Heterogeneous seeded molecular dynamics as a tool to probe the ice nucleating ability of crystalline surfaces

Philipp Pedevilla, Martin Fitzner, Gabriele C. Sosso, and Angelos Michaelides

Heterogeneous seeded molecular dynamics as a tool to probe the ice nucleating ability of crystalline surfaces

Philipp Pedevilla,[1] Martin Fitzner,[1] Gabriele C. Sosso,[2,a)] and Angelos Michaelides[1]

[1]*Thomas Young Centre, London Centre for Nanotechnology and Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom*
[2]*Department of Chemistry and Centre for Scientific Computing, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, United Kingdom*

Ice nucleation plays a significant role in a large number of natural and technological processes, but it is challenging to investigate experimentally because of the small time scales (ns) and short length scales (nm) involved. On the other hand, conventional molecular simulations struggle to cope with the relatively long time scale required for critical ice nuclei to form. One way to tackle this issue is to take advantage of free energy or path sampling techniques. Unfortunately, these are computationally costly. Seeded molecular dynamics is a much less demanding alternative that has been successfully applied already to study the homogeneous freezing of water. However, in the case of *heterogeneous* ice nucleation, nature's favourite route to form ice, an array of suitable interfaces between the ice seeds and the substrate of interest has to be built, and this is no trivial task. In this paper, we present a Heterogeneous SEEDing (HSEED) approach which harnesses a random structure search framework to tackle the ice-substrate challenge, thus enabling seeded molecular dynamics simulations of heterogeneous ice nucleation on crystalline surfaces. We validate the HSEED framework by investigating the nucleation of ice on (i) model crystalline surfaces, using the coarse-grained mW model, and (ii) cholesterol crystals, employing the fully atomistic TIP4P/ice water model. We show that the HSEED technique yields results in excellent agreement with both metadynamics and forward flux sampling simulations. Because of its computational efficiency, the HSEED method allows one to rapidly assess the ice nucleation ability of whole libraries of crystalline substrates—a long-awaited computational development in, e.g., atmospheric science. *Published by AIP Publishing.* https://doi.org/10.1063/1.5029336

I. INTRODUCTION

The nucleation of ice is the microscopic phenomenon at heart of one of the most important phase transitions on earth, that is, the freezing of water. For instance, organisms living in cold conditions need to prevent ice formation in their cells to stay alive.[1,2] The formation of ice is of relevance to atmospheric science as well: the amount of ice in clouds represents a crucial parameter in climate modeling and it also determines the extent to which solar radiation penetrates into the atmosphere.[3–5] In addition, a thorough understanding of how water freezes into ice is key to industrial applications such as cryogenic technologies,[6] fossil fuel extraction,[7] aviation,[8] and many more.

Despite its importance, it is challenging to characterise ice nucleation experimentally, due to the short time scale involved (of the order of nanoseconds), the small size of the ice nuclei (typically nanometers), and the stochastic nature of nucleation events. Molecular simulations can in principle be used to learn more about the formation of ice *in silico*, and indeed, they have recently been extensively used to get microscopic insight into the nucleation process (see, e.g., Refs. 9–16). However, the time needed for the ice nuclei to become "critical," that

is, large enough to overcome the free energy barrier preventing them to grow into actual ice crystals, is typically several orders of magnitude longer than the time scale accessible to, e.g., classical molecular dynamics (MD) simulations.[9] Direct observation of homogeneous water freezing can be achieved via brute force coarse-grained simulations (see, e.g., Ref. 17), most prominently by taking advantage of the mW model of water.[18] However, in order to nucleate ice from supercooled liquid water using fully atomistic water models, enhanced sampling methods have to be employed. Various options are available: free energy based methods such as umbrella sampling[19–23] and metadynamics[24–26] as well as path sampling methods such as transition path sampling[27,28] and forward flux sampling (FFS).[29–33]

All of these methods are computationally expensive. As an extreme example, the especially thorough investigation of homogeneous water freezing carried out by Haji-Akbari and Debenedetti[32] required ca. 21 000 000 CPU hours. This is the reason why, even by taking advantage of state-of-the-art enhanced sampling techniques, computer simulations of ice nucleation are more often than not performed only at very strong supercooling ($T_m - T = \Delta T_S \sim 40$ K, where T_m stands for the melting temperature of ice). This is sub-optimal, as making a connection between simulations and experiments requires collecting results at different temperatures—mild supercooling included. In fact, the

a)Electronic mail: G.Sosso@warwick.ac.uk

absolute values of thermodynamic and kinetic properties such as the critical nucleus size N_C^* and the ice nucleation rate,[34] respectively, are exceedingly sensitive to a number of computational details,[9] chiefly the accuracy of the water model employed,[32] so that a single absolute value of, e.g., the nucleation rate at a given supercooling is of little practical relevance.

Seeded MD (see, e.g., Ref. 35) represents one way to overcome these limitations and involves the monitoring in time of a collection of MD trajectories at different temperatures, where ice nuclei of different sizes (and possibly shapes) have been inserted into supercooled liquid water *beforehand*. At a given temperature, these ice "seeds," i.e., nuclei smaller or larger than N_C^*, would dissolve or grow, respectively, thus allowing one to pinpoint the critical nucleus size itself. This approach is computationally very efficient and thus applicable to mild supercooling. On the other hand, it does not provide direct information about the actual nucleation mechanism (how exactly water molecules come to form a critical ice nucleus), and it relies on the assumption that we can guess *a priori* the structural properties (shape, crystalline polytype, etc.) of the ice seeds. Moreover, to obtain quantities of interest to experimentalists such as the ice nucleation rate, a number of additional parameters such as the interfacial free energy between water and ice have to be used according to classical nucleation theory (CNT[34]).

The success of the seeding technique is due to the fact that the shape and the composition of the crystalline seeds is often well known *a priori*. In the case of ice, cubic ice (I_c) and hexagonal ice (I_h) are the two potential candidates, but mixtures of the two (a crystalline phase known as stacking disordered ice, I_{sd}) have also been reported in both experiments[36] and simulations.[37] Moreover, CNT assumes that the seeds have to be spherical, so as to minimise the extent of the crystal/nucleus interface. This approximation is not necessarily robust at strong supercooling.[9] However, Zaragoza *et al.*[38] found that even cubically shaped ice seeds reconstruct into a spherical morphology within a few ns of MD simulations—at mild and strong supercooling alike. In addition, I_c and I_h seeds yielded the same nucleation rate,[38] thus making seeded MD simulations a relatively straightforward computational technique to probe the homogeneous freezing of water.

In nature as well as in the laboratory, although, water almost always freezes heterogeneously, i.e., thanks to the presence of impurities that promote the kinetics of ice formation. Very diverse materials can facilitate the heterogeneous nucleation of ice, from mineral dust to birch pollen,[4] and what is it that make these substances capable of boosting the kinetics of water freezing is still not fully understood.[9] Simulations have provided useful insight into the molecular details of ice formation on a variety of different compounds, in most cases by taking advantage of the coarse-grained mW model for water.[11,12,39–45] However, addressing the freezing of water at complex interfaces, such as minerals, organic crystals, and biological matter, requires the use of fully atomistic water models in order to capture the subtleties of the hydrogen bond network in the proximity of the impurity. FFS simulations have been recently used to compute the heterogeneous ice

nucleation rate on the clay mineral kaolinite[33] using an atomistic water model, but the substantial computational cost limited the investigation to a single crystalline surface at a specific (strong) supercooling. This is why it would be desirable to extend the capabilities of seeded MD to the study of heterogeneous ice nucleation.

In this work, we present a Heterogeneous SEEDing (HSEED) approach which harnesses a random structure search (RSS) algorithm to explore the configurational space of different ice seeds sitting on arbitrary crystalline surfaces, thus enabling seeded MD simulations of heterogeneous ice nucleation. While the HSEED method does not offer the same level of detail and accuracy of free energy- and path sampling-based methods, it is orders of magnitude faster, thus allowing one to investigate different substrates at different temperatures. We demonstrate the capabilities of the HSEED method by validating its outcomes against (i) free energy (metadynamics) simulations of mW water freezing on top of Lennard-Jones (LJ) crystals and (ii) path sampling (FFS) simulations of a fully atomistic water model on cholesterol (CHL) crystals. The HSEED framework consistently pinpoints the same morphologies (in terms of, e.g., the structure, orientation, ice polytype, etc.) of the ice seeds we observe in our metadynamics (FFS) simulations of water freezing on LJ (CHL) crystals. Importantly, we show that the method allows one to obtain qualitative estimates of the critical ice nucleus size. Assuming the validity of CNT, one can thus calculate the ice nucleation rate by comparing the heterogeneous critical nucleus size with its homogeneous counterpart—albeit this comparison has to be treated with great care (as discussed in Sec. II A). Most importantly, the HSEED method can be used to rapidly screen the ice nucleating ability of whole libraries of crystalline materials and surfaces, allowing one to extract invaluable trends of practical interest for experiments and applications.

The remainder of this paper is structured as follows. The HSEED framework is illustrated in Sec. II, and we present in Sec. III the results of the method applied to mW water freezing into ice on LJ crystals (Sec. III A) and to the formation of ice (from TIP4P/ice water) on CHL crystals (Sec. III B). A discussion of the main outcomes of this work and of the potential future applications of the HSEED method can be found in Sec. IV.

II. COMPUTATIONAL METHODS

A. Heterogeneous seeded molecular dynamics

The key step of the seeded MD framework[35] is the choice/construction of the crystalline seeds. As discussed in Sec. I, this is a relatively straightforward task when dealing with homogeneous water freezing—but it becomes more challenging in the heterogeneous nucleation scenario, for the reasons outlined below.

1. The shape of the nuclei

Heterogeneous CNT relies on the assumption that crystalline nuclei of any given size are shaped as spherical caps. This is a reasonable approximation for large nuclei, where such a shape would minimise the interfacial free energy between the

ice seeds and the supercooled liquid phase. At strong super-cooling, however, where the critical nuclei can contain of the order of 10^2 molecules only, the templating effect of the substrate could lead to very anisotropic seeds. In fact, a large body of work has shown the emergence of unique water/ice-like structures forming on crystalline surfaces:[46] predicting the topology of these water clusters and/or ice-like structures on a given substrate is a challenging task. Cabriolu and Li[47] found that ice nuclei of mW water nucleating on carbonaceous surfaces can very well be approximated as spherical caps. On the other hand, we have observed a strong anisotropy in pre-critical ice nuclei forming on the clay mineral kaolinite,[33,48] albeit post-critical nuclei tended to recover the spherical cap shape. It would thus seem reasonable to build ice seeds according to the prediction of heterogeneous CNT, although nothing prevents the user from including more exotic shapes as starting points of the RSS algorithm the HSEED methodology relies upon.

2. Ice polytype and surface

According to the templating effect of a particular substrate, the heterogeneous formation of ice can proceed via I_c or I_h, and evidence of I_{sd} within the early stages of the nucleation process has also been reported.[49] Moreover, for any given polytype of ice, the particular crystalline surface with which the seed interacts with the substrate has to be chosen. Thermodynamics tells us that it is unlikely to observe high-energy (high-Miller-index) surfaces of ice forming on any crystalline substrate. Based on a comprehensive set of previous results,[43] we argue that the following surfaces are the most plausible candidates: the basal (001), primary prism (100), and secondary prism (110) of I_h and the (001) and (111) surfaces of I_c. These five options represent the starting point of our RSS algorithm. Note that one could consider including additional structures in the case of, e.g., rough crystalline surfaces or defects possibly promoting the nucleation of high-Miller-index ice surfaces. It is also worth noticing that $I_h(001)$ and $I_c(111)$ seeds expose the very same (hexagonal) plane to the substrate so that we expect the two seeds to give very similar results. However, we included them both in order to assess the impact of the structural differences between $I_h(001)$ and $I_c(111)$ which emerge within a few layers from the substrate-seed interfaces and consist in the different stacking of the said hexagonal planes [ABC for $I_c(111)$ and ABAB for $I_h(001)$[9]].

3. The ice-crystal interface

More often than not, the structure of the ice nuclei at the interface with a particular crystalline substrate has very little in common with the topology of the ice bulk phase. For instance, density functional theory calculations have shown that the layer of water molecules mediating the interaction between ice nuclei and the (001) surface of the mineral feldspar does not resemble an ice-like structure.[50] Similar results were obtained by means of classical force fields[51] and coarse-grained potentials[43,44] as well. Pinpointing the structure of the one layer or the more than one layer of water in contact with both the ice seed and the substrate is perhaps the most challenging task one has to tackle in order to extend the scope of seeded MD to heterogeneous ice nucleation. This is especially true when

specific functional groups of the substrate (such as hydroxyl groups) offer the possibility for supercooled water to form a hydrogen bond network between the substrate and ice. In this scenario, which is often observed for water in contact with a variety of potent ice nucleating agents, we have to screen as many configurations of the said hydrogen bond network as possible.

An alternative route consists of utilising the results of enhanced sampling simulations. For instance, in Ref. 48, we used metadynamics simulations to generate I_h and I_c seeds in contact with a specific crystalline surface of the clay mineral kaolinite, and the FFS simulations of Ref. 49 provided the structure of ice seeds on CHL crystals. One could thus in principle use the preliminary results of these computationally expensive methods (e.g., non-converged metadynamics runs or the initial interfaces only of the forward flux algorithm) as the starting point for seeded MD simulations, but this approach turns out to require an awful lot of computational power nonetheless.

B. The HSEED method

The HSEED methodology takes advantage instead of the RSS algorithm described in Refs. 50 and 14. A schematic of the HSEED work flow is shown in Fig. 1. We have made available via a public GitHub repository[52] a collection of (*Python*) scripts that can be used to apply the HSEED method to an arbitrary crystalline substrate.

Step 1: Spherical caps of either I_h or I_c (I_{sd} seeds can also be considered if large enough to allow for the stacking disorder to be properly represented) are built, exposing a specific low-Miller-index surface of the ice crystal (see above) to the substrate and containing a given number of water molecules. These seeds are constructed directly from bulk-ice structures fulfilling the ice rules. Seeds of different sizes can be built to study ice nucleation at different temperatures. As a rule of thumb, in the absence of any reference, the initial size of the seeds could be chosen as $\frac{N^*_{C,homo}}{2}$, i.e., half the number of water molecules contained in the homogeneous critical nucleus size at the temperature of interest; this would be the size of an ideal heterogeneous seed displaying a contact angle of ∼90° with respect to the substrate.

Step 2: The location of the seeds r_{Seed} as well as their relative orientation ϕ_{Seed} with respect to the surface of the substrate are sampled. This step is important, as specific structural features of the substrate can favor particular orientations of the ice crystals.[14,48] Then, for every $\{r_{Seed}, \phi_{Seed}\}$ combination, we generate via the RSS procedure described in Refs. 50 and 14 a substantial number (of the order of 10^3–10^4) of random configurations, varying the position and orientation of each water molecule within a certain distance (d_{HB} in Fig. 1) from the surface. This procedure allows one to explore the configurational space of the hydrogen bond network between the ice seeds and the substrate. The portion of the seed involved in the RSS typically extends up to the position of the first minimum of the density profile of water in contact with the substrate. Subsequently, the structure of the first few layers of water in contact with the surface is optimised via inexpensive algorithms such as the ℓ-BFGS,[53] keeping both the upper part of the spherical
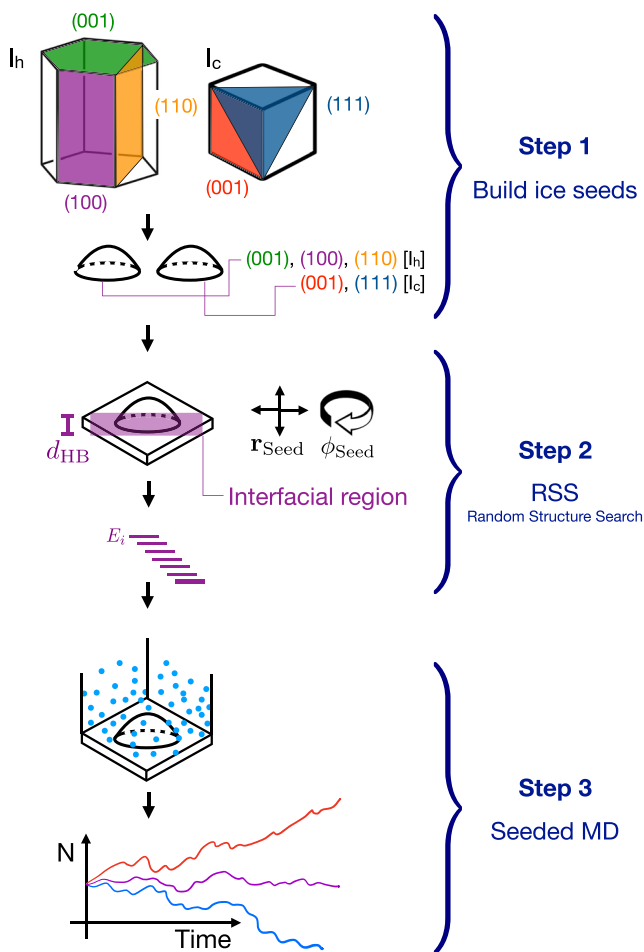
FIG. 1. Flowchart of the HSEED method. **Step 1**: From the bulk phases of I_h and I_c (I_{sd} could also be considered when building large enough seeds), spherical caps of a certain size, exposing a selection of low-Miller-index surfaces, are built. **Step 2**: By means of a random structure search (RSS) algorithm, different locations/orientations and different configurations of the ice seed-substrate interfacial region are explored. A geometry optimisation of the interfacial region of each one of these configurations is then performed, and the resulting structures are ranked according to their potential energies. **Step 3**: The "best" candidates, selected following the two criteria detailed in Sec. II A, are solvated in liquid water and then used as starting points for seeded MD simulations.

ice cap and the substrate "frozen." This is because of the large surface area of the seed-vacuum interface, which would lead to a substantial relaxation of the whole seed. Then, we select the few structures to be used as the starting point for the seeded MD runs adopting two criteria: (i) the topology of the seed should fit the structure of the surface as much as possible— i.e., the number of close contacts between the seed and surface should be kept at a minimum and (ii) the structure of the seed should be as energetically stable as possible.

Step 3: The selected configurations (the seed plus substrate) are immersed in water, and a protocol similar to the one used in the homogeneous case is used[35,54] to perform seeded MD (Step 4 in Fig. 1). This framework involves a cooling ramp, followed by an additional equilibration. Note that the entire seed is kept frozen during these preliminary MD runs, in order to equilibrate the ice/water and substrate/water interfaces without disrupting the seed-substrate interface—which we have in any case optimised beforehand. At this point, the

HSEED methodology has brought us to a situation identical to that of the homogeneous case: we are in possession of a few different ice seeds in contact with the substrate, and the time evolution of the system will be monitored by means of standard MD runs at different temperatures in order to pinpoint the critical nucleus size.

Importantly, the HSEED approach allows one to rapidly obtain information about the stability of different ice faces on a given substrate. This is crucial to heterogeneous ice formation, as being able to identify the active sites that nucleate ice on a given substrate is perhaps the most pressing challenge in the field. In fact, these active sites rarely seem to coincide with the low energy surfaces of crystalline substrates. On the mineral feldspar, for example, the active sites were recently suggested to be the high energy (100) surfaces.[14] This surface will not be exposed macroscopically on a feldspar crystal but will only be found within nanometric defects such as crystalline cracks and edges. If one wants to understand the ice nucleating efficiency of any material at a microscopic level, being able to identify where on the surface which type of ice grows is arguably the most important piece of the puzzle.

In addition, a qualitative estimate of the heterogeneous critical nucleus size can be made. However, we have recently shown[55] that CNT must be extended to take into account the heterogeneous nucleation of crystalline polytypes different from the outcome of homogeneous freezing. As such, accurate references in terms of the homogeneous critical nucleus size at different temperatures and for different polytypes are in principle needed, thus limiting the quantitative capabilities of the HSEED method. On the other hand, this technique represents a fast route toward the characterisation of the ice nucleating ability of whole libraries of crystalline compounds.

In Sec. II C, we will consider the nucleation of ice on LJ as well as CHL crystals. The former represent model substrates that allow one to extract general insight into the nucleation process, while the latter are active ice nucleating agents which have been the focus of recent experimental work.[49]

C. Molecular dynamics: Computational details

In this section, we describe the computational setup and the simulations performed on each class of substrates.

1. mW water on Lennard-Jones crystals

We considered in the first instance the heterogeneous freezing of the coarse-grained mW model for water.[18] In this case, water is represented by a single site (there are no explicit hydrogen atoms) and interacts with other water molecules via a three-body potential that favors tetrahedral order. We have taken advantage of this water model in previous studies aimed at understanding the ice nucleation capabilities of idealised[55] and hydroxylated model surfaces.[56] In order to validate our seeding approach, we have chosen two particular fcc surfaces (labeled s1 and s2) which interact with the water via a Lennard-Jones potential (details can be found in Ref. 55). In our previous work, we employed metadynamics simulations[57] to establish what sort of ice nuclei forms on the s1 and s2 surfaces at a temperature of 235 K. As we took advantage

of a collective variable (Permutation Invariant Vector, PIV[58]) which is free from bias toward any particular ice polytype or crystalline face, we have unequivocally determined that s1 and s2 promote the heterogeneous nucleation of $I_h(001)/I_c(111)$ and $I_h(100)$, respectively. We have also obtained an estimate of the critical nucleus size: 211 ± 11 and 104 ± 3 water molecules for s1 and s2, respectively.

By comparing the results of Ref. 55 to the outcomes of the HSEED approach, we will thus have the opportunity to validate both the predictive power and the accuracy of the HSEED methodology. Moreover, the mW/LJ computational setup is much less expensive compared to the simulations of ice formation on CHL crystals (see Sec. II C 2). We thus have the possibility to assess the impact on the HSEED method of intrinsic variables such as the size of the seeds and temperature. To this end, we start by performing a RSS for the five combinations of ice polytype/face considered in this study (see Sec. II A) in contact with either s1 or s2, varying the number of molecules in the seeds from 50 to 400 (in increments of 50). From the resulting dataset upon energy minimisation, we select three seeds according to the two criteria specified in Sec. II A, solvate the latter in a slab of water (so as to reach ~4000 water molecules in the whole of the simulation box), and proceed to perform twenty seeded MD runs for each seed. The production runs followed a 0.2 ns long equilibration of the systems at 273 K, where the molecules within the seeds are kept frozen, and a subsequent quenching to the target temperature. We sampled the NVT ensemble by means of a ten-fold Nosé-Hoover chain[59] with a relaxation time of 1 ps and a time step of 10 fs using the LAMMPS package.[60] As opposed to the fully atomistic water models, when dealing with mW water, the outcome of the seeding runs can almost be considered as binary in that we observe either the very rapid freezing of the whole water slab within a few nanoseconds or the complete dissolution of the seed within short time scales. We shall see in Sec. III B that in order to observe the growth of ice nuclei on CHL crystals we will need instead to monitor the seeds for as long as hundreds of nanoseconds.

2. TIP4P/ice water on cholesterol crystals

We also applied the HSEED approach to investigate ice nucleation on cholesterol monohydrate[61] (CHLM).

A single layer of CHL molecules, cleaved along the (001) plane (perpendicular to the normal to the slab), was prepared by starting from the experimental cell parameters and lattice positions.[61] Specifically, a CHLM crystal system made of two mirroring slabs (intercalated by water molecules, in a ratio of 1:1) was cleaved along the (001) plane. The triclinic symmetry of the system (space group $C1$) was preserved, and we have constructed a 3×3 supercell with in-plane dimensions of 37.17 and 36.57 Å. We positioned 1923 water molecules randomly atop this CHLM slab at the density of the TIP4P/ice model[62] at 300 K and expanded the dimension of the simulation cell along the normal to the slab to 100 Å.

10^3 structures for $I_h(001)$, $I_h(100)$, $I_h(110)$, $I_c(001)$, and $I_c(111)$ seeds were generated, each one containing ~250 water molecules. The energy minimisations were performed via the GROMACS MD package[63,64] using the CHARMM36[65,66] and TIP4P/ice[62] force fields to describe CHLM and water

molecules, respectively. A validation of this particular setup can be found in Ref. 49. According to the criteria illustrated in Sec. II A, three seeds for each ice polytype/surface [e.g., $I_h(001)$] were selected, following the outcome of the RSS procedure. These seeds have been immersed in an ~45 Å thick water slab, which resulted in simulation boxes containing ~2000 water molecules.

MD simulations have also been performed using the GRO-MACS package. The equations of motion were integrated via a leap-frog algorithm, with a time step of 2 fs. Electrostatic interactions were treated by means of a particle-mesh Ewald summation[67] with a cutoff of 12 Å. Non-bonded interactions were calculated up to 10 Å, and a switching function was used to bring them to zero at 12 Å. We sampled the NVT ensemble using a stochastic velocity rescaling thermostat[68] with a coupling constant of 2 ps. The rigid geometry of TIP4P/ice molecules was enforced, thanks to the SETTLE algorithm,[69] while additional constraints were treated via the P-LINCS algorithm.[70,71]

The equilibration of the substrate/water and ice/water interface started with a 5 ns run at 300 K, followed by a 5 ns-long cooling ramp from 300 to 200 K. A 2 ns long equilibration at 200 K followed, after which seeded MD production runs were performed at the desired target temperature by randomly selecting the initial atomic velocities according to the corresponding Maxwell-Boltzmann distribution. To verify the consistency of our results, we run simulations involving multiple seeds, different initial velocities, and different cell sizes. We shall see that the HSEED method provides a robust set of results.

III. RESULTS

A. mW water on Lennard-Jones crystals

We start by focusing on the case of mW water freezing on the LJ crystals s1 and s2 described in Sec. II C 1.

Step 1: We built ice seeds of different sizes (containing from 50 to 400 in increments of 50 water molecules) choosing five combinations of crystal polytype and face exposed to the substrate: $I_h(001)$, $I_h(100)$, $I_h(110)$, $I_c(001)$, and $I_c(111)$.

Step 2: We generated by means of our RSS algorithm between 5000 and 30 000 ice seeds for each combination of seed size, ice polytype, and ice crystalline face (see Fig. 1), exploring different locations and orientations of the ice seed on the substrate as well as optimising the geometry of the seed-substrate interface. The adsorption energy per water molecule E_{Ads} in the contact layer (i.e., within 4 Å of the substrate) for each type of ice seed (in this case, containing ~250 water molecules) as obtained upon energy minimisation is shown in Fig. 2. As expected, the spread of E_{Ads} is huge. We remark that this spread should not be considered as a source of uncertainty: on the contrary, it represents a measure of the extent to which the configurational space for a given seed has been explored. As such, a large spread is actually desirable, and the evolution of it as the RSS progresses provides an indication of the convergence of the algorithm. We also note that the lowest value of E_{Ads} found by the RSS for a given seed is the quantity that matters in determining the relative stability of different seeds—which in this respect can differ by as much
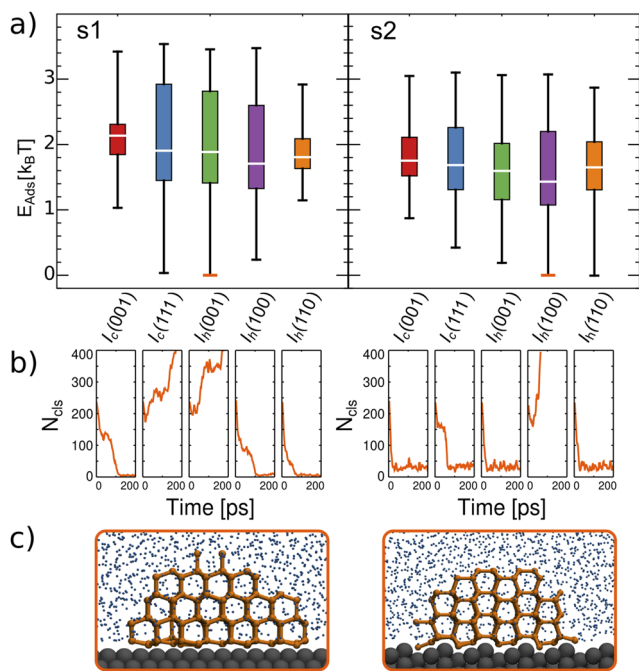
FIG. 2. (a) Adsorption energy per water molecule in the contact layer of different ice seeds (~250 molecules per seed) on the two substrates used with the mW model. The lower (upper) end of the boxes and the white line within stand for the 25th(75th) percentile and the median of the data, respectively. The lower (upper) end of the error bars corresponds instead to the energy of the most stable (least) stable structure. (b) The number of water molecules in the ice seeds N_{cls} as a function of time, as obtained in seeded MD simulations of the most stable seed found via RSS for each of the ice polytype/face combinations illustrated in panel (a). (c) Representative snapshot of the most stable ice seeds on s1 (left panel) and s2 (right panel). Substrate, ice seed, and liquid water are depicted in gray, orange, and blue, respectively.
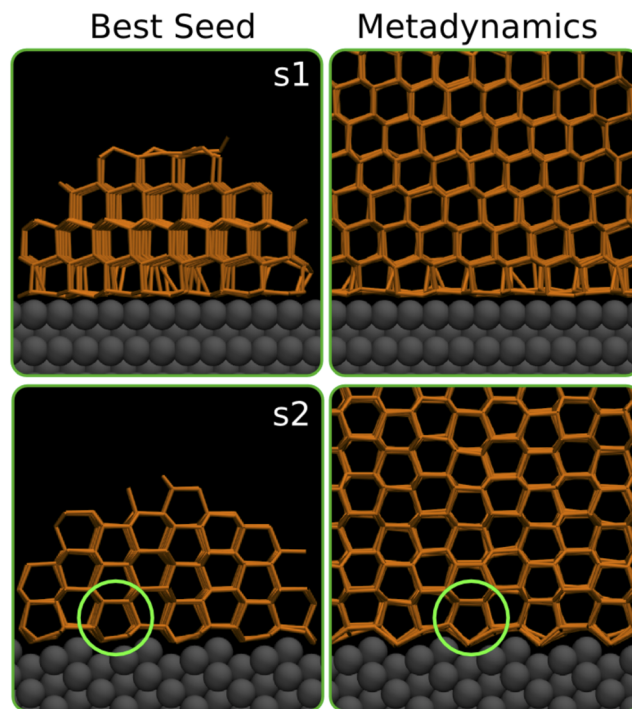

FIG. 3. The most stable seed selected via the RSS algorithm for s1 (top left) and s2 (bottom left), compared with the outcome of metadynamics simulations[55] (top right and bottom right for s1 and s2, respectively). Bonds between water molecules within the ice nuclei and s1/s2 atoms are shown in orange and gray, respectively. The green circles highlight the small difference between the two approaches in terms of the structure of the contact layer of the seeds on s2. Note that the orientation of the best seed in both cases is the same as the one found in metadynamics.

as $1 \, k_B T$ [see, e.g., $I_c(001)$ and $I_c(111)$] in Fig. 2(a). However, as illustrated in Fig. 2(a), $I_h(001)/I_c(111)$ and $I_h(100)$ seeds are amongst the most stable ones for s1 and s2, respectively. As illustrated in Fig. 3, these seeds correspond to the outcome of previous metadynamics simulations (see Sec. II C 1). The morphology of the seeds, for instance, in terms of r_{Seed} and ϕ_{Seed} (see Sec. II A), is correctly reproduced by the HSEED framework (see Fig. 3). A small mismatch between metadynamics and HSEED can be observed for the contact layer of the seeds on s2: although the network of water molecules is aligned correctly, the contact layer in the trenches does not exactly match the one obtained via metadynamics. We will make use of this observation to evaluate the overall robustness of the HSEED approach later on.

In light of the outcomes of the RSS algorithm, one could be tempted to draw the conclusion that the most stable types of seeds [e.g., $I_h(001)$ and $I_c(111)$ for s1 in Fig. 2], as obtained upon energy minimisation, would have the highest probability to grow on a given substrate. However, we shall see in Sec. III B that this is not always the case. In fact, in order to assess which particular ice polytype and face would be favored the most on a specific substrate, we have to use the seeds as the starting point for seeding MD simulations.

Step 3: We picked the three "best" structures from the RSS dataset (according to the criteria specified in Sec. II A) for each ice polytype/face and seed size, solvated them, and performed twenty MD runs at different target temperatures (see

Sec. II C 1 for further details). The results are summarised in Fig. 4: it is clear that for a low enough temperature and reasonable seed size, most of the polytype/face combinations will initiate freezing within a substantial fraction of the MD runs. At higher temperatures, however, only the "correct" (i.e., the same observed via the metadynamics simulations of Ref. 55) crystal face is capable of promoting the formation of ice. We note that in the case of s2 the secondary prism face of I_h is also a reasonable candidate—in agreement with the findings of our previous work. Importantly, the above-mentioned small metadynamics-HSEED mismatch in terms of the contact layer for $I_h(100)$ seeds on s2 does not seem to impact the outcomes of the HSEED method.

We have also found that only seeds that are substantially larger than the critical nucleus size estimates obtained in Ref. 55 induce nucleation on both s1 and s2. Specifically, according to the HSEED method N_C^* at 235 K is roughly equal (we assume N_c corresponds to a freezing probability = 50%) to 330 ± 25 and $290 \pm$ for s1 and s2, respectively—to be compared with 211 ± 11 and 104 ± 3 for s1 and s2, respectively, as obtained in Ref. 55. This discrepancy in terms of critical nucleus size is most likely to do with: (i) the structure of the ice seed; a crystalline surface interface as obtained via the HSEED method, even upon minimisation, is bound to be more defective than that obtained via conventional enhanced sampling techniques (metadynamics included); (ii) the short re-equilibration of the water-seed interface (see Sec. II C 1) negatively impacts the freezing probability of the seed; (iii) the assumption of a
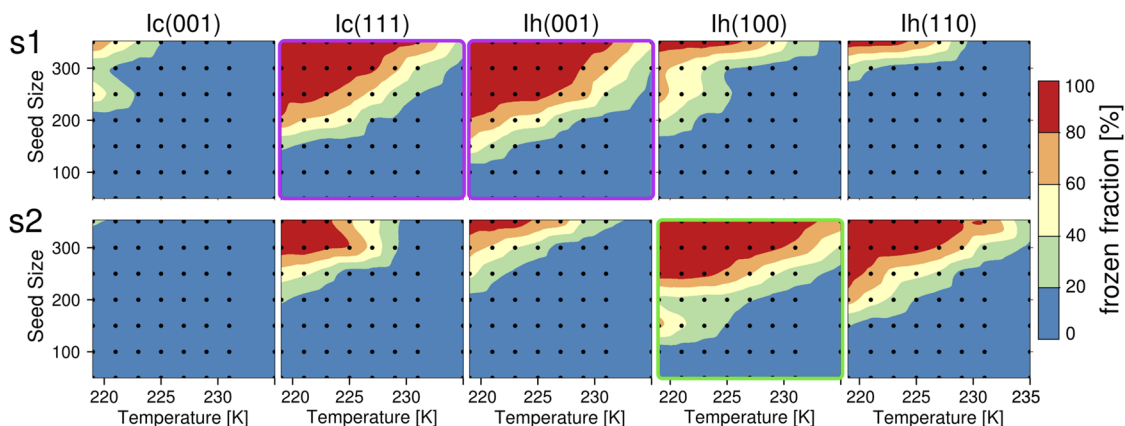
FIG. 4. Frozen percentage of simulations (of mW water freezing on LJ crystals) as a function of the temperature and seed size. Each black point indicates a set of twenty seeded MD runs at a given temperature, starting from the three best structures for a given seed size. The colour map represents the frozen percentage—i.e., the percentage of simulations where the ice seed grew to fill the whole simulation box as opposed to dissolve—for each collection of seeded MD runs. To generate smooth two-dimensional maps, we applied cubic interpolation between data points. The purple and green frames highlight the combinations of ice polytype/face we have observed nucleating on s1 and s2, respectively—by means of metadynamics simulations.[55]

contact angle that is likely to be larger than that of the nuclei obtained via, e.g., metadynamics simulations. In the case of s2, where the discrepancy in terms of N_C^* between HSEED and metadynamics amounts to almost a factor two, we have found that indeed the critical nuclei obtained via metadynamics are on average rather flat and characterised by small contact angles (of the order of ~45°). Exploring different contact angles as an additional degree of freedom within the HSEED method will be the subject of future work. However, we note that the relative trends in terms of the critical nucleus size are consistent in that $N_C^*(s1) > N_C^*(s2)$ according to both HSEED and metadynamics. Moreover, our results suggest that screening different contact angles is not necessary to establish which polytype/face will form on a particular substrate. Finally, we remark that, in the case of mW water, longer equilibration times for the seeds are difficult to deal with because the fast dynamics of the model is likely to induce heterogeneous freezing within relatively short time scales—notwithstanding the particular morphology of the seed.

Further evidence of the net preference for the s1 and s2 surfaces to promote the formation of $I_c(111)/I_h(001)$ and $I_h(100)$ is provided by the distribution of the potential energies of each one of the seeded MD runs. Specifically, we find that the systems seeded with the "correct" crystal face (see Fig. 3) are characterised on average by the lowest potential energy after freezing of all the water molecules in the simulation cell. This suggests that the $I_c(111)/I_h(001)$ and $I_h(100)$ seeds in the case of s1 and s2, respectively, led to the formation of more pristine ice if compared to the other polytype/face combinations.

For the purpose of establishing these trends, we accumulated a total of 76.8 μs of simulation time (two systems × five ice faces × eight seed sizes × eight temperatures × three seeds × twenty MD runs × 2 ns simulation time). However, if one would be interested in (i) pinpointing the most probable seed morphology and (ii) obtaining an estimate of the critical nucleus size for a given substrate at a given temperature, only a small fraction of this computational effort would be needed.

Our results suggest that in this case one would need about 0.5 μs.

B. TIP4P/ice water on cholesterol crystals

The freezing of mW water on the LJ crystals just discussed allowed us to explore the capabilities of the HSEED method for a variety of nucleation scenarios/conditions. However, the true testing ground is heterogeneous nucleation of ice from fully atomistic water models on complex/realistic crystalline surfaces, a situation where enhanced sampling simulations are necessary to observe even a single nucleation event—often requiring phenomenal computational resources. As such, we have applied the HSEED method to the formation of ice on CHLM crystals, a problem which we have recently tackled with (computationally expensive) FFS simulations.[49] Specifically, we consider the (001) hydroxylated surface of CHLM crystals ($CHLM_{001}^{-OH}$), as detailed in Sec. II C 2.

Step 1: The same five combinations of ice polytype/face detailed in Sec. III A have been considered as the starting point for the HSEED procedure. Guided by the outcome of our FFS simulations,[49] we built seeds containing 250 water molecules—roughly the dimension of $N_{C,hetero}^*$ at 230 K.

Step 2: About 2000 structures for each seed have been generated via the RSS algorithm detailed in Sec. II A. The average adsorption energy per water molecule E_{Ads} for the different ice polytype/face combinations as obtained upon energy minimisation is shown in Fig. 5. Similar to what we observed for mW water on LJ crystals, the spread of these data is huge. Interestingly, the most energetically stable seeds found expose the $I_h(100)$ and $I_h(001)$ surfaces at the ice-$CHLM_{001}^{-OH}$ interface, while our FFS simulations[49] unequivocally pinpointed $I_c(100)$ nuclei as the kinetically more favoured to form on $CHLM_{001}^{-OH}$. This is in contrast to what we have observed in the case of mW water on LJ crystals, where the most stable ice seeds displayed the same morphology as those obtained via metadynamics simulations. Is the HSEED thus incapable of dealing with complex interfaces such as the ice-$CHLM_{001}^{-OH}$ one? To answer this
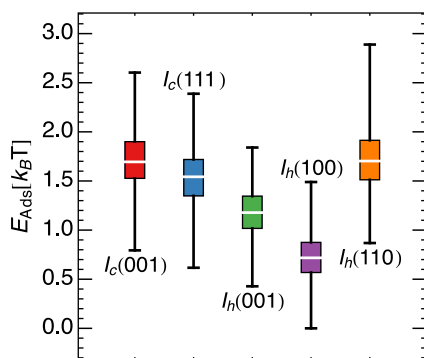
FIG. 5. Adsorption energy per water molecule in the contact layer of different ice seeds (~250 molecules per seed) on $CHLM_{001-OH}$. The lower (upper) end of the boxes and the white line within stand for the 25th(75th) percentile and the median of the data, respectively. The lower (upper) end of the error bars corresponds instead to the energy of the most (least) stable structure.

question, we kept following the work flow of the HSEED method (see Fig. 1).

Step 3: We selected three seeds for each ice polytype/face combination according to the criteria specified in Sec. II A and embedded them in a slab of liquid water. The equilibration protocol preceding the seeding MD runs is described in Sec. II C 2 and led to a substantial increase in the size of the seeds, from 250 to ~350 molecules. We have chosen to perform seeded MD simulations at 240 K, as at this temperature the dynamics of liquid water is reasonably fast—while the critical nucleus should be of the order of 200-300 water molecules, according to our FFS simulations.[49] The outcome of these simulations is summarised in Fig. 6(a): $I_c(111)$, $I_h(001)$, and $I_h(100)$ seeds dissolve within 20 ns, while $I_c(001)$ and $I_h(110)$ seeds endure. The same trend can be observed for different configurations of the initial seeds as well as for different choices of the initial velocities. As an example, we report in Fig. 6(b) additional sets of simulations for $I_c(001)$ and $I_h(110)$ seeds: despite an initial drop in the number of molecules within the seeds (which is due to the sub-optimal equilibration of the seed/water and seed/CHLM interfaces), these two combinations of ice polytype and face seem to be stable, on average, up to 40 ns. Note that, as opposed to the mW water on the LJ crystals, the time scales involved for the growth and dissolution of the seeds are much longer. Nonetheless, we were able to probe the actual growth of the stable ice seeds employing only a fraction of the computational effort of the FFS simulations of Ref. 49. We found that, consistent with the latter, $I_c(001)$ seeds do grow, as illustrated in Fig. 6(c). In addition, the HSEED result in terms of the critical nucleus size ($N_C^* = 350 \pm 50$ at 240 K) is compatible with the outcome of our FFS simulations ($N_C^* = 250 \pm 50$ at 230 K).

These results indicate that the RSS alone is not sufficient to determine which ice polytype and face would be favored on a specific substrate. Such insight has to be gained from seeded MD simulations, thus illustrating the importance of each step in the HSEED framework. In addition, the values of E_{Ads} reported in Figs. 2 and 5 originate not only from the interaction between the ice seeds and the crystalline substrate but also from the surface energies of the different ice crystalline faces. For instance, the two low energy surfaces of hexagonal
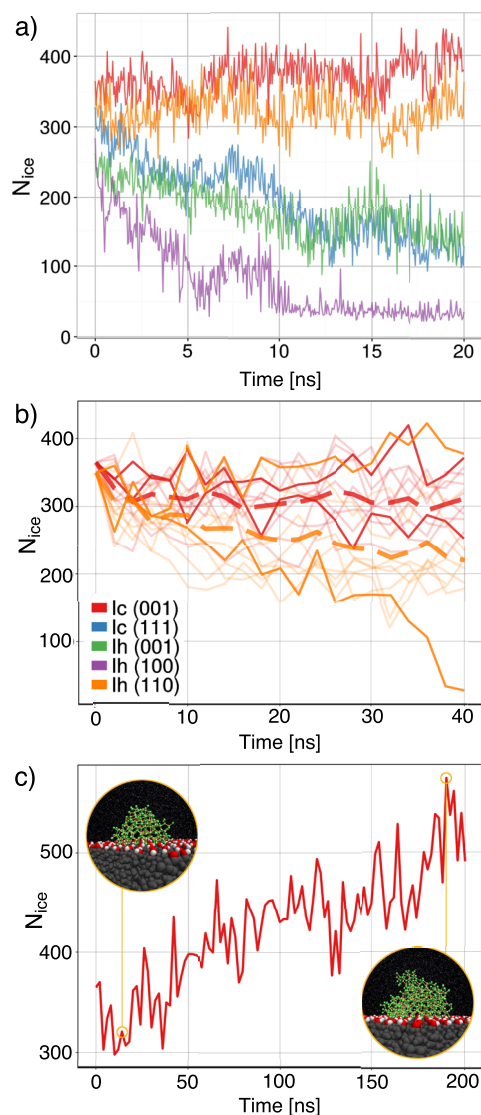

FIG. 6. (a) The number of molecules within different (see legend) ice seeds on the $CHLM_{001}^{-OH}$ surface as a function of time. (b) The same as panel (a) for ten statistically independent simulations of $I_c(001)$ and $I_h(110)$ seeds. The curves corresponding to the simulations leading to the biggest and smallest seeds are shown with thick continuous lines, while thick dashed lines correspond to the mean size of the seed at any given point in time. (c) Growth of an $I_c(001)$ seed over a longer time scale (200 ns). The insets show representative snapshots of small (left) and large (right) seeds.

ice [$I_h(001)$ and $I_h(100)$] are more stable than the secondary prism face, $I_h(110)$, of hexagonal ice.

It is also intriguing to note that, while $I_c(001)$ seeds are the most kinetically favorable at this strong supercooling, $I_h(110)$ nuclei are also possible. This is consistent with the results of Ref. 49, which have shown that CHLM crystals can promote the formation of both I_c and I_h pre-critical nuclei. In fact, our FFS simulations[49] suggest that a coexistence of the two polytypes can be expected at mild supercooling. The HSEED method thus provides further support to this hypothesis, which is in stark contrast to what has been observed in terms of ice formation on several inorganic crystals. For instance, according to both experiments and simulations, exclusively $I_h(100)$ forms on both the clay mineral kaolinite[33,72,73] and the mineral feldspar.[14]
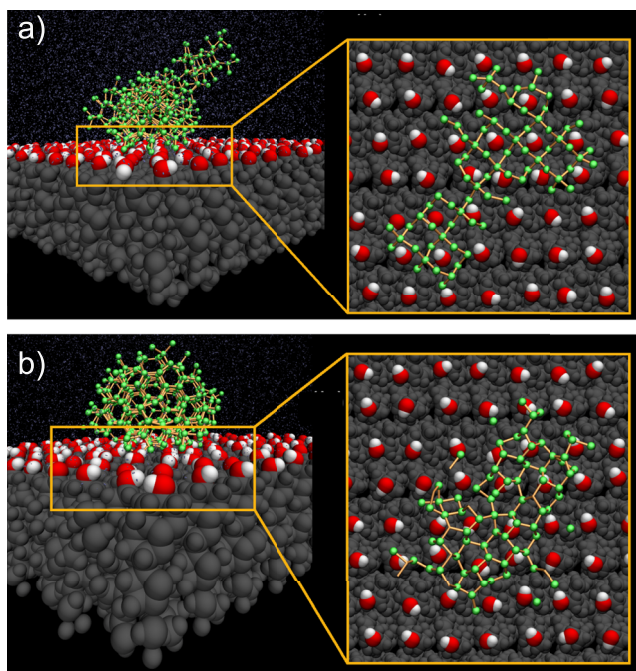
FIG. 7. Representative snapshot of (a) a $I_c(001)$ seed and (b) a $I_h(110)$ seed at 230 K (side/top view on the left/right), growing on the $CHLM_{001}^{-OH}$ surface during seeded MD simulations. Water molecules not participating in the ice nuclei are not shown. CHL molecules, the oxygen (hydrogen) atoms of their hydroxyl groups, and the oxygen atoms of ice-like molecules are depicted in gray, red (white), and green, respectively.

The rare ability of CHLM crystals to accommodate both $I_c(001)$ and $I_h(110)$ seeds could be due to the particular arrangement of the hydroxyl groups of CHL at the ice-$CHLM_{001}^{-OH}$ interface. As illustrated in Fig. 7, this seems indeed to be the case, as both $I_c(001)$ and $I_h(110)$ seeds tend to align along preferential directions leading to the relevant ice faces to grow along rows of hydroxyl groups. However, water molecules at the ice-$CHLM_{001}^{-OH}$ interface are much more ordered for $I_c(001)$ seeds if compared to the $I_h(110)$ case. We argue that $I_h(110)$ seeds can be stabilised nonetheless by the CHL surface due to the intrinsic flexibility of this substrate, which can play a significant role in the context of the kinetics of ice formation.[48] For instance, the surfaces of both feldspar and kaolinite are held together by strong covalent bonds, resulting in a rather rigid surface. On CHLM, however, weak intermolecular interactions only are responsible for the stability of the surface. This is a fundamental difference between inorganic and organic crystals, which may very well be at heart the strong ice nucleating ability of the latter.[49,74]

C. Computational cost

The challenging case of ice nucleation on CHLM represents an opportunity to compare the computational cost of the HSEED method with that of the FFS simulations reported in Ref. 49. Generating 10^3–10^4 seeds for each ice polytype/face combination required ~48 CPU hours. The geometry optimisation of the interfacial region for each one of these seeds took – on average – 0.08 CPU hours, totalling 800 CPU hours. Note that the minimisation runs can be trivially parallelised

so that this stage of the algorithm can typically be dealt with within a day. The bulk of the computational effort lies within the actual seeded MD runs. Including the equilibration stage, we estimate a cost of 40 ns × 10 seeded MD runs × 3 seeds for each ice polytype/face combination × 5 ice polytype/face combinations divided by 12 ns/day (using 8 CPUs) = 96 000 CPU hours. Overall, the HSEED algorithm thus allowed us to investigate the formation of ice on CHLM at strong supercooling using ~10^5 CPU hours.

The FFS simulations reported in Ref. 49 required ~10^6 CPU hours—taking advantage of GPU acceleration (providing an ~4× speedup). Importantly, the FFS algorithm relies on the definition of different interfaces (see, e.g., Ref. 75) along the path from water to ice, which have to be sampled one after the other. The same holds to various extents for most path sampling methods. Similarly, free energy-based enhanced sampling methods such as metadynamics can be parallelised by means of, e.g., multiple walkers[76] but still rely on the sampling of the free energy surface by means of serial production runs. On the other hand, *all* the production runs within the HSEED framework can be performed in a trivially parallel fashion so that the computational cost of the HSEED can be dealt with much more quickly than, e.g., FFS and metadynamics. To provide a practical example, the FFS simulations reported in Ref. 49 required a year-long project, while the HSEED simulations described here took one month only.

Interestingly, we observed a nominal speedup of about one order of magnitude in the case of mW water freezing on LJ crystals as well. In order to investigate a single surface at a particular temperature, the HSEED required ~10^4 CPU hours, to be compared with the ~10^5 CPU hours needed to converge the metadynamics simulations of Ref. 55 for the exact same system. We note that, despite the substantial number of different ice seeds (in terms of size/polytype) we have probed in this case, the RSS algorithm did not represent a limiting step: as an example, taking into account one substrate and 30 different combinations of ice seed size and polytype only required 1 CPU for 7 (2) days when dealing with seeds containing 400 (100) molecules. Finally, we remark that investigating ice nucleation at mild supercooling is simply not feasible by means of conventional enhanced sampling techniques, due to the low nucleation rate. The unique strength of the HSEED thus stands in the capability of the method to address this important pitfall.

IV. DISCUSSION AND CONCLUSIONS

In summary, we have presented a methodology (HSEED) to study the heterogeneous nucleation of ice via a combination of RSS algorithms and seeded MD simulations. We have made available via a public GitHub repository[52] a collection of (*Python*) scripts that can be used to apply the HSEED method to an arbitrary crystalline substrate. We validated our approach by comparing the outcomes of the HSEED method against enhanced sampling simulations of (i) coarse-grained mW water freezing on model LJ crystals[55] and (ii) fully atomistic TIP4P/Ice water turning into ice on CHLM crystals.[49] In both cases, the HSEED method is able to pinpoint the

combination of ice polytype and crystalline face which is most likely to form on the crystalline substrates. Estimates of the critical nucleus size are also in line with independent evaluations.

When dealing with computationally inexpensive simulation setups such as mW water on model surfaces, the HSEED method allows the comprehensive investigation of the ice nucleating ability of different substrates at different temperatures, including mild supercooling for which—costly—enhanced sampling simulations would be needed. Specifically, in this case, one can think about two different approaches to look for the "correct" ice seed on a given substrate:

- *Constant seed screening*: Starting from a dataset of different ice seeds of a given size, the temperature of the whole system is lowered until heterogeneous ice nucleation is observed for one (or more than one) of the ice polytype/face combinations.
- *Constant temperature screening*: At a given temperature, the size of different ice seeds is incrementally increased until heterogeneous nucleation is observed for one (or more than one) of the ice polytype/face combinations.

The former would be the method of choice when dealing with computationally inexpensive MD runs, as only one RSS has to be performed. The latter method might perform better if the seeding MD simulations turn out to be very expensive and/or if it would take longer MD runs to observe nucleation events, as it would be quicker to run multiple RSSs.

Importantly, the HSEED method performed well even in the challenging case of ice formation on CHLM. In this scenario, the hydrogen bond network between the ice seeds and the substrate had to be explicitly taken into account, and the complexity of the ice-crystal interface provided a real testing ground for the approach. We were able to identify via the HSEED approach the same combination of ice polytype/face we observed by means of forward flux sampling simulations,[49] and the structure of the seeds-substrate interface is consistent with what we have found via brute force MD simulations.[49] The specific surface of CHLM crystals we have considered herein is capable, according to previous results, to accommodate two different ice polytypes, the evidence that the HSEED method did capture as well.

In its present formulation, this method can treat relatively flat, pristine crystalline surfaces. This represents a substantial leap forward for the ice nucleation community, as we are now in a position to evaluate rapidly the ice nucleation ability of whole libraries of crystalline compounds with the same computational effort required to investigate a single substrate by means of conventional enhanced sampling methods. For instance, we have shown that in the case of ice formation of CHLM, a challenging testing ground for the HSEED method which involves a complex water-substrate interface of relevance, e.g., for cryopreservation applications, the HSEED method requires a parallel workload on the order of 10^5 CPU hours, to be compared with the only partially parallelisable 10^6 CPU hours needed to converge FFS simulations. However, it would clearly be desirable to expand the scope of the HSEED approach to non-flat, disordered, rough, and flexible interfaces. This is especially relevant to heterogeneous ice nucleation in biological matter, where most of the substrates are characterised by complex morphologies that share very little with pristine crystalline surfaces. The implementation of more sophisticated RSS algorithms could represent a first step in that direction.

The HSEED method could also be used to probe the ice nucleating ability of different nucleation sites within the same crystalline substrate. This is of paramount importance, e.g., for the atmospheric science community, as it is clear that the topology of the surface structure of ice nucleating agents such as the mineral feldspar plays a fundamental role in determining the overall kinetics of ice nucleation.[14,77] Thanks to the HSEED method, active sites such as crystalline defects on the nm scales are now within the reach of atomistic simulations of heterogeneous ice formation. We thus hope that the methodological advancement presented here will foster a new generation of MD simulations aimed at screening the ice nucleating ability of different compounds and so reducing the gap between experiments and simulations. Finally, it is worth noticing that the HSEED framework can be extended to include crystallisation scenarios other than water freezing—thus opening the possibility to accelerate the computational investigation of heterogeneous nucleation and growth of many other crystalline materials.

ACKNOWLEDGMENTS

[1] P. Mazur, "Cryobiology: The freezing of biological systems," Science **168**(3934), 939–949 (1970).

[2] A. Lintunen, T. Hölttä, and M. Kulmala, "Anatomical regulation of ice nucleation and cavitation helps trees to survive freezing and drought stress," Sci. Rep. **3**, 2031 (2013).

[3] K. A. Pratt, P. J. DeMott, J. R. French, Z. Wang, D. L. Westphal, A. J. Heymsfield, C. H. Twohy, A. J. Prenni, and K. A. Prather, "*In situ* detection of biological particles in cloud ice-crystals," Nat. Geosci. **2**(6), 398–401 (2009).

[4] B. J. Murray, D. O'Sullivan, J. D. Atkinson, and M. E. Webb, "Ice nucleation by particles immersed in supercooled cloud droplets," Chem. Soc. Rev. **41**, 6519–6554 (2012).

[5] T. Bartels-Rausch, "Chemistry: Ten things we need to know about ice and snow," Nature **494**(7435), 27–29 (2013).

[6] R. Y. Tam, C. N. Rowley, I. Petrov, T. Zhang, N. A. Afagh, T. K. Woo, and R. N. Ben, "Solution conformation of c-linked antifreeze glycoprotein analogues and modulation of ice recrystallization," J. Am. Chem. Soc. **131**(43), 15745–15753 (2009).

[7] C. A. Koh, "Towards a fundamental understanding of natural gas hydrates," Chem. Soc. Rev. **31**(3), 157–167 (2002).

[8] B. J. Murray, S. L. Broadley, and G. J. Morris, "Supercooling of water droplets in jet aviation fuel," Fuel **90**(1), 433–435 (2011).

[9] G. C. Sosso, J. Chen, S. J. Cox, M. Fitzner, P. Pedevilla, A. Zen, and A. Michaelides, "Crystal nucleation in liquids: Open questions and future challenges in molecular dynamics simulations," Chem. Rev. **116**(12), 7078–7116 (2016).

[10] S. A. Zielke, A. K. Bertram, and G. N. Patey, "A molecular mechanism of ice nucleation on model AgI surfaces," J. Phys. Chem. B **119**(29), 9049–9055 (2015).

[11] X. Zhang, M. Chen, and M. Fu, "Impact of surface nanostructure on ice nucleation," J. Chem. Phys. **141**(12), 124709 (2014).

[12] A. Reinhardt and J. P. K. Doye, "Effects of surface interactions on heterogeneous ice nucleation for a monatomic water model," J. Chem. Phys. **141**(8), 084501 (2014).

[13] G. Fraux and J. P. K. Doye, "Heterogeneous ice nucleation on silver-iodide-like surfaces," J. Chem. Phys. **141**(21), 216101 (2014).

[14] A. Kiselev, F. Bachmann, P. Pedevilla, S. J. Cox, A. Michaelides, D. Gerthsen, and T. Leisner, "Active sites in heterogeneous ice nucleation—the example of k-rich feldspars," Science **355**(6323), 367–371 (2017).

[15] Y. Bi, B. Cao, and T. Li, "Enhanced heterogeneous ice nucleation by special surface geometry," Nat. Commun. **8**, 15372 (2017).

[16] L. Lupi, A. Hudait, B. Peters, M. Grünwald, R. G. Mullen, A. H. Nguyen, and V. Molinero, "Role of stacking disorder in ice nucleation," Nature **551**(7679), 218–222 (2017).

[17] A. Hudait and V. Molinero, "What determines the ice polymorph in clouds?," J. Am. Chem. Soc. **138**(28), 8958–8967 (2016).

[18] V. Molinero and E. B. Moore, "Water modeled as an intermediate element between carbon and silicon," J. Phys. Chem. B **113**, 4008 (2008).

[19] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," J. Comput. Phys. **23**(2), 187–199 (1977).

[20] A. Warmflash, P. Bhimalapuram, and A. R. Dinner, "Umbrella sampling for nonequilibrium processes," J. Chem. Phys. **127**(15), 154112 (2007).

[21] P. R. Ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, "Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling," J. Chem. Phys. **104**(24), 9932–9947 (1996).

[22] S. Auer and D. Frenkel, "Prediction of absolute crystal-nucleation rate in hard-sphere colloids," Nature **409**(6823), 1020–1023 (2001).

[23] R. Radhakrishnan and B. L. Trout, "Nucleation of hexagonal ice (I$_h$) in liquid water," J. Am. Chem. Soc. **125**(25), 7743–7747 (2003).

[24] A. Laio and M. Parrinello, "Escaping free-energy minima," Proc. Natl. Acad. Sci. U. S. A. **99**(20), 12562–12566 (2002).

[25] F. Trudu, D. Donadio, and M. Parrinello, "Freezing of a Lennard-Jones fluid: From nucleation to spinodal regime," Phys. Rev. Lett. **97**(10), 105701 (2006), cited by 139.

[26] M. Salvalaglio, P. Tiwary, G. M. Maggioni, M. Mazzotti, and M. Parrinello, "Overcoming time scale and finite size limitations to compute nucleation rates from small scale well tempered metadynamics simulations," J. Chem. Phys. **145**(21), 211925 (2016).

[27] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, "Transition path sampling: Throwing ropes over rough mountain passes, in the dark," Annu. Rev. Phys. Chem. **53**, 291–318 (2002).

[28] W. Lechner, C. Dellago, and P. G. Bolhuis, "Role of the prestructured surface cloud in crystal nucleation," Phys. Rev. Lett. **106**(8), 085701 (2011).

[29] C. Valeriani, E. Sanz, and D. Frenkel, "Rate of homogeneous crystal nucleation in molten NaCl," J. Chem. Phys. **122**(19), 194501 (2005).

[30] L. Filion, M. Hermes, R. Ni, and M. Dijkstra, "Crystal nucleation of hard spheres using molecular dynamics, umbrella sampling, and forward flux sampling: A comparison of simulation techniques," J. Chem. Phys. **133**(24), 244115 (2010).

[31] T. Li, D. Donadio, G. Russo, and G. Galli, "Homogeneous ice nucleation from supercooled water," Phys. Chem. Chem. Phys. **13**, 19807 (2011).

[32] A. Haji-Akbari and P. G. Debenedetti, "Direct calculation of ice homogeneous nucleation rate for a molecular model of water," Proc. Natl. Acad. Sci. U. S. A. **112**(34), 10582–10588 (2015).

[33] G. C. Sosso, T. Li, D. Donadio, G. A. Tribello, and A. Michaelides, "Microscopic mechanism and kinetics of ice formation at complex interfaces: Zooming in on kaolinite," J. Phys. Chem. Lett. **7**(13), 2350–2355 (2016).

[34] V. Kalikmanov, *Nucleation Theory* (Springer, Dordrecht, The Netherlands, 2013).

[35] J. R. Espinosa, C. Vega, C. Valeriani, and E. Sanz, "Seeding approach to crystal nucleation," J. Chem. Phys. **144**(3), 034501 (2016).

[36] T. L. Malkin, B. J. Murray, C. G. Salzmann, V. Molinero, S. J. Pickering, and T. F. Whale, "Stacking disorder in ice I," Phys. Chem. Chem. Phys. **17**, 60–76 (2015).

[37] B. Emily Moore and V. Molinero, "Is it cubic? Ice crystallization from deeply supercooled water," Phys. Chem. Chem. Phys. **13**, 20008–20016 (2011).

[38] A. Zaragoza, M. M. Conde, J. R. Espinosa, C. Valeriani, C. Vega, and E. Sanz, "Competition between ices Ih and Ic in homogeneous water freezing," J. Chem. Phys. **143**(13), 134504 (2015).

[39] E. B. Moore, E. de la Llave, K. Welke, D. A. Scherlis, and V. Molinero, "Freezing, melting and structure of ice in a hydrophilic nanopore," Phys. Chem. Chem. Phys. **12**, 4124 (2010).

[40] L. Lupi, A. Hudait, and V. Molinero, "Heterogeneous nucleation of ice on carbon surfaces," J. Am. Chem. Soc. **136**, 3156 (2014).

[41] S. J. Cox, S. M. Kathmann, B. Slater, and A. Michaelides, "Molecular simulations of heterogeneous ice nucleation. II. Peeling back the layers," J. Chem. Phys. **142**(18), 184705 (2015).

[42] S. J. Cox, S. M. Kathmann, B. Slater, and A. Michaelides, "Molecular simulations of heterogeneous ice nucleation. I. Controlling ice nucleation through surface hydrophilicity," J. Chem. Phys. **142**(18), 184704 (2015).

[43] M. Fitzner, G. C. Sosso, S. J. Cox, and A. Michaelides, "The many faces of heterogeneous ice nucleation: Interplay between surface morphology and hydrophobicity," J. Am. Chem. Soc. **137**(42), 13658–13669 (2015).

[44] Y. Bi, R. Cabriolu, and T. Li, "Heterogeneous ice nucleation controlled by the coupling of surface crystallinity and surface hydrophilicity," J. Phys. Chem. C **120**(3), 1507–1514 (2016).

[45] L. Lupi, B. Peters, and V. Molinero, "Pre-ordering of interfacial water in the pathway of heterogeneous ice nucleation does not lead to a two-step crystallization mechanism," J. Chem. Phys. **145**(21), 211910 (2016).

[46] O. Björneholm, M. H. Hansen, A. Hodgson, L.-M. Liu, D. T. Limmer, A. Michaelides, P. Pedevilla, J. Rossmeisl, H. Shen, G. Tocci, E. Tyrode, M.-M. Walz, J. Werner, and H. Bluhm, "Water at interfaces," Chem. Rev. **116**(13), 7698–7726 (2016).

[47] R. Cabriolu and T. Li, "Ice nucleation on carbon surface supports the classical theory for heterogeneous nucleation," Phys. Rev. E **91**, 052402 (2015).

[48] G. C. Sosso, G. A. Tribello, A. Zen, P. Pedevilla, and A. Michaelides, "Ice formation on kaolinite: Insights from molecular dynamics simulations," J. Chem. Phys. **145**, 211927 (2016).

[49] G. C. Sosso, T. F. Whale, P. Pedevilla, and A. Michaelides, "Unraveling the origins of ice nucleation on organic crystals" (unpublished).

[50] P. Pedevilla, S. J. Cox, B. Slater, and A. Michaelides, "Can ice-like structures form on non-ice-like substrates? The example of the K-feldspar microcline," J. Phys. Chem. C **120**(12), 6704–6713 (2016).

[51] S. J. Cox, S. M. Kathmann, J. A. Purton, M. J. Gillan, and A. Michaelides, "Non-hexagonal ice at hexagonal surfaces: The role of lattice mismatch," Phys. Chem. Chem. Phys. **14**(22), 7944–7949 (2012).

[52] G. C. Sosso, Hseed-(heterogeneous) seeded molecular dynamics, https://github.com/gcsosso/HSeed.git, 2018.

[53] J. Nocedal, "Updating quasi-Newton matrices with limited storage," Math. Comput. **35**, 773–782, (1980).

[54] E. Sanz, C. Vega, J. R. Espinosa, R. Caballero-Bernal, J. L. F. Abascal, and C. Valeriani, "Homogeneous ice nucleation at moderate supercooling from molecular simulation," J. Am. Chem. Soc. **135**(40), 15008–15017 (2013).

[55] M. Fitzner, G. C. Sosso, F. Pietrucci, S. Pipolo, and A. Michaelides, "Pre-critical fluctuations and what they disclose about heterogeneous crystal nucleation," Nat. Commun. **8**(1), 2257 (2017).

[56] P. Pedevilla, M. Fitzner, and A. Michaelides, "What makes a good descriptor for heterogeneous ice nucleation on OH-patterned surfaces," Phys. Rev. B **96**, 115441 (2017).

[57] A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method," Phys. Rev. Lett. **100**(2), 020603 (2008).

[58] A. Grégoire Gallet and F. Pietrucci, "Structural cluster analysis of chemical reactions in solution," J. Chem. Phys. **139**(7), 074101 (2013).

[59] G. J. Martyna, M. L. Klein, and M. Tuckerman, "Nosé-Hoover chains: The canonical ensemble via continuous dynamics," J. Chem. Phys. **97**(4), 2635–2643 (1992).

[60] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," J. Comput. Phys. **117**(1), 1–19 (1995).

[61] B. M. Craven, "Crystal structure of cholesterol monohydrate," Nature **260**(5553), 727–729 (1976).

[62] J. L. F. Abascal, E. Sanz, R. García Fernández, and C. Vega, "A potential model for the study of ices and amorphous water: Tip4p/Ice," J. Chem. Phys. **122**(23), 234511 (2005).

[63] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulations," J. Chem. Theory Comput. **4**(3), 435–447 (2008).

[64] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, flexible, and free," J. Comput. Chem. **26**(16), 1701–1718 (2005).

[65] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, and E. Lindahl, "Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models," J. Chem. Theory Comput. **6**(2), 459–466 (2010).

[66] J. B. Lim, B. Rogaski, and J. B. Klauda, "Update of the cholesterol force field parameters in CHARMM," J. Phys. Chem. B **116**(1), 203–210 (2011).

[67] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," J. Chem. Phys. **103**(19), 8577–8593 (1995).

[68] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," J. Chem. Phys. **126**(1), 014101 (2007).

[69] S. Miyamoto and P. A. Kollman, "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models," J. Comput. Chem. **13**(8), 952–962 (1992).

[70] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, "LINCS: A linear constraint solver for molecular simulations," J. Comput. Chem. **18**(12), 1463–1472 (1997).

[71] B. Hess, "P-LINCS: A parallel linear constraint solver for molecular simulation," J. Chem. Theory Comput. **4**(1), 116–122 (2008).

[72] S. J. Cox, Z. Raza, S. M. Kathmann, B. Slater, and A. Michaelides, "The microscopic features of heterogeneous ice nucleation may affect the macroscopic morphology of atmospheric ice crystals," Faraday Discuss. **167**, 389–403 (2013).

[73] S. A. Zielke, A. K. Bertram, and G. N. Patey, "Simulations of ice nucleation by kaolinite (001) with rigid and flexible surfaces," J. Phys. Chem. B **120**, 1726–1734 (2015).

[74] N. Fukuta and B. J. Mason, "Epitaxial growth of ice on organic crystals," J. Phys. Chem. Solids **24**(6), 715–718 (1963).

[75] R. J. Allen, C. Valeriani, and P. R. ten Wolde, "Forward flux sampling for rare event simulations," J. Phys.: Condens. Matter **21**(46), 463102 (2009).

[76] P. Raiteri, A. Laio, F. Luigi Gervasio, C. Micheletti, and M. Parrinello, "Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics," J. Phys. Chem. B **110**(8), 3533–3539 (2006).

[77] B. Wang, D. A. Knopf, S. China, B. W. Arey, T. H. Harder, M. K. Gilles, and A. Laskin, "Direct observation of ice nucleation events on individual atmospheric particles," Phys. Chem. Chem. Phys. **18**(43), 29721–29731 (2016).