

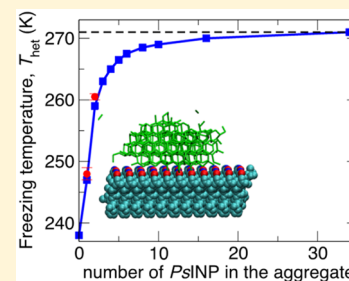
How Size and Aggregation of Ice-Binding Proteins Control Their Ice Nucleation Efficiency

Yuqing Qiu,¹ Arpa Hudait,¹ and Valeria Molinero^{1*}

Department of Chemistry, The University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112-0580, United States

Supporting Information

ABSTRACT: Organisms that thrive at cold temperatures produce ice-binding proteins to manage the nucleation and growth of ice. Bacterial ice-nucleating proteins (INP) are typically large and form aggregates in the cell membrane, while insect hyperactive antifreeze proteins (AFP) are soluble and generally small. Experiments indicate that larger ice-binding proteins and their aggregates nucleate ice at warmer temperatures. Nevertheless, a quantitative understanding of how size and aggregation of ice-binding proteins determine the temperature T_{het} at which proteins nucleate ice is still lacking. Here, we address this question using molecular simulations and nucleation theory. The simulations indicate that the 2.5 nm long antifreeze protein *TmAFP* nucleates ice at 2 ± 1 °C above the homogeneous nucleation temperature, in good agreement with recent experiments. We predict that the addition of ice-binding loops to *TmAFP* increases T_{het} but not enough to compete in efficiency with the bacterial INP. We implement an accurate procedure to determine T_{het} of surfaces of finite size using classical nucleation theory, and, after validating the theory against T_{het} of the proteins in molecular simulations, we use it to predict T_{het} of the INP of *Ps. syringae* as a function of the length and number of proteins in the aggregates. We conclude that assemblies with at most 34 INP already reach the $T_{\text{het}} = -2$ °C characteristic of this bacterium. Interestingly, we find that T_{het} is a strongly varying nonmonotonic function of the distance between proteins in the aggregates. This indicates that, to achieve maximum freezing efficiency, bacteria must exert exquisite, subangstrom control of the distance between INP in their membrane.



1. INTRODUCTION

Although ice is more stable than liquid water below 0 °C, the homogeneous nucleation of ice from micrometer-sized water droplets does not occur at temperatures above -35 °C.^{1,2} The large supercooling needed for nucleation arises from the free energy cost of the interface of the ice embryo. Surfaces that bind ice decrease that cost, promoting nucleation at warmer temperatures.³ Bacterial ice-nucleating proteins (INPs) are among the most efficient ice-nucleating materials,^{4–10} crystallizing water at temperatures as high as -2 °C.¹¹

The INPs of *Pseudomonas syringae* and *Pseudomonas borealis* bind ice through highly conserved arrays of TxT motifs, where T is threonine and x is a nonconserved amino acid.^{12–16} Hyperactive insect antifreeze proteins (AFPs) bind ice through the same TxT motifs as INPs.¹⁶ Although AFPs are very effective at inhibiting the growth of ice¹⁷ by forcing the crystal to grow with curvature,¹⁸ they are not efficient ice nucleators.^{19–21} It has been proposed that the different functions of INPs and AFPs may arise from the distinct sizes of their ice-binding surface (IBS), which are large in INPs and small in AFPs.^{15,21–24} That hypothesis is consistent with classical nucleation theory (CNT),²⁵ which predicts that the size of the critical ice nucleus is larger for nucleation at warmer temperatures, thus requiring a larger IBS to stabilize it. An increase in the ice nucleation temperature with the size of the ice-binding molecule has been reported for nanoscopic organic, biological, and inorganic ice nucleants.^{21,26,27} Never-

theless, there is not yet a quantitative, predictive understanding of how the heterogeneous ice nucleation temperature T_{het} depends on the size, shape, and strength of ice binding of the nucleating surface. Elucidating that dependence is the focus of this study.

There are two ways to modulate the size of the IBS of proteins. The first is to vary the number of ice-binding loops in the β -helix binding surface.²⁸ This changes the length but not the width of the IBS. A recent study shows that dilute solutions of a bioengineered fragment of the INP of *Ps. syringae*, *PsINP*, with 16 TxT loops (about one-quarter of the native size) nucleate ice at $T_{\text{het}} = -25 \pm 1$ °C, just 10 ± 1 °C above the homogeneous nucleation temperature T_{hom} .²⁸ The dependence of the freezing efficiency $\Delta T_f = T_{\text{het}} - T_{\text{hom}}$ with the length of the protein has not been investigated.

The second way to increase ΔT_f is to assemble a larger ice-binding site through aggregation of multiple ice-nucleating proteins.^{15,27–30} Aggregation of INPs occurs in the cell membrane of ice-nucleating bacteria under conditions of stress that require them to nucleate ice.^{11,12,31} It is not known whether the aggregation in the cell membrane is promoted by a change in the chemistry of the membrane or an increase in the concentration of proteins. The aggregation of the proteins in vitro is typically modulated by changes in protein concen-

Received: February 17, 2019

Published: April 12, 2019

tration in solution.²⁸ Increasing the concentration of oligomers of engineered INPs with 16 TxT ice-binding repeats increases T_{het} from -26 to -10 °C.²⁸ These experiments, however, cannot discard aggregation already at the lower concentrations, making it impossible to disentangle the individual effects of lengthening of the protein binding surface and formation of multimeric aggregates on the ice nucleation efficiency.

In the present study, we first use molecular dynamics simulations to elucidate the individual effect of length and aggregation on the nucleating efficiency of ice-binding proteins, including both INP and AFP that bind ice through TxT amino acid repeats. We then present an accurate implementation of heterogeneous classical nucleation theory for finite size surfaces and demonstrate that it can quantitatively represent the simulation data. We finally use the validated theory to predict how the ice nucleation temperature T_{het} of the ice-nucleating protein of *Ps. syringae* evolves with the length of the protein and the number of proteins in the aggregates that it forms in the cell membrane. We use these results to compare with and interpret experimental ice nucleation temperatures for these bacteria.

2. RESULTS AND DISCUSSION

2.1. Antifreeze Protein *Tm*AFP Nucleates Ice Close to the Homogeneous Ice Nucleation Temperature. We use molecular simulations to compute the ice nucleation efficiency ΔT_i for four related sets of ice-binding molecules:

(i) The antifreeze protein *Tm*AFP of the beetle *Tenebrio molitor*,³² shown in Figure 1a. We compute the ice nucleation efficiency of this AFP and elucidate whether the same amino acid sequence is involved in the antifreeze and ice nucleation activities of the protein.

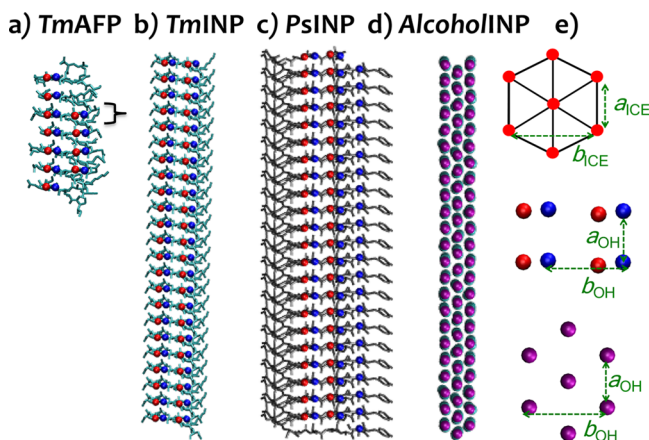


Figure 1. Ice-binding molecules of this study. (a) *Tm*AFP, (b) *Tm*INP with $N = 23$ loops, (c) model *Ps*INP with $N = 22$ loops, and (d) *Alcohol*INP with 23 alcohol molecules per row. The square bracket in (a) indicates the loop of *Tm*AFP that we repeat to produce *Tm*INP. Red and blue balls are the methyl and hydroxyl groups of the IBS. The backbones of *Tm*AFP and *Tm*INP are shown in cyan, and the backbone of *Ps*INP is in gray. Carbon tails of *Alcohol*INP are shown with cyan behind the purple beads that represent the hydroxyl groups. (e) Ice-binding motifs in the IBS of the IBP (middle row) and the alcohol monolayer (lower row) have order consistent with that of water in the basal plane of ice (top row). Note that the IBP *Tm*AFP, *Tm*INP, and *Ps*INP have two columns of OH, while *Alcohol*INP has three columns of OH as ice. Table S1 lists the distance mismatches between the OH in the IBS of these models and the basal plane of ice.

(ii) Ice-binding proteins made by repeating N times the 12-residues loop TCTNSQHCVKAN that encompasses residues 27–38 from the N-terminus of *Tm*AFP.³² We call these proteins *Tm*INP (the one with $N = 23$ is shown in Figure 1b). *Tm*INP are akin to those engineered to study thermal hysteresis in ref 33. We determine whether and how the increase in the number of loops of a model AFP produces an ice-nucleating protein.

(iii) Ice-binding proteins made by repeating N times the 16 residues loop GYGSTQTSQSESSLTA of *InaZ* INP of bacterium *Pseudomonas syringae*, built using homology^{14,23} and scaled to have mismatch to ice similar to *Tm*AFP.¹⁶ We call these proteins model *Ps*INP; the one with $N = 22$ is shown in Figure 1c. By comparing ΔT_i of the model *Ps*INP and *Tm*INP, we investigate whether the ice nucleation efficiency depends only on the number of TxT loops or also on the amino acid sequence of the non-ice-binding residues.

(iv) Rigid fragments of an ice-nucleating alcohol monolayer with three rows of hydroxyl groups that have lattice mismatch to ice identical to *Tm*AFP,³² to compare the size-dependence of ΔT_i of the proteins to a purely hydrogen-bonding IBS.³⁴ We call these molecules *Alcohol*INP; the 23 alcohol molecules-long surface is shown in Figure 1d.

*Tm*AFP is one of the most potent antifreeze molecules in nature.³⁵ Recent experiments found that *Tm*AFP nucleates ice with an efficiency that ranges from 1.3 to 4.5 K above T_{hom} for solutions with protein concentrations that range from 0.5 to 95 μM ,²¹ consistent with a previous report that found its T_{het} in mixtures of H_2O and D_2O to be 5 K above the expected T_{hom} in 2.4 mM solutions.²⁰ Our analysis in section 2.5 indicates that most of this small increase in T_{het} with concentration originates in aggregation of the proteins. Our molecular simulations predict that *Tm*AFP is a weak ice-nucleating agent, promoting the formation of ice at just 2 ± 1 K above T_{hom} . The agreement between the results of the simulations and experiments validates the accuracy of the united atom model for the prediction of the ice nucleation efficiency of proteins.

The simulations reveal that ice nucleates on the TxT binding surface of *Tm*AFP, the same that this protein uses to bind an existing ice surface to prevent its growth.^{18,32} The low freezing efficiency of *Tm*AFP may not be surprising, as *Tm*AFP evolved to bind ice at $T \approx 273$ K, and its ability to nucleate ice at temperatures close to 240 K is irrelevant for its biological function. In the next section, we show that an increase in the number of ice-binding loops can increase the ice nucleation efficiency of this antifreeze protein.

2.2. Width of Ice-Binding Site Limits T_{het} of Proteins.

We now focus on the change in ice nucleation efficiency ΔT_i with the size and shape of the ice-binding site. The β -solenoid structure of the IBS of bacterial INPs and hyperactive insect AFPs confers them a significant anisotropy in shape (Figure 1). Figure 2 presents the ice nucleation efficiency as a function of length L of the binding site for the model of bacterial INP *Ps*INP, the protein *Tm*INP made by stacking of ice-binding loops of the antifreeze protein *Tm*AFP, and the rigid fragments of alcohol monolayer, *Alcohol*INP, with the same width and lattice mismatch to ice as these proteins. The three ice-binding surfaces display the same qualitative behavior: $\Delta T_i(L)$ is zero for very short molecules, then increases sharply, and finally plateaus. In what follows, we analyze the origin and implications of these distinct regimes.

Figure 2 indicates that ice-binding molecules (IBMs) are unable to nucleate ice if they are shorter than a threshold

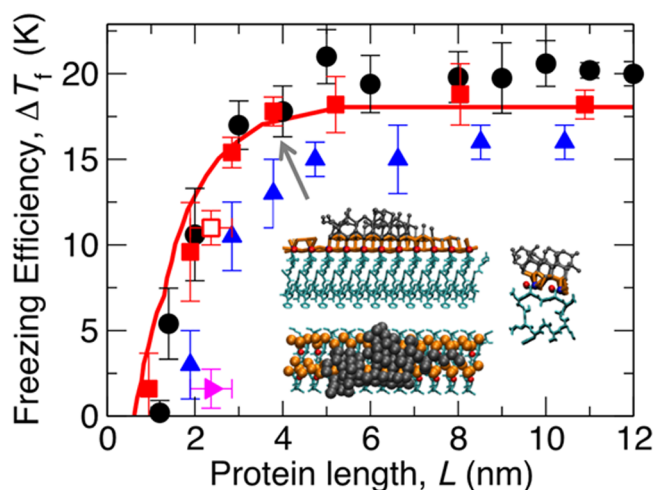


Figure 2. Ice nucleation efficiency of model proteins as a function of the length L of the ice-binding site. Symbols indicate the $\Delta T_f(L)$ computed in molecular simulations with PsINP (blue $\blacktriangle$), TmINP (red $\blacksquare$), AlcoholINP ($\bullet$), and TmAfp (magenta right $\blacktriangle$). The red $\square$ represents ΔT_f of TmINP with the same number of binding sites as TmAfp. The origin of the lower efficiency of TmAfp as compared to TmINP of the same length is discussed in Supporting Information B. The solid red line is the CNT prediction for TmINP at the nucleation rate of the simulations, with $w = 1.3$ nm as the only adjustable parameter. The calculations are performed with the parameters of the mW model at the nucleation rate of the simulations, $J = 10^{27}$ cm $^{-3}$ s $^{-1}$; i.e., $\omega = 10^9$ s $^{-1}$. Figure S4 shows the predicted $\Delta T_f(L)$ curves for the same model protein at other nucleation rates. Inset: Views of the critical ice nucleus on the 5 nm long TmINP; orange shows the anchored clathrate¹⁶ and gray the rest of the ice nucleus. The critical nucleus size is identified as that with the same probability to grow or melt^{36,37} (Figure S2). Note that AlcoholINP and TmINP have almost identical $\Delta T_f(L)$ values, as well as the same ΔT_f for unlimited surfaces. The equivalent efficiency of proteins and alcohol monolayers indicates that the IBS does not need to be amphiphilic to bind strongly to ice.^{3,16,38}

length $L_{\min}$ that is between 0.5 and 2 nm for the molecules of this study. Within the framework of classical nucleation theory, the need for a minimum size of the binding site to nucleate ice arises from the destabilizing effect of the line tension τ of the three-phase line between the ice nucleus, the liquid, and the IBM on the free energy of binding of the protein to ice. Our CNT analysis in section 4.2.2 predicts that the smallest area A_{IBS} of the binding surface that can nucleate ice is given by the condition $A_{\text{IBS}} \times \Delta\gamma_{\text{bind}} + \tau \times l_{\text{IBS}} = 0$ (see Methods, eq 9), where $\Delta\gamma_{\text{bind}} = \gamma_{\text{ice-surface}} - (\gamma_{\text{ice-liquid}} + \gamma_{\text{liquid-surface}})$ is the binding free energy of ice to the surface per unit area, and l_{IBS} is the length of the ice-liquid-IBS boundary. This predicts that surfaces that bind ice weakly require a larger threshold area to nucleate ice. Indeed, molecular simulations show that weakly binding graphitic lamellae³⁹ lose their ice nucleation ability if their ice-binding surface is lower than 4 nm 2 ,²⁶ while disks of alcohol monolayers, which strongly bind to ice,³ do not lose their ice nucleation activity until the area of their IBS is lower than ~ 1 nm 2 (Figure S3). The strong binding free energy of TmAfp to ice¹⁶ explains why, despite its very small size, this ice-binding protein is able to promote the freezing of water.

Figure 2 shows that the freezing efficiency ΔT_f of the model TmINP increases steeply as their IBS lengthens from $L_{\min}$ to the saturation length $L_{\text{sat}} \approx 5$ nm, which corresponds to 10

TxT loops. ΔT_f then plateaus upon lengthening of the protein. We note that both L_{sat} and the ΔT_f at the plateau increase with the nucleation rate (Figure S5). Although experiments have shown that short, 4-loop long, fragments of PsINP have antifreeze activity,²² the reverse transformation of an AFP into an INP by addition of ice-binding loops has not yet been demonstrated in experiments. To our knowledge, the results in Figure 2 constitute the first report of the transformation of an antifreeze protein into an efficient ice-nucleating protein by addition of ice-binding loops.

Larger mutants of TmAfp comparable to TmINP with up to 10 ice-binding loops have been produced in the lab, but only their thermal hysteresis activity has been determined.³³ Our simulations predict that if both the rigidity of the protein and the distances between the TxT repeats do not change upon addition of loops, these proteins would nucleate ice at warmer temperatures than TmAfp. However, we predict that the narrow width of the IBS of TmAfp, combined with the slow nucleation rates of experiments, will result in small gains to T_{het} upon addition of loops to this antifreeze protein (Figure S4).

The increase of ΔT_f with L in Figure 2 reflects the ability of the longer protein IBS to stabilize increasingly larger critical ice nuclei. The ice nucleation efficiency of TmINP, however, does not increase further when the protein has more than ~ 10 TxT loops in its ice-binding surface, because the width of the binding site restricts the width of the ice nucleus it can sustain. Hence, the crystal nucleus becomes more oblong with increasing L . We determine that the critical ice nucleus for the 5 nm long TmINP is as long as the protein and about 1.3 nm wide (inset of Figure 2). We conclude that ΔT_f plateaus upon further increase of the length L of the protein, because additional lengthening of the nucleus increases its area to volume ratio and does not lead to a decrease of the ice nucleation barrier.

To illustrate how the anisotropic shape of proteins limits their nucleation efficiency, we show in Figure 3a the freezing efficiency versus area of the binding surface for circular and rectangular rigid fragments of alcohol monolayers that have perfect lattice matching to ice: while the circular, isotropic surfaces increase their efficiency with area until it saturates at the ΔT_f for the macroscopic monolayer, the anisotropic rectangular surfaces plateau at a much lower freezing efficiency, limited by their width.

Surfaces that bind weaker to ice reach a lower ice nucleation efficiency than those that have more negative binding free energy $\Delta\gamma_{\text{bind}}$.³ This is the case for unlimited size surfaces (see ref 3 and section 2.3), as well as for surfaces that have a small IBS that limits nucleation (Figure S3). $\Delta\gamma_{\text{bind}}$ of surfaces that hydrogen bond to ice is modulated by their lattice mismatch to ice.³ Figure 3b shows that for surfaces, such as proteins, that have distinct lattice mismatch to ice in the two directions of the IBS, the ice nucleation efficiency is maximal when the smaller mismatch occurs along the longer direction. We conclude that both the anisotropy in shape of the IBS and its alignment with respect to the direction of minimum mismatch to ice are important for the design of efficient ice-nucleating proteins.

It has been proposed that the mass of the ice-nucleating proteins or their aggregates can be used to predict T_{het} .^{21,27} However, as ice-nucleating proteins are generally anisotropic in shape, their ΔT_f decouples from the mass of the protein (and area of the binding site) when the shape anisotropy is pronounced. This indicates that knowledge of the mass of the

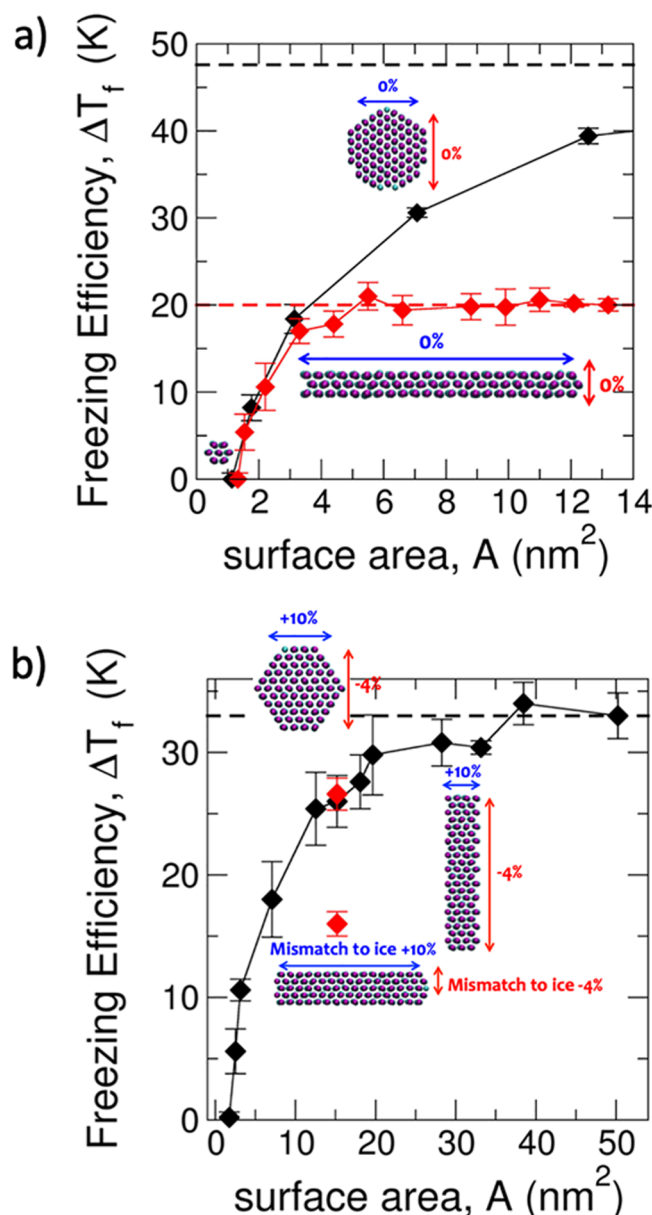


Figure 3. Effect of anisotropy in shape and in lattice mismatch to ice on the nucleation efficiency of finite surfaces. (a) Freezing efficiency versus area of the binding site determined with molecular simulations of circular (black symbols) and rectangular (red symbols) fragments of alcohol monolayers with perfect matching to ice. The rectangular surfaces are three-rows of alcohol molecules wide. The dashed black line indicates the freezing efficiency of a monolayer with the same mismatch and unlimited in size in both directions. (b) Freezing efficiency versus area for circular (black symbols) and rectangular (red symbols) fragments of an alcohol monolayer that has different anisotropic lattice mismatch to ice in the two directions, +10% and -4%. The nucleation efficiency is highest when the smaller mismatch is aligned with the long direction of the nucleating surface.

ice-nucleating molecule is, in general, not sufficient to predict its ice nucleation temperature. In the next section, we demonstrate that the ice nucleation temperature of anisotropic finite surfaces, such as that of ice-nucleating proteins, can be accurately predicted with nucleation theory.

2.3. Classical Nucleation Theory Quantitatively Predicts the Size-Dependence of T_{het} of Proteins. In what follows, we first present an implementation of classical

nucleation theory that allows us to accurately predict, for the first time, the ice nucleation temperature of surfaces of finite size and arbitrary shape and strength of interaction of the binding site, such as proteins. We then validate our implementation of the theory by comparing its predictions with T_{het} as a function of the length of the binding surface determined in molecular simulations for the *Tm*INP model. We finally use the validated implementation of the theory to predict the dependence of the ice nucleation efficiency of the INP of *Ps. syringae* as a function of the length L of its IBS and, in section 2.5, of the number N_{INP} of protein monomers in the aggregates these proteins make in the bacterial membrane at the conditions of the experiment. We use those theoretical results to interpret experimental data of ice nucleation by proteins of insect and bacteria.

Nucleation temperatures are typically determined in experiments by cooling small droplets and collecting statistics on the temperature at which they crystallize. T_{hom} is determined by both the volume of the droplets and the cooling rate. For example, microliter droplets cooled at rates of about 1 K min⁻¹ nucleate ice at $T_{hom} = 238$ K = -35 °C.⁴⁰ Under these conditions, the experimental homogeneous nucleation rate⁴¹ is $\omega_{hom} = 10^2$ s⁻¹.^{2,42} T_{het} is also controlled by the cooling rate, but is modulated by the area of the nucleating surface.^{41,43} For example, 10 μ L droplets that each contain an average of 10^4 *Ps. syringae* incubated to produce the most ice-nucleating active form of the bacteria heterogeneously nucleate ice at $T_{het} = -2$ °C when cooled at about 1 K min⁻¹.¹¹ We use this solution as reference for the calculations of heterogeneous nucleation by the bacterial ice-nucleating protein and its aggregates. It has been interpreted that just a few bacteria in these droplets are responsible for this very high T_{het} .^{11,44} As T_{hom} and T_{het} are compared using the same cooling rates (observation times) for homogeneous and heterogeneous nucleation, we here select the nucleation rate $\omega = \omega_{hom}(T_{hom})$ of the homogeneous nucleation experiments, and use classical nucleation theory to identify the temperatures T_{het} for which $\omega_{het}(T_{het}) = \omega_{hom}(T_{hom})$. Although the heterogeneous nucleation temperatures depend, in principle, on the total area that can nucleate ice in the system, the steep dependence of the nucleation barrier with temperature dwarfs changes in concentration, which modify the pre-exponent. Indeed, we show in Supporting Information E that T_{het} is quite insensitive to the concentration of proteins in the absence of aggregation.

We have previously derived a relationship between the freezing efficiency ΔT_f and the binding free energy $\Delta\gamma_{bind}$ of a nucleating surface of unlimited size using CNT and neglecting the contribution of the ice-liquid-surface line tension to the free energy of the nascent ice embryo.³ We here extend the procedure of ref 3 to first include the line tension effect on the shape and stability of the critical crystallite, and then to account for the finite size of the nucleating surface on the heterogeneous nucleation temperature. Figure 4 presents the workflow of our iterative “Heterogeneous Ice Nucleation Temperature” (HINT) procedure to solve CNT for surfaces of unlimited size. Section 4.2.1 details the HINT procedure for unlimited surfaces, and section 4.2.3 its implementation for nucleation on finite surfaces, such as proteins.

The implementation of HINT requires knowledge of properties of the nucleating surface and water. The ice-nucleating surface specific properties are the difference in surface free energy upon ice binding, $\Delta\gamma_{bind}$, and the line tension τ of the three-phase ice-liquid-surface contact line.

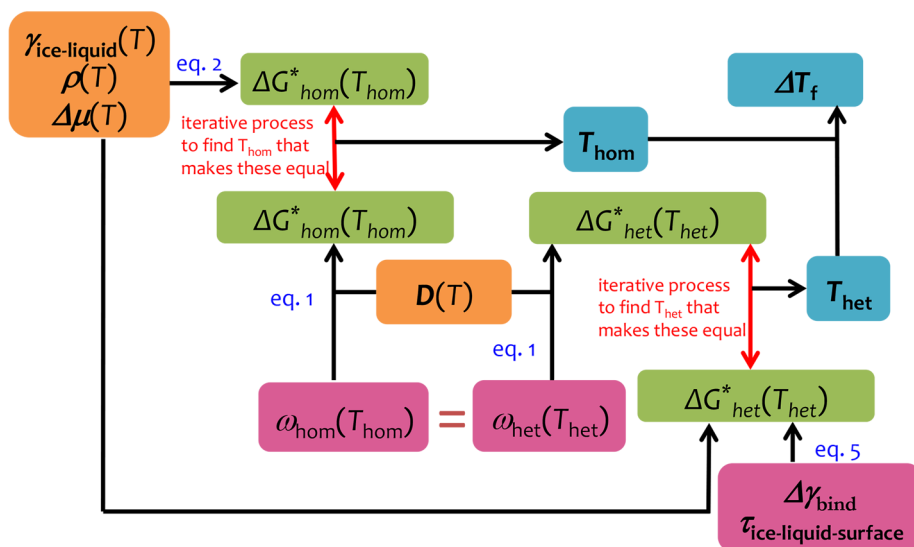


Figure 4. Schematic illustration of the HINT algorithm used for the derivation of the freezing temperature T_{het} and freezing efficiency ΔT_f from the ice-binding free energy per area $\Delta\gamma_{\text{bind}}$ and line tension τ using classical nucleation theory. Pink boxes indicate input variables: the nucleation rate ω , the $\Delta\gamma_{\text{bind}}$, and τ that control the thermodynamics of the ice embryo at the nucleating surface. Orange boxes indicate parameters intrinsic to water: the self-diffusion coefficient D , the difference in chemical potential between liquid and ice $\Delta\mu$, and the ice–liquid surface tension $\gamma_{\text{ice-liquid}}$. We run the algorithm with properties for the mW model when comparing the theory with the molecular simulations, and we implement it with properties of real water when making a prediction for proteins and their aggregates in experiments. Green boxes indicate the intermediate outputs: the free energy barriers ΔG^* for homogeneous and heterogeneous nucleation, deduced from the nucleation rate ω and the temperature dependence of the prefactor A of the rate (see section 4.2.1). The blue boxes indicate the outputs of the HINT procedure: T_{het} , T_{hom} , and their difference ΔT_f . Black arrows represent the computing processes with the corresponding equations of Methods, and red arrows represent iterative processes in which the heterogeneous rate ω_{het} is evaluated as a function of candidate T_{het} until ω_{het} becomes equal to ω_{hom} and the evaluation is converged. For finite surfaces, the iteration also scans over the contact angle of the ice nucleus, as the contact angle is not constant when the nucleus meets the boundary of the surface.⁴⁵ A detailed explanation of the method and the equations can be found in section 4.2.

Table 1. Parameters of the Classical Nucleation Theory Calculations Used to Compare with the Ice Nucleation Temperature of the Model *Tm*INP in the Simulations and with the Ice Nucleation Temperature of PsINP of *Ps. syringae* in Experiments^a

INP	ω , s ⁻¹	$T_{\text{het}}^{\text{unlim}}$, K	L_{min} , nm	w , nm	$\Delta\gamma_{\text{bind}}$, ^b mJ ⁻¹ m ⁻²	τ , pN	T_{hom} , K	T_{sat} , K
model <i>Tm</i> INP	<u>10²</u>	<u>250</u>	<u>0.47</u>	1.3	-68.1	9.5	202	220
exp PsINP	<u>10²</u>	<u>271</u>	0.51 ^c	1.8	-62.6	<u>10</u>	238	247

^aUnderlined data are input for the CNT calculation. ^b $\Delta\gamma_{\text{bind}}$ evaluated at $T_{\text{het}}^{\text{unlim}}$; the values at T_{hom} are -54.6 and -50.1 mJ⁻¹ m⁻² for the model *Tm*INP and experimental PsINP, respectively. ^cDeduced from the value of $\Delta\gamma_{\text{bind}}/\tau$.

Section 4.2.2 explains how we derive τ and $\Delta\gamma_{\text{bind}}$ from L_{min} and the freezing efficiency of the unlimiting-sized surfaces $\Delta T_f^{\text{unlim}}$. Table 1 reports these properties for the model *Tm*INP in mW water and for the bacterial PsINP in water. The water-specific properties are the temperature dependence of the excess chemical potential, ice–liquid and liquid–vapor surface tensions, and diffusion coefficients. We use HINT with the properties of the mW water model when we make theoretical predictions to compare with the molecular simulations, and we use the experimental properties of water when we make predictions or compare with the ice nucleation temperatures of proteins in experiments.

Figure 5 shows the relation between the binding free energy per area of the ice-binding surface to ice, $\Delta\gamma_{\text{bind}}$, and the freezing efficiency ΔT_f of that surface in the CNT parameterization for water (panel a) and for mW (panel b), for various values of the line tension τ of the ice–water–IBS contact line. A positive line tension destabilizes the crystal nuclei and moves down the iso-rate curves that represent the freezing efficiency for a surface of a given binding free energy, $\Delta\gamma_{\text{bind}}(\Delta T_f, \tau)$. The higher sensitivity of the freezing temperature to the line tension for real water compared to mW in the figure is due to the different nucleation rates ω we use to make the

corresponding plots, which results in smaller nucleation barriers and critical nucleus size for mW, and makes the stabilization of the nucleus by the surface more sensitive to the line tension. We have shown in ref 3 that when the rate for mW is chosen to produce the same $T_{\text{hom}} = 238$ K as in the experiments, the curves for $\Delta\gamma_{\text{bind}}$ versus ΔT_f for mW and water overlap.

We first validate the HINT implementation of CNT for the model *Tm*INP using thermodynamic and dynamic properties for the mW water model in the implementation of the algorithm. To obtain $\Delta\gamma_{\text{bind}}$, we determine the freezing efficiency of a surface of unlimiting size, $\Delta T_f^{\text{unlim}}$, from simulations of the extended TCT peptide surfaces of ref 46 and follow the procedures of section 4.2.2. The only adjustable parameter in the HINT calculation is the width w of the ice-binding surface, which we take to be 1.3 nm, the width of the critical nucleus of ice on *Tm*INP (inset of Figure 2). The HINT prediction for T_{het} of *Tm*INP as a function of length (solid red line in Figure 2) is in quantitative agreement with the one determined using molecular simulations at the same nucleation rate (red ■ in Figure 2). The agreement validates the HINT algorithm for predicting T_{het} of ice-binding surfaces of arbitrary size using classical nucleation theory.

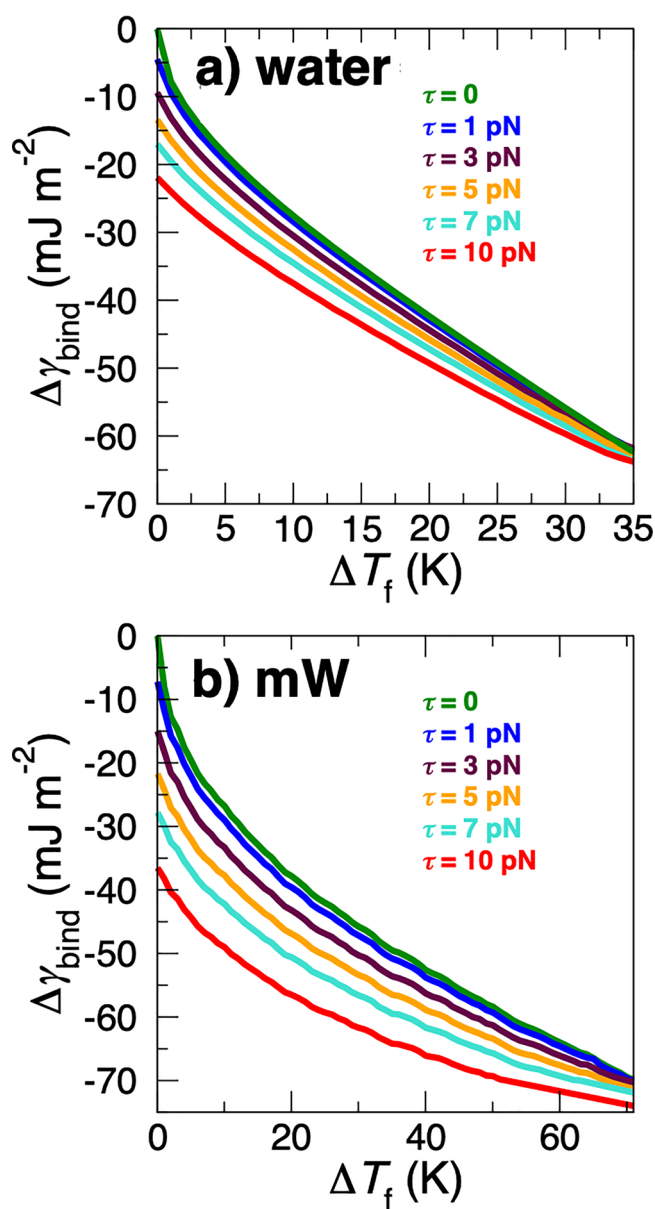


Figure 5. $\Delta\gamma_{\text{bind}}$ needed to produce a freezing efficiency $\Delta T_f = T_{\text{het}} - T_{\text{hom}}$ for an unlimited, large surface at the specified nucleation rate ω . (a) Water at the typical experimental rate $\omega = 100 \text{ s}^{-1}$ that produces $T_{\text{hom}} = 238 \text{ K}$ in microliter droplets (i.e., $J = 10^5 \text{ cm}^{-3} \text{ s}^{-1}$) and (b) mW water at the rate used in the simulations $\omega = 10^9 \text{ s}^{-1}$, which produces $T_{\text{hom}} = 202 \text{ K}$ in simulations with $\sim 10\,000$ water molecules (i.e., $J = 10^{27} \text{ cm}^{-3} \text{ s}^{-1}$). In each case, we report the results for various values of the ice–liquid surface line tension $\tau = 0, 1, 3, 5, 7$, or 10 pN (each labeled in the graphs). In both graphs, the freezing temperatures start at T_{hom} ($\Delta T_f = 0$) and end at T_{melt} .

Having validated the HINT implementation of CNT against the molecular simulations, we now use the theory to predict how the experimental freezing temperature T_{het} of the INP of the bacterium *Ps. syringae* evolves with the length of its IBS. We perform the calculations of Figure 4 using the experimental excess chemical potential, density, ice–liquid surface tension, and diffusion coefficient of water (see section 4.2). With these properties, we compute T_{het} at the experimental nucleation rate that renders the homogeneous nucleation temperature $T_{\text{hom}} = 238 \text{ K}$ for microliter droplets at cooling rates of $\sim 1 \text{ K min}^{-1}$.² We assume w of the bacterial protein to be 1.8 nm , close to the

1.6 nm distance between the serine and farthest threonine in the STxT ice-binding loop of the *Ps*INP model,¹⁶ and use the line tension $\tau = 10 \text{ pN}$ deduced from the simulations of *Tm*INP. Further considering that the maximum freezing temperature reported for *Ps. syringae* is 271 K ,⁴⁴ we deduce $\Delta\gamma_{\text{bind}} = -62.6 \text{ mJ m}^{-2}$ from the analytical CNT curves that relate T_{het} of surfaces of unlimited size to their $\Delta\gamma_{\text{bind}}$ and τ (Figure 5a). It is noteworthy that $\Delta\gamma_{\text{bind}}$ derived from experimental data for water and *Ps. syringae* is very close to $\Delta\gamma_{\text{bind}} = -68.1 \text{ mJ m}^{-2}$ of the model *Tm*INP (Table 1): both TxT-based proteins are extremely effective at binding ice. The $\Delta\gamma_{\text{bind}}$ we obtain for the bacterial protein corresponds to a zero effective contact angle of ice on the protein surface; that is, there is complete wetting of the protein surface by ice. This is consistent with a pioneering theoretical analysis by Burke and Lindow that concluded that the surface tension of the IBS of the INP of *Ps. syringae* must be essentially identical to that of ice to account for the exceptional ice nucleation efficiency of this bacterium.²⁹

Using HINT with the experimental properties of water and the -2°C maximum ice nucleation temperature of *Ps. syringae*, we predict in Figure 5 the dependence of T_{het} with the length of the bacterial INP. Our calculations indicate that the freezing temperature of the monomer saturates at 247 K when L reaches $\sim 8 \text{ nm}$ (~ 16 loops), in excellent agreement with the $248 \pm 1 \text{ K}$ measured for the 16 loop fragment of the INP in experiments with 10^4 proteins per droplet.²⁸ Table S6 shows that T_{het} of the 16-loop *Ps*INP monomer is quite insensitive to concentration. Importantly, we predict that further lengthening of the *Ps*INP monomer from 16 loops to its native length of 50–80 loops does not improve its freezing efficiency (Figure 6). We conclude that the width of the bacterial protein limits its maximum heterogeneous ice nucleation temperature. To increase the freezing efficiency, the ice-binding surface has to grow in both dimensions. This can be achieved through aggregation of monomers.

2.4. Enhancement of Ice Nucleation Efficiency upon Aggregation Is Nonmonotonous with the Separation between the Proteins. Aggregation of *Ps*INP in the membrane of *Ps. syringae* is key to the exceptional ice-nucleating ability of these bacteria.^{31,44} It has been proposed that *Ps*INP may form aggregates by interdigitation of the monomers in the membrane.⁴⁷ That model, however, assumed that the IBS of the INP adopts a β -hairpin structure, contrary to the current consensus that it is a β -solenoid.¹⁵ A more recent study proposed that *Ps*INP forms antiparallel dimers, in which the TxT binding site of one monomer is coplanar with the SLTA binding site of the other monomer.¹⁵ That mode of aggregation, however, cannot account for the formation of aggregates larger than dimers. To date, the distances and relative orientations of the monomers in the aggregates, and what holds them together, have not yet been elucidated.

Here, we use molecular simulations to determine the freezing efficiency of coplanar pairs of 12 nm long *Tm*INP (Figure 7a), as a function of the distance d between monomers. We find that ΔT_f is nonmonotonous and highly varying with d (Figure 7b). The sensitivity of ΔT_f to the distance between monomers implies that bacteria must exert accurate control of the distance between protein monomers in the membrane aggregates to maximize their ice-nucleating temperature. The predictions of the simulations are consistent with the high sensitivity of the experimental ice nucleation temperature of *Ps. syringae* to chemicals that disrupt the

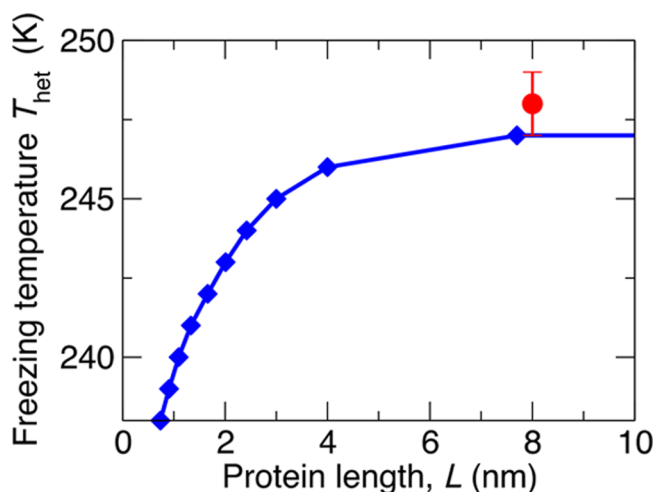


Figure 6. Ice nucleation temperature as a function of protein length for the INP monomer of *Ps. syringae* under conditions typical of laboratory experiments. Blue $\blacklozenge$ show the predictions of CNT for $T_{\text{het}}(L)$ values of the INP of *Ps. syringae* using HINT with the experimental excess chemical potential, diffusion coefficient, ice-liquid surface tension, and density of water at a nucleation rate $\omega = 100 \text{ s}^{-1}$, consistent with the $\sim 1 \text{ K min}^{-1}$ cooling rate of the experiments. The *Ps*INP surface in the HINT calculation is characterized by a width $w = 1.8 \text{ nm}$, a line tension $\tau = 10 \text{ pN}$, and $\Delta\gamma_{\text{bind}} = -62.6 \text{ mJ m}^{-2}$ derived from the maximum T_{het} of *Ps. syringae*. The maximum ΔT_f of *Ps*INP is lower than that for *Tm*INP in Figure 2, despite them having comparable $\Delta\gamma_{\text{bind}}$ values, because ΔT_f is larger for higher nucleation rates (Figure S5). Our predictions for the 16-loop *Ps*INP monomer using CNT agree with the experimental $T_{\text{het}} = 248 \pm 1 \text{ K}$ of the 16-loop variant of this protein, INpro_{16R} (red $\bullet$).²⁸

ordering and fluidity of the cell membrane,^{11,44,48,49} which may modulate the distance between the membrane-anchored INPs.

The modulation of the freezing efficiency with the distance between the monomers, $\Delta T_f(d)$, is identical for pairs of *Tm*INP and pairs of 11 nm long *Alcohol*INP monomers with almost the same lattice mismatch to ice as *Tm*INP (Figure 7a,b), although alcohol monolayers hydrogen bond directly to ice³ and the TxT surface of proteins binds ice through an anchored clathrate motif that includes both hydrogen bonding and hydrophobic groups.^{16,46,50} This indicates that the modulation of the freezing efficiency is not related to the details of how the molecules bind to ice.

Individual proteins that bind ice through TxT sequences, as well as alcohol monolayers, nucleate stacking disordered ice bound to the IBS through the basal plane,^{3,16} because that ice face provides the strongest ice-binding free energy.¹⁶ The first maximum in ΔT_f for the protein dimer occurs with the monomers at $d \approx 1.1 \text{ nm}$ (peak I in Figure 7). Ice nucleated by those dimers is also stacking disordered, but bound to the IBS through the 28° pyramidal face (10 $\bar{1}$ 1) (Figure 7c). At $d = 0.8 \text{ nm}$ (peak II in Figure 7b), the dimer has the optimum spacing to bind ice through the basal face (Figure 7d), resulting in the highest freezing efficiency. This distance already overlaps the rigid protein models of our simulations, but may be accessible to the flexible *Ps*INP in the bacteria. The dimer at $d = 0.5 \text{ nm}$ (peak III in Figure 7b) also nucleates ice bound through the basal plane, but destabilized by pentagonal defects (Figure 7d). At the distances where ΔT_f is a minimum, the ice nucleus develops destabilizing defects to simultaneously bind the two monomers.

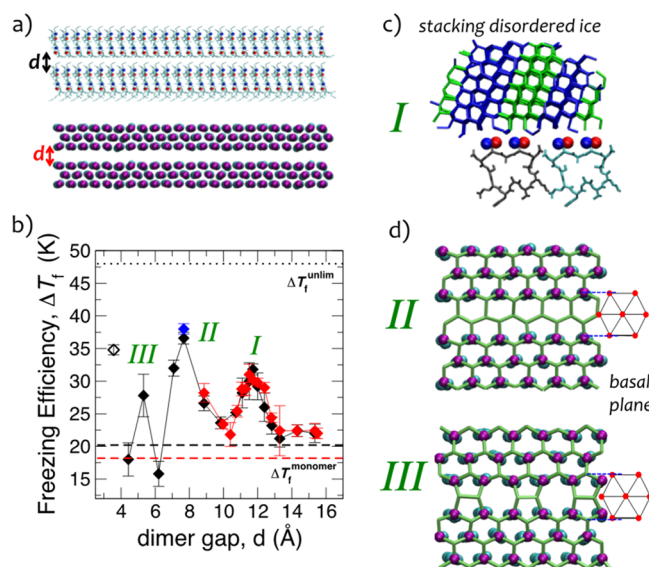


Figure 7. Freezing efficiency of a protein dimer is nonmonotonous with the distance d between monomers. (a) Snapshots of the 12 nm long coplanar *Tm*INP dimer (top, each monomer containing 25 TxT loops) and 11 nm long coplanar *Alcohol*INP dimer (bottom, each monomer containing 23 rows of alcohol molecules), with the same color coding as in Figure 1. (b) $\Delta T_f(d)$ for the *Tm*INP dimer (red $\blacklozenge$) and *Alcohol*INP dimer ($\blacklozenge$) computed in molecular simulation with nucleation rate $\omega = 10^9 \text{ s}^{-1}$; the lines through the symbols are guides to the eye. The empty $\diamond$ represents ΔT_f of two adjacent *Alcohol*INP with one monomer shifting and docking into the other (Supporting Information F). The ΔT_f of *Alcohol*INP with seven columns of hydroxyl binding sites (blue diamond) is almost the same as for the dimer that binds the basal plane, suggesting that the effectiveness of dimers is embedded in its increasing width rather than the number of ice-binding groups. The dashed lines are the efficiencies of monomers of *Tm*INP (red) and *Alcohol*INP (black). The dotted line indicates the freezing efficiency $\Delta T_f^{\text{unlim}} = 48 \text{ K}$ of *Tm*INP or *Alcohol*INP of unlimiting size (they are identical). *Tm*INP cannot approach at $d < 0.9 \text{ nm}$ in our simulations with rigid protein models. (c) Gray and cyan bonds represent the backbones of two identical INP monomers side by side at $d = 1.1 \text{ nm}$. Blue and green bonds indicate hexagonal and cubic ice layers in the stacking disordered ice. The stacking sequence varies across different nucleation trajectories, but the orientation of the ice crystal is always as shown, that is, bound to the IBS by the (10 $\bar{1}$ 1) face. (d) Top views of the *Alcohol*INP dimer (colors as in Figure 1) and first layer of ice (green) after crystallization at the dimer gap distances corresponding to peaks II and III. The sketches of water ordering on the basal plane of ice (shown with red balls and gray sticks, on the right) illustrate the matching between the dimer gap and the ice face.

Our analysis indicates that the optimal distances between INPs in the aggregates are those that allow all protein monomers to bind an ice nucleus through a strongly ice-binding face without introducing additional stress or defects in the ice lattice. We conclude that distances of water molecules in ice faces control the variation of T_{het} with the separation between monomers. Hence, we predict that ice nucleation efficiency will be a strongly varying and nonmonotonous function of the distance between proteins, irrespective of their orientation and ordering in the membrane.

2.5. Aggregates with at Most Three-Dozen *Ps*INP Monomers Suffice To Reach the Experimental Freezing Efficiency of *Ps. syringae*. Figure 7 shows that the highest ΔT_f for the dimer is still 11 K short of the freezing efficiency of an unlimiting surface, $\Delta T_f^{\text{unlim}}$, with the same strength of

binding $\Delta\gamma_{\text{bind}}$. Multimeric aggregation of the proteins is needed to produce a surface large enough that allows water to crystallize at temperatures close to the melting point.

We use the HINT implementation of CNT to predict the temperature of ice nucleation of side-by-side aggregates of the INP of *Ps. syringae* using experimental excess chemical potential, surface tensions, diffusion coefficient, and density of water, and the same ice-binding strength of the monomer ($\Delta\gamma_{\text{bind}} = -62.6 \text{ mJ m}^{-2}$ and $\tau = 10 \text{ pN}$). We assume that the width of an aggregate of N_{INP} proteins is $w = 1.8 \text{ nm} \times N_{\text{INP}}$. Figure 8 shows the ice nucleation temperature T_{het} we predict as a function of N_{INP} in the aggregates of *Ps. syringae*.

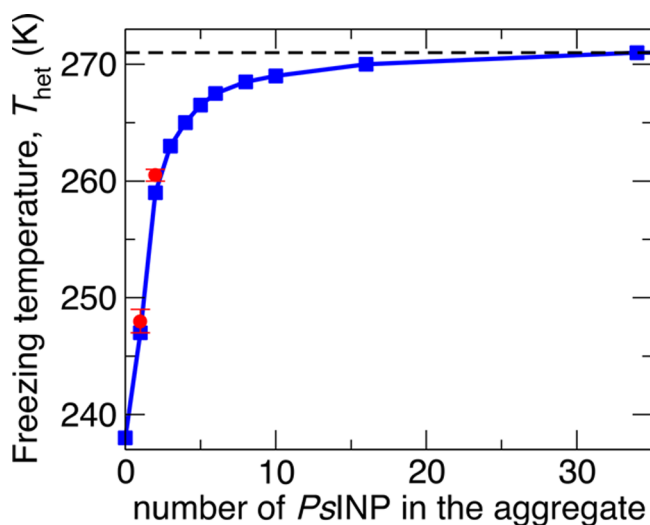


Figure 8. Freezing temperature of ice increases with the number of protein monomers in the INP aggregates. Blue points are the predictions from heterogeneous nucleation theory (see section 4.2). Red points are the experimental measurements of ice freezing temperatures, $T_{\text{het}} = 248 \pm 1 \text{ K}$ of the 16-loop variant of PsINP, INpro_{16R} taken from ref 28, and $T_{\text{het}} = 260 \pm 0.5 \text{ K}$ of minimal functional subunit of PsINP taken from ref 30. Black dashed line is $T_f = 271 \text{ K}$ for the freezing efficiency of PsINP.¹¹ We estimate from CNT that 34 protein monomers are needed to achieve $T_f = 271 \text{ K}$ (Table S7). If the monomers were not limited in size, we predict that a $50 \text{ nm} \times 50 \text{ nm}$ IBS (28 INP, each 50 nm long) would suffice to reach $T_{\text{het}} = 271 \text{ K}$.

Formation of aggregates with up to 10 PsINP results in significant gains in ice nucleation efficiency (Figure 8). Beyond that, T_{het} increases slowly upon addition of new monomers, as the driving force $\Delta\mu$ for crystallization becomes very small. Our calculations indicate that the $T_{\text{het}} = 260.5 \pm 0.5 \text{ K}$ reported in experiments³⁰ corresponds to the dimer, for which our CNT calculations predict 259 K . We further predict that a 34-mer of side-by-side PsINPs (a surface 61.2 nm wide and 40 nm long) suffices to reach the $T_{\text{het}} = 271 \text{ K}$ reported for the most active forms of *Ps. syringae*.¹¹ We note that our calculations under-predict T_{het} of the monomer and dimer by about 1 K . This may indicate that either the IBS of each monomer is $10\text{--}20\%$ wider than the 1.8 nm we assumed in our calculations, which would imply that only ~ 30 monomers are needed to reach $T_{\text{het}} = 271 \text{ K}$, or that $\tau > 10 \text{ pN}$ and, hence, $\Delta\gamma_{\text{bind}} < -62.6 \text{ mJ m}^{-2}$, which would also result in a lower number of proteins to reach the maximum nucleation efficiency of the bacterium. We conclude that the protein aggregates needed to reach the $-2 \text{ }^\circ\text{C}$ ice nucleation

temperature of *Ps. syringae* contain no more than 35 monomers, about 100 less than previously anticipated.^{27,28,30}

Aggregation can also increase the ice nucleation efficiency of antifreeze proteins. It was found in ref 21 that an increase in the concentration of TmAFP in nanoliter droplets from 0.5 to $95 \text{ } \mu\text{M}$ (i.e., $\sim 3 \times 10^8$ to 6×10^{10} proteins per droplet) results in a rise of T_{het} from -37 to $-34 \text{ }^\circ\text{C}$ ($T_{\text{hom}} = -38$ in the absence of protein in those experiments). It was not possible from the available experimental data to determine whether the rise in T_{het} was due to an increase in the active ice-nucleating area (proportional to concentration) or to aggregation to form larger ice-nucleating surfaces. To address that question, we assume that TmAFP has the ice-binding strength $\Delta\gamma_{\text{bind}} = -62.6 \text{ mJ m}^{-2}$ and $\tau = 10 \text{ pN}$ of the INP of *Ps. syringae*, and use the HINT algorithm to predict T_{het} of TmAFP at a cooling rate that produces $T_{\text{hom}} = 238 \text{ K} = -35 \text{ }^\circ\text{C}$ in microliter droplets. The calculations predict $T_{\text{het}} = 241 \text{ K} = -32 \text{ }^\circ\text{C}$ for droplets that contain $\sim 10^3\text{--}10^7$ monomers (Table S3). If we instead use the binding free energy of TmAFP to the basal plane of ice computed in simulations, $\Delta\gamma_{\text{bind}} = -57 \text{ kJ mol}^{-1}$,¹⁶ we derive $\Delta\gamma_{\text{bind}} = -61.8 \text{ mJ m}^{-2}$ for TmAFP in the model (Supp. Section E1). As the ice-liquid surface tension determines the maximum freezing efficiency of the surface (Figure 5) and $\gamma_{\text{ice-liquid}}$ at 273 K is 31.2 mJ m^{-2} for water⁴² and 35 mJ m^{-2} for the mW model,⁷⁹ we scale $\Delta\gamma_{\text{bind}} = -61.8 \times 31.2/35 \text{ mJ m}^{-2} = -55.1 \text{ mJ m}^{-2}$ to use the strength of binding in the simulations to make predictions for TmAFP in real water using CNT. Using this scaled strength of binding, we predict that the antifreeze protein would nucleate ice at $T_{\text{het}} = 240 \text{ K} = -33 \text{ }^\circ\text{C}$ (Table S3). The $\Delta T_f = 2 \text{ K}$ predicted by nucleation theory is consistent with the $2 \pm 1 \text{ K}$ we find in the molecular simulations of TmAFP (section 2.1). Table S3 shows that to raise T_{het} by a further 3 K , the number of monomers of TmAFP per droplet should increase by 10^7 to 10^8 , i.e. 5 to 6 orders of magnitude more than the range of the experiment. This suggests that the increase in ΔT_f from ~ 1 to 4 K in ref 21, as well as the $\Delta T_f = 5 \text{ K}$ of the 2.4 mM solutions of ref 20, are mostly due to partial aggregation of the proteins to produce larger ice-nucleating surfaces. Indeed, the freezing efficiency of the concentrated solutions is comparable to the $\Delta T_f = 6 \text{ K}$ we predict for aggregation of TmAFP to produce optimal coplanar dimers (Table S4). Our analysis indicates that the $T_{\text{het}} = 250 \text{ K}$ attained by functionalization of surfaces with TmAFP that expose their ice-binding surface to the solution¹⁹ require large aggregates, as we predict that coplanar trimers would nucleate, at best, at 247 K , and that the maximum ice nucleation temperature for an unlimited surface with $\Delta\gamma_{\text{bind}} = -55.1 \text{ mJ m}^{-2}$ and $\tau = 10 \text{ pN}$ is 264 K (Figure 5a). We conclude that aggregation can play a role in modulating the ice nucleation efficiency of antifreeze proteins, but also highlight that these small proteins have evolved to remain dispersed in solution, and are not prone to aggregate⁵¹ into the extended, probably coplanar ice-binding surfaces that endow bacterial INPs with their exceptional ice nucleation efficiency.

Although the present study focuses on hyperactive insect antifreeze and bacterial ice-nucleating proteins, its approach and conclusions can be generalized for other ice-binding proteins. Many freeze-tolerant insects, for example, have developed membrane ice-nucleating proteins that allow them to freeze extracellular water at temperatures that typically range from -4 to $-8 \text{ }^\circ\text{C}$.^{35,52} Like bacterial INPs, these proteins are also organized into multimeric aggregates. For example, transmission electron microscopy of the lipoprotein ice

nucleator (LPIN) from the hemolymph of the crane fly *Tipula trivittata* shows that the LPIN organize into chain structures, in which each chain is two proteins wide, about 25 nm in width.⁵³ We predict that PsINP aggregates of that width nucleate ice at about $-3.5\text{ }^{\circ}\text{C}$ (Figure 8), close to the $-6\text{ }^{\circ}\text{C}$ ice nucleation temperature of the aggregates in the crane fly. This suggests that the strength of ice binding $\Delta\gamma_{\text{bind}}$ of the insect LPIN is comparable to that of the bacterial INP.

3. CONCLUSIONS

Nature has evolved proteins that excel at either promoting the nucleation of ice or preventing its growth. The commonality to ice-nucleating and antifreeze proteins is their ability to bind ice to control the kinetics of water crystallization.¹³ A central difference is their size: ice-nucleating proteins are long and form large aggregates in the cell membrane, while antifreeze proteins are typically small and soluble in water. Experiments indicate that larger ice-binding proteins nucleate ice at warmer temperatures.^{21,27,30} In this study, we use molecular simulations and nucleation theory to elucidate how the size, shape, strength of binding to ice, and aggregation of ice-binding proteins determine the temperature at which they nucleate ice.

We demonstrate that the antifreeze protein *TmAFP* uses the same ice-binding surface to halt the growth of ice¹⁸ and to promote its nucleation. Our simulations indicate that *TmAFP* nucleates ice at $2 \pm 1\text{ }^{\circ}\text{C}$ above the homogeneous temperature, in quantitative agreement with very recent²¹ and past experimental determinations.²⁰ We predict that the proteins that result from lengthening the ice-binding surface of *TmAFP* by adding ice-binding loops³³ nucleate ice at warmer temperatures. Our analysis indicates that the ice-binding free energy per area, $\Delta\gamma_{\text{bind}}$, of *TmAFP* and PsINP are comparable, in agreement with what has been recently proposed from the analysis of experimental nucleation temperatures.²¹ However, as the binding site of the AFP is narrower, we predict that the *TmINP* set of proteins that result from addition of ice-binding loops to *TmAFP* achieves a smaller ice nucleation efficiency than the monomer of PsINP. To our knowledge, this is the first report of the transformation of an AFP into an INP through an increase in the number of ice-binding loops.

The ice-binding surfaces of hyperactive insect AFP and insect INP are not only finite in size, but also typically anisotropic in shape. We find that the ice nucleation temperature of the ice-binding proteins increases with the length of the ice-binding site, until it reaches a saturation length that depends on the nucleation rate, width, and ice-binding strength of the protein. More elongated surfaces do not further stabilize the critical ice nucleus, resulting in a plateauing of the ice nucleation temperatures with protein length. We conclude that ice-binding proteins must aggregate to reach the high ice nucleation temperatures reported for insect and bacterial INPs.

The simulations reveal that the ice nucleation temperature of protein aggregates is a nonmonotonous and strongly varying function of the distance between the proteins. This extreme sensitivity is independent of the molecular details of the ice-binding molecule, and is determined exclusively by matching between spacings in the ice lattice and the binding surfaces: the freezing efficiency of a protein aggregate is maximized when a critical nucleus can bind without defects or additional strain to all individual monomers. We conclude that bacteria have to exert subangstrom control of the distance between protein

monomers to achieve maximum ice nucleation efficiency. This may explain the high sensitivity of the ice nucleation temperature of bacteria to chemicals that modify the properties of their cell membranes.^{11,44,48}

We develop an iterative procedure, HINT, for the calculation of heterogeneous nucleation temperatures by finite surfaces of arbitrary sizes and binding strength using classical nucleation theory. After validating that HINT parametrized with data from the mW water model accurately reproduces the ice nucleation temperatures of the protein models in the simulations, we implement HINT parametrized with experimental data of water to predict the ice nucleation temperature of ice-binding proteins and their aggregates under experimental conditions. We predict that the INP monomer of *Ps. syringae* reaches its maximum efficiency $T_{\text{het}} = 247\text{ K}$ when the protein is 8 nm long (16 TxT loops), in excellent agreement with the $T_{\text{het}} = 248 \pm 1\text{ K}$ reported from experiments for the engineered 16-loop INP.²⁸ Moreover, we predict that the PsINP dimer is responsible for the $T_{\text{het}} = 260.5 \pm 0.5\text{ K}$ measured in experiments³⁰ (our calculations predict $T_{\text{het}} = 259\text{ K}$). It has been previously proposed that aggregates with ~ 130 INPs are needed to reach the full ice nucleation efficiency of *Ps. syringae*, 271 K.²⁷ Our calculations indicate that aggregates with, at most, 34 side-by-side INP monomers, each 40 nm long, suffice to nucleate ice at that temperature. The comparable width (61 nm) and length (40 nm) of these aggregates suggests that the length of the protein has evolved to allow the bacteria to reach this limit using only side-by-side aggregation of INPs.

While we have here focused on ice-binding proteins, the results we present and the tools we develop are relevant to interpret and predict the ice nucleation temperature of other finite biological, organic, or inorganic ice-nucleating surfaces. In particular, the HINT implementation of CNT we successfully use to predict the heterogeneous nucleation temperatures of the monomer and aggregates of the INP of *Ps. Syringae* and the AFP of *Tenebrio molitor* can be used to guide the optimization of surfaces designed for specific ice nucleation applications in areas as diverse as the seeding of clouds and cryopreservation of cells and tissues.

4. METHODS

4.1. Simulation Models and Methods. 4.1.1. Models. The lattice mismatch between the ice-binding molecules and ice is defined as $\delta a = (a_{\text{IBM}} - a_{\text{ice}})/a_{\text{ice}}$ and $\delta b = (b_{\text{IBM}} - b_{\text{ice}})/b_{\text{ice}}$, where a_{ice} and b_{ice} are the distances between water molecules along the two directions of the hexagonal lattice of ice, and a_{IBM} and b_{IBM} are the distances between the hydroxyl groups along the two directions of the ice-binding surface, as shown in Figure 1e. Table S1 lists the mismatch δa and δb of the surfaces considered in this study.

Water is modeled with the monatomic water model, mW,⁵⁴ which has been amply validated for the study of ice nucleation.^{3,16,18,26,38,39,55–72} Four related sets of ice-binding molecules are described in section 2.1 and shown in Figure 1. The united atom structure of *TmAFP* is built from its crystal structure from the Protein Data Bank (PDB ID code 1EZG)³² following ref 16. mW ice has lattice parameters that are 2% smaller than ice,³ so we follow ref 16 and scale down the coordinates of *TmAFP* in the crystal structure by 2% to maintain the experimental lattice mismatch of this proteins with respect to ice. *TmINP* is made by repeating the 12 residue loop sequence TCTNSQHCVKAN of *TmAFP* from the crystal structure 1EZG from ref 32. The distance between the Thr groups in this loop is 6.96 Å, corresponding to $\delta b = -9\%$ mismatch to the basal plane of ice. PsINP is made by repeating the 16 residues loop sequence GYGSTQTSGSESSLTA of InaZ as in refs 14 and 16. The lattice mismatch along the δb is scaled up to -7% , while the adjacent loop is

placed at a distance that produces $\delta a = 0\%$, as in ref 15. The force field for the interaction between mW water and *TmAFP* and *PsINP* has been presented in ref 16. We use the same force field for the interactions between mW and *TmINP*. We build *AlcoholINP* from a rigid monolayer of *n*-C₃₁H₆₃OH alcohols with $\delta a = +7\%$ and $\delta b = -7\%$. We truncate four methylene groups below the hydroxyl groups to create slabs of ice binding surfaces, and then trim the slab to create different shapes and sizes of *AlcoholINP*. The interactions between *AlcoholINP* and mW water are from ref 3, but with water–methylene interaction $\epsilon = 0.10$ kcal mol^{−1} and the strength of the water–OH interaction identical to that of the protein model.^{16,46}

4.1.2. Simulation Details. Molecular dynamics simulations of ice nucleation are performed using LAMMPS.⁷³ All four related sets of ice-binding molecules are simulated as rigid bodies at the united atom level (i.e., all atoms except H). The equations of motion are integrated with the velocity Verlet algorithm using a time step of 5 fs. The temperature and pressure are controlled with the Nose–Hoover thermostat and barostat with damping constants of 2.5 and 5 ps, respectively.^{74,75}

The nucleation temperature T_{het} is measured from the formation of ice, detected with CHILL+,⁷⁶ as the system is cooled at a rate of 1 K ns^{−1}. CHILL+ uses Steinhardt bond-order parameters to classify the water molecules as liquid, interfacial ice, cubic ice, and hexagonal ice.⁷⁶ To detect ice nucleation, we follow the total amount of cubic, hexagonal, and interfacial ice along each simulation trajectory. The homogeneous freezing temperature of mW water at this rate is $T_{\text{hom}} = 202 \pm 2$ K.⁵⁵ The freezing efficiency is computed as $\Delta T_f = T_{\text{het}} - T_{\text{hom}}$. To compute the freezing efficiency on a single ice-binding surface, we construct a periodic simulation box with dimensions 13 nm × 13 nm × 8 nm containing 42 665 water molecules and a single ice-nucleating molecule. It should be noted that aggregation does not interfere with the determinations of T_{het} through molecular simulations, because there is a single protein in the periodic simulation box. The freezing efficiency of the dimers as a function of their distance is computed in a simulation box containing 40 700 water molecules and a pair of 12 nm long *TmINP* or 11 nm long *AlcoholINP* dimers. Simulations of monomers and dimers are carried out in the *NpT* ensemble. The error bar on each reported ΔT_f is computed from five independent simulations.

We determine the critical size of the ice nucleus on the 5 nm long *TmINP* as the one that has 50% probability to commit to the crystal basin.³⁷ We use a simulation box 13 nm × 13 nm × 8 nm that contains 42 665 water molecules and the *TmINP*. To compute the committor probability, we collect 24 different configurations of the ice nucleus on this 5 nm long *TmINP* and randomize the momenta of water molecules to create for each configuration 20 1 ns long *NpT* trajectories with temperature 220 K. If the ice cluster contains more than 2500 water molecules at the end of the trajectory, we count the event as crystallization. The probability of crystallization of each ice cluster is computed from the total number of crystallization trajectory $N_{\text{crystallization}}$, $P = N_{\text{crystallization}}/20$.

4.2. Prediction of the Ice Nucleation Temperature of Proteins and Their Aggregates Using Classical Nucleation Theory. **4.2.1. Procedure To Compute the Heterogeneous Ice Nucleation Temperature (HINT) of Extended Surfaces.** Classical nucleation theory (CNT)²⁵ is a quasi-equilibrium theory that provides a relationship between the rate of nucleation and the reversible work ΔG^* required to create a critical nucleus of the new phase. CNT expresses the nucleation rate ω as^{25,41}

$$\omega(T) = A(T) \times \exp(-\Delta G^*(T)/k_B T) \quad (1)$$

where k_B is Boltzmann's constant, T is the temperature, $A(T)$ is a kinetic prefactor that depends mostly on the diffusion coefficient $D(T)$ of the liquid and the number of sites N where nucleation can occur (which is proportional to the volume of the water sample in homogeneous nucleation and to the active area of the ice nucleant in heterogeneous nucleation), and $\Delta G^*(T)$ is the nucleation barrier, which can be computed from equilibrium properties. The free energy barrier for homogeneous nucleation through a spherical nucleus is

$$\Delta G_{\text{hom}}^* = 16\pi \times \gamma_{\text{ice-liquid}}^3 / (3\rho^2 \times \Delta\mu^2) \quad (2)$$

where $\Delta\mu$ is the excess chemical potential of the liquid with respect to the crystal, ρ is the density of the crystal, and $\gamma_{\text{ice-liquid}}$ is the surface tension of the crystal–liquid interface. Each of these properties depend on temperature. The free energy barrier for heterogeneous nucleation of ice on a surface is

$$\Delta G_{\text{het}}^* = N_{\text{het}}^* \times \Delta\mu + A_{\text{ice-liquid}} \times \gamma_{\text{ice-liquid}} + A_{\text{ice-surface}} \times (\gamma_{\text{ice-surface}} - \gamma_{\text{liquid-surface}}) + \tau \times l \quad (3)$$

where N_{het}^* is the size of the critical nucleus, $A_{\text{ice-liquid}}$ and $A_{\text{ice-surface}}$ are the areas of the crystal–liquid and crystal–surface interfaces, $\gamma_{\text{ice-surface}}$ and $\gamma_{\text{liquid-surface}}$ are the surface tensions of crystal–surface and liquid–surface interfaces, τ is the line tension of the surface–crystal–liquid interface, and l is the length of the contact line of the three-phase crystal–liquid–surface interface. Figure 9 illustrates the spherical cap geometry of the ice nucleus on a large, unlimiting nucleating surface.

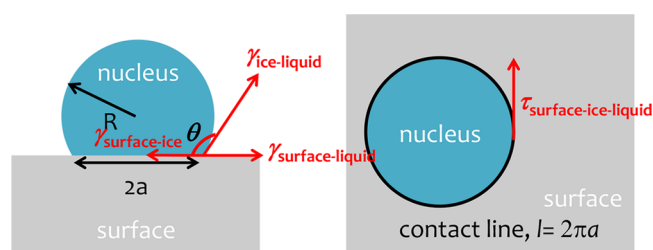


Figure 9. Sketch of the ice nucleus for heterogeneous nucleation on an unlimiting surface. The ice nucleus is shown in blue, and the nucleating surface in gray. The radius of the nucleus is R , the radius of the nucleus base is a , the contact angle of the nucleus is θ , and the contact line of crystal–liquid–surface interface l is the black perimeter. Red arrows indicate the directions of the surface tensions and the line tension.

We define the binding free energy per unit area of the crystal to the nucleating surface, $\Delta\gamma_{\text{bind}}$, as

$$\Delta\gamma_{\text{bind}} = \gamma_{\text{ice-surface}} - \gamma_{\text{liquid-surface}} - \gamma_{\text{ice-liquid}} \quad (4)$$

$\Delta\gamma_{\text{bind}}$ has units of mJ m^{−2}, and is directly related to ΔG_{bind} of ref 3, which is a free energy density for per mole of ice nuclei, in units of kJ mol^{−2} nm^{−2}. $\Delta\gamma_{\text{bind}} = 1$ mJ m^{−2} corresponds to $\Delta G_{\text{bind}} = 0.6022$ kJ mol^{−2} nm^{−2}. These two quantities contain the same information.

Replacing eq 4 into eq 3 results in a relationship between the binding free energy and the barrier for heterogeneous nucleation:

$$\Delta G_{\text{het}}^* = N_{\text{het}}^* \times \Delta\mu + (A_{\text{ice-liquid}} + A_{\text{ice-surface}}) \times \gamma_{\text{ice-liquid}} + A_{\text{ice-surface}} \times \Delta\gamma_{\text{bind}} + \tau \times l \quad (5)$$

We first derive the relation between the number of molecules N in the ice nucleus, the areas of the three interfaces, and the length of the contact line, assuming that the ice-nucleating surface is not limited in size and the geometry of the ice nucleus is a spherical cap. The contact angle θ of the spherical cap with respect to the nucleating surface is determined by the Young equation with the line tension correction:⁷⁷

$$\cos \theta = (\gamma_{\text{liquid-surface}} - \gamma_{\text{ice-surface}}) / \gamma_{\text{ice-liquid}} - \tau / (\gamma_{\text{ice-liquid}} \times a) \quad (6)$$

where $a = L/2\pi$ is the radius of the base of the ice nucleus (Figure 9).

By replacing $\gamma_{\text{liquid-surface}} - \gamma_{\text{ice-surface}}$ in eq 6 with eq 4, we rewrite the contact angle θ as

$$\cos \theta = -(\Delta\gamma_{\text{bind}} + \tau/a) / \gamma_{\text{ice-liquid}} - 1 \quad (7)$$

This set of equations indicate that, to determine the barrier for heterogeneous nucleation at a given temperature, we must know the properties needed to compute the homogeneous nucleation rate at that temperature, diffusivity, excess chemical potential, density, and crystal–liquid surface tension, plus properties specific to the nucleating surface: difference in the surface tensions upon binding (i.e., the binding free energy) and line tension.

In what follows, we explain the iterative procedure, which we call “Heterogeneous Ice Nucleation Temperature” or HINT, that we implement to compute T_{hom} and T_{het} , evaluated at the same nucleation rate, $\omega_{\text{hom}}(T_{\text{hom}}) = \omega_{\text{het}}(T_{\text{het}})$ (Figure 4) for a surface that is much larger than the critical nucleus size. In section 4.2.2, we explain how to extract $\Delta\gamma_{\text{bind}}$ and τ from simulation or experimental data, in section 4.2.3 we explain how to implement HINT for limited size surfaces (rectangular in the examples here, but trivially extendable to other shapes) and use it to compute the ice nucleation temperatures as a function of the size of the protein binding surface, and in section 4.2.4 we apply that procedure to compute the ice nucleation temperatures of protein aggregates.

We assume that the prefactor $A(T)$ is the same for homogeneous and heterogeneous nucleation at a given temperature, and is controlled by the diffusion coefficient in the liquid, $D(T)$. This condition could be relaxed without any loss of generality. We assume that the number N of sites for heterogeneous nucleation does not depend on temperature, which allows us to construct curves of T_{het} that all have the same reference state (in our simulations, that is about one protein per simulation cell; in the experiments of bacterial ice nucleation it is the number of bacteria per droplet in the experiments we take as reference for our calculation). Supporting Information E shows that the T_{het} values are quite insensitive to the concentration of proteins, in the absence of aggregation. It may be argued that the number of sites N is not the same for homogeneous and heterogeneous nucleation. We consider, however, that this issue is minimized by our choosing as references for homogeneous and heterogeneous nucleation droplets of the same size and cooled at the same rate, which is representative of the approach used to compare the freezing efficiencies in experiments as well as in simulations.⁵

We first set the rate ω and compute T_{hom} for that rate using the parametrizations of $D(T)$, $\gamma_{\text{ice-liquid}}(T)$, and $\Delta\mu(T)$ for the selected substance (which in this work are either real water or the mW model of water) following the procedure of ref 3. In a nutshell, we scan temperatures to find the one, T_{hom} , for which $\omega_{\text{hom}}(T_{\text{hom}})$ evaluated using eqs 1 and 2 matches the selected rate ω . Similarly, we define each heterogeneously nucleating surface by its $\Delta\gamma_{\text{bind}}$ and τ , and find T_{het} by scanning temperatures from T_{hom} to the equilibrium melting temperature T_{m} . As the size N^* and contact angle θ of the critical nucleus in heterogeneous nucleation are not known a priori, for each T we grow the nucleus and determine the number N of particles in the crystal nucleus, and for each N we determine the contact angle with eq 6. We then compute the free energy profile $\Delta G_{\text{het}}(N)$ with eq 5, from which we find the top of the free energy curve as a function of N , the reaction coordinate for homogeneous and heterogeneous nucleation,^{59–61} which corresponds to the nucleation barrier ΔG_{het}^* . The heterogeneous nucleation temperature T_{het} for ice on that surface is the temperature for which the free energy barrier computed through this procedure matches the one required from $\omega/A(T_{\text{het}})$. We scan temperatures by tuning the values of $\Delta\gamma_{\text{bind}}$ and τ . We neglect the temperature dependence of τ , and compute the temperature dependence $\Delta\gamma_{\text{bind}}$ using the relation derived in ref 39:

$$\begin{aligned}\Delta\gamma_{\text{bind}}(T_2) &= \Delta\gamma_{\text{bind}}(T_1) + \int_{T_1}^{T_2} (-\Delta S_{\text{bind}}) dT = \\ \Delta\gamma_{\text{bind}}(T_1) &+ \int_{T_1}^{T_2} (-S_{\text{s-ice}} + S_{\text{i-w}} \cos \theta - S_{\text{w-v}}) dT = \\ \Delta\gamma_{\text{bind}}(T_1) &+ \int_{T_1}^{T_2} (S_{\text{i-w}} - S_{\text{w-v}}) dT\end{aligned}\quad (8)$$

where we have assumed that water fully wets the IBS of the protein, that is, $\cos \theta = 1$ (which we verify in simulations), and that the surface

entropy of the ice–IBS interface is negligible, that is, $S_{\text{s-ice}} = 0$ (we have previously shown this approximation to be valid for the graphite–water interface³⁹). The procedure presented here is valid for the prediction of the nucleation temperature of any crystal from its melt.

To find the freezing efficiency of a large, unlimited surface using the equations above and the iterative HINT procedure sketched in Figure 4, we need to input the values and temperature dependences of the surface tensions, difference in chemical potential of the nucleus and liquid, and diffusion coefficients. This requires certain approximations, as, for the most part, these quantities have not been accurately measured for water or water models in the supercooled region. We here follow the approximations of ref 3 to compute the freezing efficiency from the binding free energy $\Delta\gamma_{\text{bind}}$ and line tension τ for the crystallization of ice with (a) water at the nucleation rate of $\omega_{\text{hom}} = 10^2 \text{ s}^{-1}$, corresponding to $T_{\text{hom}} = 238 \text{ K}$ in μL droplets; and (b) mW water models at the nucleation rate measured in the simulations with a cooling ramp of 1 K ns^{-1} used in the present study, $\omega_{\text{hom}} = 10^9 \text{ s}^{-1}$, which results in $T_{\text{hom}} = 202 \text{ K}$ in simulation cells with $\sim 10\,000$ molecules:

(i) We approximate that the critical nucleus is made of hexagonal ice. This neglects the size-dependent entropic stabilization arising from stacking disorder.⁵⁹

(ii) The difference in chemical potential between hexagonal ice and liquid, $\Delta\mu(T)$, is taken from ref 2 for water and from refs 55 and 78 for mW; the density of ice $\rho(T)$ is taken from ref 42 for water and from ref 54 for mW.

(iii) We consider that the ice–water surface tension of water at the melting temperature is $\gamma_{\text{ice-water}}(T_{\text{m}}) = 31.2 \text{ mJ m}^{-2}$, following ref 3, and for mW $\gamma_{\text{ice-water}}(T_{\text{m}}) = 35 \text{ mJ m}^{-2}$, determined by the thermodynamic integration with the Mold method.^{39,79} We note that the parametrization of ref 3 assumed $\gamma_{\text{ice-water}}(T_{\text{m}}) = 30.8 \text{ mJ m}^{-2}$ for mW water because that is the value that reproduces the rate of ice nucleation determined with forward flux simulations at 240 K in ref 80.

(iv) We approximate that the temperature dependence of the ice–liquid surface tension $\gamma_{\text{ice-water}}(T)$ is given by Turnbull’s relation,⁸¹ $\gamma_{\text{ice-water}}(T)/\gamma_{\text{ice-water}}(T_{\text{m}}) = \Delta H_{\text{m}}(T)/\Delta H_{\text{m}}(T_{\text{m}})$, where T_{m} is the equilibrium melting point of ice and ΔH_{m} is the excess enthalpy of liquid to ice. This relation has been validated for mW in ref 82.

(v) We take the temperature dependence of the diffusion coefficient of the liquid, $D(T)$, from ref 42 for water and from ref 54 for mW; we compute the prefactor $A(T)$ using eq 1 of ref 42 and eq 4 of ref 80 for mW.

(vi) The dependence of liquid–vapor surface tension with temperature, needed for the calculation of the temperature dependence of $\Delta\gamma_{\text{bind}}$, is taken from ref 83 for water and from ref 84 for mW.

4.2.2. Procedure To Determine the Line Tension and Ice-Binding Free Energy for an IBS. To determine τ and $\Delta\gamma_{\text{bind}}$, we need to know the freezing efficiency of an unlimited size surface $\Delta T_{\text{f}}^{\text{unlim}}$ in conjunction with the data of L_{min} for a limited surface of the same binding efficiency.

To find the relationship between line tension and binding free energy, we consider that if the sum of the last two terms in eq 5 is positive, the ice nucleus is less stable at the surface than fully immersed in liquid water, and the nucleation cannot proceed heterogeneously. This indicates the condition for which the surface heterogeneously nucleates ice is given by

$$A_{\text{ice-surface}} \times \Delta\gamma_{\text{bind}} + \tau \times l < 0 \quad (9)$$

When ice nucleates on a surface, such as a protein, which has a narrow ice-binding site, the width w of the base of the nucleus is the width of the ice-binding site. In that case, the minimum length L_{min} of the IBS needed to promote heterogeneous nucleation is

$$\Delta\gamma_{\text{bind}}/\tau = -l/A_{\text{min}} = -2 \times (L_{\text{min}} + w)/(L_{\text{min}} \times w) \quad (10)$$

where we have considered that the area $A_{\text{ice-surface}} = L_{\text{min}} \times w$, and the three-phase line $l = 2 \times (L_{\text{min}} + w)$. Equation 10 establishes a relationship between $\Delta\gamma_{\text{bind}}$ and τ from the width w and minimum length L_{min} of the surface that promotes ice nucleation.

$L_{\min}$, w , τ , and $\Delta\gamma_{\text{bind}}$ for the ice-nucleating molecules of this study are listed in Table 1. The width w of the IBS of *Tm*INP is assumed to be the width of the critical ice nucleus (see Figure 2 and Supp. Info. C) on the 5 nm long protein, ~ 1.3 nm. The width of the IBS in *Ps*INP in experiments is taken to be $w = 1.8$ nm, 0.2 nm larger than the distance between serine and the last threonine in the STQT binding site of the model *Ps*INP.

Equation 10 is insufficient to find the absolute values of $\Delta\gamma_{\text{bind}}$ and τ . We derive the values of both variables by combining the relation provided by eq 10 with the relations derived from the freezing efficiency of a surface large enough that does not limit the nucleus size.

To solve the individual values of line tension τ and $\Delta\gamma_{\text{bind}}$, we need the freezing efficiency of a surface that exposes an unlimited ice-binding site with the same chemistry as the protein. We use the 10 nm $\times$ 10 nm periodic threonine-cysteine-threonine (TCT) peptide surface of ref 46 to represent an infinite surface of *Tm*INP. The peptide surface is composed of 294 TCT units with the mismatch to ice of *Tm*AFP in experiments and in our simulations. The infinite surface for this peptide does not have the full backbone of the *Tm*INP protein. We call this surface “unlimited *Tm*INP”. We place a slab of liquid water containing 23 400 water molecules on top of each surface in a periodic cubic simulation cell, with the other side of the water slab exposed to vacuum, and determine from cooling ramps at 1 K ns^{-1} its freezing temperature to be $T_{\text{het}}^{\text{unlim}} = 250$ K; that is, its ice nucleation efficiency is $\Delta T_{\text{f}}^{\text{unlim}} = 48$ K.

We use the freezing efficiency $\Delta T_{\text{f}}^{\text{unlim}}$ of the unlimited *Tm*INP surfaces to read the values of $\Delta\gamma_{\text{bind}}$ (ΔT_{f} , τ) for each possible value of τ using the parametric curves shown in Figure 5. For each of these values of $\Delta\gamma_{\text{bind}}$ at $T_{\text{het}}^{\text{unlim}}$, we obtain $\Delta\gamma_{\text{bind}}$ at T_{hom} using eq 8. With these values, we compute $L_{\min}$ for *Tm*INP at T_{hom} using eq 10. If the prediction matches the $L_{\min}$ for *Tm*INP in the simulations, then the procedure is complete. We report the converged values of τ , $\Delta\gamma_{\text{bind}}$ at T_{hom} and at $T_{\text{het}}^{\text{unlim}}$ in Table 1.

For the bacterial *Ps*INP in experiments, $T_{\text{het}}^{\text{unlim}} = 271$ K, but we do not know τ nor $L_{\min}$. Hence, we assume that the line tension for *Ps*INP in experiments is 10 pN, the value we deduce for *Tm*INP using simulations, and we follow the same procedure described above for *Tm*INP to determine $\Delta\gamma_{\text{bind}} = -62.6$ mJ m^{-2} for *Ps*INP in experiments at $T_{\text{het}}^{\text{unlim}}$ (Table 1).

4.2.3. Prediction of the Saturation Length L_{sat} and Corresponding Heterogeneous Nucleation Temperature T_{sat} for Ice-Nucleating Proteins in Simulations and Experiments. We extend here the HINT procedure explained for unlimiting surfaces in section 4.2.1, to predict the length of the protein L_{sat} for which T_{het} reaches its maximum value T_{sat} for IBS of arbitrary (here exemplified with rectangular) shape. First, we assume that the shape of the ice nucleus is a cylinder with two half spherical caps at its ends (Figure 10). That

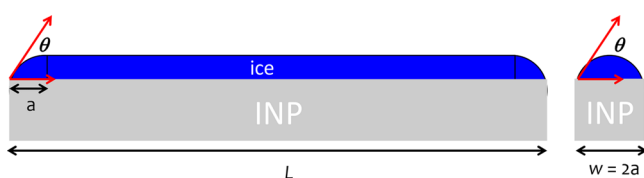


Figure 10. Sketch of the ice geometry on the protein that limits the nucleus size. We assume that the contact angle θ is the same in the two directions of the protein.

figure shows the case when $L > w$. If $L < w$, we assume the shape of the ice nucleus is a partial cylinder with the two ends formed by half spherical caps along the width of the INP, in which $L = 2a$. The width w , the binding free energy $\Delta\gamma_{\text{bind}}$, and the line tension τ of the model *Tm*INP and of *Ps*INP using experimental data are listed in Table 1. For each length L of the protein binding site, we vary the contact angle from 0 to π , and track the reversible work of forming an ice nucleus as a function of N_{ice} to find the ice nucleation barrier $\Delta G^*(T)$. We compute $\Delta G^*(T)$ for all temperatures in the range

between T_{m} and T_{hom} , until this computed $\Delta G^*(T)$ matches that we derived from the nucleation rate (see section 4.2.1). This procedure is the same sketched in Figure 4 for an unlimited surface, except for the following two conditions. First, the geometry of the ice nucleus is not a spherical cap as on the infinite surface (Figure 9), but the elongated geometry shown in Figure 10. Second, the contact angle of ice on IBS is not determined with eq 7, because in principle it can take any value between 0 and π when the edge of the ice nucleus is touching the boundary of the IBS, as we have previously deduced for pore-condensation freezing.⁴⁵ We find that the heterogeneous ice nucleation temperature T_{het} for *Tm*INP increases with L (Figure 2), until it saturates at $T_{\text{het}} = 220$ K when the length of the protein reaches $L_{\text{sat}} = 5.3$ nm. The width of the IBS of *Tm*INP is 1.3 nm. It might be possible to consider that the two directions of propagation of the nucleus have different contact angles, but that complicates the calculation of the volumes and areas, and we find already excellent agreement between theory and simulations for *Tm*INP (see Figure 2) assuming that the contact angles in the two directions are identical.

We use the same procedure to predict the maximum freezing temperature by a monomer of the bacterial INP *Ps*INP using the τ and $\Delta\gamma_{\text{bind}}$ we deduced in section 4.2.2 and listed in Table 1 (the results are shown in Figure 6), and for *Tm*AFP using two sets of binding free energies (Supp. Info. E), using HINT with experimental data for water and the experimental nucleation rates.

4.2.4. Prediction of the Maximum Nucleation Temperature of Aggregates. To compute the maximum efficiency of the aggregates with N_{INP} side-by-side INP of *Ps*INP, we repeat the same procedure assuming that the width of the IBS is proportional to the number of monomers, and we grow the length of the IBS until we find that either the freezing temperature does not increase, or the length reaches 40 nm, the maximum length of monomers for *Ps. syringae* INP.²⁸ Table S7 lists the saturation temperature T_{sat} as a function of number of monomers in the ice-nucleating aggregate.

■ ASSOCIATED CONTENT

● Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.9b01854.

Six supporting figures, seven supporting tables, and six sections that discuss: the structures of the ice-binding entities (Section A), the freezing efficiency of chimeric ice-binding proteins (Section B), committor probability of ice nuclei as a function of size (Section C), nucleation temperatures versus area of binding site for strongly and weakly binding ice-nucleating surfaces (Section D), the heterogeneous nucleation temperatures of *Tm*AFP and *Ps*INP and their aggregates predicted with classical nucleation theory (section E), and structures of *Alcohol*INP dimers (Section F) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: valeria.molinero@utah.edu.

ORCID

Yuqing Qiu: 0000-0002-0745-3155

Arpa Hudait: 0000-0002-4279-4000

Valeria Molinero: 0000-0002-8577-4675

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We gratefully acknowledge insightful comments by Claudia Marcolli, Steven E. Lindow, Michael J. Burke, and Thomas Hill. This study was supported by the National Science Foundation through award CHE-1305427 “Center for

Aerosols Impacts on Climate and the Environment". We thank the Center for High Performance Computing at The University of Utah for technical support and a grant of computing time.

REFERENCES

- (1) Murray, B. J.; Broadley, S. L.; Wilson, T. W.; Bull, S. J.; Wills, R. H.; Christenson, H. K.; Murray, E. J. Kinetics of the Homogeneous Freezing of Water. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10380–10387.
- (2) Koop, T.; Luo, B.; Tsias, A.; Peter, T. Water Activity as the Determinant for Homogeneous Ice Nucleation in Aqueous Solutions. *Nature* **2000**, *406*, 611–614.
- (3) Qiu, Y.; Odendahl, N.; Hudait, A.; Mason, R. H.; Bertram, A. K.; Paesani, F.; DeMott, P. J.; Molinero, V. Ice Nucleation Efficiency of Hydroxylated Organic Surfaces Is Controlled by Their Structural Fluctuations and Mismatch to Ice. *J. Am. Chem. Soc.* **2017**, *139*, 3052–3064.
- (4) Hartmann, S.; Augustin, S.; Clauss, T.; Wex, H.; Šantl-Temkiv, T.; Voigtländer, J.; Niedermeier, D.; Stratmann, F. Immersion Freezing of Ice Nucleation Active Protein Complexes. *Atmos. Chem. Phys.* **2013**, *13*, 5751–5766.
- (5) Amato, P.; Joly, M.; Schaupp, C.; Attard, E.; Möhler, O.; Morris, C. E.; Brunet, Y.; Delort, A. M. Survival and Ice Nucleation Activity of Bacteria as Aerosols in a Cloud Simulation Chamber. *Atmos. Chem. Phys.* **2015**, *15*, 6455–6465.
- (6) Wolber, P.; Warren, G. Bacterial Ice-Nucleation Proteins. *Trends Biochem. Sci.* **1989**, *14*, 179–182.
- (7) Warren, G.; Wolber, P. Molecular Aspects of Microbial Ice Nucleation. *Mol. Microbiol.* **1991**, *5*, 239–243.
- (8) Gurian-Sherman, D.; Lindow, S. E. Bacterial Ice Nucleation: Significance and Molecular Basis. *FASEB J.* **1993**, *7*, 1338–1343.
- (9) Wolber, P. K.; Deininger, C. A.; Southworth, M. W.; Vandekerckhove, J.; Van Montagu, M.; Warren, G. J. Identification and Purification of a Bacterial Ice-Nucleation Protein. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 7256–7260.
- (10) Hew, C. L.; Yang, D. S. Protein Interaction with Ice. *Eur. J. Biochem.* **1992**, *203*, 33–42.
- (11) Maki, L. R.; Galyan, E. L.; Chang-Chien, M.-M.; Caldwell, D. R. Ice Nucleation Induced by *Pseudomonas Syringae*. *Appl. Microbiol.* **1974**, *28*, 456–459.
- (12) Davies, P. L. Ice-Binding Proteins: A Remarkable Diversity of Structures for Stopping and Starting Ice Growth. *Trends Biochem. Sci.* **2014**, *39*, 548–555.
- (13) Bar Dolev, M.; Braslavsky, I.; Davies, P. L. Ice-Binding Proteins and Their Function. *Annu. Rev. Biochem.* **2016**, *85*, 515–542.
- (14) Pandey, R.; Usui, K.; Livingstone, R. A.; Fischer, S. A.; Pfaendtner, J.; Backus, E. H.; Nagata, Y.; Fröhlich-Nowoisky, J.; Schmäser, L.; Mauri, S. Ice-Nucleating Bacteria Control the Order and Dynamics of Interfacial Water. *Sci. Adv.* **2016**, *2*, No. e1501630.
- (15) Garnham, C. P.; Campbell, R. L.; Walker, V. K.; Davies, P. L. Novel Dimeric B-Helical Model of an Ice Nucleation Protein with Bridged Active Sites. *BMC Struct. Biol.* **2011**, *11*, 1–12.
- (16) Hudait, A.; Odendahl, N.; Qiu, Y.; Paesani, F.; Molinero, V. Ice-Nucleating and Antifreeze Proteins Recognize Ice through a Diversity of Anchored Clathrate and Ice-Like Motifs. *J. Am. Chem. Soc.* **2018**, *140*, 4905–4912.
- (17) Fletcher, G. L.; Hew, C. L.; Davies, P. L. Antifreeze Proteins of Teleost Fishes. *Annu. Rev. Physiol.* **2001**, *63*, 359–390.
- (18) Naullage, P. M.; Qiu, Y.; Molinero, V. What Controls the Limit of Supercooling and Superheating of Pinned Ice Surfaces? *J. Phys. Chem. Lett.* **2018**, *9*, 1712–1720.
- (19) Liu, K.; Wang, C.; Ma, J.; Shi, G.; Yao, X.; Fang, H.; Song, Y.; Wang, J. Janus Effect of Antifreeze Proteins on Ice Nucleation. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 14739–14744.
- (20) Modig, K.; Qvist, J.; Marshall, C. B.; Davies, P. L.; Halle, B. High Water Mobility on the Ice-Binding Surface of a Hyperactive Antifreeze Protein. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10189–10197.
- (21) Eickhoff, L.; Dreischmeier, K.; Zipori, A.; Sirotsinskaya, V.; Adar, C.; Reicher, N.; Braslavsky, I.; Rudich, Y.; Koop, T. Contrasting Behavior of Antifreeze Proteins: Ice Growth Inhibitors and Ice Nucleation Promoters. *J. Phys. Chem. Lett.* **2019**, *10*, 966–972.
- (22) Kobashigawa, Y.; Nishimiya, Y.; Miura, K.; Ohgiya, S.; Miura, A.; Tsuda, S. A Part of Ice Nucleation Protein Exhibits the Ice-Binding Ability. *FEBS Lett.* **2005**, *579*, 579.
- (23) Graether, S. P.; Jia, Z. Modeling *Pseudomonas Syringae* Ice-Nucleation Protein as a Beta-Helical Protein. *Biophys. J.* **2001**, *80*, 1169–1173.
- (24) Charpentier, T. V.; Neville, A.; Millner, P.; Hewson, R.; Morina, A. An Investigation of Freezing of Supercooled Water on Anti-Freeze Protein Modified Surfaces. *J. Bionic. Eng.* **2013**, *10*, 139–147.
- (25) Turnbull, D.; Fisher, J. C. Rate of Nucleation in Condensed Systems. *J. Chem. Phys.* **1949**, *17*, 71–73.
- (26) Lupi, L.; Hudait, A.; Molinero, V. Heterogeneous Nucleation of Ice on Carbon Surfaces. *J. Am. Chem. Soc.* **2014**, *136*, 3156–3164.
- (27) Pummer, B. G.; Budke, C.; Augustin-Bauditz, S.; Niedermeier, D.; Felgitsch, L.; Kampf, C. J.; Huber, R. G.; Liedl, K. R.; Loerting, T.; Moschen, T.; Schauerperl, M.; Tollinger, M.; Morris, C. E.; Wex, H.; Grothe, H.; Pöschl, U.; Koop, T.; Fröhlich-Nowoisky, J. Ice Nucleation by Water-Soluble Macromolecules. *Atmos. Chem. Phys.* **2015**, *15*, 4077–4091.
- (28) Ling, M.; Wex, H.; Grawe, S.; Jakobsson, J.; Löndahl, J.; Hartmann, S.; Šantl-Temkiv, T. Effects of Ice Nucleation Protein Repeat Number and Oligomerization Level on Ice Nucleation Activity. *J. Geophys. Res.* **2018**, *123*, 1802–1810.
- (29) Burke, M. J.; Lindow, S. E. Surface Properties and Size of the Ice Nucleation Site in Ice Nucleation Active Bacteria: Theoretical Considerations. *Cryobiology* **1990**, *27*, 80–84.
- (30) Govindarajan, A. G.; Lindow, S. E. Size of Bacterial Ice-Nucleation Sites Measured in Situ by Radiation Inactivation Analysis. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1334–1338.
- (31) Mueller, G. M.; Wolber, P. K.; Warren, G. J. Clustering of Ice Nucleation Protein Correlates with Ice Nucleation Activity. *Cryobiology* **1990**, *27*, 416–422.
- (32) Liou, Y. C.; Tocilj, A.; Davies, P. L.; Jia, Z. Mimicry of Ice Structure by Surface Hydroxyls and Water of a Beta-Helix Antifreeze Protein. *Nature* **2000**, *406*, 322–324.
- (33) Marshall, C. B.; Daley, M. E.; Sykes, B. D.; Davies, P. L. Enhancing the Activity of a B-Helical Antifreeze Protein by the Engineered Addition of Coils. *Biochemistry* **2004**, *43*, 11637–11646.
- (34) Popovitz-Biro, R.; Wang, J. L.; Majewski, J.; Shavit, E.; Leiserowitz, L.; Lahav, M. Induced Freezing of Supercooled Water into Ice by Self-Assembled Crystalline Monolayers of Amphiphilic Alcohols at the Air-Water Interface. *J. Am. Chem. Soc.* **1994**, *116*, 1179–1191.
- (35) Duman, J. G. Antifreeze and Ice Nucleator Proteins in Terrestrial Arthropods. *Annu. Rev. Physiol.* **2001**, *63*, 327–357.
- (36) Peters, B.; Trout, B. L. Obtaining Reaction Coordinates by Likelihood Maximization. *J. Chem. Phys.* **2006**, *125*, 054108.
- (37) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.
- (38) Naullage, P. M.; Lupi, L.; Molinero, V. Molecular Recognition of Ice by Fully Flexible Molecules. *J. Phys. Chem. C* **2017**, *121*, 26949–26957.
- (39) Qiu, Y.; Lupi, L.; Molinero, V. Is Water at the Graphite Interface Vapor-Like or Ice-Like? *J. Phys. Chem. B* **2018**, *122*, 3626–3634.
- (40) Pruppacher, H. R.; Klett, J. D. *Microphysics of Clouds and Precipitation*; Springer Science & Business Media: Dordrecht, 2010.
- (41) Kaufmann, L.; Marcolli, C.; Luo, B.; Peter, T. Refreeze Experiments with Water Droplets Containing Different Types of Ice Nuclei Interpreted by Classical Nucleation Theory. *Atmos. Chem. Phys.* **2017**, *17*, 3525–3552.

- (42) Koop, T.; Murray, B. J. A Physically Constrained Classical Description of the Homogeneous Nucleation of Ice in Water. *J. Chem. Phys.* **2016**, *145*, 211915.
- (43) Zobrist, B.; Koop, T.; Luo, B. P.; Marcolli, C.; Peter, T. Heterogeneous Ice Nucleation Rate Coefficient of Water Droplets Coated by a Nonadecanol Monolayer. *J. Phys. Chem. C* **2007**, *111*, 2149–2155.
- (44) Lindow, S.; Lahue, E.; Govindarajan, A.; Panopoulos, N.; Gies, D. Localization of Ice Nucleation Activity and the Ice Gene Product in *Pseudomonas Syringae* and *Escherichia Coli*. *Mol. Plant-Microbe Interact.* **1989**, *2*, 262–272.
- (45) David, R. O.; Marcolli, C.; Fahrni, J.; Qiu, Y.; Sirkin, Y. A. P.; Molinero, V.; Mahrt, F.; Brühwiler, D.; Lohmann, U.; Kanji, Z. A. Pore Condensation and Freezing Is Responsible for Ice Formation Below Water Saturation for Porous Particles. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, 201813647.
- (46) Hudait, A.; Moberg, D. R.; Qiu, Y.; Odendahl, N.; Paesani, F.; Molinero, V. Preordering of Water Is Not Needed for Ice Recognition by Hyperactive Antifreeze Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 8266–8271.
- (47) Kajava, A. V.; Lindow, S. E. A Model of the Three-Dimensional Structure of Ice Nucleation Proteins. *J. Mol. Biol.* **1993**, *232*, 709–717.
- (48) Wolber, P. K.; Deininger, C. A.; Southworth, M. W.; Vandekerckhove, J.; Van Montagu, M.; Warren, G. J. Identification and Purification of a Bacterial Ice-Nucleation Protein. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 7256–7260.
- (49) Kozloff, L.; Turner, M.; Arellano, F.; Lute, M. Phosphatidylinositol, a Phospholipid of Ice-Nucleating Bacteria. *J. Bacteriol.* **1991**, *173*, 2053–2060.
- (50) Hudait, A.; Qiu, Y.; Odendahl, N.; Molinero, V. Hydrogen-Bonding and Hydrophobic Groups Contribute Equally to the Binding of Hyperactive Antifreeze and Ice Nucleating Proteins to Ice. *J. Am. Chem. Soc.* **2018**, *140*, 4905.
- (51) Mochizuki, K.; Matsumoto, M. Collective Transformation of Water between Hyperactive Antifreeze Proteins: Riafcs. *Crystals* **2019**, *9*, 188.
- (52) Lee, R. E. Principles of Insect Low Temperature Tolerance. In *Insects at Low Temperatures*; Lee, R. E., Denlinger, D. L., Eds.; Chapman and Hall: London, 1991.
- (53) Yeung, K. L.; Wolf, E. E.; Duman, J. G. A Scanning Tunneling Microscopy Study of an Insect Lipoprotein Ice Nucleator. *J. Vac. Sci. Technol., B: Microelectron. Process. Phenom.* **1991**, *9*, 1197–1201.
- (54) Molinero, V.; Moore, E. B. Water Modeled as an Intermediate Element between Carbon and Silicon. *J. Phys. Chem. B* **2009**, *113*, 4008–4016.
- (55) Moore, E. B.; Molinero, V. Structural Transformation in Supercooled Water Controls the Crystallization Rate of Ice. *Nature* **2011**, *479*, 506–508.
- (56) Moore, E. B.; Molinero, V. Is It Cubic? Ice Crystallization from Deeply Supercooled Water. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20008–20016.
- (57) Moore, E. B.; Molinero, V. Ice Crystallization in Water's "No-Man's Land". *J. Chem. Phys.* **2010**, *132*, 132.
- (58) Moore, E. B.; de la Llave, E.; Welke, K.; Scherlis, D. A.; Molinero, V. Freezing, Melting and Structure of Ice in a Hydrophilic Nanopore. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4124–4134.
- (59) Lupi, L.; Hudait, A.; Peters, B.; Grünwald, M.; Gotchy Mullen, R.; Nguyen, A. H.; Molinero, V. Role of Stacking Disorder in Ice Nucleation. *Nature* **2017**, *551*, 218–222.
- (60) Lupi, L.; Hanscam, R.; Qiu, Y.; Molinero, V. Reaction Coordinate for Ice Crystallization on a Soft Surface. *J. Phys. Chem. Lett.* **2017**, *8*, 4201–4205.
- (61) Lupi, L.; Peters, B.; Molinero, V. Pre-Ordering of Interfacial Water in the Pathway of Heterogeneous Ice Nucleation Does Not Lead to a Two-Step Crystallization Mechanism. *J. Chem. Phys.* **2016**, *145*, 211910.
- (62) Hudait, A.; Qiu, S.; Lupi, L.; Molinero, V. Free Energy Contributions and Structural Characterization of Stacking Disordered Ices. *Phys. Chem. Chem. Phys.* **2016**, *18*, 9544–9553.
- (63) Lupi, L.; Molinero, V. Does Hydrophilicity of Carbon Particles Improve Their Ice Nucleation Ability? *J. Phys. Chem. A* **2014**, *118*, 7330–7337.
- (64) Lupi, L.; Kastelowitz, N.; Molinero, V. Vapor Deposition of Water on Graphitic Surfaces: Formation of Amorphous Ice, Bilayer Ice, Ice I, and Liquid Water. *J. Chem. Phys.* **2014**, *141*, 18C508.
- (65) Bi, Y.; Cao, B.; Li, T. Enhanced Heterogeneous Ice Nucleation by Special Surface Geometry. *Nat. Commun.* **2017**, *8*, 8.
- (66) Mochizuki, K.; Qiu, Y.; Molinero, V. Promotion of Homogeneous Ice Nucleation by Soluble Molecules. *J. Am. Chem. Soc.* **2017**, *139*, 17003–17006.
- (67) Bullock, G.; Molinero, V. Low-Density Liquid Water Is the Mother of Ice: On the Relation between Mesosstructure, Thermodynamics and Ice Crystallization in Solutions. *Faraday Discuss.* **2014**, *167*, 371–388.
- (68) Fitzner, M.; Sosso, G. C.; Cox, S. J.; Michaelides, A. The Many Faces of Heterogeneous Ice Nucleation: Interplay between Surface Morphology and Hydrophobicity. *J. Am. Chem. Soc.* **2015**, *137*, 13658–13669.
- (69) Johnston, J. C.; Molinero, V. Crystallization, Melting, and Structure of Water Nanoparticles at Atmospherically Relevant Temperatures. *J. Am. Chem. Soc.* **2012**, *134*, 6650–6659.
- (70) Kastelowitz, N.; Johnston, J. C.; Molinero, V. The Anomalous High Melting Temperature of Bilayer Ice. *J. Chem. Phys.* **2010**, *132*, 124511.
- (71) Johnston, J. C.; Kastelowitz, N.; Molinero, V. Liquid to Quasicrystal Transition in Bilayer Water. *J. Chem. Phys.* **2010**, *133*, 154516.
- (72) Kastelowitz, N.; Molinero, V. Ice–Liquid Oscillations in Nanoconfined Water. *ACS Nano* **2018**, *12*, 8234–8239.
- (73) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
- (74) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (75) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695–1697.
- (76) Nguyen, A. H.; Molinero, V. Identification of Clathrate Hydrates, Hexagonal Ice, Cubic Ice, and Liquid Water in Simulations: The Chill+ Algorithm. *J. Phys. Chem. B* **2015**, *119*, 9369–9376.
- (77) Young, T. An Essay on the Cohesion of Fluids. *Philos. Trans. R. Soc. London* **1805**, *95*, 65–87.
- (78) Jacobson, L. C.; Hujo, W.; Molinero, V. Thermodynamic Stability and Growth of Guest-Free Clathrate Hydrates: A Low-Density Crystal Phase of Water. *J. Phys. Chem. B* **2009**, *113*, 10298–10307.
- (79) Espinosa, J. R.; Vega, C.; Sanz, E. Ice–Water Interfacial Free Energy for the Tip4p, Tip4p/2005, Tip4p/Ice, and Mw Models as Obtained from the Mold Integration Technique. *J. Phys. Chem. C* **2016**, *120*, 8068–8075.
- (80) Li, T.; Donadio, D.; Russo, G.; Galli, G. Homogeneous Ice Nucleation from Supercooled Water. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19807–19813.
- (81) Turnbull, D. Formation of Crystal Nuclei in Liquid Metals. *J. Appl. Phys.* **1950**, *21*, 1022–1028.
- (82) Limmer, D. T.; Chandler, D. Phase Diagram of Supercooled Water Confined to Hydrophilic Nanopores. *J. Chem. Phys.* **2012**, *137*, 044509.
- (83) Floriano, M.; Angell, C. Surface Tension and Molar Surface Free Energy and Entropy of Water to 272. Degree. *C. J. Phys. Chem.* **1990**, *94*, 4199–4202.
- (84) Baron, R.; Molinero, V. Water-Driven Cavity-Ligand Binding: Comparison of Thermodynamic Signatures from Coarse-Grained and Atomic-Level Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 3696–3704.