How Do Size and Aggregation of Ice-binding Proteins Control their Ice Nucleation Efficiency

Yuqing Qiu, Arpa Hudait, and Valeria Molinero*

Department of Chemistry, The University of Utah

315 South 1400 East, Salt Lake City, Utah 84112-0580, USA

corresponding author, email: Valeria.Molinero@utah.edu

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 do 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 *Tm*AFP 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 *Tm*AFP increases T_{het} until the length of the binding-site becomes ~4 times its width, beyond which T_{het} plateaus. 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_{het} = -2 °C characteristic of this bacterium. Interestingly, we find that T_{het} is a strongly varying non-monotonic function of the distance between proteins in the aggregates. This indicates that to achieve maximum freezing efficiency, bacteria must exert exquisite, sub-angstrom control of the distance between INP in their membrane.

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 a non-conserved 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 ice[17] 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] Nevertheless, there is not yet a quantitative, predictive understanding of how does the heterogeneous ice nucleation temperature T_{het} depend on the size, shape, and strength of ice binding of the nucleating surface. Elucidating that dependence is the focus of the present 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*, *Ps*INP, with 16 TxT loops (about a ¼ of the native size) nucleate ice at T_{het} = -25±1 °C, just 10±1 °C above the homogeneous nucleation temperature T_{hom}.[28] The dependence of the freezing efficiency $\Delta T_f = T_{het} - T_{hom}$ 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 concentration 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 does 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.

METHODS

Models. The lattice mismatch between IBM and ice is defined as $\delta a = (a_{IBM}-a_{ice})/a_{ice}$ and $\delta b = (b_{IBM}-b_{ice})/b_{ice}$, where a_{ice} and b_{ice} are the distances between water molecules along the two direction of the hexagonal lattice of ice, and a_{IBM} and b_{IBM} are the distances between the hydroxyl groups along the two direction of the hexagonal lattice of IBS (Figure 1e).

Water is modeled with monoatomic water model, mW,[32] which has been amply validated for the study of ice nucleation.[3, 16, 18, 26, 33-52] Four related sets of ice-binding molecules are described in Section A and shown in Figure 1. The united atom structure of *Tm*AFP is built from its crystal structure from the Protein Data Bank (PDB ID code1EZG)[53] following ref [16]. mW ice has a lattice that is 2% smaller than ice,[3] so we follow ref [16] and scale down the coordinates of *Tm*AFP crystal structure by 2% to maintain the experimental lattice mismatch of this proteins with respect to ice. *Tm*INP is made by repeating the 12 residue loop sequence TCTNSQHCVKAN of *Tm*AFP from the crystal structure 1EZG from ref [53]. The distance between the Thr groups in this loop is 6.96 Å, corresponding to δb = -9% mismatch to basal plane of ice. *Ps*INP is made by repeating the 16 residues loop sequence GYGSTQTSGSESSLTA of InaZ as in refs [14, 16]. The lattice mismatch along the δb is scaled up to -7% as in ref [15], while the adjacent loops is placed at a distance that produces δa = +7%. The force field for the interaction between mW water and *Tm*AFP and *Ps*INP has been presented in ref. [16]. We use the same force field for the interactions between mW and *Tm*INP. We build *Alcohol*INP from rigid monolayer of *n*-$C_{31}H_{63}OH$ alcohols with δa = +7% and δb = -7%. We truncate four methylene groups below the hydroxyl groups to create slabs of ice binding surfaces and trim the slab to create different shapes and sizes of *Alcohol*INP. The interactions between *Alcohol*INP and mW water are from ref. [3], but with water-methylene interaction ε = 0.10 kcal mol^{-1} and the strength of the water-OH interaction identical to that of the protein model.[16, 54]

Simulation details. Molecular dynamics simulations of ice nucleation are performed using LAMMPS.[55] 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 2.5 and 5 ps, respectively.[56-57]

The nucleation temperature T_{het} is measured from the formation of ice, detected with CHILL+,[58] 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.[58] 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_{hom} =202 ± 2 K.[33] The freezing efficiency is computed as $\Delta T_f = T_{het} - T_{hom}$. 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 *Tm*INP or 11 nm long *Alcohol*INP dimer. 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 *Tm*INP as the one that has 50% probability to commit to the crystal basin.[59] We use a simulation box 13 nm × 13 nm × 8 nm that contains 42665 water molecules and the *Tm*INP. To compute the committor probability, we collect 24 different configurations of the ice nucleus on this 5 nm long *Tm*INP and randomize the momenta of water molecules to create for each configuration 20 1ns long *NpT* tra-

jectories with temperature 220 K. If the ice cluster is larger than 2500 water molecule 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 $N_{\text{crystallization}}$, $P = N_{\text{crystallization}}/20$.

RESULTS AND DISCUSSION

A. Antifreeze protein *Tm*AFP nucleates ice close to the homogeneous ice nucleation temperature.

We use molecular simulations to compute the ice nucleation efficiency ΔT_f for four related sets of ice-binding molecules:

i) The antifreeze protein *Tm*AFP of the beetle *Tenebrio molitor*,[53] , 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 activity of the protein.

ii) Ice-binding proteins made by repeating N times the 12 residues loop TCTNSQHCVKAN that encompasses residues 27 to 38 from the N-terminus of *Tm*AFP.[53] 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. [60]. 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 GYGSTQTSGSESSLTA 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_f 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[53], to compare the size-dependence of ΔT_f of a purely hydrogen-bonding IBS.[61] We call these molecules *Alcohol*INP; the 23 alcohol molecules-long surface is shown in Figure 1d.

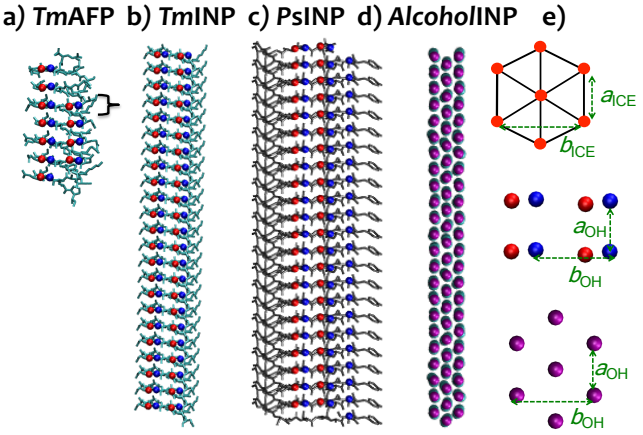


Figure 1. Ice-binding molecules of this study. a) *Tm*AFP, b) *Tm*INP with $N = 23$ loops, c) model *Ps*INP with $N = 22$ loops, and d) *Alcohol*INP with 23 alcohol molecules per row. The square bracket in a) indicates the loop of *Tm*AFP that we repeat to produce *Tm*INP. the Red and blue balls are the methyl and hydroxyl groups of the IBS. The backbones of *Tm*AFP and *Tm*INP are shown in blue, and the backbone of *Ps*INP 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 to 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 of the Supporting Information lists the distance mismatches between the OH in the IBS of these models and ice.

*Tm*AFP is one of the most potent antifreeze molecules in nature.[62] Recent experiments found that *Tm*AFP nucleates ice with an efficiency that ranges from 1.3 K to 4.5 K above T_{hom} for solutions 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 E below indicates that most of this small increase in T_{het} with concentration originates from aggregation of the proteins. The 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 predictions 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, 53] 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 next section we show that an increase in the number of ice-binding loops can transform *Tm*AFP into an ice nucleating protein with efficiency comparable to the INP of *Ps. Syringae*.

B. Width of ice binding site limits T_{het} of proteins.

We now focus on the change in ice nucleation efficiency ΔT_f 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 than these proteins. The three ice-binding surfaces display the same qualitative behavior: $\Delta T_f(L)$ 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.

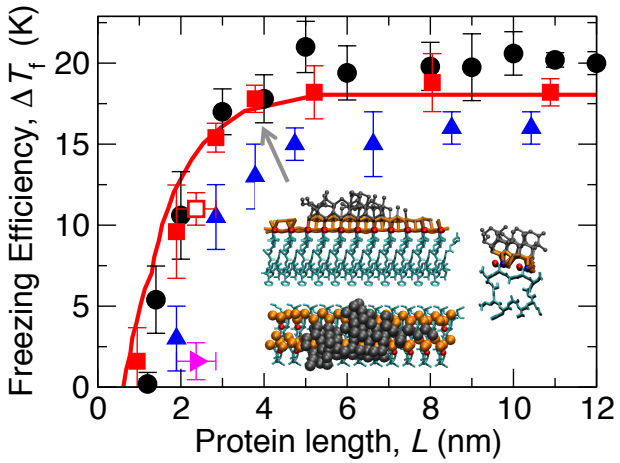


Figure 2. Ice nucleation efficiency of model proteins as a function of the length L of the ice-binding site. Symbols indicate the $\Delta T_f(L)$ computed in molecular simulations with *Ps*INP (blue up triangles), *Tm*INP (red squares), *Alcohol*INP (black circles), and *Tm*AFP (magenta right triangle). The red empty square represents ΔT_f of *Tm*INP with the same number of binding sites as *Tm*AFP. The origin of the lower efficiency of *Tm*AFP compared to *Tm*INP of the same length is discussed in Supporting Information B. The solid red line is the CNT predictions for *Tm*INP 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, $J = 10^{23}$ cm^{-3} s^{-1}. Inset: Views of the critical ice nucleus on the 5 nm long *Tm*INP: orange shows the anchored clathrate[16] and gray the rest of the ice nucleus. The critical nucleus size is identified as that with same probability to grow or melt.[59, 63] (Supporting Figure S2). Note that *Alcohol*INP and *Tm*INP have almost identical $\Delta T_f(L)$, as well as same ΔT_f for unlimited surfaces (see caption of Figure 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]

Figure 2 indicates that ice-binding molecules (IBMs) are unable to nucleate ice if they are shorter than a threshold 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 Appendix II predicts that the smallest area A_{IBS} of the binding surface that can nucleate ice is given by the condition $A_{IBS} \times \Delta\gamma_{bind} + \tau \times l_{IBS} = 0$ (Appendix eq. a9), where $\Delta\gamma_{bind} = \gamma_{ice\text{-}surface} - (\gamma_{ice\text{-}liquid} + \gamma_{liquid\text{-}surface})$ 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 show that weakly binding graphitic lamellae[37] 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 (Supporting Figure S3). The strong binding free energy of *Tm*AFP to ice[16] 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 *Tm*INP increases steeply as their IBS lengthens from L_{min} to the saturation length $L_{sat} \approx 5$ nm, which corresponds to 10 TxT loops. ΔT_f then plateaus upon lengthening of the protein. Larger mutants of *Tm*AFP comparable to *Tm*INP with up to 10 ice-binding loops have been produced in the lab, but only their thermal hysteresis activity has been determined.[60] 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 *Tm*AFP, making them comparable in efficiency to *Ps*INP.

Although experiments have shown that short, 4-loop long, fragments of *Ps*INP 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.

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 *Tm*INP, 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 *Tm*INP 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 *vs* 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 increases its 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.

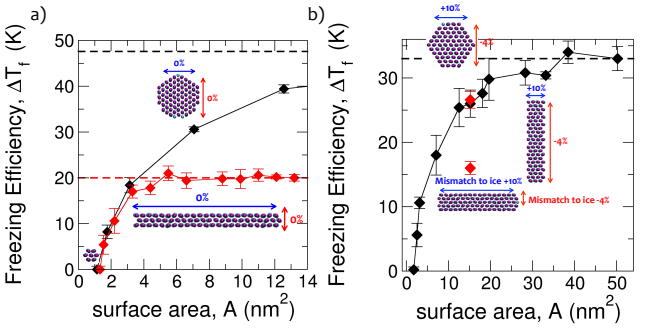


Figure 3. Effect of anisotropy in shape and in lattice mismatch to ice on the nucleation efficiency of finite surfaces. a) Freezing efficiency vs area of the binding site determined with molecular simulations of circular (black symbols) and rectangular (red symbols) fragments of alcohol monolayers with perfect matching to ice. The rectangular surfaces are three-row 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 vs area for circular (black symbols) and rectangular (red symbols) fragments of an alcohol monolayers that has different anisotropic lattice mismatch to ice in the two directions, +10% and -4%, as it is the case for *Tm*AFP in experiments[53] and in our simulations. The nucleation efficiency is highest when the smaller mismatch is aligned with the long direction of the nucleating surface.

Surfaces that bind weaker to ice reach a lower ice nucleation efficiency than those that have more negative binding free energy $\Delta\gamma_{bind}$.[3] This is the case for unlimited size surfaces (Appendix Fig. A2), as well as for surfaces that have a small IBS that limits nucleation (Supp. Fig. S3). $\Delta\gamma_{bind}$ 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 along the parallel and perpendicular axes 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 ice-nucleating molecule is, in general, not sufficient to predict its ice nucleation temperature. In 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.

C. 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 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 E, 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 *Ps. syringae.*

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, μL droplets cooled at rates of about 1 K min^{-1} nucleate ice at T_{hom} = 238 K = -35 °C.[64] Under these conditions, the experimental homogeneous nucleation rate[65] is ω_{hom} = 10^2 s^{-1}.[2, 66] T_{het} is also controlled by the cooling rate, but is modulated by the area of the nucleating surface.[65, 67] For example, 10 μl 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, 68] As T_{hot} 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, that modify the pre-exponent. Indeed, we show in Supporting Section C 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. Appendix I details the HINT procedure for unlimited surfaces, and Appendix III its implementation for nucleation on finite surfaces, such as proteins.

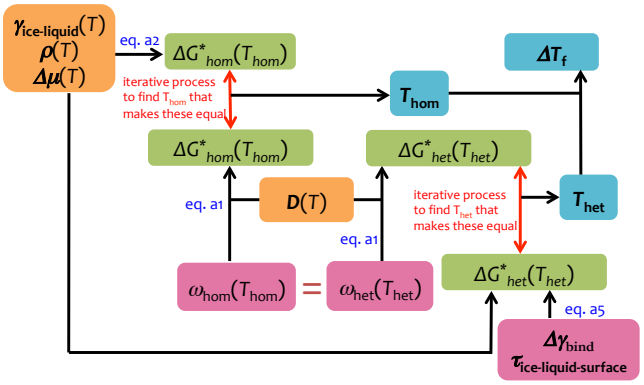


Figure 4. Schematic illustration the HINT algorithm used for the derivation of the freezing temperature ΔT_f from the binding free energy of $\Delta\gamma_{bind}$ and line tension τ. Pink boxes indicate input variables: the nucleation rate ω, the $\Delta\gamma_{bind}$ and τ that control the thermodynamics 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 $\Delta\mu$, the ice-liquid surface tension $\gamma_{ice\text{-}liquid}$. 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 bacterial *Ps*INP 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 Appendix I). 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 the Appendix, 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 contact angle of the ice nucleus, as these are not constant when the nucleus meets the boundary of the surface.[69] A detailed explanation of the method and the equations can be found in the Appendix.

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. Appendix II explains how we derive τ and $\Delta\gamma_{bind}$ from L_{min} and the freezing efficiency of the unlimiting-sized surfaces ΔT_f^{unlim}. Table 1 reports these properties for the model *Tm*INP in mW water and for the bacterial *Ps*INP 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 to compare with the ice nucleation temperatures of proteins in experiments.

Table 1. CNT predictions for ice nucleation by monomers of the model *Tm*INP using simulation data, and for the *Ps*INP protein of *Ps. syringae* using experimental data. Underlined data is input for the CNT calculation.

INP	J $cm^{-3}s^{-1}$	T_{het}^{u} K	L_{min} nm	w nm	$\Delta\gamma_{bind}$ [a] mJ^{-1} m^{-2}	τ pN	T_{hom} K	T_{sat} K
model *Tm*INP	$\underline{10^{23}}$	<u>250</u>	<u>0.47</u>	1.3	-68.1	9.5	202	220
exper. *Ps*INP	$\underline{10^{5}}$	<u>271</u>	0.51[b]	1.8	-62.6	<u>10</u>	238	247

[a] $\Delta\gamma_{bind}$ evaluated at T_{het}^{unlim}; the values at T_{hom} are -54.6 and -50.1 mJ^{-1} m^{-2} for the model *Tm*INP and experimental *Ps*INP, respectively. [b] Deduced from the value of $\Delta\gamma_{bind}/\tau$.

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 $\Delta\gamma_{bind}$, we determine the freezing efficiency of a surface of unlimiting size, ΔT_f^{unlim}, from simulations of the extended TCT peptide surfaces of ref [54]. 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 squares in Figure 2). The agreement validates the HINT algorithm for predicting T_{het} of ice-binding surfaces of arbitrary size using classical nucleation theory.

Having validated the HINT implementation of CNT against the molecular simulations, we now use the theory to predict how does 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 Appendix). With these properties, we compute T_{het} at the experimental nucleation rate that renders the homogeneous nucleation temperature T_{hom} = 238 K for microliter droplets at cooling rates of ~1 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 τ = 10 pN deduced from the simulations of *Tm*INP. Further considering that the maximum freezing temperature reported for *Ps. syringae* is 271 K,[68] we deduce $\Delta\gamma_{bind}$ = -62.6 mJ m^{-2} from the analytical CNT curves that relate T_{het} of surfaces of unlimited size to their $\Delta\gamma_{bind}$ and τ (Appendix Fig. A2a). It is noteworthy that $\Delta\gamma_{bind}$ derived from experimental data for water and *Ps. syringae* is very close to $\Delta\gamma_{bind}$ = -68.1 mJ m^{-2} of the model *Tm*INP (Table 1): both TxT-based proteins are extremely effective at binding ice. The $\Delta\gamma_{bind}$ we obtain for the bacterial protein corresponds to a zero effective contact angle of ice on the protein surface, *i.e.* 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 ~8 nm (~16 loops), in excellent agreement with the 248 ± 1 K measured for the 16 loop fragment of the INP in experiments with 10^4 proteins per droplet.[28] Supp. 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 to 80 loops does not improve its freezing efficiency (Figure 5). 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.

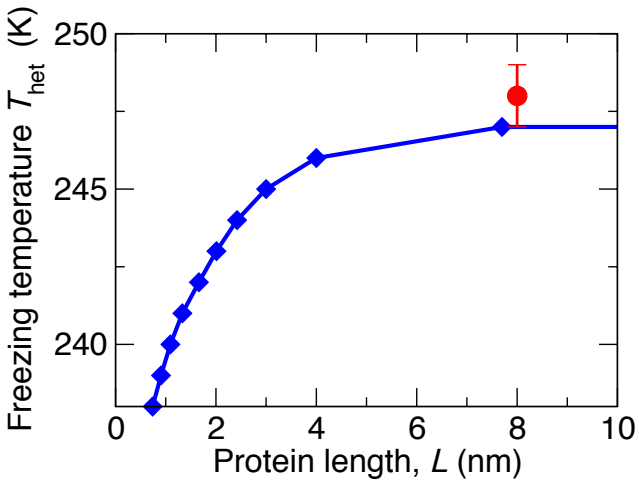


Figure 5. Ice nucleation temperature as a function of protein length for the INP monomer of *Ps. syringae* under conditions typical of laboratory experiments. Blue diamonds show the predictions of CNT using HINT for $T_{het}(L)$ of the INP of *Ps. syringae* using HINT with the experimental excess chemical potential, diffusion coefficient, ice-liquid surface tension and density of water at a nucleation rate consistent with the ~1 K min^{-1} of the experiments. The *Ps*INP surface in the HINT calculation is characterized by a width w = 1.8 nm, a line tension τ = 10 pN and $\Delta\gamma_{bind}$ = -62.6 mJ m^{-2} derived from the maximum T_{het} of *Ps. syringae*. The maximum ΔT_f of *Ps*INP is lower than for *Tm*INP in Figure 2, despite them having comparable $\Delta\gamma_{bind}$, because ΔT_f is larger for higher nucleation rates.[3] Our predictions for the 16-loop *Ps*INP monomer using CNT agree with the experimental T_{het} = 248 ± 1 K of the 16-loop variant of this protein, INpro$_{16R}$ (red circle).[28]

D. Enhancement of ice nucleation efficiency upon aggregation is non-monotonous 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, 68] It has been proposed that *Ps*INP may form aggregates by interdigitation of the monomers in the membrane.[70] 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

6a), as a function of the distance d between monomers. We find that ΔT_f is non-monotonous and highly varying with d (Figure 6b). 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 ordering and fluidity of the cell membrane,[11, 68, 71-72] which may modulate the distance between the membrane-anchored INPs.

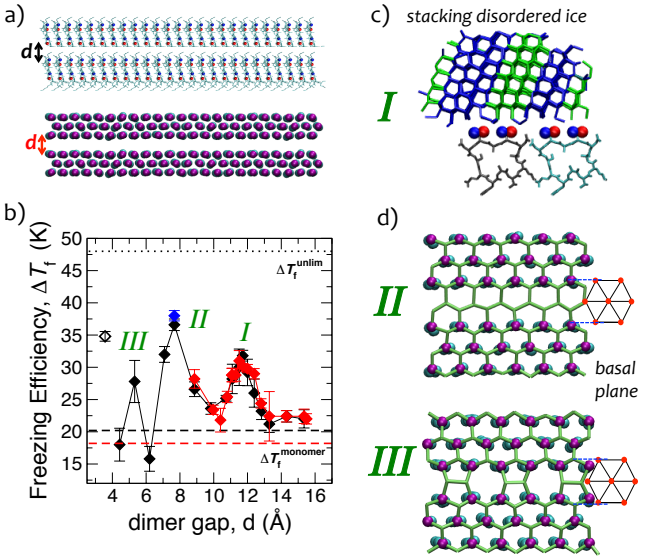


Figure 6. The freezing efficiency of a protein dimer is non-monotonous with the distance d between monomers. a) Snapshots of the 12 nm long coplanar *Tm*INP dimer (top, each monomer containing 25 TxT loops) and 11 nm long coplanar *Alcohol*INP dimer (bottom, each monomer containing 23 rows of threonines.), with same color coding as in Figure 1. b) $\Delta T_f(d)$ for the *Tm*INP dimer (red diamonds) and *Alcohol*INP dimer (black diamonds) computed in molecular simulation with nucleation rate 10^{23} cm^{-3} s^{-1}; the lines through the symbols are guides to the eye. The empty black diamond represents ΔT_f of two adjacent *Alcohol*INP with one monomer shifting and docking into the other (Supp. Information E). The ΔT_f of *Alcohol*INP with seven columns of hydroxyl binding sites (blue diamond) is almost 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 ΔT_f^{unlim} = 48 K of *Tm*INP or *Alcohol*INP of unlimiting size (they are identical). *Tm*INP cannot approach at d < 0.9 nm in our simulations with rigid protein models. c) Gray and cyan bonds represent two identical INP monomers side by side at d = 1.1 nm. Blue and green bonds are 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, *i.e.* 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 illustrate the matching between the dimer gap and the ice face.

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 the same lattice mismatch to ice as the INP (Figure 6), although alcohol monolayers hydrogen bond directly to ice[3] and the TxT binding site of proteins binds ice through an anchored clathrate motif that includes both hydrogen bonding and hydrophobic groups.[16, 54, 73] This indicates that the modulation of the freezing efficiency is not related to the details of how to 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 $d \approx 1.1$ nm (peak *I* in Figure 6). Ice nucleated by those dimers is also stacking disordered, but bound to the IBS through the 28° pyramidal face $(10\bar{1}1)$ (Figure 6c). At d = 0.8 nm (peak *II* in Figure 6b) the dimer has the optimum spacing to bind ice through the basal face (Figure 6d), 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 nm (peak *III* in Figure 6b) also nucleates ice bound through the basal plane, but destabilized by pentagonal defects (Figure 6d). At the distances where ΔT_f is a minimum, the ice nucleus develops destabilizing defects to simultaneously bind the two monomers.

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 non-monotonous function of the distance between proteins, irrespective of their orientation and ordering in the membrane.

E. Aggregates with at most three-dozen *Ps*INP monomers suffice to reach the experimental freezing efficiency of *Ps. syringae*.

Figure 6 shows that the highest ΔT_f for the dimer is still 11 K short of the freezing efficiency of an unlimiting surface, ΔT_f^{unlim}, with the same strength of binding $\Delta\gamma_{bind}$. 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 ($\Delta\gamma_{bind}$ = -62.6 mJ m^{-2} and τ = 10 pN). We assume that the width of an aggregate of N_{INP} proteins is w = 1.8 nm × N_{INP}. Figure 7 shows the ice nucleation temperature T_{het} we predict as a function of N_{INP} in the aggregates of *Ps. syringae.*

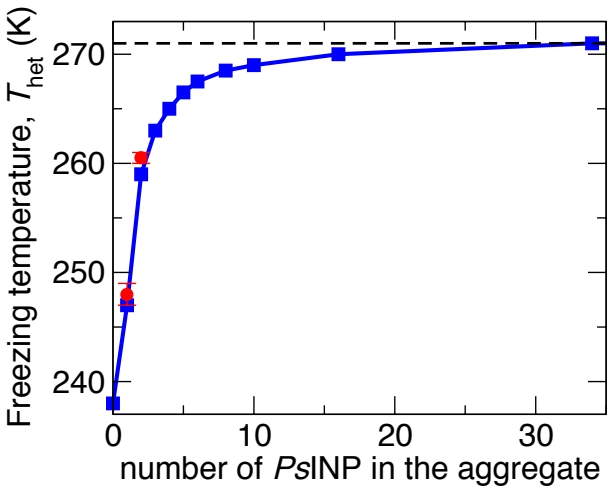


Figure 7. The freezing temperature of ice increases with the number of protein monomers in the INP aggregates. Blue points are the predictions from heterogeneous nucleation theory (see Appendix). Red points are the experimental measurements of ice freezing temperatures, T_{het} = 248 ± 1 K of the 16-loop variant of *Ps*INP, INpro$_{16R}$ taken from ref.[28] and T_{het} = 260 ± 0.5 K of minimal functional subunit of *Ps*INP taken from ref.[30]. Black dashed line is T_f = 271 K for the freezing efficiency of *Ps*INP.[11] We estimate from CNT that 34 protein monomers is needed to achieve T_f = 271 K (Supp. Table S7). If the monomers were not limited in size, we predict that a 50 nm x 50 nm IBS (28 50 nm long INP) would suffice to reach T_{het} = 271 K.

Formation of aggregates with up to 10 *Ps*INP results in significant gains in ice nucleation efficiency (Figure 7). 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_{het} = 260.5 ± 0.5 K reported in experiments[30] corresponds to the dimer, for which our CNT calculations predict 259 K. We further predict that 34-mer side-by-side *Ps*INP (a surface 61.2 nm wide and 40 nm long) suffice to reach the T_{het} = 271 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 to 20% wider than the 1.8 nm we assumed in our calculations, which would imply than only ~30 monomers are needed to reach T_{het} = 271 K, or that τ > 10 pN and, hence, $\Delta\gamma_{bind}$ < - 62.6 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 full ice nucleation efficiency 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 of ref. [21] that an increase in the concentration of *Tm*AFP in nL droplets from 0.5 to 95 μM (*i.e.* ~3 × 10^8 to 6 × 10^{10} proteins per droplet) results in a rise of T_{het} from -37 to -34°C (T_{hom} = -38 in the absence of protein in those experiments). It was not possible from the available experimental data to determine whether the raise 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 *Tm*AFP has the ice-binding strength $\Delta\gamma_{bind}$ = -62.6 mJ m^{-2} and τ = 10 pN of the INP of *Ps. syringae*, and use the HINT algorithm to predict T_{het} of *Tm*AFP at a cooling rate that produces T_{hom} = 238 K = -35°C in μL droplets (Supp. Section C1). The calculations predict T_{het} = 240 K = -33°C for droplets that contain ~10^3 to 10^7 monomers. The ΔT_f = 2 K predicted by nucleation theory is consistent with the 2 ± 1 K we find in the molecular simulations of *Tm*AFP (section A). Supp. Table S3 shows that to raise T_{het} by further 3 K, the number of monomers of *Tm*AFP per droplet should increase by 10^7, 5 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 ΔT_f = 5 K of the 2.4 mM solutions of Modig et al. [20] are mostly due to partial aggregation of the proteins to produce larger ice-nucleating surfaces. The freezing efficiencies of these concentrated solutions, however, are rather modest compared to the T_{het} = 247 and 253 K, we predict for optimal coplanar dimers and trimers, respectively, of *Tm*AFP (Supp. Table S4). Interestingly, the T_{het} predicted for the trimer is close to the maximum T_{het} = 250 K attained by functionalization of surfaces with *Tm*AFP that expose their ice-binding surface to the solution.[19] Our analysis indicates 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 disperse in solution, and are not prone to aggregate[74] 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 the extracellular water at temperatures that typically range from -4 to -8°C.[62, 75] Like bacterial INPs, these proteins are also organized in multimeric aggregates. For example, transmission electron microscopy of the lipoprotein ice nucleator (LPIN) from the hemolymph of the cranefly *Tipula trivittata*, shows that the LPIN organize into chain structures, in which strain is two protein wide, about 25 nm in width.[76] We predict that *Ps*INP aggregates of that width nucleate ice at about -3.5 °C (Figure 7), close to the -6°C ice nucleation temperature of these aggregates in the cranefly. This suggests that strength of ice binding $\Delta\gamma_{bind}$ of the insect LPIN is comparable to that of the bacterial INP.

CONCLUSIONS

Nature has evolved proteins that excel at either promoting the nucleation of ice or at 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 do 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 *Tm*AFP uses the same ice-binding surface to halt the growth of ice[18] and to promote its nucleation. Our simulations indicate that *Tm*AFP nucleates ice at 2 ± 1 °C above the homogeneous temperature, in quantitative agreement with very recent[21] and past experimental determinations.[20] We predict that the proteins that result from lengthening the ice-binding surface of *Tm*AFP by adding ice-binding loops[60] nucleate ice at warmer temperatures. Our analysis indicates that the ice-binding free energy per area, $\Delta\gamma_{bind,}$ of *Tm*AFP and *Ps*INP 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 expect that the *Tm*INP set of proteins that result from addition of ice-binding loops to *Tm*AFP achieves a comparable, but smaller ice nucleation efficiency than the monomer of *Ps*INP. To our knowledge, this is the first report of the transformation of an AFP into an efficient INP by increase in the number of ice-binding loops.

The ice-binding surface of hyperactive insect AFPs and bacterial INPs is 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 the length is almost 4 times the width of the IBS. More elongated surfaces do not further stabilize the critical ice nucleus, resulting in a plateauing of the ice nucleation temperatures with protein length. Ice-binding proteins must aggregate to reach the high ice nucleation temperatures reported for bacterial INPs.

The simulations reveal that the ice nucleation temperature of protein aggregates is a non-monotonous 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 sub-angstrom 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, 68, 71]

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 parameterized with data from the mW water model accurately reproduces the ice nucleation temperatures predicted by the simulations, we implement it parameterized with experimental data of water to predict the ice nucleation temperature of ice-binding proteins and their aggregates. We predict that the INP monomer of *Ps. syringae* reaches its maximum efficiency T_{het} = 247 K when the protein is 8 nm long (16 TxT loops), in excellent agreement with the T_{het} = 248 ± 1 K reported from experiments for this engineered 16-loop INP.[28] Moreover, we predict that the *Ps*INP dimer is responsible for the T_{het} = 260.5 ± 0.5 K measured in experiments[30] (our calculations predict T_{het} = 259 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 in-

terpret 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 *Ps. syringae* and *Tm*AFP 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.

APPENDIX. PREDICTION OF THE ICE NUCLEATION TEMPERATURE OF PROTEINS AND THEIR AGGREGATES USING CLASSICAL NUCLEATION THEORY.

I. 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 of a crystal and the reversible work ΔG^* required to create a critical nucleus of the new phase. CNT expresses the nucleation rate ω as[25, 65]

$$\omega(T) = A(T) \times \exp(-\Delta G^*(T)/k_B T), \qquad \text{(a1)}$$

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

$$\Delta G^*_{hom} = 16\,\pi \times \gamma_{ice\text{-}liquid}^3/(3\,\rho^2 \times \Delta\mu^2), \qquad \text{(a2)}$$

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

$$\Delta G^*_{het} = N^*_{het} \times \Delta\mu + A_{ice\text{-}liquid} \times \gamma_{ice\text{-}liquid} + A_{ice\text{-}surface} \times (\gamma_{ice\text{-}surface} - \gamma_{liquid\text{-}surface}) + \tau \times l, \qquad \text{(a3)}$$

where N^*_{het} is the size of the critical nucleus, $A_{ice\text{-}liquid}$ and $A_{ice\text{-}surface}$ are the areas of the crystal-liquid and crystal-surface interfaces, $\gamma_{ice\text{-}surface}$ and $\gamma_{liquid\text{-}surface}$ 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 A1 illustrates the spherical cap geometry of the ice nucleus on an unlimiting large nucleating surface.

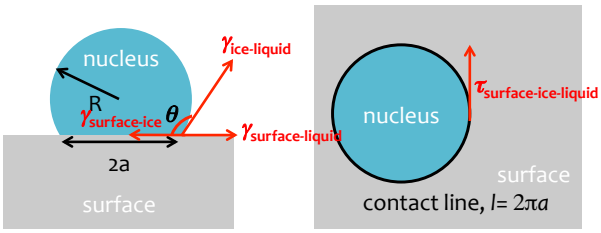


Figure A1. Sketch of the ice nucleus for heterogeneous nucleation on an unlimiting surface. The ice nucleus is shown in blue, and the nucleating surface in gray. The radius of the nucleus is R, the radius of the nucleus base is a, the contact angle of the nucleus is θ, and the contact line of crystal-liquid-surface 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 $\Delta\gamma_{bind}$ as,

$$\Delta\gamma_{bind} = \gamma_{ice\text{-}surface} - \gamma_{liquid\text{-}surface} - \gamma_{ice\text{-}liquid}. \qquad \text{(a4)}$$

$\Delta\gamma_{bind}$ 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 the unit of kJ mol^{-2} nm^{-2}. $\Delta\gamma_{bind}$ =1 mJ m^{-2} corresponds to ΔG_{bind} = 0.6022 kJ mol^{-2} nm^{-2}. These two quantities contain the same information.

Replacing eq. a4 in eq. a3 results in a relationship between the binding free energy and the barrier for heterogeneous nucleation:

$$\Delta G^*_{het} = N^*_{het} \times \Delta\mu + (A_{ice\text{-}liquid} + A_{ice\text{-}surface}) \times \gamma_{ice\text{-}liquid} + A_{ice\text{-}surface} \times \Delta\gamma_{bind} + \tau \times l, \qquad \text{(a5)}$$

We first derive the relation between the number of molecules N in the ice nucleus, the areas of the three interfaces and the length contact line, assuming that 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 Young equation with the line tension correction[77]

$$\cos\theta = (\gamma_{liquid\text{-}surface} - \gamma_{ice\text{-}surface})/\gamma_{ice\text{-}liquid} - \tau/(\gamma_{ice\text{-}liquid} \times a), \qquad \text{(a6)}$$

where $a = L/2\pi$ is the radius of the base of the ice nucleus (Figure A1).

By replacing $\gamma_{liquid\text{-}surface} - \gamma_{ice\text{-}surface}$ in eq. a6 with eq. a4, we rewrite the contact angle θ as:

$$\cos\theta = -(\Delta\gamma_{bind} + \tau/a)/\gamma_{ice\text{-}liquid} - 1. \qquad \text{(a7)}$$

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, we implement to compute T_{hom} and T_{het}, evaluated at the same nucleation rate, $\omega_{hom}(T_{hom}) = \omega_{het}(T_{het})$ (Figure 4) for a surface that is much larger than the critical nucleus size. In section II of this Appendix we explain how to extract $\Delta\gamma_{bind}$ and τ from simulation or experimental data, in section III 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 site, and in section IV we apply that procedure to compute the ice nucleation temperature of protein aggregates.

We assume that the prefactor $A(T)$ is controlled by the diffusion in the liquid, $D(T)$, and is the same for homogeneous and heterogeneous nucleation at a given temperature. The assumption that the number N of sites for heterogeneous nucleation does not depend on temperature is appropriate and allows us to construct curves of T_{het} that all build on the same reference state (in our simulations that is about one protein per simulation cell, in the experiments of bacterial ice nucleation is the number of bacteria per droplet in the experiments we take as reference for our calculation). Supporting Section C shows that the T_{het} 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 how the comparison of freezing efficiencies is performed in experiments as well as in simulations.[3]

We first set the rate ω and compute T_{hom} for that rate using the parameterizations of $D(T)$, $\gamma_{ice\text{-}liquid}(T)$, and $\Delta\mu(T)$ for the selected substance (in this work, these 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 $\omega_{hom}(T_{hom})$ evaluated using eq. a1 and a2, matches the selected rate ω. Similarly, we define each heterogeneously nucleating surface by its $\Delta\gamma_{bind}$ 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. a6. We then compute the free energy profile $\Delta G_{het}(N)$ using with eq. a5, from which we find the top of the free energy curve as a function of N –the reaction coordinate for homogeneous and heterogeneous nucleation[39-41]- 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 $\omega/A(T)$. We scan surfaces by tuning the values of $\Delta\gamma_{bind}$ and τ. We neglect the temperature dependence of τ, and compute the temperature dependence $\Delta\gamma_{bind}$ using the relation derived in ref. [37],

$$\Delta\gamma_{bind}(T_2) = \Delta\gamma_{bind}(T_1) + \int_{T_1}^{T_2}(-\Delta S_{bind})dT =$$
$$\Delta\gamma_{bind}(T_1) + \int_{T_1}^{T_2}(-S_{s-i} + S_{i-w}cos\theta - S_{w-v})dT =$$
$$\Delta\gamma_{bind}(T_1) + \int_{T_1}^{T_2}(S_{i-w} - S_{w-v})dT, \qquad (a8)$$

where we have assumed that water fully wets the IBS of the protein, *i.e.* $\cos\theta = 1$ (which we verify in simulations) and that the surface entropy of the ice-IBS interface is negligible, *i.e.* $S_{s\text{-}i} = 0$ (we have previously shown this approximation to be valid for the graphite-water interface[37]). The procedure presented here is valid for the prediction of the freezing efficiency of any crystal from its melt.

To find the freezing efficiency of a large, unlimiting 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 the follow the approximations of ref [3] to compute the freezing efficiency from the binding free energy $\Delta\gamma_{bind}$ and line tension τ for the crystallization of ice with i) water at the nucleation rate of $\omega_{hom} = 10^2$ s^{-1}, corresponding to $T_{hom} = 238$ K; and ii) mW water models at the nucleation rate measured in the simulations with a cooling ramp of 1 Kns^{-1} used in the present study, $\omega_{hom} = 10^9$ s^{-1}, which results in $T_{hom} = 202$ K:

i) We approximate that the critical nucleus is made of hexagonal ice. This neglects the size-dependent entropic stabilization arising from stacking disorder.[39]

ii) The difference in chemical potential between hexagonal ice and liquid, $\Delta\mu(T)$, is taken from ref. [2] for water and from refs. [33, 78] for mW; the density of ice $\rho(T)$ is taken from ref.[66] for water and from ref. [32] for mW.

iii) We consider that the ice-water surface tension of water at the melting temperature is $\gamma_{ice\text{-}water}(T_m) = 31.2$ mJ m^{-2}, following ref [3], and for mW $\gamma_{ice\text{-}water}(T_m) = 35$ mJ m^{-2}, determined by the thermodynamic integration with the Mold method.[37, 79] We note that the parameterization of ref. [3] assumed $\gamma_{ice\text{-}water}(T_m) = 30.8$ mJ m^{-2} 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 $\gamma_{ice\text{-}water}(T)$ is given by Turnbull's relation,[81] $\gamma_{ice\text{-}water}(T)/\gamma_{ice\text{-}water}(T_m) = \Delta H_m(T)/\Delta H_m(T_m)$,

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.[66] for water and from ref.[32] for mW; we compute the pre-factor $A(T)$ using eq. 1 in ref. [66] and eq. 4 in ref. [80] for mW.

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

Figure A2 shows the relation between the binding free energy per area of the ice-binding surface to ice, $\Delta\gamma_{bind}$, and the freezing efficiency ΔT_f of that surface for water (panel a) and 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, $\Delta\gamma_{bind}(\Delta T_f, \tau)$. The higher sensitivity of the freezing temperature to the line tension for 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 a 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] than when the rate for mW is chosen to produce the same T_{hom} = 238 K as in the experiments, the curves for $\Delta\gamma_{bind}$ vs ΔT_f for mW and water overlap.

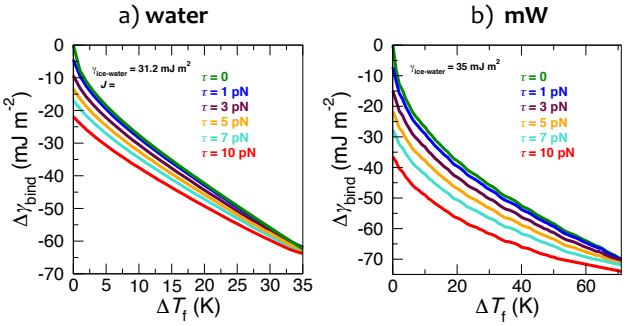


Figure A2. The curves indicate the $\Delta\gamma_{bind}$ needed to produce a freezing efficiency $\Delta T_f = T_{het} - T_{hom}$ for an unlimited, large surface at the specified nucleation rate J a) water at the typical experimental rate ω = 100 s^{-1} that produces T_{hom} = 238 K in microliter droplets and b) mW water at the rate used in the simulations ω =10^9 s^{-1} which produces T_{hom} = 202 K in simulations with ~10,000 water molecules. In each case, we report the results for various values of the ice-liquid-surface line tension τ = 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}.

II. Procedure to determine the line tension and ice-binding free energy for an IBS.

To determine τ and $\Delta\gamma_{bind}$ we need to know the freezing efficiency of an unlimited size surface ΔT_f^{unlim} in conjunction with the data of L_{min} for a limited surface of the same binding efficiency.

To find the first relationship between line tension and binding free energy, we consider that if the sum of the last two terms in eq. a5 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 that the condition for which the surface heterogeneously nucleate ice is given by

$$A_{ice\text{-}surface} \times \Delta\gamma_{bind} + \tau \times l < 0. \qquad (a9)$$

When ice nucleates on a surface -such as a protein- that 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

$$\Delta\gamma_{bind} / \tau = - l / A_{min} = - 2 \times (L_{min} + w) / (L_{min} \times w), \quad (a10)$$

where we have considered that the area $A_{ice\text{-}surface} = L_{min} \times w$, and the three-phase line $l = 2 \times (L_{min} + w)$. Eq. a6 establishes a relationship between $\Delta\gamma_{bind}$ and τ from the width w and minimum length L_{min} of the surface that promotes ice nucleation.

L_{min}, w, τ, and $\Delta\gamma_{bind}$ 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 section IV below) 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.

Eq. a6 is insufficient to find the absolute values of $\Delta\gamma_{bind}$ and τ. We derive the values of both variables by combining the relation provided by eq. a6 with the relations derived from the freezing efficiency of a surface large enough that does not limit the nucleus size, as explained section III below.

To solve the individual values of line tension τ and $\Delta\gamma_{bind}$ 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 × 10 nm periodic threonine-cysteine-threonine (TCT) peptide surface of ref. [54] 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 *Tm*INP 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 23400 water molecules on top of each that surface in a periodic cubic simulation cell, with the other side of the water slab exposed to vac-

uum, and determine from cooling ramps at 1 Kns^{-1} its freezing temperature to be T_{het}^{unlim} = 250 K, *i.e.* its ice nucleation efficiency is ΔT_f^{unlim} = 48 K.

We use the freezing efficiency ΔT_f^{unlim} of the unlimited *Tm*INP surfaces to read the values of $\Delta\gamma_{bind}(\Delta T_f, \tau)$ for each possible value of τ using the parametric curves shown in Figure A2,. For each of these values of $\Delta\gamma_{bind}$ at T_{het}^{unlim}, we obtain $\Delta\gamma_{bind}$ at T_{hom} using eq. a8. With these values we compute L_{min} for *Tm*INP at T_{hom} using eq. a10. 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, $\Delta\gamma_{bind}$ at T_{hom} and at T_{het}^{unlim} in Table 1.

For the bacterial *Ps*INP in experiments T_{het}^{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 $\Delta\gamma_{bind}$ = -62.6 mJ m^{-2} for *Ps*INP in experiments at T_{het}^{unlim} (Table 1).

III. 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 I, 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 A3). That 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 form half spherical caps along the width of the INP, in which $L = 2a$. The width w, the binding free energy $\Delta\gamma_{bind}$ 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 II above). 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 in the infinite surface (Fig. A1), but the elongated geometry shown in Fig. A3. Second, the contact angle of ice on IBS is not determined with eq. a7, because in principle it can take any value between 0 to π when the edge of the ice nucleus is touching the boundary of the IBS, as we have previously deduced for pore-condensation freezing.[69] We find that the heterogeneous ice nucleation temperature T_{het} for *Tm*INP increases with L (Figure 2), until it saturates to at T_{het} = 220 K when the length of the protein reaches L_{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.

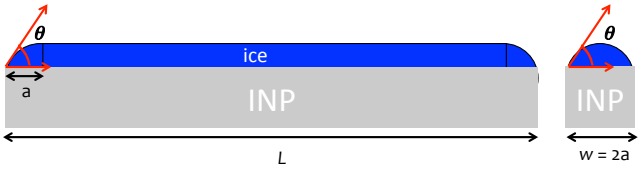


Figure A3. Sketch of the ice geometry on the protein that limits the nucleus size. We assume that the contact angle θ is the same in the two directions of the protein.

We use the same procedure to predict the maximum freezing temperature by a monomer of the bacterial INP *Ps*INP using the τ and $\Delta\gamma_{bind}$ we deduced in section II and listed in Table 1. The results are shown in Figure 5 and the first row of Supporting Table S3.

IV. Prediction of the maximum nucleation temperature of aggregates

To compute the maximum efficiency of the aggregates with N_{INP} side-by-side 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 that the length reaches 40 nm, the maximum length of monomers for *Ps. Syringae* INP.[28] Supporting Table S3 lists the saturation temperature T_{sat} as a function of number of monomers in the ice-nucleating aggregate.

Supporting Information.

The file contains seven Supporting Figures, four Supporting Tables and five sections that discuss the structures of the ice-binding entities (Section A), the freezing efficiency of chimeric ice-binding proteins (Section B), the heterogeneous nucleation temperatures of *Tm*AFP and *Ps*INP and their aggregates predicted with classical nucleation theory implemented with the HINT procedure (Section C), committor analysis of the critical nuclei (Section D), ice nucleation temperatures vs area for strongly and weakly binding ice-nucleating surfaces (Section E), and structures of *Alcohol*INP dimers (Section F).

Author Information.

Corresponding Author

*E-mail: Valeria.Molinero@utah.edu

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 awards CHE-1305427 "Center for Aerosols Impacts on Chemistry and the Environment". We thank the Center for High Performance Computing at the University of Utah for technical support and a grant of computer 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. *The FASEB journal* **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 USA* **1986,** *83*, 7256-7260.

(10) Hew, C. L.; Yang, D. S., Protein Interaction with Ice. *FEBS J.* **1992,** *203*, 33-42.

(11) R. Maki, L.; 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.; Schmüser, L.; Mauri, S., Ice-Nucleating Bacteria Control the Order and Dynamics of Interfacial Water. *Sci. Adv.* **2016,** *2*, 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 USA* **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.; Sirotinskaya, V.; Adar, C.; Reicher, N.; Braslavsky, I.; Rudich, Y.; Koop, T., Contrasting Behavior of Antifreeze Proteins: Ice Growth Inhibitors and Ice Nucleation Promotors. *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*.

(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.; Schauperl, 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 USA* **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) Molinero, V.; Moore, E. B., Water Modeled as an Intermediate Element between Carbon and Silicon. *J. Phys. Chem. B* **2009,** *113*, 4008-4016.

(33) Moore, E. B.; Molinero, V., Structural Transformation in Supercooled Water Controls the Crystallization Rate of Ice. *Nature* **2011,** *479*, 506-508.

(34) Moore, E. B.; Molinero, V., Is It Cubic? Ice Crystallization from Deeply Supercooled Water. *Phys. Chem. Chem. Phys.* **2011,** *13*, 20008-20016.

(35) Moore, E. B.; Molinero, V., Ice Crystallization in Water's "No-Man's Land". *J. Chem. Phys.* **2010,** *132*.

(36) 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.

(37) 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.

(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) 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.

(40) 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.

(41) 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.

(42) 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.

(43) Lupi, L.; Molinero, V., Does Hydrophilicity of Carbon Particles Improve Their Ice Nucleation Ability? *J. Phys. Chem. A* **2014,** *118*, 7330-7337.

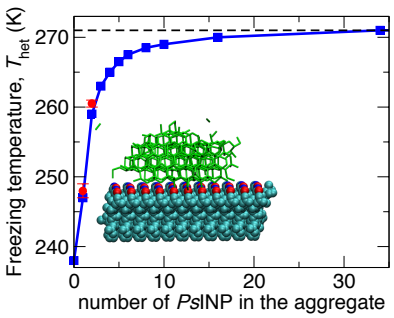
(44) Lupi, L.; Kastelowitz, N.; Molinero, V., Vapor Deposition of Water on Graphitic Surfaces: Formation of Amorphous Ice, Bilayer Ice, Ice I, and Liquid Water. *J. Chem. Phys.* **2014,** *141*, 18C508.

(45) Bi, Y.; Cao, B.; Li, T., Enhanced Heterogeneous Ice Nucleation by Special Surface Geometry. *Nat Commun.* **2017,** *8*.

(46) Mochizuki, K.; Qiu, Y.; Molinero, V., Promotion of Homogeneous Ice Nucleation by Soluble Molecules. *J. Am. Chem. Soc.* **2017,** *139*, 17003-17006.

(47) Bullock, G.; Molinero, V., Low-Density Liquid Water Is the Mother of Ice: On the Relation between Mesostructure, Thermodynamics and Ice Crystallization in Solutions. *Farad. Discuss.* **2013,** *167*, 371-388.
(48) 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.
(49) Johnston, J. C.; Molinero, V., Crystallization, Melting, and Structure of Water Nanoparticles at Atmospherically Relevant Temperatures. *J. Am. Chem. Soc.* **2012,** *134*, 6650-6659.
(50) Kastelowitz, N.; Johnston, J. C.; Molinero, V., The Anomalously High Melting Temperature of Bilayer Ice. *J. Chem. Phys.* **2010,** *132*, 124511.
(51) Johnston, J. C.; Kastelowitz, N.; Molinero, V., Liquid to Quasicrystal Transition in Bilayer Water. *J. Chem. Phys.* **2010,** *133*, 154516.
(52) Kastelowitz, N.; Molinero, V., Ice–Liquid Oscillations in Nanoconfined Water. *ACS nano* **2018,** *12*, 8234-8239.
(53) 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.
(54) 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 USA* **2018,** *115*, 8266-8271.
(55) Plimpton, S., Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995,** *117*, 1-19.
(56) Nosé, S., A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984,** *81*, 511-519.
(57) Hoover, W. G., Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985,** *31*, 1695-1697.
(58) 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.
(59) 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.
(60) 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.
(61) 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.
(62) Duman, J. G., Antifreeze and Ice Nucleator Proteins in Terrestrial Arthropods. *Annu. Rev. Physiol.* **2001,** *63*, 327-357.
(63) Peters, B.; Trout, B. L., Obtaining Reaction Coordinates by Likelihood Maximization. *J. Chem. Phys* **2006,** *125*, 054108.
(64) Pruppacher, H. R.; Klett, J. D., *Microphysics of Clouds and Precipitation*. Springer Science & Business Media: Dordrecht, 2010.
(65) Kaufmann, L.; Marcolli, C.; Luo, B.; Peter, T., Refreeze Experiments with Water Droplets Containing Different Types of Ice Nuclei Interpreted by Classical Nucleation Theory. *Atmospheric Chem. Phys.* **2017,** *17*, 3525-3552.
(66) Koop, T.; Murray, B. J., A Physically Constrained Classical Description of the Homogeneous Nucleation of Ice in Water. *J. Chem. Phys.* **2016,** *145*, 211915.
(67) 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.
(68) Lindow, S.; Lahue, E.; Govindarajan, A.; Panopoulos, N.; Gies, D., Localization of Ice Nucleation Activity and the Icec Gene Product in Pseudomonas Syringae and Escherichia Coli. *Mol. Plant-Microbe Interact* **1989,** *2*, 262-272.
(69) David, R. O.; Marcolli, C.; Fahrni, J.; Qju, 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 USA* **2019**, https://doi.org/10.1073/pnas.1813647116.
(70) Kajava, A. V.; Lindow, S. E., A Model of the Three-Dimensional Structure of Ice Nucleation Proteins. *J Mol Biol* **1993,** *232*, 709-717.
(71) 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. *Proceedings of the National Academy of Sciences* **1986,** *83*, 7256-7260.
(72) Kozloff, L.; Turner, M.; Arellano, F.; Lute, M., Phosphatidylinositol, a Phospholipid of Ice-Nucleating Bacteria. *J. Bacteriol.* **1991,** *173*, 2053-2060.
(73) 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,** *under review*.
(74) Mochizuki, K.; Matsumoto, M., Collective Transformation of Water between Hyperactive Antifreeze Proteins: Riafps. *Ctystals* **2019,** *9*, 188.
(75) 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.
(76) 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* **1991,** *9*, 1197-1201.
(77) Young, T., An Essay on the Cohesion of Fluids. *Phil. Trans. R. Soc. Lond.* **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. B* **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-27.2. 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.

TOC Figure



Supporting Information for

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

Yuqing Qiu, Arpa Hudait, and Valeria Molinero*
Department of Chemistry, 315 South 1400 East,
The University of Utah, Salt Lake City, Utah 84112-0580, USA

*corresponding author, email: Valeria.Molinero@utah.edu

A. Lattice mismatch to ice of the ice-binding molecules of this study.

Table S1 shows the lattice mismatch to ice, defined as in refs. [1-3], of the four sets of ice-binding molecules of this study. In the case of *Tm*AFP, the mismatch corresponds to an average over the different pairs of neighboring OH groups at the binding site. In the other ice-binding molecules, the mismatches are the same for all pairs of OH groups.

Table S1. Average lattice mismatch of the hydroxyl groups of the IBS with respect to the positions of water molecules in the basal plane of ice.

Protein	δa%	δb%
*Tm*AFP	+7	-7
*Tm*INP	+7	-9
*Ps*INP	+7	-7
*Alcohol*INP	+7	-7

B. Role of the backbone and OH groups at the IBS on the ice nucleation efficiency of model proteins.

To understand the gap in ice nucleation efficiency between *Tm*AFP and *Tm*INP with the same size of binding site, we produce two chimeric versions of ice nucleating proteins that allow us to evaluate the separate contributions of the chemical heterogeneity of the protein backbone and the slight positional disorder of the OH groups of the threonine residues of *Tm*AFP on the low freezing efficiency of this antifreeze protein. *Tm*AFP has two rows of threonine residues at its binding site, one has 4 Thr and the other has 7 (Figure S1), but the closest to the C-terminus does not bind to ice.[2] To make a fair comparison between *Tm*AFP and *Tm*INP, we build a protein that repeats the loop we use to produce TmINP (residues 27 to 38 of TmAFP) six times, preserving the average lattice mismatch to ice of TmAFP, but replacing the two Thr on two loops close to the N-terminus with Ala. We refer to this backbone as *Tm*INP* (Figure S1). For the first chimera, we take the IBS of *Tm*AFP and the backbone of *Tm*INP* with six repeating loops, align the binding sites Thr65 and Thr63 on the first loop of *Tm*INP* to make a chimeric INP of *Tm*AFP IBS + *Tm*INP* backbone.

For the second chimera, we take the regular IBS of *Tm*INP* containing ten binding sites (four in one column and six in the other) and the backbone of *Tm*AFP, align the first row of Thr in the place of Thr65 and Thr63 *Tm*AFP to make a chimeric INP of *Tm*INP* backbone + *Tm*AFP IBS. Figure S1 shows snapshots of the two chimeras, together with *Tm*AFP and *Tm*INP. We solvat`e each of these four ice-binding proteins in a 13 nm × 13 nm × 8 nm periodic simulation cell containing 42665 water molecules, and measure their ice freezing temperature T_{het} in a cooling ramp at the rate of 1 K ns^{-1} in the NpT ensemble at 1 bar. Table S2 reports the freezing efficiency $\Delta T_f = T_{het} - T_{hom}$, for each of the proteins (T_{hom} = 202 K at this cooling rate,[4]). The error bar of ΔT_f is computed from five independent simulations of ice nucleation with each protein.

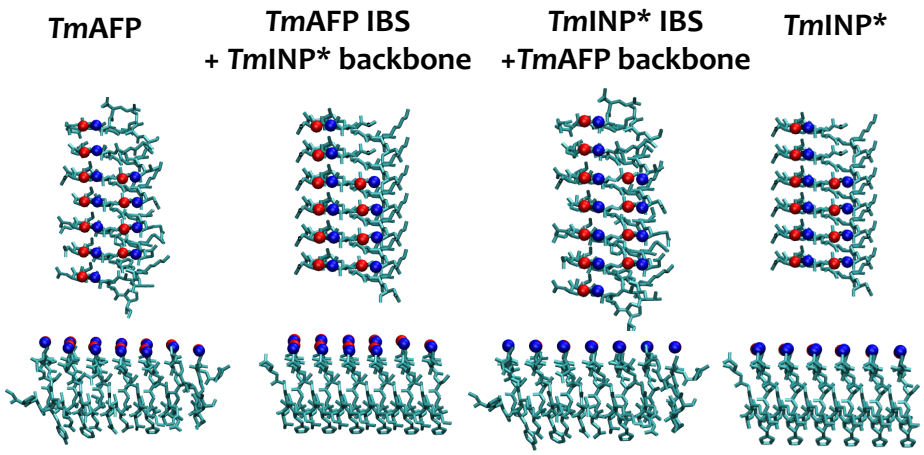


Figure S1. Snapshots of *Tm*AFP, *Tm*INP* and the mix-and-match chimeric INPs. Color code is the same as Figure 1. Cyan bonds represent the backbones. Red and blue balls are the methyl and hydroxyl groups of IBS. The sources of backbones and IBS are labeled.

*Tm*AFP nucleates ice only 2 K above of the homogeneous limit, much lower than the *Tm*INP* that exposed the same size of IBS. Table S2 suggests that is the synergism between the unevenness of the IBS of *Tm*AFP and the chemical heterogeneity in the backbones of *Tm*AFP that is responsible for the relatively low freezing efficiency of that antifreeze protein.

Table S2. The freezing efficiency ΔT_f for the four chimeric INPs in Figure S1

IBS	Backbone	ΔT_f (K)
*Tm*AFP		2 ± 1
*Tm*AFP (uneven)	*Tm*INP (repeating 6 loops)	8 ± 1
*Tm*INP (even)	*Tm*AFP (heterogeneous backbone)	9 ± 1
*Tm*INP		11 ± 1

Table S2 suggests that is the synergism between the unevenness of the IBS of *Tm*AFP and the chemical heterogeneity in the backbones of *Tm*AFP that is responsible for the relatively low freezing efficiency of that antifreeze protein. The dependence of ΔT_f on the protein backbone sequence also explains the lower ice nucleation efficiency of *Ps*INP compared to *Tm*INP (Figure 2), despite the

better matching of *Ps*INP to ice (Table S1). The results are consistent with those of a previous study of ice nucleation by alcohol monolayers, which demonstrated that both the positioning of the alkyl chains of the alcohols and the distribution and structural fluctuations of the OH groups have a large effect on the ice nucleation temperature of the monolayers.[1]

C. Ice nucleation temperatures of TmAFP, PsINP, and their aggregates predicted by classical nucleation theory implemented in HINT.

In Tables S3-36 below, $A(T)$ is the pre-exponent -which includes the number of proteins- and $\Delta G^*(T)$ the barrier at different temperatures but same nucleation rate ω, computed using CNT with the HINT algorithm. Changes in the concentration of protein are included into the pre-exponent $A(T) = A_o(T)N/N_o$, where $A_o(T)$ is the value of the pre-exponent for the reference with N_o ice nucleanting particles. Our reference number N_o of proteins per droplet is between 10^4, the density of Ps. Syringae bacteria in the experiment of ref. [5] from which we derive the maximum ice nucleation efficiency of these ice nucleating proteins, and ~1, the number of bacteria that and other studies estimate that reach this ice nucleation temperature per droplet.[5-6]

C.1. Effect of concentration and aggregation on ice nucleation by TmAFP.

The evolution of T_{het} with number N of individual *Tm*AFP per droplet shown Table S3 indicates that the 4 K increase of T_{het} upon 200-fold increase in concentration of *Tm*AFP observed in the experiments of ref. [7] cannot be explained by a mere increase in total nucleating area, and is mostly due to aggregation of *Tm*AFP to more slightly more efficient ice-nucleating surfaces (see Table S4).

We model *Tm*AFP as 1.3 nm wide and 2 nm long (4 ice-binding TxT loops), and with the same thermodynamics of ice binding as *Ps*INP deduced from experiments (τ = 10 pN and $\Delta\gamma_{bind}$ = -62.6 mJ m^2 at 272 K) at the nucleation rate $\omega = 10^2$ s^{-1} that leads to T_{hom} = 238 K in μL droplets. As the area of the 34-mer of PsINP has an area that is ~1000 times that of TmAFP, and our reference for the calculation of Thet of the proteins in experiment are droplets with 10^4 *Ps syringae*, or which only few are expected to be able to nucleate ice at -2°C,[5] we assign N_o here to a range of 10^3 to 10^7 *Tm*AFP per droplet.

Table S3. Effect of number N of *Tm*AFP per droplet on its heterogeneous nucleation temperature. The results for the reference number of proteins per droplet, N_o that ranges 10^3 to 10^7, are shown with bold font.

T_{het} (K)	$\ln(A(T_{het}))$	$\Delta G^*(T_{het})/RT$	N/N_0
240	**79.0**	**55**	**1**
241	79.3	59	50
242	79.5	65	10^4
243	79.8	71	10^7

Table S4. Heterogeneous nucleation temperature predicted with HINT for *Tm*AFP and maximum T_{het} for the dimer and the trimer The HINT calculation assumes that the model is 2 nm long and 1.3 nm wide, the dimer is 2 nm long and 2.6 nm wide, the trimer is 2 nm long and 3.9 nm wide, and that

the binding efficiency of all these proteins and assemblies are the same as those deduced from experiments of *Ps*INP (τ = 10 pN and $\Delta\gamma_{bind}$ = -62.6 mJ m^2 at 272 K). All calculations assume the same number of proteins per droplet, N_o between 10^3 to 10^7, of Table S3.

*Tm*AFP	T_{het} (K)	$\ln(A(T_{het}))$	$\Delta G^*(T_{het})/RT$
monomer	240	79.0	55
dimer	247	80.8	57
trimer	253	82.1	58

The calculations in Table S4 assume that *Tm*AFP dimers and trimmers are flat and at distances that produce a good matching to ice. Hence, they are probably an upper limit to the T_{het} of dimers and trimers of *Tm*AFP in solution. Note that the T_{het} we predict for the trimer is comparable –albeit still warmer – than the maximum T_{het} reported for an Al_2O_3 surface functionalized with *Tm*AFP exposing the ice-binding site to the solution, 250 K.[8]

C.2. Effect of concentration and aggregation on ice nucleation by PsINP

Table S5 shows that the T_{het} predicted for *Ps*INP 34-mer is not sensitive to the amount of bacteria with 34-mer aggregates per droplet. The T_{het} are computed with HINT using the geometry of the 34-mer of *Ps*INP (40 nm long by 61.2 nm wide) and the ice-binding thermodynamics we deduced for PsINP from the experimental data: τ = 10 pN and $\Delta\gamma_{bind}$ = -62.6 mJ m^2 at 272 K, and nucleation rate is $\omega = 10^2$ s^{-1} that leads to T_{hom} = 238 K in μL droplets. The reference experiment from which the T_{het} = 271 K reported 10^4 bacteria per 10 mL droplet, of which the authors conclude that just a few are able to nucleate ice at these high temperatures.[5] Hence, the reference concentration of nucleating aggregate N_o can be considered to be at least 1 and at most 10^4.

Table S5. Effect of number of bacteria per droplet with 34-mer PsINP on the heterogeneous nucleation temperature is negligible. The results for the reference concentration are shown with bold font.

T_{het} (K)	$\ln(A(T_{het}))$	$\Delta G^*(T_{het})/RT$	N/N_0
270	84.8	No barrier	Any amount of this protein aggregate nucleate ice.
271	**85.0**	**61**	**1**
272	85.1	> 5000	Nucleation is impossible at this T for any amount of this protein aggregate; a larger one is required

Table S6. Effect of number of 16-loop long *Ps*INP per droplet on the heterogeneous nucleation temperature. This protein is 8 nm long and 1.8 nm wide; the other parameters are same as in the calculations of Table S3. The results for the reference concentration are shown with bold font.

T_{het} (K)	$\ln(A(T_{het}))$	$\Delta G^*(T_{het})/RT$	N/N_0
246	80.6	48	10^{-4}
247	**80.8**	**57**	**1**
248	81.0	71	10^{6}
249	81.2	84	5×10^{11}

Table S7. Maximum nucleation temperature T_{sat} predicted with HINT for the aggregates of *Ps*INP with binding free energy deduced from the maximum freezing temperature of *Ps. syringae* in experiments assuming that each monomer contributes 1.8 nm to the width and that their length is 40 nm.

Number of *Ps*INP monomers	Width of IBS (nm)	T_{sat} (K)
1	1.8	247
2	3.6	259
3	5.4	263
4	7.2	265
5	9.0	266.5
6	10.8	267.5
8	14.4	268.5
10	18.0	269
16	28.8	270
34	61.2	271

D. Determination of the size of the critical ice nucleus using committor analyses.

We compute the committor probabilities for the ice nuclei on the L = 5 nm long *Tm*INP using the procedure described in Methods. To find the critical size we first compute the probability P_B that ice nuclei with N_{ice} water molecules crystalize (circles in Figure S5), and then fit these points with the inverse trigonometric function, *committor* $P_B = 2^*\arctan(0.012^*(N_{ice} - 90))/\pi + 0.5$ (red line in Figure S5). We find that the critical ice nucleus at $P_B = 0.5$ contains about 90 water molecules. Snapshots in Figure 2a are taken from configurations with an ice cluster of 86 water molecules (blue point of Figure S2).

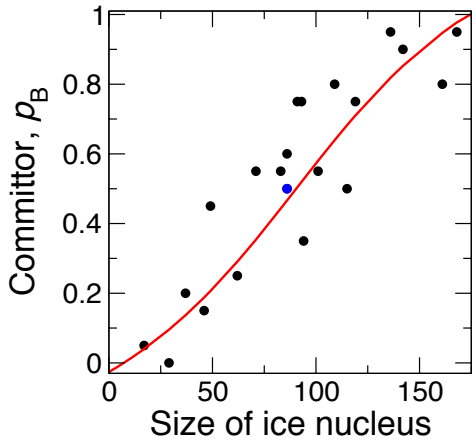


Figure S2. Committor probability P_B of the 5 nm long model *Tm*INP at 220 K increases with the size of the ice cluster. Black points are the results from the calculations and the red curve is the best fit, P_B = 2*atan(0.012*(N_{ice} - 90))/π+0.5. The blue point is the size for which the probability to commit to the liquid or ice states are identical, *i.e.*, the critical size of the ice nucleus at 220 K.

E. Ice nucleation temperature vs area for strongly and weakly ice-binding surfaces.

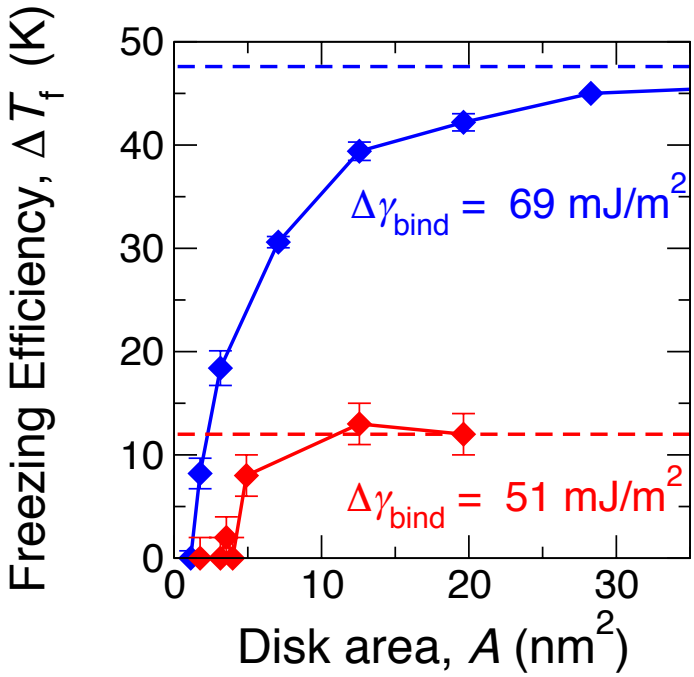


Figure S3. Freezing efficiency for disks of a strongly binding alcohol monolayer (blue diamonds) and graphite (red diamonds) as a function of the area of their ice-binding surfaces. The freezing temperatures were determined in simulations with mW water at a cooling rate of 1 K/ns with the alcohol monolayer model of ref. [1] and the graphite model of ref. [9]. The values of binding free energy to ice per area, $\Delta\gamma_{bind}$, reported for each surface in the figure were determined using the procedures explained in the Appendix. A stronger binding free energy to ice results in ice nucleation efficiency

for smaller surfaces and also higher efficiencies of the surfaces of unlimited sizes (shown with dashed lines in the same color as the data).

F. Configurations of Dimer of *Alcohol*INP.

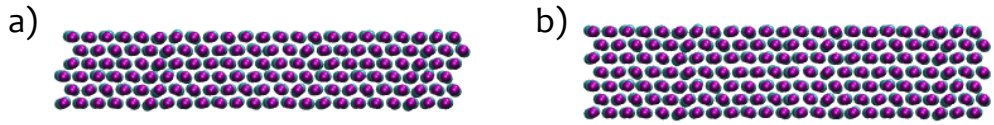


Figure S4. the snapshots of *Alcohol*INP dimers a) with one monomer shifted and b) dimers at peak II with the gap in the dimer filled with an extra column of binding sites. Color code is the same as Figure 1. Carbon tails are in cyan and hydroxyl groups are in purple.

References.

(1) 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

(2) 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 USA* **2018,** *115,* 8266-8271.

(3) 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.

(4) Moore, E. B.; Molinero, V., Structural Transformation in Supercooled Water Controls the Crystallization Rate of Ice. *Nature* **2011,** *479,* 506-508.

(5) R. Maki, L.; Galyan, E. L.; Chang-Chien, M.-M.; Caldwell, D. R., Ice Nucleation Induced by Pseudomonas Syringae. *Appl. Microbiol* **1974,** *28,* 456-459.

(6) Lindow, S.; Lahue, E.; Govindarajan, A.; Panopoulos, N.; Gies, D., Localization of Ice Nucleation Activity and the Icec Gene Product in Pseudomonas Syringae and Escherichia Coli. *Mol. Plant-Microbe Interact* **1989,** *2,* 262-272.

(7) Eickhoff, L.; Dreischmeier, K.; Zipori, A.; Sirotinskaya, V.; Adar, C.; Reicher, N.; Braslavsky, I.; Rudich, Y.; Koop, T., Contrasting Behavior of Antifreeze Proteins: Ice Growth Inhibitors and Ice Nucleation Promotors. *J. Phys. Chem. Lett.* **2019,** *10,* 966–972.

(8) 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 USA* **2016,** *113,* 14739-14744.

(9) Lupi, L.; Hudait, A.; Molinero, V., Heterogeneous Nucleation of Ice on Carbon Surfaces. *J. Am. Chem. Soc.* **2014,** *136,* 3156-3164.