecule-neighbor angles (denoted $\psi_{j,k}$ between neighbors j and k) from the value of 109.5° they would take in a tetrahedral arrangement

$$q_8 = \frac{1}{3} \sum_{k=1}^{4} \frac{\left(|r_{\text{O}}^{j} - r_{\text{O}}^{a}| - \bar{r}\right)^2}{4\bar{r}^2}$$

$$\text{with} \quad \bar{r} = \frac{1}{4} \sum_{k=1}^{4} |r_{\text{O}}^{j} - r_{\text{O}}^{a}| \quad \text{dist. tetrahed. meas.,}$$

$$q_9 = \frac{3}{32} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos\psi_{j,k} + \frac{1}{3}\right)^2$$

$$\text{with} \quad \cos\psi_{j,k} = \frac{(r_{\text{O}}^{j} - r_{\text{O}}^{a}) \cdot (r_{\text{O}}^{k} - r_{\text{O}}^{a})}{|r_{\text{O}}^{j} - r_{\text{O}}^{a}|\,|r_{\text{O}}^{k} - r_{\text{O}}^{a}|} \quad \text{angular tetrahed. meas.}$$

[1]T. Annable, R. Cordwell, and P. Ewing, Eur. Coat. J. **44**, 44 (2006).
[2]Y. Holl, J. L. Keddie, P. McDonald, and W. Winnik, *Film Formation in Coatings: Mechanisms, Properties and Morphology*, ACS Symposium Series Vol. 790, edited by T. Provder and M. W. Urban (American Chemical Society, 2001), Chap. 1, pp. 2–29.
[3]N. Chayen, J. Appl. Crystallogr. **30**, 198 (1997).
[4]S. Talreja, D. Kim, A. Mirarefi, C. Zukoski, and P. Kenis, J. Appl. Crystallogr. **38**, 988 (2005).
[5]N. E. Chayen and E. Saridakis, Nat. Methods **5**, 147 (2008).
[6]M. Knudsen, Ann. Phys. **352**, 697 (1915).
[7]E. M. Mortensen and H. Eyring, J. Phys. Chem. **64**, 846 (1960).
[8]As a monograph by Frank E. Jones points out, assuming a constant surface temperature, the water in Lake Mead in Nevada would evaporate in less than a day at this theoretical rate;[75] obviously, heat and mass transfer play a limiting role in macroscopic systems.

[9]E. Rideal, J. Phys. Chem. **29**, 1585 (1925).
[10]T. Alty, Philos. Mag. **15**, 82 (1933).
[11]T. Alty and C. Mackay, Proc. R. Soc. London, Ser. A **149**, 104 (1935).
[12]R. Marek and J. Straub, Int. J. Heat Mass Transfer **44**, 39 (2001).
[13]P. Davidovits, C. E. Kolb, L. R. Williams, J. T. Jayne, and D. R. Worsnop, Chem. Rev. **111**, PR76 (2011).
[14]Y. Li, P. Davidovits, Q. Shi, J. Jayne, C. Kolb, and D. Worsnop, J. Phys. Chem. A **105**, 10627 (2001).
[15]C. Cappa, W. Drisdell, J. Smith, R. Saykally, and R. Cohen, J. Phys. Chem. B **109**, 24391 (2005).
[16]J. D. Smith, C. D. Cappa, W. S. Drisdell, R. C. Cohen, and R. J. Saykally, J. Am. Chem. Soc. **128**, 12892 (2006).
[17]P. M. Winkler, A. Vrtala, R. Rudolf, P. E. Wagner, I. Riipinen, T. Vesala, K. E. J. Lehtinen, Y. Viisanen, and M. Kulmala, J. Geophys. Res. [Atmos] **111**, D19202, doi:10.1029/2006JD007194 (2006).
[18]P. Davidovits, C. Kolb, L. Williams, J. Jayne, and D. Worsnop, Chem. Rev. **106**, 1323 (2006).
[19]M. Faubel and T. Kisters, Nature (London) **339**, 527 (1989).
[20]M. Wilson, A. Pohorille, and L. Pratt, J. Phys. Chem. **91**, 4873 (1987).
[21]M. Wilson, A. Pohorille, and L. Pratt, J. Chem. Phys. **88**, 3281 (1988).
[22]M. Matsumoto and Y. Kataoka, J. Chem. Phys. **88**, 3233 (1988).
[23]J. Alejandre, D. Tildesley, and G. Chapela, J. Chem. Phys. **102**, 4574 (1995).
[24]R. Taylor, L. Dang, and B. Garrett, J. Phys. Chem. **100**, 11720 (1996).
[25]B. Shi, S. Sinha, and V. K. Dhir, J. Chem. Phys. **124**, 204715 (2006).
[26]P. Vassilev, C. Hartnig, M. Koper, F. Frechard, and R. van Santen, J. Chem. Phys. **115**, 9815 (2001).
[27]I. Kuo and C. Mundy, Science **303**, 658 (2004).
[28]I. Kuo, C. Mundy, B. Eggimann, M. McGrath, J. Siepmann, B. Chen, J. Vieceli, and D. Tobias, J. Phys. Chem. B **110**, 3738 (2006).
[29]C. Wick, I. Kuo, C. Mundy, and L. Dang, J. Chem. Theory Comput. **3**, 2002 (2007).
[30]R. Taylor, D. Ray, and B. Garrett, J. Phys. Chem. B **101**, 5473 (1997).
[31]R. Taylor and B. Garrett, J. Phys. Chem. B **103**, 844 (1999).
[32]R. Vacha, P. Slavicek, M. Mucha, B. Finlayson-Pitts, and P. Jungwirth, J. Phys. Chem. A **108**, 11573 (2004).
[33]J. Vieceli, M. Roeselova, and D. Tobias, Chem. Phys. Lett. **393**, 249 (2004).
[34]L. Dang and B. Garrett, Chem. Phys. Lett. **385**, 309 (2004).
[35]B. Garrett, G. Schenter, and A. Morita, Chem. Rev. **106**, 1355 (2006).
[36]P. E. Mason, J. Phys. Chem. A **115**, 6054 (2011).
[37]T. Tsuruta and G. Nagayama, J. Phys. Chem. B **108**, 1736 (2004).
[38]A. Morita, M. Sugiyama, H. Kameda, S. Koda, and D. Hanson, J. Phys. Chem. B **108**, 9111 (2004).
[39]J. Vieceli, M. Roeselova, N. Potter, L. Dang, B. Garrett, and D. Tobias, J. Phys. Chem. B **109**, 15876 (2005).
[40]T. Ishiyama, T. Yano, and S. Fujikawa, Phys. Fluids **16**, 4713 (2004).
[41]T. Yang and C. Pan, Int. J. Heat Mass Transfer **48**, 3516 (2005).
[42]M. Matsumoto, Fluid Phase Equilib. **144**, 307 (1998).
[43]C. Caleman and D. van der Spoel, Phys. Chem. Chem. Phys. **9**, 5105 (2007).
[44]C. Caleman and D. van der Spoel, J. Chem. Phys. **125**, 154508 (2006).
[45]L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, J. Chem. Phys. **125**, 024106 (2006).
[46]H. Berendsen, J. Grigera, and T. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[47]F. Chen and P. Smith, J. Chem. Phys. **126**, 221101 (2007).
[48]J. G. Kirkwood and F. P. Buff, J. Chem. Phys. **17**, 338 (1949).
[49]J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kale, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).
[50]T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[51]U. Essmann, L. Perera, M. Berkowitz, T. Darden, H. Lee, and L. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[52]A. Wynveen and F. Bresme, J. Chem. Phys. **124**, 104502 (2006).
[53]J. Alejandre and R. M. Lynden-Bell, Mol. Phys. **105**, 3029 (2007).
[54]R. Sakamaki, A. K. Sum, T. Narumi, R. Ohmura, and K. Yasuoka, J. Chem. Phys. **134**, 144702 (2011).
[55]W. E, W. Ren, and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002).
[56]W. E, W. Q. Ren, and E. Vanden-Eijnden, J. Phys. Chem. B **109**, 6688 (2005).
[57]W. Ren, E. Vanden-Eijnden, P. Maragakis, and W. E, J. Chem. Phys. **123**, 134109 (2005).
[58]W. E and E. Vanden-Eijnden, Annu. Rev. Phys. Chem. **61**, 391 (2010).
[59]P. Chau and A. Hardwick, Mol. Phys. **93**, 511 (1998).
[60]E. Vanden-Eijnden, M. Venturoli, G. Ciccotti, and R. Elber, J. Chem. Phys. **129**, 174102 (2008).
[61]E. Vanden-Eijnden and M. Venturoli, J. Chem. Phys. **130**, 194101 (2009).
[62]E. Vanden-Eijnden and M. Venturoli, J. Chem. Phys. **130**, 194103 (2009).
[63]L. Maragliano, E. Vanden-Eijnden, and B. Roux, J. Chem. Theory Comput. **5**, 2589 (2009).
[64]T. Eiter and H. Mannila, "Computing discrete Frechet distance," Technical report CD-TR 94/64 (Christian Doppler Laboratory for Expert Systems, TU Vienna, Vienna, Austria, 1994).
[65]Geometrically, the maximum value of $\cos\omega$, and thus q_5^{omeg} given a certain value of q_4^{thet} or θ, is $\cos\omega^{max} = \cos(\pi/2 - \theta)$.
[66]A. Ben-Naim and Y. Marcus, J. Chem. Phys. **81**, 2016 (1984).
[67]K. V. Mardia and P. E. Jupp, *Directional Statistics* (Wiley, 2000).
[68]In Ref. 60, the committor function, denoted $q(\mathbf{x})$, is a function of system coordinates, not collective variables, and has a single value at each point $\mathbf{x}$ which is conceptually measurable. The committor function at a particular value of *collective variables* has a range of values in a distribution, relating to the set of microstates that exhibit those particular CV values. For purposes of discussion in this subsection, we will consider the mean value of the committor probability p_B at collective variable values.
[69]R. E. Welsch, "Influential data," in *Encyclopedia of Statistical Sciences*, 2nd ed., edited by N. Balakrishnan, C. B. Read, and B. Vidakovic (Wiley, 2006).
[70]P. Prescott, "Influential observations," in *Encyclopedia of Statistical Sciences*, 2nd ed., edited by N. Balakrishnan, C. B. Read, and B. Vidakovic (Wiley, 2006).
[71]A. P. Willard and D. Chandler, J. Phys. Chem. B **114**, 1954 (2010).
[72]E. E. Santiso and B. L. Trout, J. Chem. Phys. **134**, 064109 (2011).
[73]P. E. Mason and J. W. Brady, J. Phys. Chem. B **111**, 5669 (2007).
[74]W. Humphrey and A. Dalke, J. Mol. Graphics **14**, 33 (1996).
[75]F. Jones, *Evaporation of Water: With Emphasis on Applications and Measurements* (CRC Press, 1992).
[76]See supplementary material at http://dx.doi.org/10.1063/1.4798458 for more details on the computational methods and analysis.