# Direct calculation of ice homogeneous nucleation rate for a molecular model of water

**Amir Haji-Akbari and Pablo G. Debenedetti**[1]

Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544

Ice formation is ubiquitous in nature, with important consequences in a variety of environments, including biological cells, soil, aircraft, transportation infrastructure, and atmospheric clouds. However, its intrinsic kinetics and microscopic mechanism are difficult to discern with current experiments. Molecular simulations of ice nucleation are also challenging, and direct rate calculations have only been performed for coarse-grained models of water. For molecular models, only indirect estimates have been obtained, e.g., by assuming the validity of classical nucleation theory. We use a path sampling approach to perform, to our knowledge, the first direct rate calculation of homogeneous nucleation of ice in a molecular model of water. We use TIP4P/Ice, the most accurate among existing molecular models for studying ice polymorphs. By using a novel topological approach to distinguish different polymorphs, we are able to identify a freezing mechanism that involves a competition between cubic and hexagonal ice in the early stages of nucleation. In this competition, the cubic polymorph takes over because the addition of new topological structural motifs consistent with cubic ice leads to the formation of more compact crystallites. This is not true for topological hexagonal motifs, which give rise to elongated crystallites that are not able to grow. This leads to transition states that are rich in cubic ice, and not the thermodynamically stable hexagonal polymorph. This mechanism provides a molecular explanation for the earlier experimental and computational observations of the preference for cubic ice in the literature.

nucleation | ice | molecular simulations | statistical mechanics | water

Ice nucleation affects the behavior of many systems (1–6). For example, the formation of ice crystals inside the cytoplasm can damage living cells (1). The amount of ice in a cloud determines both its light-absorbing properties (5) and its precipitation propensity (6), and is therefore an important input parameter in many meteorological models (7, 8). However, current experiments are incapable of uncovering the kinetics and the molecular mechanism of freezing due to their limited spatiotemporal resolution. The ice that nucleates homogeneously in the atmosphere and vapor chamber experiments is predominantly comprised of the cubic-rich stacking-disordered polymorph, not the thermodynamically stable hexagonal polymorph (9, 10). This observation has been rationalized invoking the Ostwald step rule (11). However, the molecular origin of this preference is unknown, due to the limited spatiotemporal resolution of existing experimental techniques. Furthermore, experimental measurements of nucleation rates are only practical over narrow ranges of temperatures (12), with any extrapolation being prone to large uncertainties.

Computer simulations are attractive alternatives in this quest, as they make it possible to obtain, at any given thermodynamic condition, a statistically representative sample of nucleation events that can then be used to estimate the rates and identify the mechanism of nucleation. This, however, has only been achieved (13–15) for coarse-grained representations of water, such as the monoatomic water (mW) model (16). For the more realistic molecular force fields, all of the existing studies have relied either on launching a few-microseconds-long molecular dynamics (MD)

trajectories (17, 18), or on applying external fields (19), or biasing potentials along prechosen reaction coordinates (20) to drive nucleation, and the generation of statistically representative nucleation trajectories that can allow direct and accurate rate predictions has so far been beyond reach.

In this work, we achieve this goal in a system of 4,096 water molecules at 230 K and 1 bar by introducing a novel coarse-graining modification to the path sampling method known as forward-flux sampling (FFS) (21). In the FFS approach, the nucleation process is sampled in stages defined by an order parameter, $\lambda$. In crystallization studies, $\lambda$ is typically chosen as the size of the largest crystalline nucleus in the system (13–15). Individual molecules are labeled as solid- or liquid-like based on the Steinhardt order parameters (22), and the neighboring solid-like molecules are connected to form a cluster (for further details, refer to *SI Text*, Table S1, and Fig. S1). The cumulative probability of growing a crystallite with $\lambda$ molecules is then computed from the success probabilities at individual stages (e.g., Fig. 1). If a sufficiently large number of trajectories are sampled at each stage, the nucleation mechanism can be accurately determined by inspecting the ensemble of pseudotrajectories that connect the liquid and crystalline basins. We use the term "pseudotrajectory" as, during FFS, all velocities are randomized at any given milestone.

In conventional FFS, the underlying MD trajectories are monitored as frequently as possible, usually every single MD step. In the TIP4P/Ice system, however, this approach is unsuccessful, as the cumulative growth probability never converges (plateaus) and instead plummets unphysically (Fig. S2A). Because of the five-orders-of-magnitude separation between the structural relaxation time, $\tau_r$ (Fig. 2A), and the sampling time, $\tau_s$, the

**Significance**

Despite the importance of ice nucleation in a wide range of natural and technological systems, it is not trivial to predict its kinetics or unravel its mechanism experimentally. It is also challenging to study the stochastic nature of ice nucleation in computer simulations. The direct calculation of the nucleation rate for realistic molecular models of water is commonly regarded as one of the most challenging open problems in computational studies of water. By applying a novel coarse-graining modification to the forward-flux sampling technique, we perform the first direct rate calculation for TIP4P/Ice, one of the best existing molecular models of water. We also identify the molecular origin of the preference for the cubic polymorph during the homogeneous nucleation process.
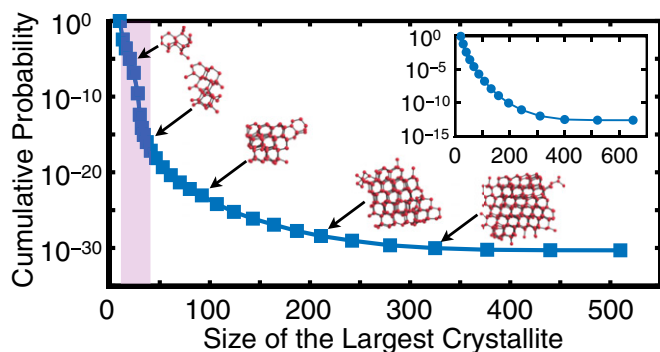
**Fig. 1.** Cumulative transition probability vs. size of the largest crystalline nucleus in the TIP4P/Ice system at 230 K and 1 bar. The inflection region is shown in shaded purple. Several representative crystallites are also depicted. The cumulative probability curve for the LJ system simulated at $k_B T/\epsilon = 0.82$ and $\rho\sigma^3/\epsilon = 5.68$ is shown in the *Inset* with $\epsilon$ and $\sigma$ the LJ energy and size parameters. No inflection region is observed in the LJ system.

high-frequency fluctuations in $\lambda(t)$ do not reflect physically relevant structural transformations. We therefore filter such high-frequency fluctuations by computing the order parameter along MD trajectories less frequently. We choose $\tau_s = 1$ ps, which is still around three orders of magnitude smaller than the hydrogen bond relaxation time (23) (Fig. 2C). By decreasing the separation between $\tau_s$ and $\tau_r$, the FFS calculation converges and the cumulative probability eventually plateaus (Fig. 1). The computed nucleation rate is $\log_{10} R = 5.9299 \pm 0.6538 - R$ in nucleation events per cubic meter per second. This implies, statistically, one nucleation event per $9 \times 10^{18}$ s in the 4,096-molecule system considered in this work, which has an average volume of ∼125 nm³. Note the astronomical separation of time scales between structural relaxation ($\tau_r = 0.6$ ns) and ice nucleation. This rate is placed in the context of earlier experimental estimates (12, 24) below (see *Comparison with Experimental Rate Measurements*). We confirm the accuracy of the coarse-grained FFS by observing that the computed crystallization rates in the Lennard–Jones (LJ) system are insensitive to $\tau_s$ if $\tau_s/\tau_r < 10^{-1}$ (Figs. 3 and 2B).

For most materials, the probability of adding a certain number of molecules to a crystallite of $\lambda$ molecules increases with $\lambda$. This leads to a consistent positive curvature in the cumulative probability curve, e.g., in the crystallization of the LJ system (Fig. 1, *Inset*, and Fig. S3A). For water, however, the cumulative probability curve has a pronounced inflection at $\lambda \approx 30$, where the probability of growing an average crystallite decreases significantly with $\lambda$ before rebounding again at larger $\lambda$. The inflection is accompanied by nonmonotonicities in several other mechanical observables. For instance, in the inflection region, the average density increases with $\lambda$ (Fig. 4D), even though there is an overall decrease in density upon crystallization. We observe

similar nonmonotonicities in the longest principal axes (Fig. 4A) and the asphericity (Fig. 4B) of the largest crystallite, as well as the number of five-, six- and seven-member rings in the system (Fig. 4C). The nonmonotonicity in ring size distribution has also been observed in the freezing of ST2, another molecular model of water (25). In the LJ system, however, all of these quantities evolve monotonically from their averages in the liquid to their averages in the crystal (Fig. 4, *Insets*, and Fig. S3). In the coarse-grained mW system, this inflection is present, but is very mild, and the nonmonotonicities are much weaker (Fig. S4).

To understand the origin of this inflection, we examine all of the configurations in the shaded purple regions of Figs. 1 and 4, and identify those that survive the inflection region by giving rise to a progeny at $\lambda = 41$. Visual inspection of these configurations reveals an abundance of double-diamond cages (DDCs) in their largest crystallites. DDCs (Fig. 5A) are the basic building blocks of cubic ice ($I_c$), and are topologically identical to the carbon backbone of the polycyclic alkane diamantane (26). The largest crystallites of the vanishing configurations, however, are rich in hexagonal cages (HCs) (Fig. 5B), the basic building blocks of hexagonal ice ($I_h$). We then use a topological criterion to detect DDCs and HCs (see *SI Text*). In this approach, all primitive hexagonal rings in the nearest-neighbor network are identified, and DDCs and HCs are detected based on the connectivity of the neighboring hexagonal rings (see *SI Text* for further details). We identify several isolated cages even in the supercooled liquid. Due to their distorted geometries, however, such cages can only be detected topologically, and not through conventional order parameters such as $q_3$ (13). Similar to the crystallites that are clusters of neighboring molecules with local solid-like environments (see *SI Text*), the cages that share molecules can also be clustered together to define interconnected DDC/HC networks. With their constituent cages detected topologically, such networks can contain both solid- and liquid-like molecules. We observe that almost all of the molecules of the largest crystallites participate in DDC/HC networks. This is consistent with earlier experimental and computational observations (10, 27) that the ice that nucleates from supercooled water is a stacking-disordered mixture of both $I_c$ and $I_h$ polymorphs.

Consistent with our visual observation, a stark difference exists between the DDC makeup of the surviving and vanishing configurations. In the surviving configurations, the water molecules of the largest crystallite are more likely to participate in DDCs than in HCs (Fig. 5 C and D), making the corresponding crystallites more cubic than the average. Such cubic-rich configurations are scarce at the beginning and only grow in number toward the end of the inflection region. Conversely, the majority of configurations, which are HC rich, become extinct toward the end of the inflection region. This preference can be explained by comparing the geometric features of the HC-rich and DDC-rich crystallites. Although the DDC-rich crystallites are comparatively uniform in shape (Fig. 5H), the HC-rich crystallites are
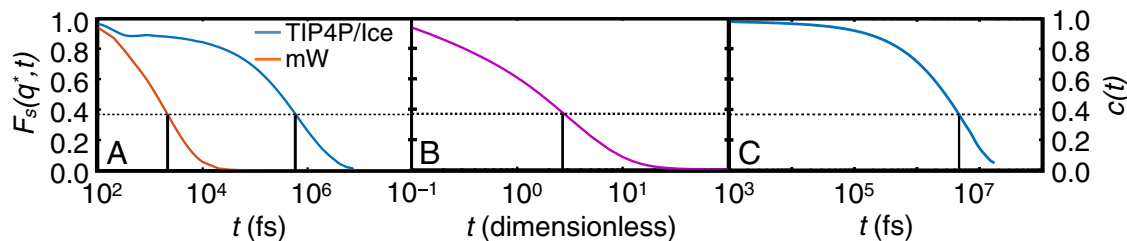


**Fig. 2.** Structural relaxation in the supercooled liquid. Self-intermediate scattering functions computed from MD simulations of (A) the TIP4P/Ice (blue) and the mW (orange) systems at 230 K and 1 bar and (B) the LJ system at $k_B T/\epsilon = 0.82$ and $\rho_N \sigma^3 = 0.974$. In each case, $q^*$ is in close correspondence with the first peak of $S(q)$, the structure factor, in the corresponding system. The structural relaxation time, $\tau_r$, is defined as the time at which $F_s(q^*, t) = 1/e$. (C) The $c(t)$, the hydrogen bond correlation function, computed in *NpT* simulations of a system of 216 TIP4P/Ice molecules at 230 K and 1 bar; $\tau_h$ is defined as $c(\tau_h) = 1/e$.
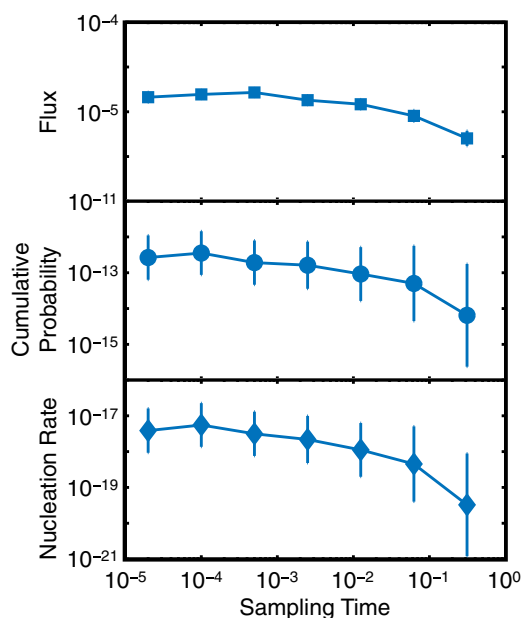
**Fig. 3.** Effect of $\tau_s$, the sampling time, on fluxes, cumulative probabilities, and nucleation rates computed from a series of FFS calculations conducted for a system of 4,096 LJ atoms at $k_B T/\epsilon = 0.82$ and $p\sigma^3/\epsilon = 5.68$. Divergence only occurs when $\tau_s$ becomes comparable to $\tau_r$. Computed quantities are insensitive to $\tau_s$ for $\tau_s \ll \tau_r$. All quantities are in the LJ dimensionless units.

nucleation has been observed in previous studies of ice formation in different water models (27, 29, 30). To the best of our knowledge however, the molecular origin of this preference had not been identified before this work. Indeed, the nonmonotonicities in the shape and asphericity of the largest crystallite almost disappear when only the surviving configurations are considered (Fig. 5 *E* and *F*). A similar correlation exists between the DDC makeup of a configuration and its density and ring size distribution (Fig. S6).

Fig. 6 depicts the fate of the cubic-rich crystallites that survive the inflection region. Due to the thermodynamic stability of $I_h$ relative to $I_c$, one expects the surviving cubic-rich crystallites to eventually transform into $I_h$. We observe no such transformation during the nucleation process, and the crystallites retain their high DDC content (Fig. 6*A*) even after they are postcritical (Fig. 6*G*). (For a discussion of criticality, see *SI Text* and Fig. S7*B*.) This suggests the need for caution in the interpretation of earlier indirect calculations of nucleation rate (17) in which the critical nuclei are assumed to be exclusively hexagonal. We also observe no tendency for the hexagonal polymorph to prefer the core of the crystallite. This is in contrast to the traditional picture of nucleation in which the more thermodynamically stable phase concentrates at the core, with a shell of the less stable phase shielding it from the liquid (31). Instead, we observe a large number of exposed HCs at the surface (Fig. 6 *B*–*G*), with attrition tendencies similar to the HCs in the inflection region (e.g., the HC appendages in Fig. 6*D* and the large prismatic-to-basal ratio in Fig. S5*F*). The propensity to grow more cubic stacks even after the inflection region is consistent with the proposed mechanism, as the addition of new HCs to a large crystallite is more likely to lead to chain-like appendages at the surface, henceforth making it less stable than an equal-sized crystallite grown via the addition of DDCs. Indeed, the propensity to form thicker cubic stacks has been observed in the growth and consolidation of postcritical crystallites in the growth-limited freezing of the mW system (27).
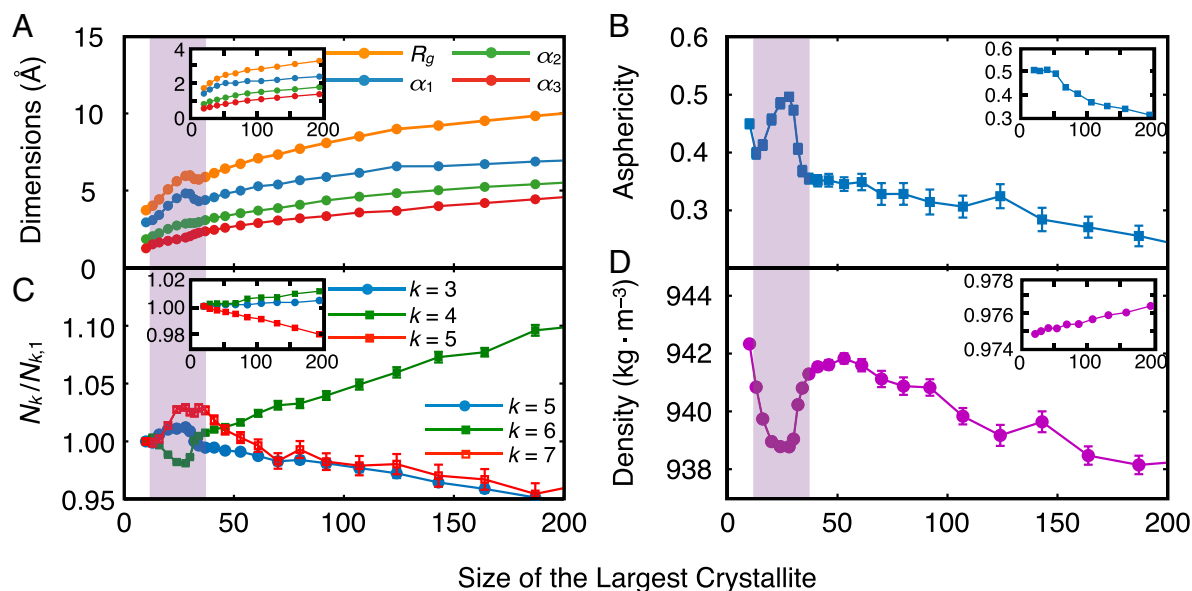
more aspherical (Fig. 5*G*), and therefore less likely to grow and survive the inflection region. This higher asphericity arises from the preferential addition of new HCs to the prismatic faces of the existing HCs, as evident in the abrupt increase in the ratio of prismatic to basal HC–HC connections in the inflection region (Fig. S5*F*). This is qualitatively consistent with earlier observations showing that the growth of bulk $I_h$ is faster along its prismatic plane (28). The preference for $I_c$ in the early stages of



**Fig. 4.** Nonmonotonicities in average mechanical observables for the configurations obtained from the FFS calculation. *Insets* correspond to the FFS calculation in the LJ system. (*A*) Radius of gyration ($R_g$), principal axes ($\alpha_1 \geq \alpha_2 \geq \alpha_3$), and (*B*) asphericity of the largest crystallite. (*C*) Ring statistics and (*D*) density of the system. $N_k(\lambda)$ corresponds to the average number of $k$-member rings at $\lambda$, with $N_{k,1} = N_k(\lambda_1)$. For water, five-, six-, and seven-member rings are enumerated, whereas for the LJ system, three-, four-, and five-member rings are enumerated. The shaded purple region corresponds to the inflection region. All quantities are in dimensionless units for the LJ system.
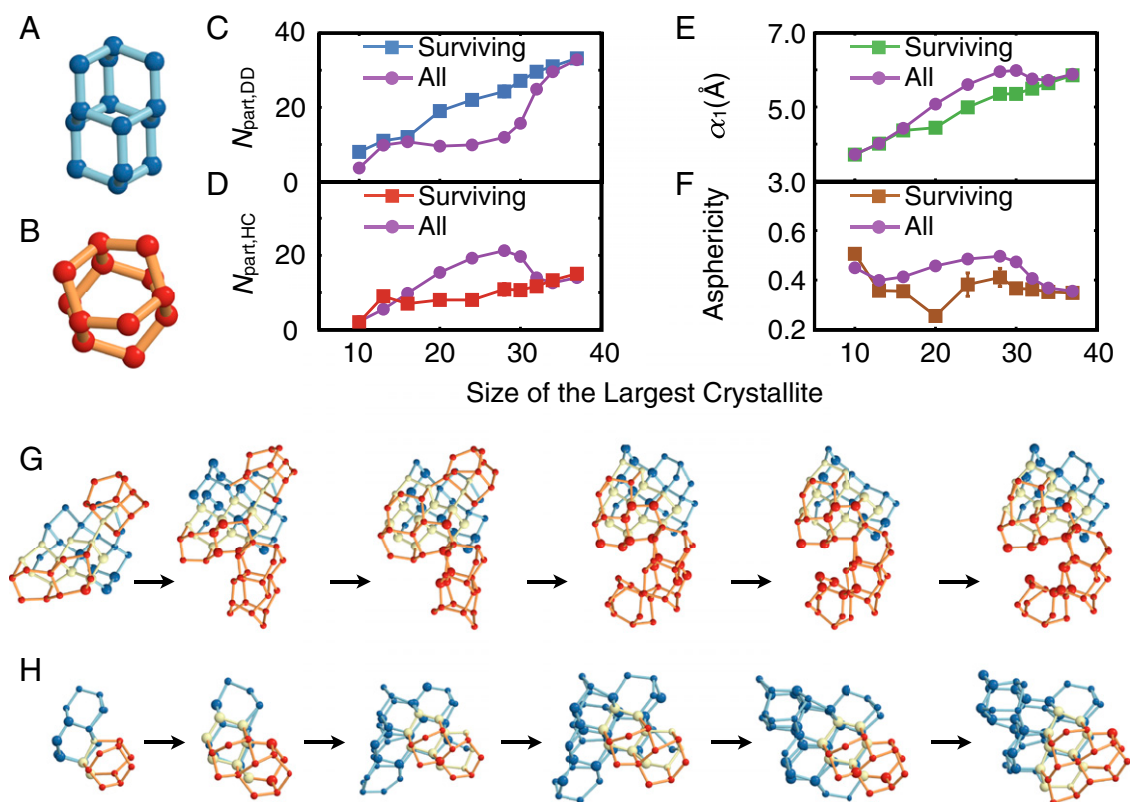
**Fig. 5.** Competition between cubic and hexagonal ice in the inflection region. (*A*) DDC and (*B*) HCs. (*C* and *D*) Number of water molecules in the largest crystallite that participate in (*C*) a DDC and (*D*) an HC. (*E*) The longest principal axis and (*F*) asphericity of the largest crystallite. (*G*) A pseudotrajectory that does not survive the inflection region. DDC and HC shown in blue and red, respectively. Yellow particles belong to both a DDC and an HC. Note the abundance of HCs. (*H*) A pseudotrajectory that survives the inflection region. Note the abundance of DDCs. Molecules that are part of the largest crystallite (based on $q_6$) are shown larger than liquid-like molecules that participate in the topological DDC/HC network that encompasses the largest crystallite.

## Comparison with Experimental Rate Measurements

As mentioned above, experimental measurements of nucleation rate are only practical over a narrow range of thermodynamic conditions. This is because of the fundamental limitation of existing experimental techniques, which are based on probing the temporal evolution of the number of freezing events that are detected in a small population of supercooled microdroplets (32). Therefore, the nucleation rates that can be measured from the existing experimental techniques can span a few orders of magnitude only, confining the range of thermodynamic conditions over which nucleation rates are measurable. For temperatures that are outside this range, the nucleation rate is either so small that none of the microdroplets would freeze during the timescale of the experiment or so large that all droplets would freeze immediately. For droplets as small as a few micrometers in diameter, nucleation rates have been measured for temperatures as low as 234 K, which corresponds to a supercooling of 39 K, 3 K smaller than the supercooling considered in this work. (The melting temperature of the TIP4P/Ice model is 272.2 K (33) vs. the experimental melting temperature of 273 K. Henceforth, our temperature of 230 K corresponds to a supercooling of 42 K.) Therefore, no direct comparison can be made between our computed nucleation rate and any actual experimental measurement, without extrapolating to lower temperatures. These extrapolations are typically based on classical nucleation theory, and are prone to large uncertainties, leading to large variations in the extrapolated nucleation rates. In particular, such extrapolations fail to take into account the transition to the transport-controlled nucleation at low temperatures, which is responsible for the appearance of a maximum in the nucleation rate with respect to temperature. For

real water, the temperature of maximum crystallization rate has been estimated to be ~225 K (34), which is very close to the temperature considered in this work. Such extrapolations yield a wide range of nucleation rates at a supercooling of 42 K, from $10^{18} \mathrm{m}^{-3} \cdot \mathrm{s}^{-1}$ in refs 12 and 35 to $10^{24} \mathrm{m}^{-3} \cdot \mathrm{s}^{-1}$ in ref. 36. Another potential source of error, which can lead to a systematic overestimation of rates at lower temperatures, is the possibility of surface-dominated nucleation in smaller droplets that are typically used for rate measurements at lower temperatures (37).

Thanks to the superior temporal resolution of new experimental techniques, direct measurements of nucleation rates at larger supercoolings will be possible in the near future. One such technique is the femtosecond X-ray laser pulsing that was recently used by Sellberg et al. (24) to probe the structural transformation of a population of evaporatively cooled microdroplets of supercooled water. Although no nucleation rates are reported in their work, it is possible to obtain an approximate estimate using figure 2 of ref. 24, which depicts the temporal profiles of the temperature and the ice fraction of evaporatively cooled 12-μm droplets. The first frozen droplets are detected approximately four milliseconds after they enter the chamber and when they reach a temperature of 229 K. The average freezing time of 4 ms can be used to obtain an upper bound of $J_v \approx 2.7631 \times 10^{17} \mathrm{m}^{-3} \cdot \mathrm{s}^{-1}$ for the homogeneous nucleation rate at the supercooling of 42 K. This is around 11 orders of magnitude larger than the rate computed in this work. As we will discuss below, however, this discrepancy is reasonable considering the sensitivity of the nucleation rate to different thermodynamic features of the system. According to the classical nucleation theory, the nucleation rate is proportional to $\exp[-\Delta G_c/k_B T]$ with $\Delta G_c$, the
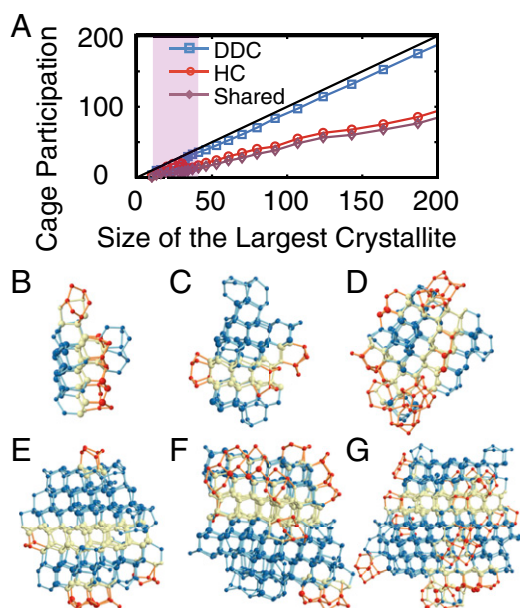
**Fig. 6.** Nucleation beyond the inflection region. (*A*) Average cage participation of the molecules in the largest crystallite. The solid black line has a slope of unity. The molecules that participate in a DDC (or HC) are included in the corresponding count even if they also participate in a neighboring cage of the other type. The overwhelming majority of molecules are at least part of a DDC, whereas very few molecules are only a part of an HC. (*B–G*) Several representative configurations obtained at different milestones after the inflection region. *B–E* are precritical, *F* is critical, and *G* is postcritical. Molecules that are a part of a DDC, an HC, or both are depicted in dark blue, dark red, and light yellow, respectively. Here, we use the same size convention used in Fig. 5.

free-energy barrier associated with the formation of a critical nucleus, given by

$$\Delta G_c = \frac{16\pi\gamma^3}{3\rho_s^2|\Delta\mu|^2}.$$  [1]

Here $\gamma$ is the solid−liquid surface tension, $\rho_s$ is the number density of the solid, and $\Delta\mu$ is the free-energy difference between the crystalline and liquid phases. The exponential dependence of the nucleation rate on $\Delta G_c$, and the sensitivity of $\Delta G_c$ to $\gamma$ and $\Delta\mu$ implies that only a slight deviation of any of these quantities from the experimental value can shift the nucleation rate by several orders of magnitude. Both these quantities are difficult to measure at large supercoolings, mostly because of the difficulty of stabilizing supercooled water at such low temperatures.

**Free-Energy Difference.** If $\Delta H_f$, the latent heat of fusion, is not a strong function of temperature, $\Delta\mu$ can be approximated as $\Delta\mu \approx \Delta H_f(T_f - T)/T_f$. This approach, which yields a value of $\Delta\mu \approx 0.2215$ kcal·mol$^{-1}$ at a supercooling of 42 K, is, however,

**Table 1. Numerical correlations used for fitting the experimental heat capacity measurements of refs. 38 and 39**

| Phase | Correlation |
|---|---|
| Supercooled water (39) | $C_p(T) = 4 \times 10^{15}T^{-5.824} + 0.7131T - 202$ |
| $I_h$ (38) | $C_p(T) = 0.032T + 0.3252$ |

Units are in calories per mole per Kelvin. For $I_h$, we use a linear fit, whereas, for supercooled water, we use a combination of a power law and a linear fit. The actual experimental data for supercooled water are for $T \geq 236$ K. We thus use the numerical fit provided above to extrapolate $C_p$ at 231 K $\leq T \leq$ 236 K.

not very accurate for water due to its heat capacity anomaly. To obtain a more accurate estimate, we take the heat capacity measurements for $I_h$ (38) and supercooled water (39) (Table 1), and use thermodynamic integration to obtain a more accurate estimate of $\Delta\mu_{\text{exp}} = 0.1855$ kcal·mol$^{-1}$. Similarly, we use MD simulations in the isothermal isobaric ($NpT$) ensemble to compute enthalpies of $I_h$ and supercooled water at 230 K$\leq T \leq$ 272 K and use those enthalpies to compute $\Delta\mu$ using thermodynamic integration. We obtain a value of $\Delta\mu_{\text{TIP4P/Ice}} = 0.147$ kcal·mol$^{-1}$ for the TIP4P/Ice system, which is around 20% smaller than $\Delta\mu_{\text{exp}}$. This discrepancy alone can lead to an overestimation of the nucleation barrier by as much as 60% if everything else is identical. To be more quantitative, the classical nucleation theory predicts a nucleation barrier of $\Delta G_c = (1/2)|\Delta\mu|N_c \approx 51 \; k_BT$ for the TIP4P/Ice system at $T = 230$ K. However, if we use $\Delta\mu_{\text{exp}}$ instead of $\Delta\mu_{\text{TIP4P/Ice}}$, and $\rho_{s,\text{exp}} = 0.922$g·cm$^{-3}$ (40) instead of $\rho_{s,\text{TIP4P/Ice}} = 0.908$g·cm$^{-3}$ (obtained from $NpT$ MD simulation of $I_h$ at 230 K and 1 bar) in Eq. **1**, we obtain a barrier of ~31 $k_BT$, which corresponds to an increase in the nucleation rate by eight to nine orders of magnitude. This is very close to the discrepancy between our calculation and the experimental estimates of rate.

**Surface Tension.** At temperatures below $T_f$, the supercooled water that is in contact with ice is not stable and will immediately freeze. This makes experimental measurements of $\gamma$ in the supercooled regime extremely challenging. Therefore, $\gamma$ is typically estimated indirectly from the nucleation data assuming the validity of the classical nucleation theory. Consequently, there is a large variation in the reported estimates of $\gamma$ for supercooled water that span between 25 mJ·m$^{-2}$ and 35 mJ·m$^{-2}$ (41). Similarly, it is very challenging to compute $\gamma$ directly from molecular simulations at $T < T_f$, and all of the existing estimates are obtained from nucleation calculations (13, 17, 42). The existing direct calculations have all been performed at coexistence conditions (43, 44). The computed numbers cover even a wider range, from 20.4 mJ·m$^{-2}$ in ref. 17 to 35 mJ·m$^{-2}$ in ref. 42. This large variability underscores the sensitivity of the computed value to the particulars of the water model, and to the thermodynamic conditions at which the calculation has been made. In light of the mechanism that is proposed for freezing in this work, the problem of determining $\gamma$ is further compounded by the stacking disorder nature of the critical nucleus. Considering the cubic dependence of the nucleation barrier on $\gamma$, even the slightest deviation from the experimental value can shift the nucleation rate by several orders of magnitude. For instance, a 7% deviation can change the nucleation barrier by as much as 22%, which can shift the nucleation rate by several orders of magnitude.

Overall, the existing classical models of water inevitably predict certain thermodynamic properties of water to be at variance with experiments by a significant margin, and a model that predicts all thermodynamic properties accurately has yet to be developed (45). Therefore, the agreement between the orders of magnitude of the computational estimate of the nucleation rate in a classical model of water, like TIP4P/Ice, and the corresponding experimental value is difficult to achieve with the existing models due to strong sensitivity of the nucleation rate to the particular thermodynamic features of the water model used (e.g., the free energies of the liquid and the solid, and the liquid−solid surface tension).

**Earlier Computational Studies of Nucleation Rate.** It is inherently problematic to compare computational estimates of nucleation rate obtained for different force fields using different methodologies. This is not only because of the large uncertainties associated with the methods used (e.g., the validity of classical nucleation theory in the seeding technique) but also due to the empirical and approximate nature of the force fields used, which, as shown above, can shift the computed rates by several orders of

magnitude. This difficulty becomes apparent upon observing the spread of the reported computational estimates of the homogeneous ice nucleation rate in the literature. Li et al. (13, 14) and Haji-Akbari et al. (15) used FFS to compute the nucleation rate in the mW system over a wide range of temperatures. Their computed rates are lower than the corresponding experimental values, but are still a few orders of magnitude higher than the rate computed in this work. This cannot only be attributed to the inherently faster dynamics of the mW system but also to the higher $|\Delta\mu|$ of the mW model at deep supercoolings, as depicted in figure 2a of ref. 46. Recently, Sanz and coworkers (17, 46) used the seeding technique to compute the nucleation rates for several water models. This interesting approach assumes the validity of the classical nucleation theory and the precise crystallographic nature of the critical nucleus (e.g., $I_h$). The estimated uncertainties associated with these—and other—assumptions are very large (e.g., error bars in figure 7 of ref. 17). Nevertheless, a comparison at the same reduced conditions suggests that the present rates are lower than the those estimated by Sanz and coworkers. In particular, the rates reported in refs. 17 and 46 are very sensitive to the a priori definition of what constitutes a nucleus, as the size of the critical nucleus is directly used for estimating the nucleation barriers and the nucleation rates. Our approach, however, does not rely on determining the size of the critical nucleus as a prerequisite for computing the nucleation rate.

## Conclusions

In this work, we establish the feasibility of computing the rate of homogeneous ice nucleation for realistic molecular models of water. This is significant considering the difficulties associated with measuring nucleation rates in experiments. However, the computed rates for the TIP4P/Ice model are several orders of magnitude smaller than the experimental estimates at comparable conditions. This discrepancy is attributed to the smaller thermodynamic driving force for the freezing of the TIP4P/Ice system relative to experiment. Nevertheless, the ability to directly compute rates for a molecular model makes it possible, in principle, to parameterize molecular force fields with an eye toward accurate prediction of nucleation rates. In addition, this paves the way for studying the kinetics and mechanism of ice nucleation across a wide range of environments, such as the atmospherically relevant films, droplets, and aerosols. Finally, the coarse-grained FFS used in this work can prove useful in studying disorder−order transitions in other slowly relaxing systems, such as water/gas mixtures, ionic liquids, and macromolecular and biomolecular systems. In addition to being able to compute nucleation rates, we obtain valuable mechanistic information that is not attainable in experiments. In particular, we provide a molecular explanation for the initial formation of cubic-rich ice in homogeneous nucleation of supercooled water.

## Methods

Individual MD simulations are performed using large-scale atomic/molecular massively-parallel simulator (LAMMPS) (47). The size of the largest crystalline nucleus is chosen as the order parameter. The largest crystallites are detected using the Steinhardt $q_6$ order parameter (22) and the chain exclusion algorithm of Reinhardt et al. (48). Technical specifications of the MD simulations and the order parameter can be found in SI Text and Table S1. Rings are detected using the King criteria (49), and DDCs and HCs are identified using a novel topological approach, with the detection algorithms thoroughly mentioned in SI Text.

Rate calculations are performed applying a novel coarse-graining to the forward-flux sampling algorithm (21). The FFS technique is based on sampling the nucleation process in stages, by staging milestones between the liquid and crystalline basins. (See SI Text for further explanation.) The essence of FFS is thus to identify first passage events between the absorbing milestones of each iteration. In principle, this should be done by monitoring all of the time-continuous trajectories originating at any given milestone and by determining the exact times at which they cross any of the two absorbing milestones. In reality, however, these time-continuous trajectories are approximated by solving the discretized versions of the equations of motion. As a result, the order parameter can only be computed as frequently as every single MD step, and any crossings that might occur at intermediate times will be inevitably ignored. Historically, this has been the approach taken in all reported applications of the FFS algorithm, with some authors using larger sampling times (up to a few MD steps) only for convenience (50). In the context of crystallization, however, it is reasonable to argue that fluctuations in the order parameter are only meaningful if they occur at time scales that are not significantly smaller than the structural relaxation time, $\tau_r$, or the hydrogen bond relaxation time, $\tau_h$. One can therefore coarse-grain the FFS algorithm by using a larger sampling time and ignoring any high-frequency oscillations in the order parameter that occur at intermediate times. To test the validity of this argument, we carry out a series of FFS calculations of the rate of homogeneous crystal nucleation in the LJ system, with sampling times spanning over four orders of magnitude. We confirm that the cumulative probabilities and nucleation rates are virtually insensitive to the selection of $\tau_s$ unless $\tau_s/\tau_r > 10^{-1}$ (Fig. 3). This approach, however, leads to considerable errors when $\tau_r \approx \tau_s$, as the system starts losing some of its memory between successive samplings of the trajectory. No loss of physically relevant information occurs when $\tau_s \ll \tau_r$, as the fluctuations of the order parameter at times smaller than $\tau_s$ are not representative of physically relevant structural transformations. However, these high-frequency fluctuations can become an issue at extremely small sampling times, as in the TIP4P/Ice system. Choosing a large sampling time is, thus, conceptually similar to applying a low-pass filter to the order parameter time series. In the TIP4P/Ice system, we choose a sampling time of 1 ps, which is two to three orders of magnitude smaller than both $\tau_r = 0.6$ ns (Fig. 2A) and $\tau_h = 4.0$ ns (Fig. 2C). By doing this, we manage to turn an otherwise diverging unsuccessful FFS calculation (Fig. S2A) into a converging successful calculation presented in Fig. 1. Further technical details about the method, as well the computational cost of the calculations, are included in SI Text.

1. Padayachee K, Watt MP, Edwards N, Mycock DJ (2009) Cryopreservation as a tool for the conservation of Eucalyptus genetic variability: Concepts and challenges. *South For* 71(2):165–170.
2. Chamberlain EJ, Gow AJ (1979) Effect of freezing and thawing on the permeability and structure of soils. *Eng Geol* 13(1-4):73–92.
3. Potapczuk MG (2013) Aircraft icing research at NASA Glenn Research Center. *J Aerosp Eng* 26(2):260–276.
4. Ye Z, Wu J, Ferradi NE, Shi X (2013) Anti-icing for key highway locations: Fixed automated spray technology. *Can J Civ Eng* 40(1):11–18.
5. Baker MB (1997) Cloud microphysics and climate. *Science* 276(5315):1072–1078.
6. Carslaw KS, Harrison RG, Kirkby J (2002) Cosmic rays, clouds, and climate. *Science* 298(5599):1732–1737.
7. Fowler LD, Randall DA, Rutledge SA (1996) Liquid and ice cloud microphysics in the CSU general circulation model. Part 1: Model description and simulated microphysical processes. *J Clim* 9(3):489–529.
8. Herbert RJ, Murray BJ, Dobbie SJ, Koop T (2015) Sensitivity of liquid clouds to homogenous freezing parameterizations. *Geophys Res Lett* 42(5):1599–1605.
9. Murray BJ, Knopf DA, Bertram AK (2005) The formation of cubic ice under conditions relevant to Earth's atmosphere. *Nature* 434(7030):202–205.
10. Malkin TL, Murray BJ, Brukhno AV, Anwar J, Salzmann CG (2012) Structure of ice crystallized from supercooled water. *Proc Natl Acad Sci USA* 109(4):1041–1045.
11. Malkin TL, et al. (2015) Stacking disorder in ice I. *Phys Chem Chem Phys* 17(1): 60–76.
12. Taborek P (1985) Nucleation in emulsified supercooled water. *Phys Rev B Condens Matter* 32(9):5902–5906.
13. Li T, Donadio D, Russo G, Galli G (2011) Homogeneous ice nucleation from supercooled water. *Phys Chem Chem Phys* 13(44):19807–19813.
14. Li T, Donadio D, Galli G (2013) Ice nucleation at the nanoscale probes no man's land of water. *Nat Commun* 4:1887.
15. Haji-Akbari A, DeFever RS, Sarupria S, Debenedetti PG (2014) Suppression of subsurface freezing in free-standing thin films of a coarse-grained model of water. *Phys Chem Chem Phys* 16(47):25916–25927.
16. Molinero V, Moore EB (2009) Water modeled as an intermediate element between carbon and silicon. *J Phys Chem B* 113(13):4008–4016.

17. Sanz E, et al. (2013) Homogeneous ice nucleation at moderate supercooling from molecular simulation. *J Am Chem Soc* 135(40):15008–15017.
18. Matsumoto M, Saito S, Ohmine I (2002) Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing. *Nature* 416(6879):409–413.
19. Svishchev IM, Kusalik PG (1994) Crystallization of liquid water in a molecular dynamics simulation. *Phys Rev Lett* 73(7):975–978.
20. Radhakrishnan R, Trout BL (2003) Nucleation of hexagonal ice (*I_h*) in liquid water. *J Am Chem Soc* 125(25):7743–7747.
21. Allen RJ, Frenkel D, ten Wolde PR (2006) Forward flux sampling-type schemes for simulating rare events: Efficiency analysis. *J Chem Phys* 124(19):194111.
22. Steinhardt PJ, Nelson DR, Ronchetti M (1983) Bond-orientational order in liquids and glasses. *Phys Rev B* 28:784–805.
23. Luzar A, Chandler D (1996) Hydrogen-bond kinetics in liquid water. *Nature* 379(6560): 55–57.
24. Sellberg JA, et al. (2014) Ultrafast X-ray probing of water structure below the homogeneous ice nucleation temperature. *Nature* 510(7505):381–384.
25. Palmer JC, et al. (2014) Metastable liquid–liquid transition in a molecular model of water. *Nature* 510(7505):385–388.
26. Courtney T, Johnston DE, McKervey MA, Rooney JJ (1972) The chemistry of diamantane. Part I. Synthesis and some functionalisation reactions. *J Chem Soc Perkin Trans 1* (0):2691–2696.
27. Moore EB, Molinero V (2011) Is it cubic? Ice crystallization from deeply supercooled water. *Phys Chem Chem Phys* 13(44):20008–20016.
28. Seo M, Jang E, Kim K, Choi S, Kim JS (2012) Understanding anisotropic growth behavior of hexagonal ice on a molecular scale: A molecular dynamics simulation study. *J Chem Phys* 137(15):154503.
29. Brukhno AV, Anwar J, Davidchack R, Hande R (2008) Challenges in molecular simulation of homogeneous ice nucleation. *J Phys Condens Matter* 20(49):494243.
30. Quigley D, Rodger PM (2008) Metadynamics simulations of ice nucleation and growth. *J Chem Phys* 128(15):154518.
31. ten Wolde PR, Ruiz-Montero MJ, Frenkel D (1996) Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling. *J Chem Phys* 104(24):9932–9947.
32. Koop T (2004) Homogeneous ice nucleation in water and aqueous solutions. *Z Phys Chem* 218(11):1231–1258.
33. Abascal JLF, Sanz E, García Fernández R, Vega C (2005) A potential model for the study of ices and amorphous water: TIP4P/Ice. *J Chem Phys* 122(23):234511.
34. Moore EB, Molinero V (2011) Structural transformation in supercooled water controls the crystallization rate of ice. *Nature* 479(7374):506–508.
35. DeMott PJ, Rogers DC (1990) Freezing nucleation rates of dilute solution droplets measured between 30° and 40°C in laboratory simulations of natural clouds. *J Atmos Sci* 47(9):1056–1064.
36. Heymsfield AJ, Miloshevich LM (1993) Homogeneous ice nucleation and supercooled liquid water in orographic wave clouds. *J Atmos Sci* 50(15):2335–2353.
37. Tabazadeh A, Djikaev YS, Reiss H (2002) Surface crystallization of supercooled water in clouds. *Proc Natl Acad Sci USA* 99(25):15873–15878.
38. Giauque WF, Stout JW (1936) The entropy of water and the third law of thermodynamics. The heat capacity of ice from 15 to 273°K. *J Am Chem Soc* 58:1144–1150.
39. Angell C, Shuppert JC, Tucker J (1973) Anomalous properties of supercooled water. heat capacity, expansivity, and proton magnetic resonance chemical shift from 0 to −38%. *J Phys Chem* 77(26):3092–3099.
40. Melinder A (2010) Properties and other aspects of aqueous solutions used for single-phase and ice slurry applications. *Int J Refrig* 33(8):1506–1512.
41. Gránásy L, Pusztai T, James PF (2002) Interfacial properties deduced from nucleation experiments: A Cahn–Hilliard analysis. *J Chem Phys* 117(13):6157.

42. Limmer DT, Chandler D (2012) Phase diagram of supercooled water confined to hydrophilic nanopores. *J Chem Phys* 137(4):044509.
43. Handel R, Davidchack RL, Anwar J, Brukhno A (2008) Direct calculation of solid-liquid interfacial free energy for molecular systems: TIP4P ice-water interface. *Phys Rev Lett* 100(3):036104.
44. Hudait A, Molinero V (2014) Ice crystallization in ultrafine water-salt aerosols: Nucleation, ice-solution equilibrium, and internal structure. *J Am Chem Soc* 136(22): 8081–8093.
45. Vega C, Abascal JLF (2011) Simulating water with rigid non-polarizable models: A general perspective. *Phys Chem Chem Phys* 13(44):19663–19688.
46. Espinosa JR, Sanz E, Valeriani C, Vega C (2014) Homogeneous ice nucleation evaluated for several water models. *J Chem Phys* 141(18):18C529.
47. Plimpton SJ (1995) Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys* 117(1):1–19.
48. Reinhardt A, Doye JPK, Noya EG, Vega C (2012) Local order parameters for use in driving homogeneous ice nucleation with all-atom models of water. *J Chem Phys* 137(19):194504.
49. King SV (1967) Ring configurations in a random network model of vitreous silica. *Nature* 213(5081):1112–1113.
50. Li T, Donadio D, Ghiringhelli LM, Galli G (2009) Surface-induced crystallization in supercooled tetrahedral liquids. *Nat Mater* 8(9):726–730.
51. Lennard-Jones JE (1924) On the determination of molecular fields. *Proc R Soc London Ser A* 106(738):463–477.
52. Abascal JLF, Vega C (2005) A general purpose model for the condensed phases of water: TIP4P/2005. *J Chem Phys* 123(23):234505.
53. Lekner J (1997) Electrostatics of proton arrangements in ice *I_c*. *Physica B* 240(3): 263–272.
54. Hayward JA, Reimers JR (1997) Unit cells for the simulation of hexagonal ice. *J Chem Phys* 106(4):1518–1529.
55. Alder BJ, Wainwright TE (1959) Studies in moleculardDynamics. I. General method. *J Chem Phys* 31(2):459–466.
56. Swope WC, Andersen HC, Berens PH, Wilson KR (1982) A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J Chem Phys* 76(1):637–649.
57. Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys* 52(2):255–268.
58. Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A* 31(3):1695–1697.
59. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52(12):7182–7190.
60. Hockney RW, Eastwood JW (1989) *Computer Simulation Using Particles* (CRC Press, New York).
61. Ryckaert J, Ciccotti G, Berendsen H (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J Comput Phys* 23(3):327–341.
62. Sciortino F, Fornili SL (1989) Hydrogen bond cooperativity in simulated water: Time dependence analysis of pair interactions. *J Chem Phys* 90(5):2786–2792.
63. Chandra A (2003) Dynamical behavior of anion-water and water-water hydrogen bonds in aqueous electrolyte solutions: A molecular dynamics study. *J Phys Chem B* 107(16):3899–3906.
64. Lechner W, Dellago C (2008) Accurate determination of crystal structures based on averaged local bond order parameters. *J Chem Phys* 129(11):114707.
65. Borrero EE, Escobedo FA (2007) Reaction coordinates and transition pathways of rare events via forward flux sampling. *J Chem Phys* 127(16):164101.

Haji-Akbari and Debenedetti