Supporting Information

Tinti et al. 10.1073/pnas.1714796114

CNT Calculations

CNT calculations shown in Fig. 3 were performed using the software Surface Evolver (65), which performs energy minimizations subject to constraints. To reproduce the nucleation theory presented in refs. 4 and 5, we constrained the bubble volume to "drive" the system along the thermally activated event. This is analogous to performing a (continuum) string calculation in a single (collective) variable (25). Minimization of the surface free energy for a given volume yielded the most probable configurations for that volume and the related free energy. Then, by changing the bubble volume, it was possible to reconstruct putative segments of the transition path (Fig. 3 and Fig. S2): Since several distinct bubble shapes can be enclosed within the nanopore, we ran separate Surface Evolver calculations for each of them (Fig. S3). It is important to remark that, since the bubble volume is an insufficient order parameter (25), the identified path differs from the one found with the string method in the density field (Fig. 2).

Euclidean distances in Fig. 3 were obtained by extracting the refined triangulated mesh used in the Surface Evolver calculation, discretizing the space in an array of parallelepipeds, and computing the vapor occupation of each cell. We then constructed the vectors containing the vapor occupations for each Surface Evolver configuration; Fig. 3 shows the euclidean distance between two vectors corresponding to successive bubble volumes spanning the whole path: This distance is an estimator of the continuity of the transition path. An identical approach was used for the atomistic data, starting from the target value of the CVs, which is, by definition, a coarse-grained density field.

Simulation Details

The MD simulations discussed in the main text belong to two main types, (*i*) string simulations and (*ii*) intrusion–extrusion experiments. The LAMMPS (large-scale atomic/molecular massively parallel simulator) (66) MD software was used to perform these computations.

The String Method in CVs. The string simulations were performed on the largest system we investigated (2.6-nm pore) which consists of 12,748 TIP4P/2005 (36) water molecules and *ca.* 77,168 solid particles. The shake algorithm is used for simulating water molecules (67). These simulations were performed at constant temperature $T = 300$ K and at constant liquid pressure $P \approx 0$. The system was thermostated using Nosé–Hoover chains (68). The pressure was kept constant by using two pistons (22); in practice, a force is applied to two rigid walls far away from the pore such that the total force applied to the wall counteracts the liquid pressure. For the string calculations, this force was set to 0. Periodic boundary conditions were enforced in the y and z directions, which are orthogonal to the axis of the pore.

We used a rare event technique known as the string method in CVs (34) to capture the most probable—or minimum free-energy—path for the thermally activated transition from the fully wet metastable state to the stable state which is characterized by vapor filling the hydrophobic pore. This is a local iterative method which allows for identifying the minimum free-energy path (MFEP) in the reduced space of CVs. The string method also allows for computing the free energy along the MFEP. The interested reader is referred to the original paper (34) for a thorough derivation of the method and to ref. 69 for the optimized algorithm implemented here; in this *Supporting Information*, we

only report the few details which are instrumental to understand the results of the present work.

It is often convenient to describe, in macroscopic terms, the rare event of interest by defining a set of CVs. We used the discretized density field as CV; this choice is known to render the correct path for hydrophobically confined liquids (25). This CV can be represented as a vector of components θ_i, each rendering the number density inside the ith cell of a tessellation of the 3D space; see Fig. 1. The expression for θ_i is

$$\theta_i(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n_w}) = \frac{1}{\Delta V_i} \sum_{j=1}^{n_w} \bar{\chi}_i(\boldsymbol{r}_j), \qquad \textbf{[S1]}$$

where $\bar{\chi}_i(\boldsymbol{r}_j)$ is a smoothed approximation to the characteristic function of the ith cell, $\boldsymbol{r}_j \in \mathcal{R}^3$ is the position vector of the oxygen of the jth water molecule, ΔV_i is the cell volume, and the sum at the right-hand side runs over the n_w water molecules.

In practice, the functions defined in Eq. **S1** count the number of water molecules on a Cartesian grid comprising $n_{cv} = 10 \times 4 \times 4 = 160$ cells, corresponding to cubic cells of *ca.* 7 Å. This level of discretization is sufficient to ensure that the path is continuous (Fig. 3) and, at the same time, that there are sufficient statistics for computing the observables required in the string method and defined in Eq. **S3**. Coarser discretized density fields, such as, e.g., the number of particles within the cavity (corresponding to the case $n_{cv} = 1$) used in CNT and analogous atomistic calculations, may yield discontinuous nucleation pathways because they are not able to distinguish between distinct morphologies of the meniscus (see the discussion in ref. 25).

Within the definitions given in Eq. **S1**, the MFEP, namely, the final result of the string method, corresponds to a parametric curve $\zeta(\lambda)$ associating the most probable value of the n_{cv}-dimensional density field with a given value of the progress parameter λ of the rare event. In the limit in which $k_B T$ is lower than the typical variations of the free energy, the MFEP $\zeta(\lambda)$ is identified by the condition (34)

$$\left[M(\zeta(\lambda)) \nabla_\zeta F(\zeta(\lambda)) \right]^\perp = 0, \qquad \textbf{[S2]}$$

where $(M)_{ij} \equiv \langle \nabla_{\boldsymbol{r}} \theta_i \cdot \nabla_{\boldsymbol{r}} \theta_j \rangle$, with $\boldsymbol{r} = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n_w}\}$, are the components of the metric matrix M accounting for the change in variables and the reduction in dimensionality associated with expressing the free energy as a function of the n_{cv} CVs (for a thorough derivation, see ref. 34). $\nabla_\zeta F$ is the vector containing the derivatives of the free energy with respect to ζ, and $\perp$ stands for the component of the n_{cv}-dimensional vector which is perpendicular to the MFEP.

The actual algorithm used for computing the MFEP, i.e., a path obeying condition (2), consists in finding the steady-state solution of the first-order pseudodynamics,

$$\frac{\partial \zeta(\lambda, t)}{\partial t} = -\left[M(\zeta(\lambda, t)) \nabla_\zeta F(\zeta(\lambda, t)) \right]^\perp, \qquad \textbf{[S3]}$$

converging for $t \to \infty$ to Eq. **S2**.

A tentative path $\zeta(\lambda, t)$—the string—is used as the initial guess in Eq. **S3**. Here the string is initialized from the path followed by the system during a barrierless transition happening at high pressure $\Delta P \approx \Delta P_{int}^{sp}$. At steady state, the string converges to the solution of Eq. **S2**, i.e., to the MFEP. For our system, the string is discretized in 64 points, each of which is an independent replica of the system—a nanopore with a different filling level.

Restrained MD (RMD) (52) was used to estimate accurately the relevant quantities in Eq. **S3**, i.e., the mean force $\nabla_\zeta F$

and the metric matrix M in correspondence to the 64 points along the string. RMD consists in adding a biasing potential $U_k(\mathbf{r}) \equiv \sum_i k/2(\theta_i(\mathbf{r}) - z_i)^2$ with $i = 1, \ldots, n_{cv}$ to the physical one, $U(\mathbf{r})$; this bias enables the system to explore regions of the phase space close to a specific realization of the CVs, including regions with very low probability. It can be demonstrated (34, 52) that, within the biased MD, the following time averages yield an approximation to the desired quantities:

$$\frac{\partial F}{\partial z_i} = -\lim_{\tau \to \infty} \frac{k}{\tau} \int_0^\tau (\theta_i(\mathbf{r}(t)) - z_i)\, dt, \quad \text{[S4]}$$

$$(M(z))_{i,j} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau (\nabla_\mathbf{r}\theta_i \cdot \nabla_\mathbf{r}\theta_j)dt, \quad \text{[S5]}$$

where the components of the gradient and of the metric matrix have been used. Strictly speaking, the equality holds only in the limit $k \to \infty$.

The 64 RMD simulations were run for each evolution step of the pseudodynamics (3) for about 300,000 MD steps, each of 1 fs. The string is then evolved according to Eq. **S3** without projecting on the orthogonal subspace. Subsequently, the points along the string are reparametrized to preserve the same distance between them. This latter step is equivalent to, but computationally more convenient than, solving directly Eq. **S3** with an explicit projection step (69). After reparametrization, new initial conditions are available for simulating another set of 64 replicas. The procedure was repeated until convergence, which took about 20 steps of the pseudodynamics. Overall, the string calculations required *ca.* $64 \times 20 = 1,280$ independent MD simulations of a nanopore.

Intrusion/Extrusion Experiments. A slightly smaller system was used to perform a thought intrusion/extrusion experiment, in which the pressure was first increased until complete intrusion and then decreased until extrusion. The objective of these simulations was to reproduce the experimental intrusion/extrusion curves (5) and not to reconstruct free energy and the mechanism as in the more expensive string simulations reported in *Results*. (In all MD simulations, the somewhat idealized case of a single pore, or a coordinate array of pores, is simulated.) In these in silico experiments, the system was equilibrated at the desired pressure, which was subsequently varied in steps by changing the force applied on the pistons. This variation was achieved gradually via ramps of ca. 400,000 steps. The system was then allowed enough time to equilibrate, typically around 2 million MD steps of 1.5 fs (Fig. S7). Pressure was monitored by probing the stress tensor computed in two symmetrical boxes on the two sides of the nanoporous matrix, far enough from the solid walls to be considered a good estimator of the liquid pressure. The observable P in Fig. S7 was computed from the mean diagonal stress tensor terms in the two directions parallel to the nanoporous matrix. To smooth out the highly fluctuating instantaneous stress tensor, a moving average was performed in time with a window of *ca.* 400,000 steps.

Calculation of Material Properties. The interaction between the solid and the oxygen atoms is of the Lennard-Jones type,

$$V(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - c\left(\frac{\sigma}{r_{ij}}\right)^6 \right],$$

where c tunes the attractive part of the potential to render materials that span from hydrophobic to hydrophilic behavior. Hydrogens do not interact with the solid.

In macroscopic terms, we characterized these interactions via the contact angle θ_Y; see Fig. S1. The angle was estimated by simulating a cylindrical droplet over a flat surface made of the same solid material from which the nanopores are excavated; this procedure avoids contributions of the line tension to the contact angle. The average density field is then computed, allowing us to identify density isolines. A least-squares regression analysis allowed us to identify the contact angle corresponding to each value of c. In the final setup, we used $\theta_Y = 119°$. We remark that this value was computed for an atomically smooth surface as would be measured in a (macroscopic) sessile drop experiment.

The value of the LV surface tension of TIP4P/2005 water was taken from the literature (37): $\gamma_{lv} = 0.069$ N/m.

Confinement Effects in the Pore. To study the effect of confining the liquid inside the pore discussed in the main text, we performed density averages in the azimuthal direction. The resulting radial density profile of the water confined inside the pore is presented in Fig. S6. The effect of confinement due to the pore size is evident: The bulk liquid density $\rho_l = 0.0333$ Å^{-3} is recovered in the middle for the pore with $D = 2.6$ nm, while the density profile is much more peaked in the $D = 1.2$ nm pore. Only one layer is observed in the latter case. In Fig. S6, we also show the radial dependence of the wall potential felt by a single probe oxygen inside the nanopore. This value corresponds to the potential at the section in the center of the pore.

Postprocessing of String Method Results. Density fields were computed along the final step of the string evolution (the MFEP) by averaging over the atomistic configurations. Density isosurfaces were extracted from such a discretized density field using a marching-cube-like algorithm (70). This procedure allowed us to visualize the liquid density isosurfaces which are shown in Movie S1. These isosurfaces were also used to extract all the relevant quantities (LV and LS areas and triple line length; see Fig. S5) to be plugged into the modified CNT equations.

In particular, from the triangulated mesh constituting the isosurface, it was possible to discriminate LS and LV interfaces based on the distance from the wall of the centroid of the triangles. This also resulted in the definition of the separatrix curve at the intersection of the LV and LS meshes, i.e., the triple line. To obtain a robust indicator for the length of the triple line, we used the length of the line connecting the middle points of the triangle's edges; see Fig. S5. The procedure was repeated for each configuration along the MFEP.
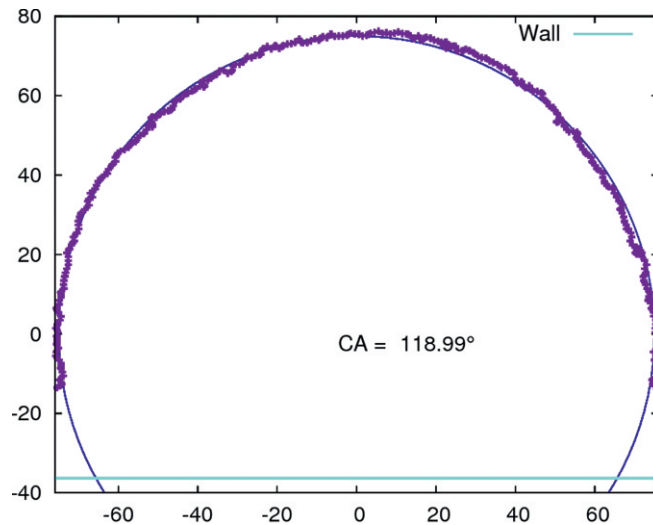
Fig. S1. Calculations of the Young contact angle were performed on an atomically flat substrate with a cylindrical droplet deposited on top. The material is the same as that from which the nanopore is excavated. The straight line at the bottom indicates the position of the last wall atoms.
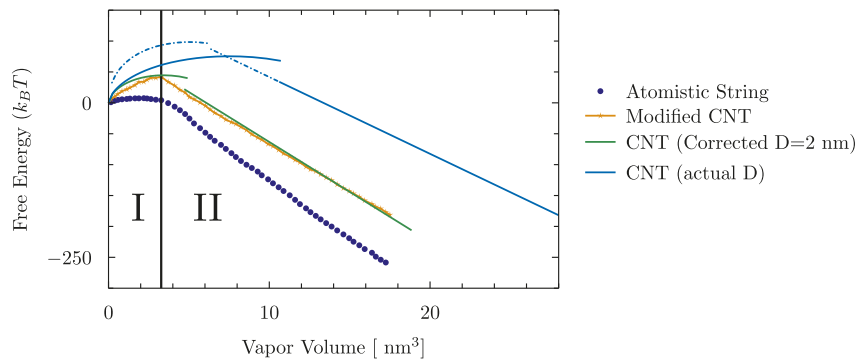

Fig. S2. Comparison of free-energy profiles deriving from atomistic simulations (blue dots), CNT Surface Evolver calculations (light blue line), and Modified CNT (green and orange lines). In orange, we plot the Modified CNT data computed via Eq. **2** using the quantities (LS area and LV area) derived from the postprocessing of atomistic data. The light blue lines show the standard CNT solution obtained via the Surface Evolver for the cases shown in Fig. S3: two symmetric menisci (straight line on the right), an asymmetric bubble (solid curved line), and a vapor annulus (dot-dashed curved line); the solution devised in ref. 4 is made up by the two solid lines; these calculations refer to a pore with $D = 2.6$ nm, i.e., the nominal pore diameter. In green, we show Surface Evolver calculations with a smaller effective diameter ($D = 2.06$ nm) which accounts for the excluded volume at the walls.


Fig. S3. Using the Surface Evolver, three morphologically different bubble configurations were studied: a droplet, an annulus, and a bubble enclosed between two facing menisci. The CNT path proposed in refs. 4 and 5 involves a morphological transition from the droplet to the two-menisci configuration. The annulus configuration is never associated with the lowest free energy, for all values of the bubble volume.
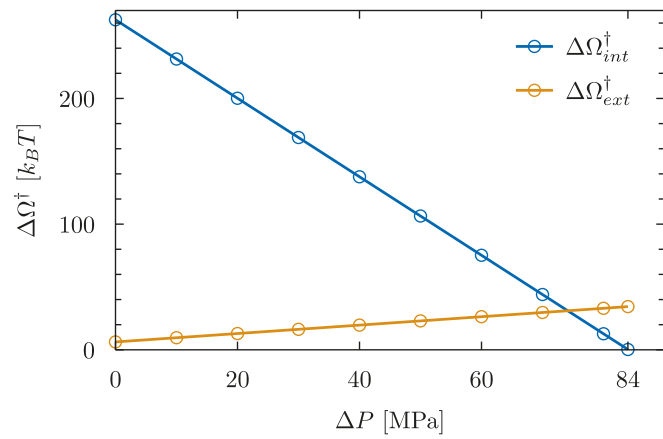
**Fig. S4.** Intrusion (blue) and extrusion (orange) free-energy barriers as a function of the pressure ΔP.
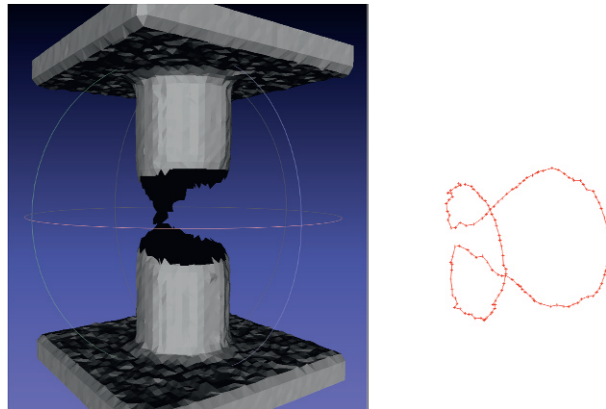
Fig. S5. (*Left*) Example of the extraction procedure of LV and LS surfaces from the liquid isodensity surface $\rho = 0.5\rho_l$. (*Right*) The corresponding calculation of the length of the triple line.
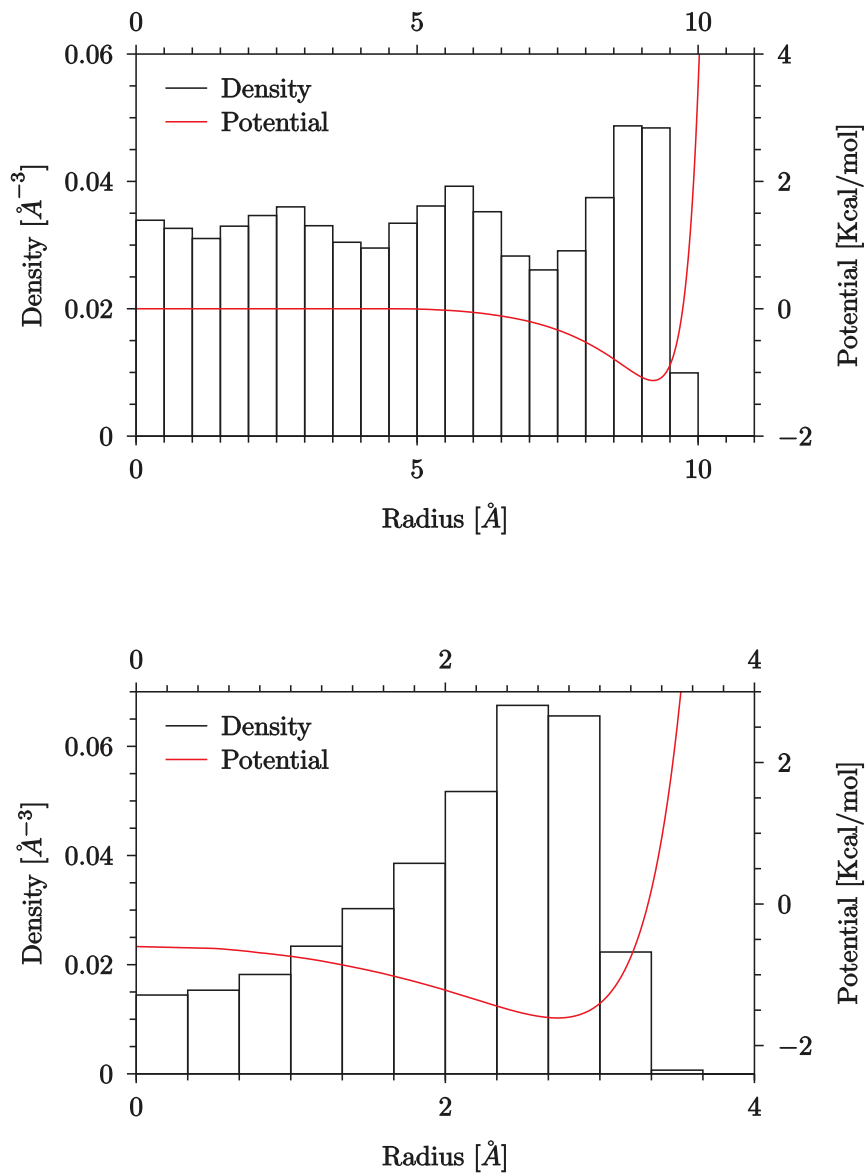
Fig. S6. (*Upper*) Radial density profile for the $D = 2.6$ nm pore, measured as the number of water molecules per unit volume. For reference, the bulk liquid density of water at the same conditions is $\rho_l = 0.0333$ Å^{-3}. (*Lower*) The same observable is plotted for the $D = 1.2$ nm pore. The wall potential felt by a single oxygen atom at the middle section of the pore is also shown in the plots (red lines).
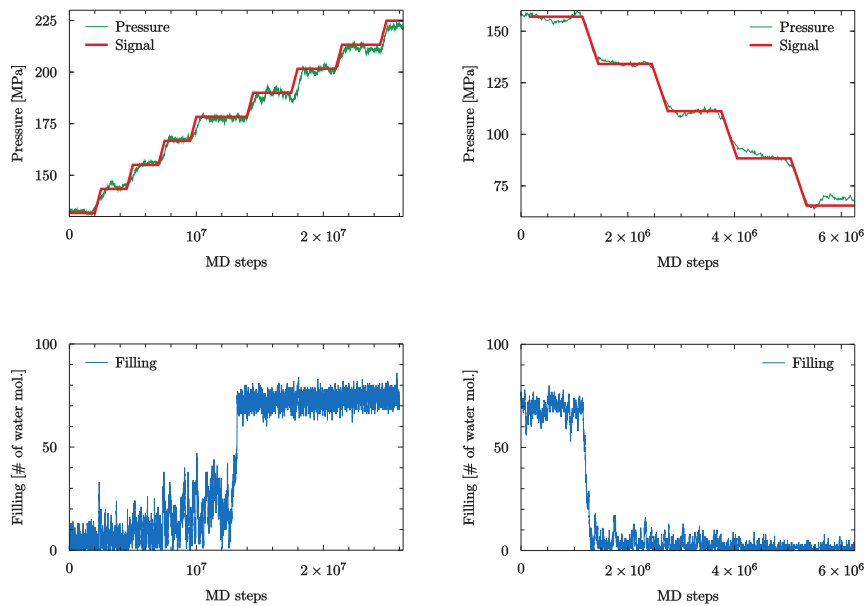
Fig. S7. (*Left*) Intrusion MD experiment for a nanopore with $D = 1.2$ nm. The imposed (red) and measured (green) pressures are plotted as a function of the simulated time. The pressure signal in green is obtained via a running average with a 400,000-steps window. The stress tensor was probed inside two independent boxes located in the bulk water, far from the solid walls. (*Right*) Extrusion curves for the same system.
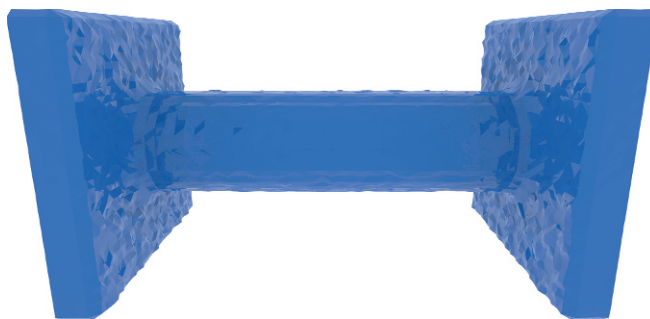
Movie S1. Transition path for the extrusion process in a nanopore. The shown fluid density isosurface corresponds to $\rho = 0.5\rho_l$, where ρ_l is the bulk liquid density. Since the string method implies that the process is quasi-static, the same movie, played backward, corresponds to the most probable intrusion process.

Movie S1