

Solvent coarse-graining and the string method applied to the hydrophobic collapse of a hydrated chain

Thomas F. Miller III[†], Eric Vanden-Eijnden[‡], and David Chandler^{†§}

[†]Department of Chemistry, University of California, Berkeley, CA 94720; and [‡]Courant Institute of Mathematical Sciences, New York University, New York, NY 10012

Contributed by David Chandler, June 25, 2007 (sent for review February 14, 2007)

With computer simulations of >100,000 atoms, the mechanism for the hydrophobic collapse of an idealized hydrated chain was obtained by tiling space with (0.2 nm)³ cubes and projecting the atomistic water molecule positions onto this grid. With the coarse-grained field thus defined, the string method in collective variables was used to compute a minimum free-energy pathway (MFEP) for the collapsing chain. These calculations provide a proof of principle for a coarse-grained description of water solvent. Furthermore, the calculated MFEP characterizes the mechanism for the collapse of the hydrated chain by providing a path of maximum likelihood for dynamical trajectories. The reliability of the calculated MFEP was confirmed with the use of conventional molecular dynamics trajectories. Analysis of the MFEP provides atomistic confirmation for the mechanism of hydrophobic collapse proposed by ten Wolde and Chandler. In particular, we show that length-scale-dependent hydrophobic dewetting is the rate-limiting step in the hydrophobic collapse of the considered chain.

rare-event dynamics | water

This paper applies the string method (1–3) to the phenomenon of hydrophobic collapse. We show that the method can describe complex dynamics in large atomistic systems, ones for which other currently available rare-event methods would seem intractable. Furthermore, we use the string method to demonstrate that atomistic dynamics can be usefully projected onto that of a coarse-grained field. The specific application of the string method considered herein finds results that are consistent with the mechanism of hydrophobic collapse put forward by ten Wolde and Chandler (tWC) (4).

The hydrophobic effect, or the tendency of oil and water to separate on mesoscopic length scales, has long been recognized as an important driving force in molecular assembly (5). Recent theoretical developments have helped establish a quantitative understanding of the thermodynamics of hydrophobicity (6, 7), but the dynamics of hydrophobic collapse remain poorly understood because it couples a large range of length and time scales. Relevant processes include the atomic-scale motions of individual water molecules, collective solvent density fluctuations, and the nanometer-scale movements of the hydrophobic solutes. Bridging these dynamical hierarchies and addressing the problem of complex dynamics in large systems are fundamental challenges for computational methods.

We address this challenge by using the string method in collective variables (1–3). We consider the collapse of a chain composed of 12 spherical hydrophobes in an explicit solvent of $\approx 34,000$ water molecules. We studied the system by coarse-graining the water molecule positions onto a set of 129,000 collective variables that represent the solvent density field and then using the string method in these variables to compute the minimum free-energy path (MFEP) for the hydrophobic collapse of the chain. Conventional molecular dynamics (MD) simulations were subsequently performed to confirm that this coarse-grained description adequately describes the mechanism of hydrophobic collapse.

tWC have previously reported simulations of an idealized hydrophobic chain solvated by a coarse-grained model of liquid water (4). For this model, they found that the key step in the collapse dynamics is a collective solvent density fluctuation that is nucleated at the hydrophobic surface of the chain. However, it was not clear whether this mechanism was an artifact of their coarse-grained model of water. Atomistic simulations are needed to resolve the issue.

Previous efforts to characterize the mechanism of hydrophobic collapse using atomistic computer simulations neither confirm nor disprove the mechanism proposed by tWC. In recent work, for example, MD simulations showed that dewetting accompanies the collapse of hydrophobes in water (8–11). However, the rate-limiting step and, thus, the mechanism for hydrophobic collapse were not characterized. Specifically, in ref. 10, MD trajectories were initiated at various separation distances for a pair of meletin protein dimers. Collapse dynamics were observed only when the initial dimer configuration was on the product side of the free-energy barrier, but the actual nature of that barrier and the dynamics of crossing it were not studied. In ref. 11, the thermodynamics and solvation of a hydrophobic chain were studied as a function of its radius of gyration, but, again, the dynamics of collapse was not characterized.

Our atomistic simulations explicitly confirm the mechanism of hydrophobic collapse put forward by tWC (4). In particular, we show that the rate-limiting step in the collapse of their hydrophobic chain coincides with a collective solvent motion and that performing the rate-limiting step involves performing work almost exclusively in the solvent coordinates. We further show that the solvation free energy along the MFEP can be decomposed into small- and large-length-scale contributions. This analysis demonstrates that atomistic solvent energetics can be quantitatively modeled by using a grid-based solvent density field, and it suggests that the rate-limiting step for hydrophobic collapse coincides with length-scale-dependent hydrophobic dewetting.

Atomistic Model, Coarse-Graining, and the String Method

The System. As indicated, we considered the atomistic version of the hydrated hydrophobic chain studied by tWC (4), a model that exhibits the length-scale-dependent hydrophobic effect (6). More complicated models can be considered, but we limited this study to this particular chain. The chain is composed of 12 spherical hydrophobes, each with a diameter of 7.2 Å and a mass of 71 atomic

Author contributions: T.F.M., E.V.-E., and D.C. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Abbreviations: tWC, ten Wolde and Chandler; MFEP, minimum free-energy path; MD, molecular dynamics.

[§]To whom correspondence should be addressed. E-mail: chandler@cchem.berkeley.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0705830104/DC1.

© 2007 by The National Academy of Sciences of the USA

mass units. These are ideal hydrophobes, because they exert purely repulsive interactions with water that expel the centers of the oxygen atoms from the volume within 0.5 nm of the center of the hydrophobe. The volume and mass are typical of amino acid residues (12, 13). Consecutive hydrophobes in the unbranched chain interact via harmonic bonds, and the chain is made semirigid by a potential energy term that penalizes its curvature. The rigidity is chosen so that the extended configurations of the chain are dominant in vacuum. It is only solvent (and thus the hydrophobic interaction) that stabilizes the collapsed globule configurations. The chain is hydrated with 33,912 rigid water molecules interacting with the SPC/E (simple point charges/extended model) potential (14) in an orthorhombic simulation box with periodic boundary conditions. Electrostatic interactions were included by using the smoothed particle mesh Ewald method (15), and all simulations were performed at 300 K. Full details of the simulation protocol are provided in [supporting information \(SI\) Text: “Details of the System.”](#)

To describe hydrophobic collapse, it is necessary to choose an appropriate thermodynamic ensemble for the simulations. Nanometer-scale fluctuations in solvent density are suspected to play a key role in these dynamics. Use of the NVT ensemble (with a 1-g/cm³ density of water) might suppress these density fluctuations and bias the calculated mechanism. We avoided this problem with a simple technique that is based on the fact that, under ambient conditions, liquid water is close to phase coexistence. Indeed, it is this proximity that leads to the possibility of large-length-scale hydrophobicity (6, 16). By placing a fixed number of water molecules at 300 K in a volume that corresponds to an average density of <1 g/cm³, we obtained a fraction of the system at the density of water vapor and the majority at a density of bulk water. Because we are not concerned with solvent fluctuations on macroscopic length scales, the difference between simulating bulk water at its own vapor pressure compared with atmospheric pressure is completely negligible. This strategy has been used to study the dewetting transition between solvophobic surfaces (17, 18). To ensure that the liquid-vapor interface remains both flat and well distanced from the location of the chain, we repelled particles from a thin layer at the top edge of the simulation box, as is discussed in [SI Text: “Details of the System.”](#)

Coarse-Graining of Solvent. Throughout this study, we simulated water and the hydrated chain by using atomistic MD. To apply the string method to the dynamics of this system, we employed a choice of collective variables that describe the density field of water. In particular, a coarse-graining algorithm was developed to connect the atomistic and collective variable representations of the solvent. By following tWC (4), the simulation box was partitioned into a three-dimensional lattice (48 × 48 × 56) of cubic cells with a side length of $l = 2.1 \text{ \AA}$. We labeled the cells with the vector $\mathbf{k} = (k_x, k_y, k_z)$, where each k_ζ takes on integer values bounded as follows: $1 \leq k_x \leq 48$, $1 \leq k_y \leq 48$, and $1 \leq k_z \leq 56$. On this grid, the molecular density $\rho(\mathbf{r})$, as determined by the positions of the water oxygen atoms, is coarse-grained into the field $P_{\mathbf{k}}$, where

$$P_{\mathbf{k}} = \int d\mathbf{r} \rho(\mathbf{r}) \prod_{\zeta=x,y,z} \Phi_{k_\zeta}(\mathbf{r} \cdot \mathbf{1}_\zeta). \quad [1]$$

Here, the integral extends over the volume of the system, and $\mathbf{1}_\zeta$ denotes the unit vector in the ζ th Cartesian direction. The coarse-graining function, $\Phi_{k_\zeta}(x)$, is normalized, ensuring that $\sum_{\mathbf{k}} P_{\mathbf{k}} = N$ is the total number of water molecules. The particular function that we have chosen to use is

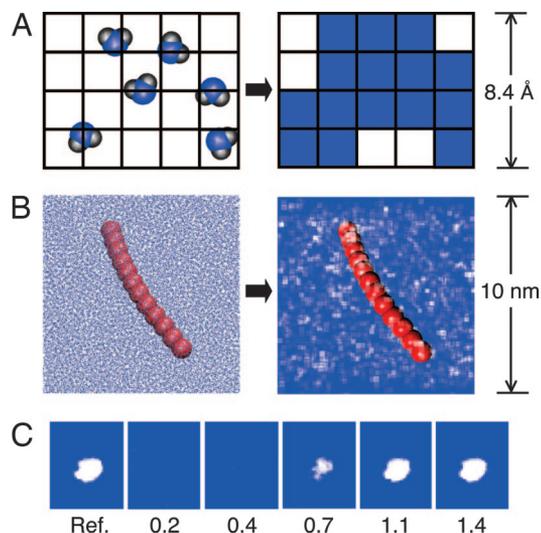


Fig. 1. Solvent coarse graining. (A) The coarse-graining procedure is schematically shown to project the atomistic solvent density onto a discrete grid. (B) The same procedure is shown for an instantaneous solvent configuration of the actual system. Grid cells containing less solvent density than $c/2$ are colored white; the remaining cells are left transparent against the blue background. The hydrophobic chain is shown in red. (C) Atomistic solvent density is restrained to the reference distribution at the far left. With larger restraint force constant κ reported numerically in units of $2 k_B T/c^2$, the average atomic solvent density reproduces the reference distribution in detail (see text for notation).

$$\Phi_j(x) = \int dy \phi(x-y) \left[h_j(x) h_j(y) + h_{j+1}(x) \sum_{i \leq j} h_i(y) + h_{j-1}(x) \sum_{i \geq j} h_i(y) \right], \quad [2]$$

where $\phi(x) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2)$, $\sigma = 1 \text{ \AA}$, and $h_i(x)$ is unity when x is in the i th interval and zero otherwise. In effect, this choice spreads the atomistic density field $\rho(\mathbf{r})$ over the length scale σ and bins it into a grid of length scale l in such a way as to preserve normalization. While we have found this choice of coarse-graining function to be convenient, others are possible.

Fig. 1A and B illustrates the coarse-graining procedure. In Fig. 1A, the solvent is schematically shown before and after coarse-graining. Fig. 1B, the same mapping is shown for the actual system considered here. Cells are shaded white when their solvent occupation, $P_{\mathbf{k}}$, is less than half of the bulk average value of $\langle P \rangle_{\text{bulk}} \equiv c = 0.3$ molecules. Small local density fluctuations are seen throughout the simulation box, as is expected for an instantaneous solvent configuration.

In addition to visualizing solvent density, the coarse-graining algorithm is useful for controlling the solvent density in MD simulations. For example, if it is desired that a particular cell \mathbf{k} exhibit a solvent occupation $P_{\mathbf{k}}^*$, the potential energy term $\kappa(P_{\mathbf{k}} - P_{\mathbf{k}}^*)^2/2$ can be used to derive the appropriate forces on the atoms in the simulation, where κ is a force constant. This technique is illustrated in Fig. 1C. The leftmost image shows a density distribution that is exceedingly unlikely for a simulation of ambient liquid water. Images to the right show the average solvent density distribution from MD simulations that are restrained to this unlikely reference distribution with increasing κ . Force constant values of $>2 k_B T/c^2$, in which the simulation incurs an energetic penalty of at least $k_B T$ for placing the average bulk density in a cell that is

restrained to be empty, ensure that the reference distribution is recovered in detail. Here, $k_B T$ denotes Boltzmann's constant times temperature T .

Defining the MFEP. We used the coarse-graining procedure to obtain collective variables that describe the position of the hydrophobic chain and the water density field. Specifically, let $\mathbf{x} = (\mathbf{x}_c, \mathbf{w})$ be the position vector of length $n = 3 \times 12 + 3 \times 3 \times N$ for the atomistic representation of the entire system, where \mathbf{x}_c is the position vector for the hydrophobes in the chain and \mathbf{w} is the position vector for the atoms in the water molecules. Then, $\mathbf{z}(\mathbf{x}) = (\mathbf{x}_c, \mathbf{P})$ is the vector of length $N = 3 \times 12 + 48 \times 48 \times 56$ for the collective variable representation of the system, where the elements of \mathbf{P} are defined in Eq. 1.

The MFEP is a curve in the space of collective variables. It is represented by $\mathbf{z}^*(\alpha)$, where $\alpha = 0$ corresponds to the collapsed chain and $\alpha = 1$ corresponds to the extended chain. For intermediate values $\alpha \in (0, 1)$, the MFEP obeys the condition

$$\frac{dz_i^*(\alpha)}{d\alpha} \text{ parallel to } \sum_{j=1}^N M_{ij}(\mathbf{z}^*(\alpha)) \frac{\partial F(\mathbf{z}^*(\alpha))}{\partial z_j}, \quad [3]$$

where $F(\mathbf{z}) = -\beta^{-1} \ln \langle \Pi_{i=1}^n \delta(z_i - z_i(\mathbf{x})) \rangle$ is the free-energy surface defined in the collective variables, and

$$M_{ij}(\mathbf{z}) = \exp[\beta F(\mathbf{z})] \times \left\langle \sum_{k=1}^n m_k^{-1} \frac{\partial z_i(\mathbf{x})}{\partial x_k} \frac{\partial z_j(\mathbf{x})}{\partial x_k} \prod_{i=1}^n \delta(z_i - z_i(\mathbf{x})) \right\rangle. \quad [4]$$

Here, angle brackets indicate equilibrium expectation values, $\beta^{-1} = k_B T$, and m_k is the mass of the atom corresponding to coordinate x_k . The matrix $M_{ij}(\mathbf{z})$, which arises from projecting the dynamics of the atomistic coordinates onto the collective variables (and, thus, in general, curving the coordinate space) (3, 19), ensures that the various collective variables evolve on a consistent time scale. As is seen in Eq. 10, the matrix scales the relative diffusion coefficients for the collective variables.

If the used collective variables are adequate to describe the mechanism of the reaction (here, the hydrophobic collapse), then it can be shown that the MFEP is the path of maximum likelihood for reactive MD trajectories that are monitored in the collective variables (3). For the current application, we checked the adequacy of the collective variables *a posteriori* by running MD trajectories that are initiated from the presumed rate-limiting step along the MFEP (i.e., the configuration of maximum free energy) and verified that these trajectories led with approximately equal probability to either the collapsed or extended configurations of the chain (see *Hydrophobic Collapse of a Hydrated Chain: "The Committor Function and a Proof of Principle for Course-Graining"*).

String Method in Collective Variables. The string method yields the MFEP by evolving a parameterized curve (i.e., a string) according to the dynamics (3, 20)

$$\begin{aligned} \frac{\partial z_i^*(\alpha, t)}{\partial t} = & - \sum_{j=1}^N M_{ij}(\mathbf{z}^*(\alpha, t)) \frac{\partial F(\mathbf{z}^*(\alpha, t))}{\partial z_j} \\ & + \lambda(\alpha, t) \frac{\partial z_i^*(\alpha, t)}{\partial \alpha}, \end{aligned} \quad [5]$$

where the term $\lambda(\alpha, t) \partial z_i^*(\alpha, t) / \partial \alpha$ enforces the constraint that the string remain parameterized by normalized arc length. The end-

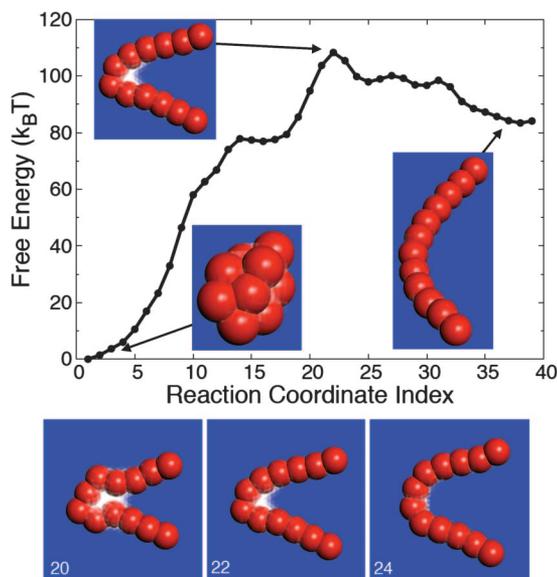


Fig. 2. The minimum free-energy path obtained by using the string method. (Upper) The free-energy profile exhibits a single peak at configuration 22. (Lower) The configurations of the path in the vicinity of the free-energy peak are shown with configuration numbers indicated in white text.

points of the string evolve by steepest descent on the free-energy surface,

$$\frac{\partial z_i^*(\alpha, t)}{\partial t} = - \frac{\partial F(\mathbf{z}^*(\alpha, t))}{\partial z_i}, \quad [6]$$

for $\alpha = 0$ and $\alpha = 1$. These artificial dynamics of the string yield the MFEP, which satisfies Eq. 3.

In practice, the string is discretized by using N_d configurations of the system in the collective variable representation. The dynamics in Eqs. 5 and 6 are then accomplished in a three-step cycle, where (i) the endpoint configurations of the string are evolved according to Eq. 6 and the rest of the configurations are evolved according to the first term in Eq. 5, (ii) the string is (optionally) smoothed, and (iii) the string is reparameterized to maintain equidistance of the configurations in the discretization. This cycle is repeated until the discretized version of Eq. 3 is satisfied. Step i requires evaluation of the mean force elements $\partial F(\mathbf{z}) / \partial z_i$ and the tensor elements $M_{ij}(\mathbf{z})$ at each configuration. These terms are obtained by using restrained atomistic MD simulations of the sort illustrated for the solvent degrees of freedom in Fig. 1C. The details of the string calculation are provided in *SI Text: "String Method in Collective Variables."*

Hydrophobic Collapse of a Hydrated Chain

MFEP. Fig. 2 shows the MFEP for the hydrophobic collapse of the hydrated chain. It was obtained by using the string method in the collective variables for the chain atom positions and the grid-based solvent density field. The converged MFEP was discretized by using $N_d = 40$ configurations of the system, and we shall hereafter refer to these configurations by their index number, $s = 1, \dots, N_d$. The free-energy profile was obtained by integrating the projection of the mean force along the MFEP, using

$$F^*(\alpha) = \int_0^\alpha \nabla F(\mathbf{z}^*(\alpha')) \cdot d\mathbf{z}^*(\alpha'). \quad [7]$$

The resolution of $F^*(\alpha)$ in Fig. 2 could be improved by employing a larger N_d but at larger computational cost.

The statistical errors in the free-energy profile between consecutive configurations are approximately the size of the plotted circles, and the small features in the profile at configurations 27 and 31 are due to noise in the convergence of the string calculation. Further discussion of the statistical error in the string calculation is provided in *SI Text: "String Method in Collective Variables."* Fig. 2 Lower presents configurations along the MFEP in the region of the free-energy barrier. As in Fig. 1, lattice cells with less than half of the bulk solvent occupation number fade to white.

The free-energy profile in Fig. 2 is dominated by a single barrier at configuration 22, where a liquid–vapor interface is formed at a bend in the hydrophobic chain. The sharply curved chain geometry presents an extended hydrophobic surface to molecules located in the crook of the bend, an environment that is analogous to that experienced by water trapped between hydrophobic plates and known to stabilize large-length-scale solvent density fluctuations (16, 21, 22). The barrier in the calculated free-energy profile clearly coincides with a collective motion in the solvent variables.

The string method characterizes the most likely member among a local channel of reactive trajectories. A reasonable concern, however, is the extent to which other channels (such as those that might correspond to forming bends at other points along the hydrophobic chain) are accessible and important. However, if other bends in the chain did correspond to well separated transition channels, then it seems unlikely, given the arbitrary manner in which the string was initialized, that we could locate a MFEP that exhibits a totally symmetric bend in the chain. Furthermore, an estimate for the energetic cost of translating the bend along the chain, which is described in *SI Text: "Estimation of the Barrier Between Reactive Barriers,"* suggests that trajectories that collapse by forming asymmetric bends in the chain do not belong to a distinct reactive channel from those that form a symmetric bend. That said, the less symmetric transition state configuration illustrated in figure 2 of ref. 4 is but one member of the transition-state ensemble, the most likely of which is pictured here in Fig. 2.

Solvation Free Energy. A simple theory was constructed to understand the contributions to the free-energy profile in Fig. 2 and to test the validity of coarse-graining solvent interactions. Noting that our choice of collective variables for the string calculation neglects the configurational entropy of the chain, the free-energy profile $F(\alpha)$ can be decomposed into the configurational potential energy of the chain $E_c(\alpha)$ and the free energy of hydrating the chain, $F_h(\alpha)$.

$$F^*(\alpha) = E_c(\alpha) + F_h(\alpha). \quad [8]$$

The term $E_c(\alpha)$ is easily evaluated from the chain potential energy term, yielding the components of the free-energy profile shown in Fig. 3A.

Now we will focus our attention on F_h . The solvation free energy for hydrophobes of idealized geometry is well understood. For small-length-scale hydrophobes (<1 nm), this energy scales linearly with the solute volume, whereas, for large-length-scale hydrophobes, it scales linearly with the solute surface area (16, 23). However, theoretical prediction of the solvation free energy for more complicated solute geometries is not necessarily trivial. For example, depending on its configuration, some parts of the chain considered here might be solvated like a large-length-scale hydrophobe, whereas other parts might behave like a small-length-scale hydrophobe.

Here, we present a technique for separating a general hydrophobic solute into components that belong to either the small-length-scale or large-length-scale regimes. We define the solvent-depleted volume as the continuous set of lattice cells that are, on average, occupied by $<50\%$ of the average bulk solvent value (this set includes both the volume that is excluded by the hard-sphere-like interactions between the solute and solvent interactions and any additional volume in the vicinity of the solute that is not substan-

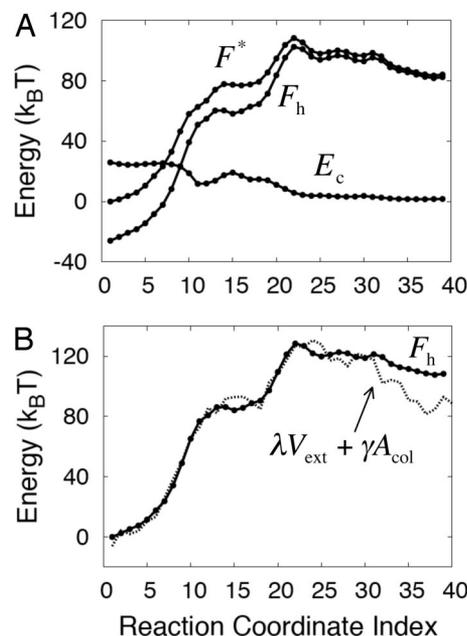


Fig. 3. A simple theory describes the solvation free energy of irregular hydrophobic solute geometries. (A) The free-energy profile F^* from simulation, decomposed into contributions from the chain configurational potential energy E_c and the free energy of solvation F_h . (B) The free energy of solvation obtained from Eq. 9 (dotted line) and from simulation (solid line).

tially occupied by the solvent). We also introduce a probe volume that is a cubic $4 \times 4 \times 4$ set of cells, which is approximately the size at which the water solvation structure changes from the small-length-scale to the large-length-scale regime.

The probe volume is used to determine whether a given section of the solvent-depleted volume is in the small- or large-length-scale regime. This is done by moving the probe volume throughout the simulation box and determining whether it can be fit into different portions of the solvent-depleted volume. If at a given position the vast majority of this probe volume (specifically, 59 of 64 cells) fits within the solvent-depleted volume, then the cells in that portion of the solvent-depleted volume are included in the large-length-scale component, V_{col} . Portions of the solvent-depleted volume that do not meet this criterion for any position of the probe volume are included in the small-length-scale component, V_{ext} . The volumes V_{col} and V_{ext} , which are plotted in *SI Fig. 8A*, primarily include contributions from the collapsed portions and extended portions of the chain, respectively. The total volume V_{tot} is obtained by counting the total number of cells in the solvent-depleted volume, such that $V_{tot} = V_{ext} + V_{col}$.

The total surface area, A_{tot} , and its large-length-scale component, A_{col} , were obtained by counting the number of external faces on the cells that comprise V_{tot} and V_{col} , respectively. The surface area of the small-length-scale component was obtained by using $A_{ext} = A_{tot} - A_{col}$. To account for the fact that we are calculating the area of a smooth surface by projecting it onto a cubic lattice, each surface area term also is multiplied by a $2/3$, a correction factor that is exact for an infinitely large sphere. The surface area components are plotted in *SI Fig. 8A*.

We use the separated components of the solute volume and surface area to estimate the solvation free energy for the chain along the minimum free-energy path with the relationship.

$$F_h(\alpha) = \lambda V_{ext}(\alpha) + \gamma A_{col}(\alpha). \quad [9]$$

The coefficients λ and γ are, respectively, the prefactors for the linear scaling of the solvation free energy in the small-length-scale

and large-length-scale regimes, as obtained by calculations on a spherical hydrophobic solute (23). Their values are taken to be $\lambda = 8 \text{ mJ}/(\text{m}^2\text{\AA})$ and $\gamma = 45 \text{ mJ}/\text{m}^2$.

The results of this calculation are presented in Fig. 3*B*. The dotted line indicates our theoretical estimate of the solvation free energy using Eq. 9, and the solid line indicates the data from the atomistic simulations (from Fig. 3*A*). The agreement is very good, and, as is shown in *SI Text: "Alternate Models for the Solvation Free Energy,"* it is better than can be obtained from solvation free-energy estimates that exclusively consider either the solute volume or the solute surface area. This result demonstrates that atomistic solvent energetics can be quantitatively modeled by using a grid-based solvent density field. Furthermore, note in *SI Fig. 8* that the collapsing chain first develops large-length-scale components in the vicinity of configuration 22, which suggests that both the peak in the free energy profile in Fig. 2 and the corresponding collective solvent density motion coincide with the crossover from small- to large-length-scale solvation (i.e., dewetting).

The Committor Function and a Proof of Principle for Coarse-Graining.

The various assumptions that are used in our implementation of the string method, including our choice of collective variables and the corresponding neglect of the chain configurational entropy, raise the possibility that the barrier in the free-energy profile in Fig. 2 does not correspond to the dynamical bottleneck for hydrophobic collapse. This is a general concern in trying to relate free-energy calculations to dynamical quantities, such as the reaction mechanism and the reaction rate. To confirm that the calculated free-energy profile is dynamically relevant, we evaluated the committor function at several configurations along the MFEP. The committor function reports the relative probability that MD trajectories that are initialized from a particular collective variable configuration first proceed to the extended configurations of the chain, as opposed to the collapsed configurations. For initial collective variable configurations that coincide with the true dynamical bottleneck, the committor function assumes a value of exactly 0.5.

We first evaluated the committor function at configuration 22, which corresponds to the peak in the calculated free energy profile in Fig. 2. The committor function is obtained by performing straightforward MD trajectories from initial conditions that are consistent with this collective variable configuration and then tallying the fraction of those trajectories whose endpoints are "extended," as opposed to "collapsed." The initial conditions for these trajectories are obtained from a 20-ps MD trajectory that is restrained to the collective variables for configuration 22; the atomistic coordinates of the restrained trajectory are recorded every picosecond. From each set of atomistic coordinates, an unrestrained trajectory was run for 150 ps forwards and backwards in time with the initial atomistic velocity vector drawn from the Maxwell-Boltzmann distribution at 300 K. To decide whether a given unrestrained trajectory terminates in either an extended or collapsed configuration of the chain, we employed an order parameter based on the number of chain atoms that are within $d = 1.25 \text{ \AA}$ of another chain atom to which it is not directly bonded, where d refers to the separation between the surfaces of spheres with 3.6-\AA radii surrounding the chain atoms. If the number of nonbonded contacts exceeds three, the chain is considered to be in a collapsed configuration; otherwise it is considered extended. This order parameter need only distinguish between collapsed and extended configurations of the chain; it need not be (and in fact is not) a good reaction coordinate. Fig. 4 illustrates representative forward and backward unrestrained MD trajectories. Fig. 4*B* reminds the reader of the atomistic basis for the coarse-grained field.

The committor function at configuration 22 is 0.3 ± 0.1 , suggesting that the calculated free-energy barrier is slightly biased toward the basin of stability for the collapse chain. However, this deviation from the ideal value of 0.5 is only marginally statistically

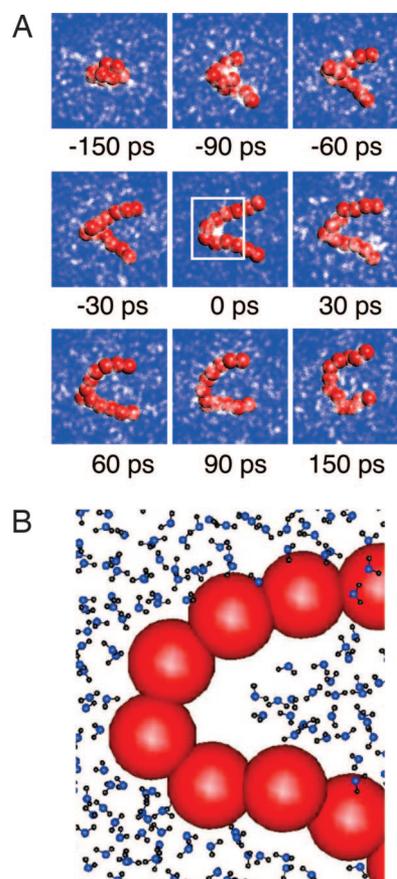


Fig. 4. Snapshots from unrestrained MD trajectories that are initiated from the calculated free-energy barrier at configuration 22. (A) The instantaneous configurations are visualized in the collective variable representation by using the same technique that was introduced in Fig. 1. (B) The boxed area from the 0-ps snapshot in A, with a planar cross-section of water with a thickness of 7.2 \AA , is shown in atomistic detail.

significant, and we emphasize that the committor function is exponentially sensitive in the region of the dynamical bottleneck. To illustrate this point, we repeat the evaluation of the committor function at configuration 20, which is seen in Fig. 2 to be near the barrier peak but slightly closer to the collapsed configurations of the chain, as well as at configuration 24, which is slightly closer to the extended configurations. We found the committor function at configuration 20 to be 0.00 ± 0.05 and the value at configuration 24 to be 0.90 ± 0.05 . Only very minor shifts along the MFEP dramatically change the value of the committor function.

Given the extreme sensitivity of the committor function near the dynamical bottleneck, the fact that configuration 22 gives rise to a significant fraction of trajectories that proceed to both the collapsed and extended basins is very significant. Furthermore, the direction in which the committor function changes as a function of the configuration number is as is expected in the vicinity of the dynamical bottleneck. We conclude that the barrier in the calculated free-energy profile correctly characterizes the true free-energy barrier, thus identifying the rate-limiting step for the collapse dynamics.

The committor function calculations, which are based on straightforward MD simulations, indicate that our choice of collective variables provides a reasonable description of the mechanism of hydrophobic collapse. This result yields atomistic support for the strategy of coarse-graining solvent dynamics. Because it is straightforward (3) to obtain the stochastic dynamics in the collective variables for which the most likely reaction pathway is the same

as the mechanism obtained by using the string method, and further, because the mechanism for hydrophobic collapse obtained in the used collective variables agrees with the mechanism obtained by using atomistic dynamics, we obtain a proof of principle that atomistic solvent dynamics can be usefully projected onto a coarse-grained field. The Langevin equation for the collective variables in this description is (3)

$$\begin{aligned} \bar{\gamma} \frac{\partial z_i(t)}{\partial t} = & - \sum_{j=1}^N \left(M_{ij}(\mathbf{z}(t)) \frac{\partial F(\mathbf{z}(t))}{\partial z_j} - \beta^{-1} \frac{\partial M_{ij}(\mathbf{z}(t))}{\partial z_j} \right) \\ & + (2\bar{\gamma}/\beta)^{1/2} \sum_{j=1}^N M_{ij}^{1/2}(\mathbf{z}(t)) \eta_j(t), \end{aligned} \quad [10]$$

where $\eta_i(t)$ is a white noise satisfying $\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t')$ and t is an artificial time that is scaled by the friction coefficient $\bar{\gamma}$.

The computational feasibility of directly integrating these coarse-grained dynamics, however, hinges on the cost of calculating the mean force elements $\partial F(\mathbf{z})/\partial z_j$ and the tensor elements $M_{ij}(\mathbf{z})$ and $\partial M_{ij}(\mathbf{z})/\partial z_j$, because these terms are required at every coarse-grained time step. It is therefore encouraging that the free-energy surface in the collective variables can be quantitatively modeled in lieu of atomistic simulations, by using Eqs. 8 and 9. Similar approximations for the tensor elements might also be possible.

The Rate-Limiting Step. Using the calculated MFEP and committor function values, we established that the rate-limiting step for the hydrophobic collapse of the hydrated chain coincides with a collective solvent motion. Using a simple analysis of the solvation free energy, we found that this collective solvent motion is consistent with length-scale-dependent dewetting. However, it remains to be shown whether the rate-limiting step involves performing work in the solvent or chain degrees of freedom. This is an important distinction. If the latter case is true, then dewetting merely accompanies hydrophobic collapse as a spectator, but, if the former case is true, then dewetting is the rate-limiting step to hydrophobic collapse.

To address this issue, we again decomposed the free-energy profile, this time into contributions from work performed along the solvent and the chain collective variables. The definition of the free-energy profile in Eq. 7 can be written more explicitly as

$$\begin{aligned} F^*(\alpha) = & \int_0^\alpha \nabla_{\mathbf{c}} F(\mathbf{z}^*(\alpha')) \cdot d\mathbf{x}_{\mathbf{c}}(\alpha') \\ & + \sum_{\mathbf{k} \in V} \int_0^\alpha \nabla_{P_{\mathbf{k}}} F(\mathbf{z}^*(\alpha')) dP_{\mathbf{k}}(\alpha') \end{aligned} \quad [11]$$

where $\nabla_{\mathbf{c}} F(\mathbf{z}^*(\alpha'))$ is the vector of mean forces acting on the chain atom positions at configuration $s = N_d \times \alpha$ along the MFEP, and $\nabla_{P_{\mathbf{k}}} F(\mathbf{z}^*(\alpha'))$ is the corresponding mean force on the solvent collective variable $P_{\mathbf{k}}$. The full free-energy profile is obtained by setting the volume V equal to the entire simulation

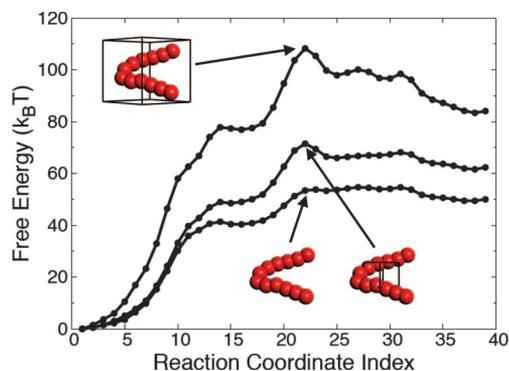


Fig. 5. The free-energy profile from Fig. 2 is separated into contributions from the chain and solvent coordinates. In the bottom curve, only the first term in Eq. 11, which corresponds to work performed in the chain coordinates, is included in the free-energy profile. Solvent contributions additionally are included in the middle and top curves. The solvent volume included for each curve is indicated with a wire box. Inclusion of larger solvent volumes does not substantially change the profile.

box. To understand the role of solvent in hydrophobic collapse, however, it is informative to calculate the free-energy profile by using various smaller solvent volumes.

The bottom curve in Fig. 5 is obtained from Eq. 11 by letting V be the empty set, thus eliminating the second term. The middle curve is obtained by letting V be the set of $8 \times 8 \times 8$ lattice cells at the middle of the simulation box, as is indicated in the corresponding image, and the top curve is obtained by letting V be the middle set of $20 \times 20 \times 20$ lattice cells. Consideration of larger sets of lattice cells does not further alter the free-energy profile, as is shown in *SI Text*: “Convergence of the Free-Energy Profile with Increasing Solvent Box Sizes.” The solvent variables for regions of space that are distant from the collapsing chain remain constant along the path and do not contribute to the free-energy profile.

The bottom curve in Fig. 5 shows only the work performed on the chain atom positions during hydrophobic collapse. Remarkably, it lacks almost any free-energy barrier, indicating that essentially no work is performed in the chain degrees of freedom in crossing the dynamical bottleneck. Instead, we see that the barrier emerges only upon inclusion of the work performed in the solvent collective variables. Performing the hydrophobic collapse involves traversing a free-energy barrier that exists only in the solvent coordinates. Fig. 5 shows that the dewetting transition not only accompanies the rate-limiting step for hydrophobic collapse, but that it is the rate-limiting step.

We thank Giovanni Ciccotti, Mauro Ferrario, Ray Kapral, and Luca Maragliano for useful discussions and the National Energy Research Scientific Computing Center for computing resources. T.F.M. was supported by Department of Energy Grants CHE-0345280 and DE-FG03-87ER13793. E.V.-E. was supported by the Miller Institute at University of California (Berkeley, CA), Office of Naval Research Grant N00014-04-1-0565, and National Science Foundation Grants DMS02-09959 and DMS02-39625. D.C. was supported by National Science Foundation Grant CHE-0543158.

1. E W, Ren W, Vanden-Eijnden E (2002) *Phys Rev B* 66:52301.
2. E W, Ren W, Vanden-Eijnden E (2005) *J Phys Chem B* 109:6688–6693.
3. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G *J Chem Phys* 125:024106, 2006.
4. ten Wolde PR, Chandler D (2002) *Proc Natl Acad Sci USA* 99:6539–6543.
5. Kauzmann W (1959) *Adv Prot Chem* 14:1–63.
6. Chandler D (2005) *Nature* 437:640–647.
7. Hummer G, Garde S, Garcia AE, Pratt LR (2000) *Chem Phys* 258:349–370.
8. Huang X, Margulis CJ, Berne BJ (2003) *Proc Natl Acad Sci USA* 100:11953–11958.
9. Huang Q, Ding S, Hua C-Y, Yang H-C, Chen C-L (2004) *J Chem Phys* 121:1969–1977.
10. Liu P, Huang X, Zhou R, Berne BJ (2005) *Nature* 437:159–162.
11. Athawale MV, Goel G, Ghosh T, Truskett TM, Garde S (2007) *Proc Natl Acad Sci USA* 104:733–738.

12. Huang DM, Chandler D (2000) *Proc Natl Acad Sci USA* 97:8324–8327.
13. Mishra AK, Ahluwalia JC (1984) *J Phys Chem* 88:86–92.
14. Berendsen HJC, Grigera JR, Straatsma TP (1987) *J Chem Phys* 91:6269–6271.
15. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) *J Chem Phys* 103:8577–8593.
16. Lum K, Chandler D, Weeks JD (1999) *J Phys Chem B* 103:4570–4577.
17. Bolhuis PG, Chandler D (2000) *J Chem Phys* 113:8154–8160.
18. Huang X, Zhou R, Berne BJ (2005) *J Phys Chem B* 109:3546–3552.
19. Fukui K (1981) *Acc Chem Res* 14:363–368.
20. E W, Ren W, Vanden-Eijnden E (2007) *J Chem Phys* 126:164103.
21. Luzar A, Leung K (2000) *J Chem Phys* 113:5836–5844.
22. Leung K, Luzar A (2000) *J Chem Phys* 113:5845–5852.
23. Huang DM, Geissler PL, Chandler D (2001) *J Phys Chem B* 105:6704–6709.