

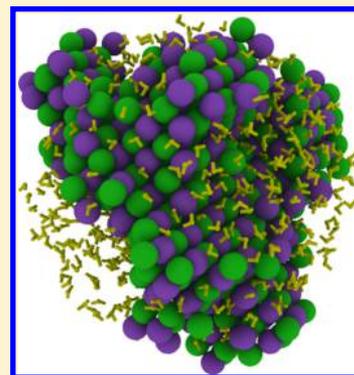
Birth of NaCl Crystals: Insights from Molecular Simulations

G. Lanaro* and G. N. Patey*

Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1

S Supporting Information

ABSTRACT: Molecular dynamics simulations are used to investigate the factors that influence the nucleation of NaCl crystals in a supersaturated aqueous solution. We describe a methodology for detecting solidlike NaCl clusters (potential nuclei) and following their evolution in time until they achieve nucleation (which is very rare) or dissolve back into solution. Through an analysis of cluster lifetimes and multiple nucleation events, we demonstrate that cluster size is not the only property that influences cluster stability and the probability of achieving nucleation. We introduce a parameter called cluster crystallinity, which is a measure of the solidlike order in a particular cluster. We show that cluster order (as measured by this parameter) has a strong influence on the lifetime and nucleation probability of clusters of equal sizes, with the lifetime and probability of nucleation increasing with increasing crystallinity. These observations remain true for clusters as small as six ions, showing that the structural factors are important even at the earliest stages of crystal birth.



1. INTRODUCTION

The initial process by which one phase transitions to another is termed nucleation. Nucleation has a central role in a wide range of physically interesting processes, such as crystal and rock formation,^{1,2} drug development,^{3,4} and the formation of atmospheric aerosols.^{5,6} There have been a number of interesting experimental studies on crystal nucleation,^{7–11} but there remain significant challenges for current experimental methods, particularly for the analysis of early prenucleation events. Therefore, to augment experimental studies, molecular simulations are being increasingly used to gain physical insight into crystal nucleation and growth.^{12–19}

The usual approach to nucleation, that is, classical nucleation theory (CNT), is a subject of current debate.^{20–22} In particular, the assumption that early-stage potential nuclei can be characterized by properties of the nascent bulk phase (plus a surface term) is being challenged by both experiments and molecular simulations. One such example is provided by CaCO₃, for which experiments suggest that prenucleation clusters serve as precursors to nucleation.²³ There is also evidence that crystallization can occur via a two-step mechanism consisting of an initial density “transition”, followed by a slower ordering transition, whereby the dense liquid region becomes geometrically ordered.^{13,24,25}

In the present article, we use direct molecular dynamics (MD) simulations to investigate the nucleation of NaCl nanocrystals. Supersaturated aqueous NaCl solutions are a good choice for detailed study because relatively accurate models exist,²⁶ and previous work has shown that spontaneous nucleation occurs on simulation time scales.^{13,14,18,19} Earlier direct simulation studies of NaCl crystallization from a supersaturated solution suggest a process by which less-ordered and more-hydrated NaCl clusters evolve with time into a largely anhydrous crystalline arrangement.^{13,27} There is also some

evidence that the two-step mechanism applies, with a large density (concentration) fluctuation preceding any spatial ordering.¹³

Nucleation and crystal growth have also been considered employing indirect simulation techniques. In an early investigation, Zahn¹⁸ approached the problem using trajectory sampling methods. He noticed that small NaCl clusters tended to prefer a Na⁺ ion at the center and suggested that such clusters might be important in nucleation. A very recent study reported by Zimmerman et al.¹⁹ used a seeded-trajectory approach to determine the sizes of critical nuclei, ion-attachment frequencies, and nucleation rates for NaCl in supersaturated solutions.

However, despite earlier efforts, the factors that influence the formation of critical nuclei have not been determined, and this is particularly true for very early stages or the birth of the crystal. For example, it is interesting to ask whether the size alone determines the relative stability of small NaCl clusters or whether other factors, such as cluster geometry, are important. This is one of the questions addressed in the present article.

Spontaneous nucleation events for NaCl are typically rare (on simulation time scales) except at a very high solute concentration, and multiple nucleation events must be observed to draw any meaningful conclusion. Additionally, it is necessary to define, detect, and follow in time a great many NaCl clusters that do not achieve nucleation. Here, we carry out multiple direct simulations, sufficient to generate a number of nucleation events. We develop a methodology to define and detect crystal-like NaCl clusters and to follow them in time from very early stages (~6 ions) until nucleation is achieved or, much more

Received: May 25, 2016

Revised: July 20, 2016

Published: July 25, 2016

frequently, until the cluster dissolves back into solution. An important conclusion of our analysis is that cluster size is not the only factor influencing cluster lifetime and the probability of nucleation. By introducing a new parameter called cluster “crystallinity”, we show that cluster geometry is also a very significant factor influencing cluster lifetime and nucleation probability; moreover, this is true for clusters as small as six ions.

The remainder of this article is divided into three parts. The models and simulation methods are described in Section 2, the results are presented in Section 3, and our conclusions are summarized in Section 4.

2. MODELS AND METHODS

2.1. Simulation Details. In our simulations, we adopt the Joung–Cheatham parameter set²⁶ for the Na⁺ and Cl[−] ions, paired with the SPC/E water model.²⁸ For this parameter set, the saturation mole fraction of NaCl, x_{NaCl} , at 298 K was calculated^{29–31} to be ~ 0.06 using the chemical potential route, whereas results from direct coexistence methods^{26,32–34} report values of ~ 0.09 – 0.11 (experimentally,³⁵ the value is ~ 0.10). With this force field, all nonbonded interactions consist of Lennard-Jones (LJ) plus electrostatic terms, such that the site–site pair potentials have the form

$$u(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (1)$$

where ϵ_0 is the permittivity of free space, q_i and q_j are the partial charges on sites i and j , and σ_{ij} and ϵ_{ij} are the usual LJ length and energy parameters, respectively. The Joung–Cheatham parameters for Na⁺ and Cl[−] are given in Table 1, and σ_{ij} and ϵ_{ij} are obtained from these and the SPC/E parameters²⁸ using the Lorenz–Berthelot combining rules, $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ and $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$.

Table 1. Joung–Cheatham LJ Parameters for NaCl²⁶

	σ (nm)	ϵ (kJ mol ^{−1})
Na ⁺	0.2160	1.4754533
Cl [−]	0.4830	0.0534924

Simulations were carried out under NPT conditions employing the GROMACS³⁶ MD package, version 4.5.4. The temperature was controlled using a velocity-rescale thermostat,³⁷ with a relaxation time of 0.1 ps, and the pressure was kept constant at 1 bar using a Berendsen barostat,³⁸ with a compressibility of 4.5×10^{-5} bar^{−1} and a relaxation time of 1.0 ps. A timestep of 2 fs was used in all simulations. A spherical cutoff of 0.9 nm was applied to the pair potentials, and long-range electrostatic interactions were calculated using the particle mesh Ewald method.³⁹

The details of all simulations performed are summarized in Table 2. Initial simulations were carried out for x_{NaCl} ranging from 0.20 to 0.30, and the lowest concentration at which a nucleation event was observed within 200 ns at 300 K was $x_{\text{NaCl}} = 0.22$. This time frame is feasible for an investigation employing direct MD simulations, so we focus on this concentration. More nucleation events occur at higher concentrations, but the occurrence of many simultaneous nucleations can complicate the analysis. Note that, on the basis of the estimates mentioned above, $x_{\text{NaCl}} = 0.22$ is between two

Table 2. Summary of the Simulations Performed^a

water	ions	x_{NaCl}	length (ns)	nuclei
25 600	12 800	0.20	157	0
24 960	14 080	0.22	187	3
24 000	16 000	0.25	515	5
22 400	19 200	0.30	202	n.a.
24 960	14 080	0.22	523	1
			547	2
			532	3
			530	4
			550	3

^aThe number of water molecules (water), number of ions (ions), mole fraction (x_{NaCl}), run length (length), and number of clusters that achieved nucleation (nuclei) are indicated. At $x_{\text{NaCl}} = 0.3$, multiple nucleations occurred, resulting in a polycrystal. The five simulations at $x_{\text{NaCl}} = 0.22$ listed at the bottom of the table were used for the statistical analysis given in the text.

and four times the saturation value at 298 K. The atomic coordinates for the system were collected at intervals of 0.04 ns, and the coordinate sets collected with this time interval are referred to as time frames elsewhere in the article.

The analysis given in this article is based on five “replica” simulations at $x_{\text{NaCl}} = 0.22$, carried out as follows. Initially, 7040 NaCl pairs and 24 960 water molecules were distributed randomly on a lattice within the simulation cell. The system was then “equilibrated” at 400 K, and five configurations were extracted at intervals of ~ 10 ns. Five simulations, initiated with these configurations, were cooled to 300 K and allowed to evolve for ~ 500 ns, as summarized in Table 2. The number of nucleation events observed in each simulation is also given in Table 2.

2.2. Cluster Detection. The aim of the present work is to detect NaCl clusters and follow their evolution in time to understand the factors influencing (or not) the probability of achieving crystal nucleation. From the perspective of CNT, these clusters can be viewed as subcritical nuclei unless and until nucleation is achieved. As we are trying to observe the very early stages of a phase transition, it is necessary to find some measure to distinguish local liquidlike and solidlike structures and to determine which ions are part of the same solidlike cluster. Generally, such measures are based on local environments, and several approaches to this problem have been employed. These include methods based on ion connectivity,^{27,40} local ion density,⁴¹ local solvent density,¹⁸ and local bond order parameters.^{42,43} The last approach has been used extensively in studies of ice nucleation.^{44–46}

In the present article, the computational procedures used to detect (define) ion clusters and follow their time evolution consist of three steps: filtering, clustering, and entity resolution, which are described in detail below.

2.2.1. Filtering. The purpose of this step is to identify for each configuration along a trajectory the ions that are most likely to be part of a local solidlike structure. To do this, we follow the general bond-orientational order parameter approach of Steinhardt et al.⁴⁷ This method defines potential order parameters of the form

$$q_l = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^l |q_{lm}|^2 \right]^{1/2}$$

where

$$q_{lm} = \frac{1}{N} \sum_{r_i=r_l}^{r_N} Y_l^m(\theta(r_i), \phi(r_i))$$

In the above equation, the number of neighbors $N = 12$, Y_l^m is a spherical harmonic, r_i is the position vector of the i th neighbor with respect to a central ion, and $\theta(r_i)$ and $\phi(r_i)$ are, respectively, the polar and azimuthal angles with respect to an arbitrary frame of reference. The actual choice of reference frame is irrelevant as the order parameter q_l is rotationally invariant.

The objective is to find the q_l that best distinguishes between ions with liquidlike and solidlike local environments. To do this, we examined q_l distributions for $l = 2, 4, 6, 8$, and 10 , using a spherical crystal of ~ 2000 ions to represent the solid phase and a supersaturated solution at $x_{\text{NaCl}} = 0.20$ (that did not nucleate) as representative of the liquid phase. Both reference systems were held at 300 K. On the basis of this investigation, order parameter q_8 provides the best separation between liquid and solid phase distributions (Figure 1). Also, by testing a

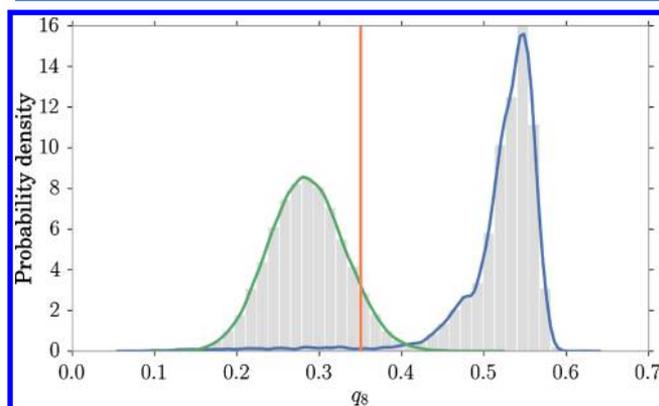


Figure 1. Density of q_8 values in a sample crystal (~ 2000 ions) (blue) and a supersaturated solution ($x_{\text{NaCl}} = 0.20$) (green). The two distributions show excellent separation. The vertical orange line indicates the selected threshold ($q_8 = 0.35$).

reference solution of lower concentration ($x_{\text{NaCl}} \approx 0.025$), we found that the q_8 distribution is largely independent of concentration (unlike order parameters based on local ion densities⁴¹). In the remainder of this article, we refer to q_8 as the bond order parameter.

Order parameter q_8 is used to filter out ions that are a part of liquidlike disordered structures. From Figure 1, we see that the best separation point between liquidlike and solidlike structures occurs at $q_8 \approx 0.4$. However, we select $q_8 = 0.35$ as the threshold for our analysis to capture structures from the upper end of the solution distribution. As clusters that achieve nucleation form spontaneously from solution, selecting a lower threshold value allows us to trace the nucleation processes from the very initial stages. A very low threshold would include too many transient, short-lived structures, whereas a very high threshold would filter out interesting prenucleation stages. The selected value of 0.35 proved to be a good compromise between these two factors. We verified that our results and conclusions are not affected by small variations in the selected threshold value. Note that with the 0.35 threshold only $\sim 10\%$ of the ions are classified as solidlike for each configuration. It is also important to emphasize that in our simulations all clusters that eventually achieved nucleation originated within the set of

ions identified as solidlike; thus, our filtering process does not eliminate any interesting events.

2.2.2. Cluster Identification. Although order parameter filtering efficiently detects ions in solidlike local environments, another algorithm is necessary to identify which of the selected (solidlike) ions are part of the same cluster or aggregate. For this purpose, we adopted the density-based spatial clustering of applications with noise (DBSCAN) algorithm,⁴⁸ as implemented in the Python library scikit-learn.⁴⁹ In the DBSCAN algorithm, clusters are identified by means of two parameters, which in the present context are a distance (ϵ) and the number of neighboring ions (η) within that distance. To find ions connected to each other, an ion is selected at random and the number of ions within ϵ are counted. If the number of ions within ϵ is equal to or greater than η , then the central ion is labeled as *core* and its neighbors are labeled as *border*; otherwise, the ion is labeled as *noise* (Figure 2). The procedure

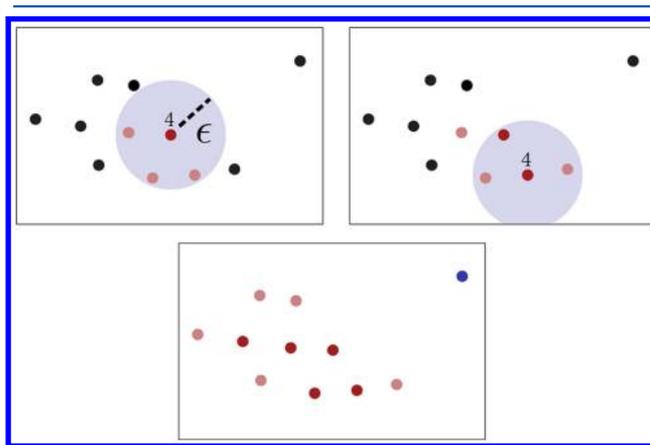


Figure 2. Schematic diagram illustrating the DBSCAN algorithm for $\eta = 4$. Top left panel: an ion is chosen at random and ions within ϵ are counted; the value is 4, so the ion is marked as *core* (red). Its neighbors are labeled as *border* (pink). Top right panel: the procedure is repeated for one of the border ions, which becomes a core ion. Bottom panel: all ions in the cluster have been processed, and no more ions are reachable from the core ions. The isolated ion is labeled as *noise* (blue).

is repeated by picking an ion from the *border*. If that ion is a *core* ion, it will generate a new *border*, and the procedure is repeated recursively. Eventually, no more points are reachable, and we label all of the collected *core* and *border* ions as part of the same cluster. The procedure is then repeated by choosing another ion at random (which might belong to a new cluster), until all ions are processed.

The value of ϵ is the minimum distance for which two ions are considered a part of the same aggregate. A high value of ϵ causes some clusters that are close but not connected to each other to be aggregated into a single cluster. On the other hand, a small value of ϵ would fail to connect ions that are effectively bonded (every ion would be detected as a single cluster). The value of η acts as a filter for very small clusters, as every cluster that has fewer than η ions is discarded. A very low value of η detects too many short-lived fluctuations, which are difficult to follow and analyze, whereas a high value would detect only large clusters, providing no opportunity to investigate early stages of cluster formation. In the present analysis the values $\epsilon = 0.3$ nm and $\eta = 6$ ions are used. The value 0.3 nm is approximately the NaCl bond length, and $\eta = 6$ ions was

chosen to be the minimum cluster size. These values proved to work well in practice, avoiding the problems noted above.

2.2.3. Time Resolution of Clusters. We wish to follow the time evolution of all clusters identified, as described above. Therefore, it is necessary to determine how clusters are connected in time. This would be a simple question if we had only a single cluster that grew or dissolved as time progressed. However, in the present case, we have many clusters that continually change (divide, grow, dissolve) as the simulation advances in time. Therefore, we need to clearly define how we identify and follow clusters in time. Our general approach is to find “similar” clusters in successive configurations (using an appropriate similarity index) and build a time connectivity graph using the following procedure.

First, we initialize an empty graph, in which each cluster detected appears as a node labeled $c_i(s)$, where i indicates the cluster and s , the time frame (configuration). Then, for all consecutive frames, pairwise similarities, $d_{ij}(s) = \text{sim}(c_i(s), c_j(s + 1))$, are calculated between clusters using the similarity measure defined below. If two clusters in different frames are detected to be sufficiently similar (satisfy an appropriate threshold), then they are deemed to be the same cluster at different points in time, and the nodes representing these clusters are connected through an edge. The outcome of this procedure is a graph that connects clusters in different time frames, such that every connected component of this graph is the trajectory of a particular cluster, as illustrated in Figure 3.

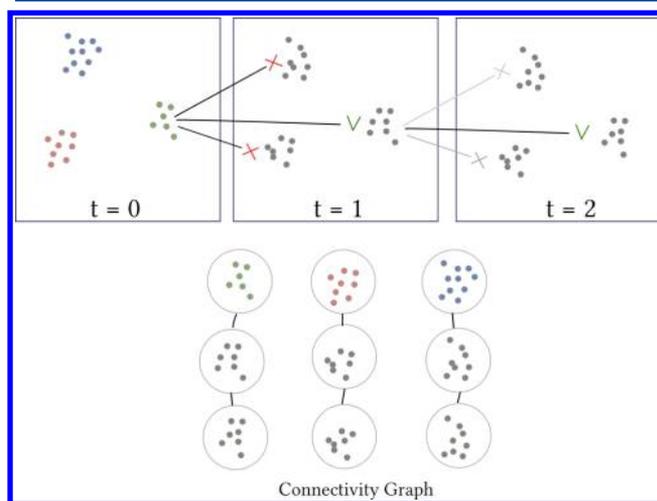


Figure 3. Schematic diagram illustrating the cluster time resolution algorithm. At $t = 0$ the clusters are labeled employing the DBSCAN algorithm. A cluster is compared to clusters in the following frame, and a connection is made if the clusters are sufficiently similar, judged by the Jaccard index as described in the text. The procedure is repeated recursively for the following frames. By performing this procedure for all clusters, it is possible to obtain a connectivity graph (bottom), in which each connected component represents the time evolution of a single cluster.

Several similarity measures could be used to identify matching clusters, and here, we employ the Jaccard index,⁵⁰ defined as the intersection of two sets divided by the union of two sets

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Here, A and B represent two clusters in consecutive time frames, and clearly, $J(A, B)$ is 1 if both clusters contain exactly the same ions and 0 if they do not share any ions. The threshold value of $J(A, B)$ must be selected with care to avoid ambiguities in cluster identity. For example, if two clusters, a and b , detected in frame one merge into a single cluster in frame two, it is not clear whether the merged cluster is a , b , or an entirely new cluster, c . Similar ambiguities can arise from other possible cluster evolution scenarios. To assign a unique well-defined timeline to a cluster, it is necessary to remove all ambiguities. A natural way to do this is to connect clusters in consecutive frames only if they satisfy the threshold $J(A, B) > 0.5$. By definition of the Jaccard index, this threshold ensures that a cluster will match at most one cluster in the following frame, giving a set of unique cluster trajectories.

One potential issue with this method of obtaining cluster trajectories is how to take account of splits and merges that happen over multiple time frames. For example, if we have a large cluster (e.g., 500 ions) and a piece (e.g., 100 ions) becomes temporarily detached, the connectivity with the original cluster would be lost and the detached piece would appear as a newborn cluster of 100 ions. This problem is handled by detecting abnormally large newborn clusters and merging them with the original cluster. Fortunately, such events are rare (it happened once in our five simulations), and the postprocessing step was sufficient to solve the problem.

Another possible issue is that small clusters can partially redissolve (in the sense that some of their ions become less ordered), fall below the six-ion threshold, and go undetected for a few time frames. To solve this problem, we modified the algorithm to look for a match in 10 consecutive frames following the initial cluster detection. Newborn clusters that had no match in the following 10 frames were assumed to be short-term fluctuations and discarded from the analysis.

3. RESULTS AND DISCUSSION

Ionic clusters form and dissipate continuously in the super-saturated solution. These clusters are detected and followed in time, as described above. We are interested in understanding what features (if any) of ionic clusters influence their ability to survive and eventually achieve nucleation. We considered several cluster attributes or properties, but only two had a significant, easily observable influence on the nucleation probability of small clusters. One of those, cluster size, is of rather obvious importance and is in fact the crucial parameter in the usual application of CNT, which assumes spherical nuclei that become critical at a certain radius. The other cluster property that proved very influential, especially so in smaller clusters, we call \bar{q}_8 or cluster *crystallinity*. This property is simply the average value of the q_8 bond order parameter defined above, taken over all ions in a cluster.

Other cluster properties considered include volume, surface area, sphericity, average neighbor count, hydration, and radius of gyration. These properties are precisely defined and discussed in the [Supporting Information](#). For smaller clusters (e.g., 10 ions), none of these properties showed a significant correlation with the probability of nucleation. For larger clusters (e.g., 30 ions), we do find noticeable correlations for three properties, with smaller surface areas, higher sphericities, and smaller radii of gyration all favoring nucleation. These attributes are all measures of cluster compactness, and it is not surprising that they are correlated with nucleation probability for larger clusters.

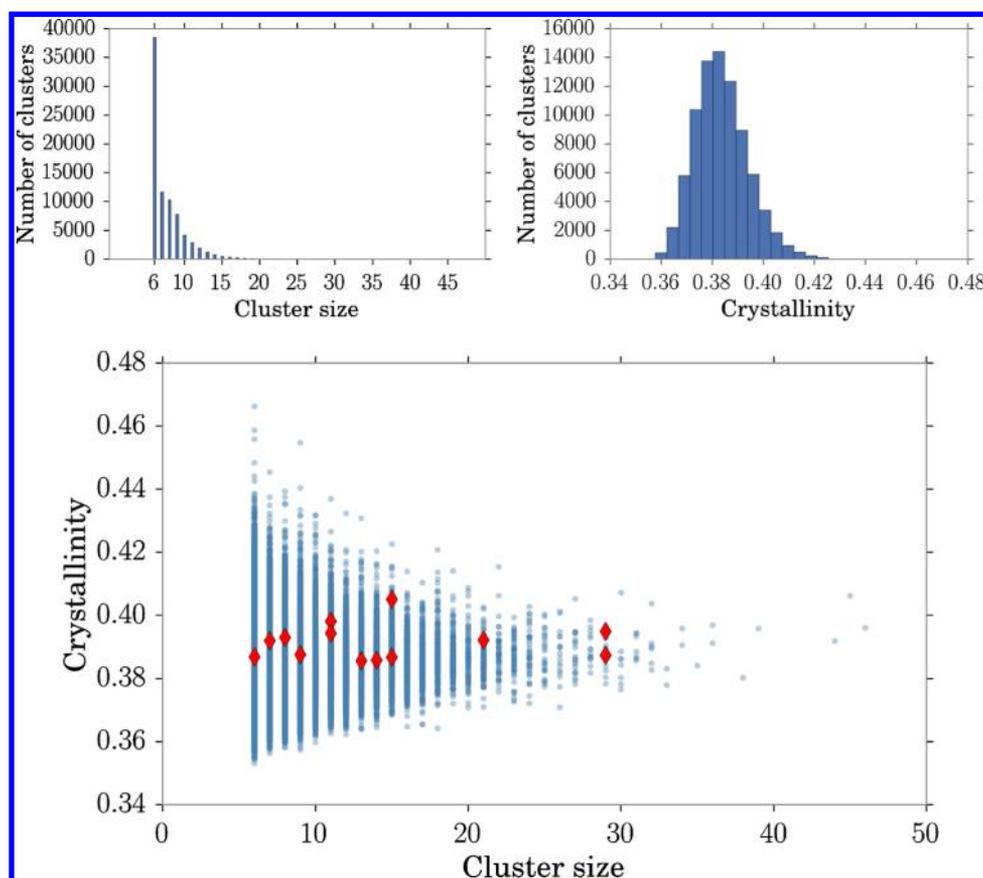


Figure 4. Size (top left) and crystallinity (top right) histograms for clusters at first detection. The bottom panel shows the joint distribution of size and crystallinity. Note that very few clusters begin with both high size and high crystallinity, suggesting that this feature develops as clusters evolve in time. The red points indicate the 13 clusters that achieve nucleation, and we note that these show no obvious preference for any region of the joint distribution.

We note that all results reported below are based on the five simulations of 523–550 ns at $x_{\text{NaCl}} = 0.22$, as listed in Table 2. In our five simulations, a total of 13 nucleation events were observed. In these simulations, we noticed that clusters with a lifetime of at least 30 ns that reached a size greater than ~ 50 ions never redissolved and continued to grow during the simulation. These clusters were deemed to have achieved nucleation. Note that we would expect this observational “definition” of nucleation to be concentration dependent.

At inception (defined as the earliest time a cluster is detected), the ionic clusters span a rather wide range of sizes and crystallinities. The size and crystallinity distributions of all clusters detected are shown in the top left and top right panels of Figure 4, respectively. Note that on average our algorithm detects approximately two new clusters per time frame, giving a total of $\sim 80\,000$ clusters detected over the course of the five simulations. It is also worth noting that the filtering step of the algorithm removes $\sim 90\%$ of the ions, indicating that only $\sim 10\%$ of the ions present in solution belong to a cluster of any sort.

From the histogram of initial sizes (Figure 4, top left panel), we see that the peak occurs at six ions (the smallest cluster detected by our algorithm) and the frequency of larger sizes follows an exponential-like decay. Although very rare, clusters of more than 30 ions are found; however, as noted below, these correspond to elongated, amorphous-looking structures. The distribution of crystallinities (Figure 4, top right panel) is skewed to the right, but one does not observe the high values

characteristic of a bulk crystal (Figure 1). A scatter plot of cluster crystallinity versus size is shown in the bottom panel of Figure 4. It is apparent that at cluster inception the correlation between these variables is modest at most. For smaller clusters, there are large fluctuations of crystallinity for a given cluster size, with large and small values occurring with high frequency. For larger cluster sizes, the crystallinity fluctuations decrease markedly, which is not surprising, as, statistically, variance is expected to decrease with increasing sample size, which here is the number of ions in the cluster.

Ionic clusters exhibit a range of crystallinities, and the examples given in Figure 5 provide an idea of how cluster structure and crystallinity are related. Newly detected small clusters can have low (bottom left panel) or high (bottom right panel) crystallinities, with high values being associated with more regular structures. For larger clusters (more than 30 ions), we do not find any very high crystallinities at the point of initial detection, as such clusters tend to have elongated shapes (Figure 5, top left panel).

3.1. Cluster Lifetimes. As discussed above, clusters are born with certain characteristics, and after some time, they will either achieve nucleation (very rare) and continue to grow as a crystal or, much more commonly, dissolve back into solution. Given that one would expect a connection between cluster lifetime and nucleation, it is of interest to investigate which cluster characteristics influence their lifetimes, defined here as the total time a cluster is detected (by the algorithm described above) as a distinct entity in solution. As, by definition, clusters

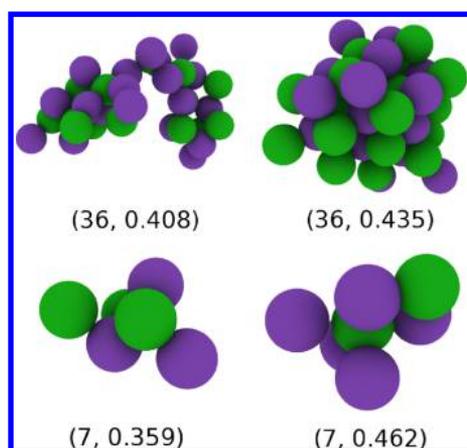


Figure 5. Examples of clusters of different sizes and crystallinities (given in parentheses below each cluster). Na^+ and Cl^- ions are colored purple and green, respectively. At inception, larger clusters tend to be elongated and of lower crystallinity (top left), whereas larger high-crystallinity clusters (not observed at early stages) tend to be more compact (top right). For small clusters, structural features associated with low (bottom left) and high crystallinities (bottom right) are not so apparent.

that nucleate have infinite lifetimes, we excluded these from our statistical analysis. Note that in the five simulations (Table 2) included in the analysis only 13 clusters achieved nucleation, whereas $\sim 80\,000$ failed to nucleate. Focusing on the large number of failures provides us with excellent statistics, and clearly, the very few that did nucleate would have had a negligible effect on the lifetime and survival distributions discussed below.

The overall lifetime distribution of failed clusters (uncontrolled for size or crystallinity) is shown in the left panel of Figure 6. This plot shows that the great majority of clusters are very short lived (lifetimes less than a nanosecond), but there are a significant number of clusters that live for more than 10 ns.

A more instructive way to represent lifetime information is to construct survival functions. The survival function, $S(t)$, is defined as the probability that a cluster has a lifetime, τ , larger than t , or

$$S(t) = P(\tau > t)$$

The survival function starts at 1 (all clusters have a nonzero lifetime) and decreases with t , depending on the surviving fraction of the population. The survival function can be written as a cumulative product

$$S(t) = R(1)R(2)\dots R(t)$$

with

$$R(t) = \frac{\text{population}(t)}{\text{population}(t-1)}$$

where $\text{population}(t-1)$ represents the population at frame $t-1$ and $\text{population}(t)$ is the surviving population at the next frame, t . $R(t)$ is the probability of surviving for one time interval at t . We note that this method of calculating $S(t)$ is commonly used and is known as the Kaplan–Meier estimator.⁵¹ The overall survival function, including all clusters, is shown in the right panel of Figure 6, and we note that over 90% of all clusters survive for less than 2 ns.

Survival functions can be used to investigate whether and how particular cluster characteristics influence their probability of survival. To isolate the influence of crystallinity, we divide clusters of fixed size into groups of high (>0.40) and low (≤ 0.40) crystallinities and compare the survival functions. Note that survival functions are constructed using the size and crystallinity of a cluster at first detection. Obviously, neither of these cluster characteristics remains fixed as the cluster evolves in time. Results for clusters of 10 ions are shown in Figure 7,

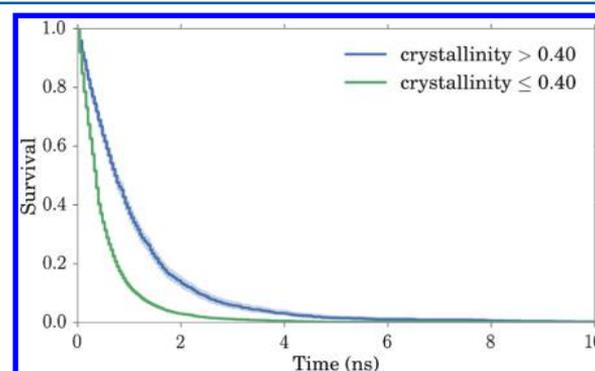


Figure 7. Survival curves for clusters of size 10, grouped by high (>0.40) and low (≤ 0.40) crystallinities.

and it is obvious that crystallinity has a large effect on the survival probability. Clusters in the higher crystallinity group have a substantially higher survival probability compared to clusters in the lower crystallinity group. This is true for all cluster sizes, remarkably, even in the six-ion case, which is the smallest cluster size we consider. This is illustrated in Figure 8, in which we plot the ratio (high/low) of median lifetimes for clusters of high and low crystallinities as a function of cluster size. We see that the ratio is significant (~ 1.5) for clusters of six

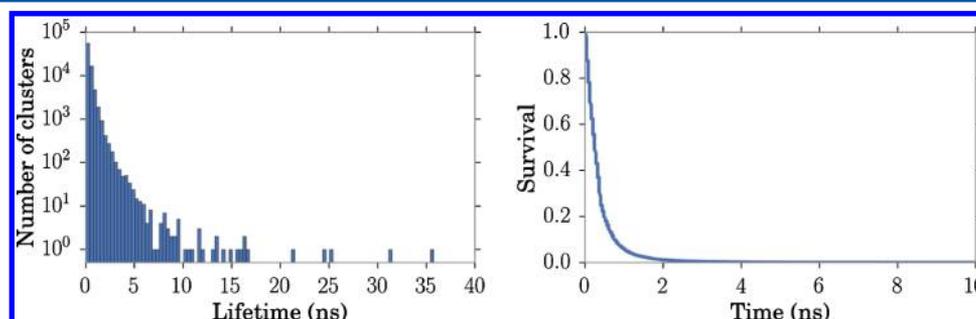


Figure 6. Histograms of “failed” cluster lifetimes (left) and the overall survival function (right) for all clusters detected in the simulations.

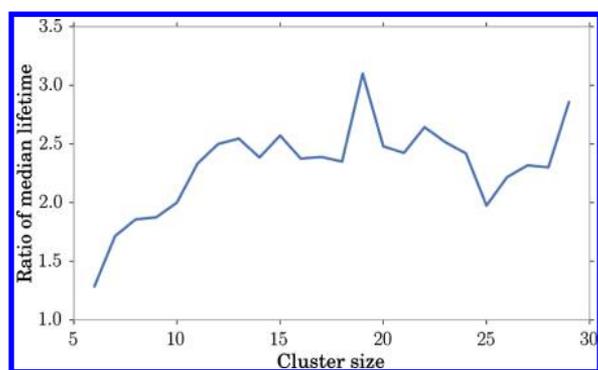


Figure 8. Ratio of median lifetimes for high vs low crystallinity clusters of different sizes. The high-crystallinity clusters always have a higher median lifetime.

ions; it rises rapidly with increasing cluster size, reaches a plateau at clusters of ~ 10 ions, and then rises again for clusters containing more than ~ 18 ions. Results for clusters larger than 30 ions are not included, as for larger clusters the statistical analysis is less reliable due to small sample sizes. In an analogous manner, one can fix crystallinity and calculate how survival functions vary with size. This is less interesting, simply showing that, as expected, larger clusters have a larger survival probability.

Although cluster survival is not a direct measure of the nucleation probability, one would expect both phenomena to be related to cluster stability. Therefore, on the basis of the above analysis, we would expect cluster crystallinity to be an important factor influencing the probability of a cluster achieving nucleation. Specifically, for fixed cluster sizes, higher crystallinities should favor higher nucleation probabilities. The relationship between cluster crystallinity and nucleation is treated directly in the following section, and we show that the behavior is indeed as we would anticipate on the basis of the survival analysis. We remark that the survival analysis has a great advantage of the large number of clusters (events) included providing convincing statistics. On the other hand, we observe only a few nucleation events, so a statistical analysis based on nucleation alone would be less convincing.

3.2. Crystallinity and Nucleation. To get a qualitative idea of the influence of cluster size and crystallinity on nucleation, it is useful to plot cluster trajectories in (size, crystallinity) space. This is done in Figure 9 for a cluster that nucleates (blue) and one that does not (red). It is easy to see that after a brief residency at low sizes and low crystallinities the cluster that nucleates appears to surpass a “critical region” in (size, crystallinity) space and never falls back. This suggests that both size and crystallinity influence the probability of nucleation.

This can also be seen in the early-stage time evolution profiles shown in Figure 10 for the nucleating and failing clusters. We see that although relatively long lived (~ 12 ns), and sometimes exceeding 30 ions in size, for most of the trajectory the crystallinity of the failing cluster lies below that in the nucleating case. Snapshots (not shown) suggest that the failing cluster grows in an elongated manner, with a resulting lower crystallinity. The growth profile for a nucleating cluster is shown in Figure 11. It is interesting to note that the growth rate increases with size, attributable to the growing surface area of the nucleus. The crystallinity profile initially grows but plateaus

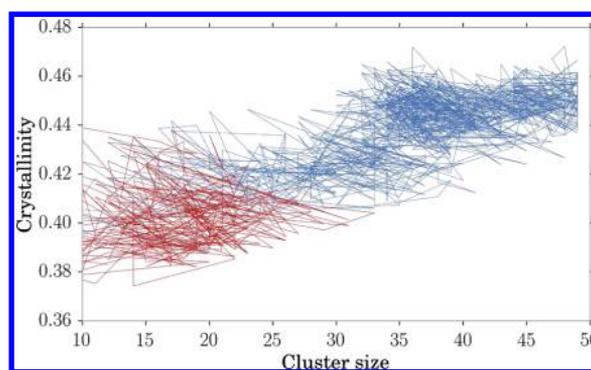


Figure 9. Trajectory shown in (size, crystallinity) space for nucleated (blue) and failed (red) clusters as they evolve in time. In early stages, both clusters oscillate in size up to ~ 30 ions, but the crystallinity of the cluster that eventually achieves nucleation reaches higher values than those in the failed case. This suggests that it is a combined effect of both size and crystallinity that promotes nucleation.

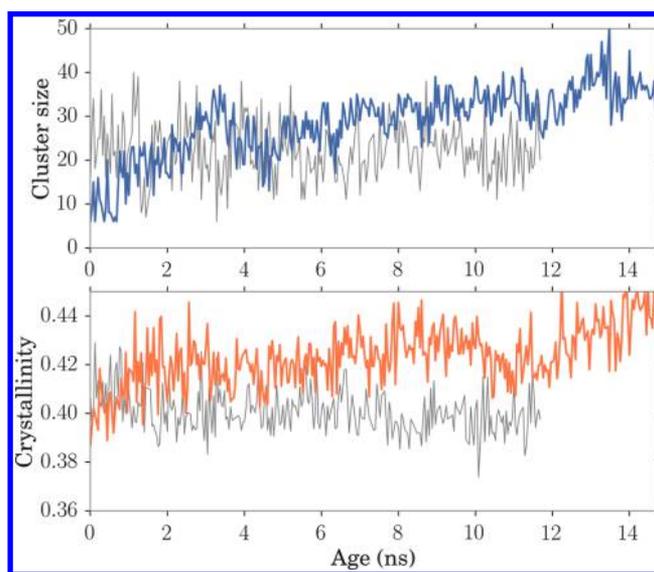


Figure 10. Comparison of the growth and crystallinity profiles for nucleated (colored) and failed (gray) clusters. After a short period, the nucleated cluster manages to reach quite a high crystallinity and maintain its size. In contrast, the failed cluster, although maintaining its size, experiences a steady decrease in crystallinity.

as the cluster becomes larger in size, and the crystallinity approaches that of a perfect crystal.

The influence of cluster size and crystallinity on nucleation probability can be explored more quantitatively by statistically comparing these properties for clusters that nucleate with those for clusters that fail. If cluster size were the only factor influencing nucleation, then two clusters of the same size would have the same probability of nucleation. Therefore, if we fix the cluster size and observe a difference in crystallinity between nucleations and failures, we can isolate the effect of crystallinity.

One way to fix cluster size is to follow the trajectory of each cluster in time until it reaches a particular size, at which point the cluster crystallinity is recorded. Values obtained in this way can be reasonably assumed to be independent because the clusters exist at different points in time or in different simulations, and we register only a single crystallinity value for each cluster. Crystallinity distributions obtained as described are shown in Figure 12 for clusters of different sizes. The

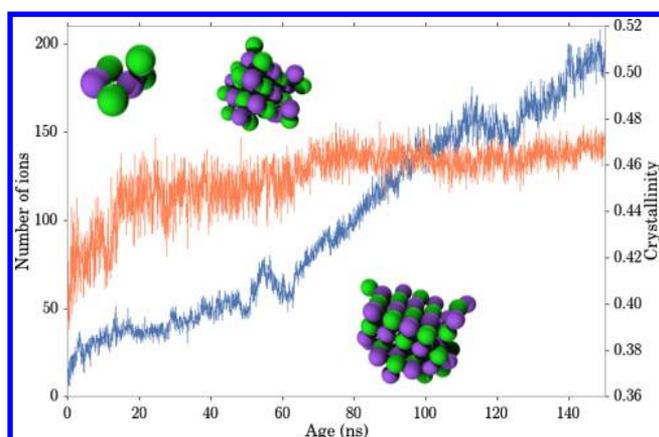


Figure 11. Size (blue) and crystallinity (orange) profiles for a nucleated cluster. The fluctuation in crystallinity is fairly high, and as the crystal increases in size, the crystallinity reaches a plateau at ~ 0.47 . The snapshots represent the cluster at 0 ns (left), 40 ns (center), and 80 ns (right).

crystallinities of the clusters that achieved nucleation are also indicated in the plots, and we see that for all cluster sizes the

crystallinities of the nucleated clusters fall mainly in the upper half of the crystallinity distribution. In other words, the clusters that eventually nucleate have, on average, higher crystallinities than those of clusters that fail, and it is remarkable that this distinction exists even for clusters as small as six ions. For comparison, similar plots for other properties are given in the Supporting Information (Figure S1).

The influence of cluster crystallinity on nucleation is further demonstrated in Figure 13, in which the probability of nucleation, $P(N)$, for clusters of N ions is shown for clusters of “high” and “low” crystallinities. If the cluster crystallinity is >0.40 , it is labeled high, otherwise it is labeled low. The nucleation probabilities, $P(N|high)$ and $P(N|low)$, are shown in Figure 13 for N ranging from 6 to 35 ions. Results for larger clusters are not shown because the sample size is too small to give good estimates of the probabilities. We note that $P(N|high)$ is much higher than $P(N|low)$ for all cluster sizes. The effect is most pronounced for small clusters (which have a very low total probability of nucleation), where $P(N|high)$ is up to 8 times larger than $P(N|low)$. The ratio $P(N|high)/P(N|low)$ generally decreases with increasing N , but it remains at ~ 3 for $N = 35$.

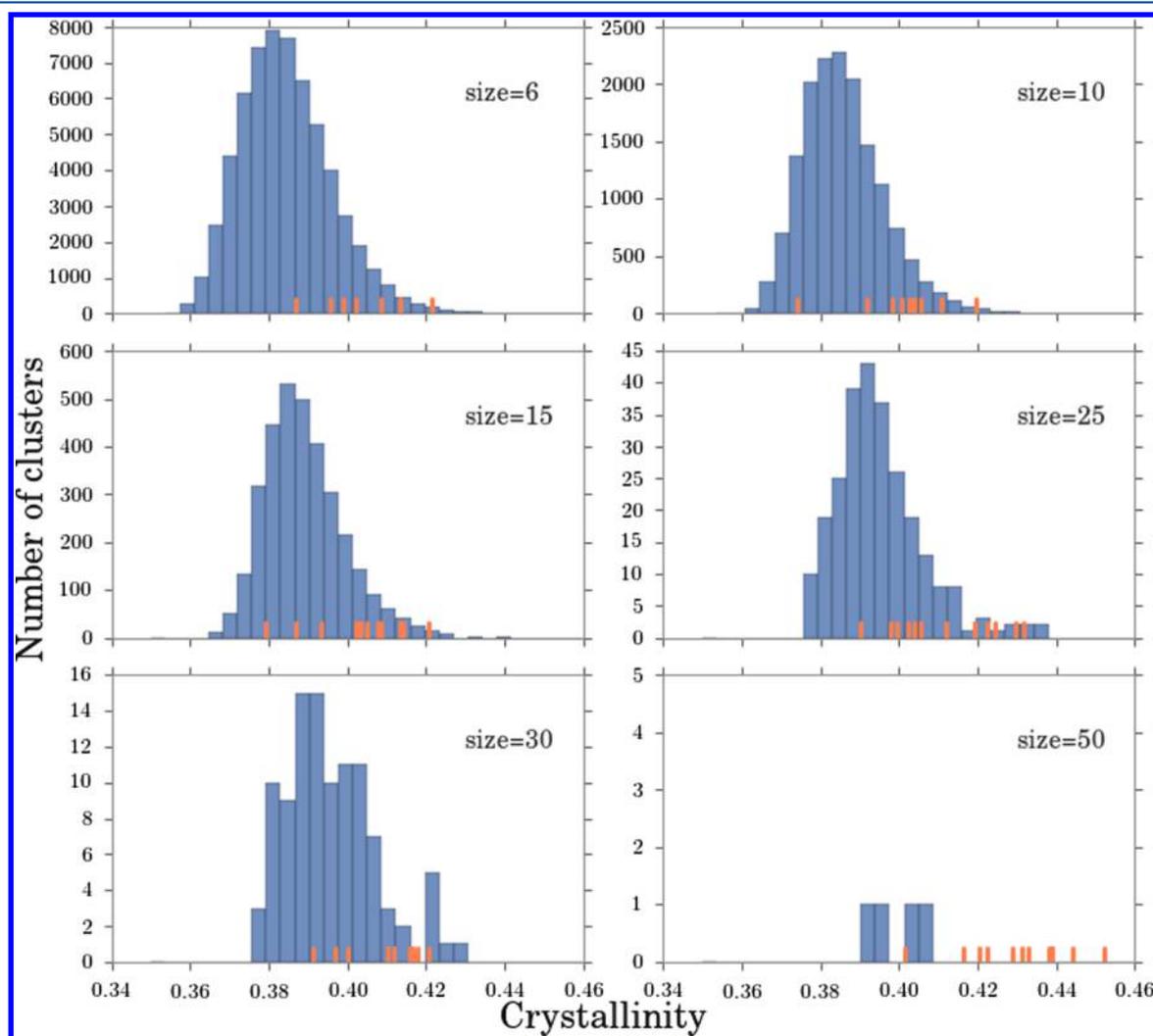


Figure 12. Crystallinity distributions (blue histogram) of failed clusters of different sizes (number of ions). Clusters that achieved nucleation are indicated by single orange lines. Note that clusters that achieve nucleation come preferentially from the upper part of the crystallinity distribution.

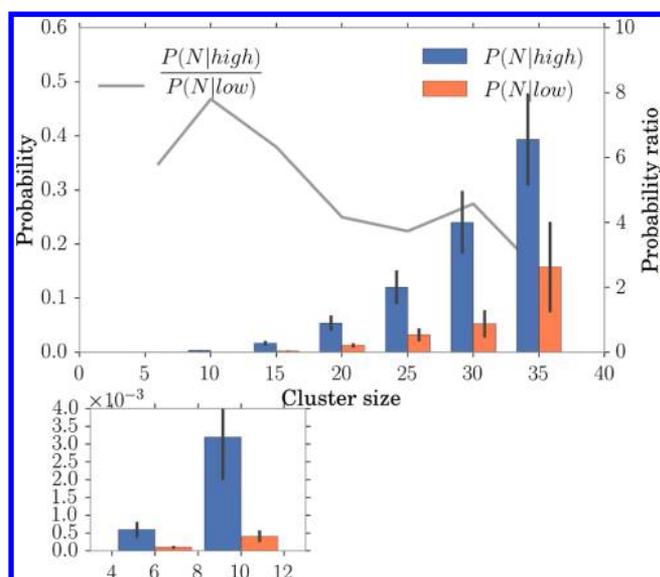


Figure 13. Probabilities of high ($\bar{q}_8 > 0.40$) and low ($\bar{q}_8 < 0.40$) crystallinities achieving nucleation. The error bars (vertical black lines) represent one standard deviation. The lower panel shows the results for small clusters on an expanded scale. Note that for all cluster sizes $P(N|high)$ is always higher than $P(N|low)$, and the high/low ratio is shown as a gray line in the upper panel.

This analysis highlights the fact that nucleation is substantially influenced by factors other than cluster size. For a given cluster size, geometrical order, as captured by the crystallinity parameter, increases the cluster lifetime and the probability of achieving nucleation.

3.2.1. Cluster Binding Energy. It is interesting to ask why high crystallinity leads to increased cluster lifetimes and influences the probability that a cluster will achieve nucleation. One possibility that immediately comes to mind is that in high-crystallinity clusters the direct ion–ion interactions lead to lower cluster energies and hence increased stabilities. To investigate this possibility, we calculated the energies

$$U_k = \frac{1}{2} \sum_i^{N_k} \sum_j^{N_k} u(r_{ij}) \quad (2)$$

where N_k is the total number of ions in cluster k and $u(r_{ij})$ is the ion–ion interaction as defined in Section 2. One potential problem with this analysis is that large fluctuations can occur in the apparent cluster energy depending on whether or not a cluster has a net charge. These fluctuations can be created by a single ion moving in or out across the defined cluster boundary and hence can be rather arbitrary and artificial. Therefore, to avoid this problem, we include only electrically neutral clusters in this analysis.

Joint distributions of crystallinity and cluster energy for clusters of 6 and 10 ions are shown in Figure 14. For clusters of six ions, correlation coefficient R is -0.074 , indicating only a very weak correlation between energy and crystallinity. For clusters of 10 ions, $R = -0.22$, showing that the correlation increases with cluster size but still remains rather weak. Moreover, we see from Figure 14 that the clusters that eventually nucleate come mainly from the high-crystallinity tail of the crystallinity distribution, but they are nearly evenly spread over the energy distribution. These observations show that crystallinity is not a measure of increased cluster stability due to lower cluster energies coming through the direct ion–ion interactions. Clearly, other factors must be involved.

3.2.2. Role of Na^+ and Cl^- Ions in Small Crystals. It is also of interest to more closely examine the structural nature of small NaCl clusters, as these clusters represent a fundamental step in the nucleation process. To do this, we divide all ions in the system into two classes, high q_8 (≥ 0.4) and low q_8 (< 0.4), and calculate the running counterion coordination numbers for Na^+ and Cl^- .

Results for single frames at 40 and 240 ns are shown in Figure 15. For the high q_8 class, we see that both Na^+ and Cl^- have more counterions (three to four) in the first coordination shell (~ 0.35 nm) than the solution average (~ 1). This is not surprising because high q_8 ions generally belong to ionic clusters. However, from Figure 15, we also notice that high q_8 Na^+ ions have more counterion neighbors than high q_8 Cl^- ions, suggesting that Na^+ ions tend to lie more within the “interior” of small clusters than Cl^- ions, consistent with an earlier observation of Zahn.¹⁸ A possible reason for this is that the smaller Na^+ ions interact more strongly with water molecules and hence more first-shell counterions are needed to compensate for water molecules lost from the first hydration shell when ionic clusters are formed. Note that the gap between

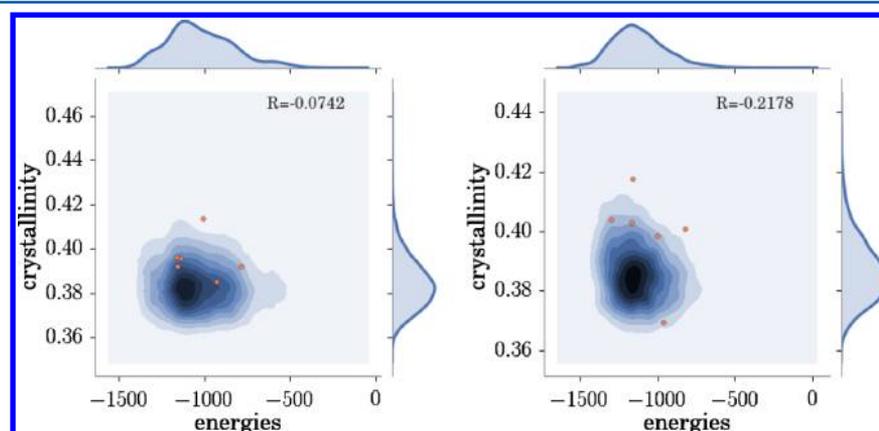


Figure 14. Joint distribution of crystallinity and cluster energy (kJ mol^{-1}) for cluster sizes of 6 (left) and 10 (right) ions. Values corresponding to clusters that eventually nucleated are displayed in orange. The correlation coefficient is negative and very low at size 6 and moderately low at size 10, indicating that crystallinity carries different information from that carried by energetic stability.

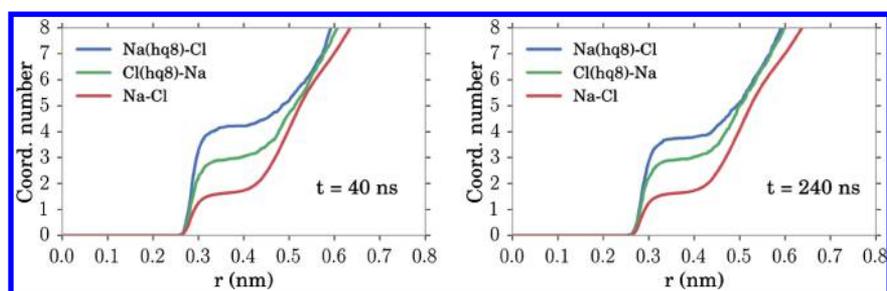


Figure 15. Running coordination numbers for the following: high q_8 Na^+ with all Cl^- (blue), high q_8 Cl^- with all Na^+ (green), and all Na^+ and Cl^- (red). In the first coordination shell (~ 0.35 nm), both high q_8 ions are surrounded by more counterions (three or four) than the average (one). Also, high q_8 Na^+ are surrounded by more counterions than those around high q_8 Cl^- , suggesting that, on average, they occupy positions deeper within the clusters. The difference in the first-shell coordination number decreases at longer times, as larger crystals develop in the simulation, and the internal preference of Na^+ becomes less noticeable.

the coordination numbers of high q_8 Na^+ and Cl^- decreases in later frames, as larger clusters develop in the simulation.

To obtain additional evidence that Na^+ and Cl^- do not contribute symmetrically to the structure of small clusters, for each cluster we calculate the parameter

$$\Delta q_8 = \langle q_8^+ \rangle - \langle q_8^- \rangle$$

where $\langle q_8^+ \rangle$ and $\langle q_8^- \rangle$ are the average q_8 values for the Na^+ and Cl^- ions in the cluster.

Figure 16 shows Δq_8 as a function of cluster size. These results are from the second replica simulation (Table 2). We

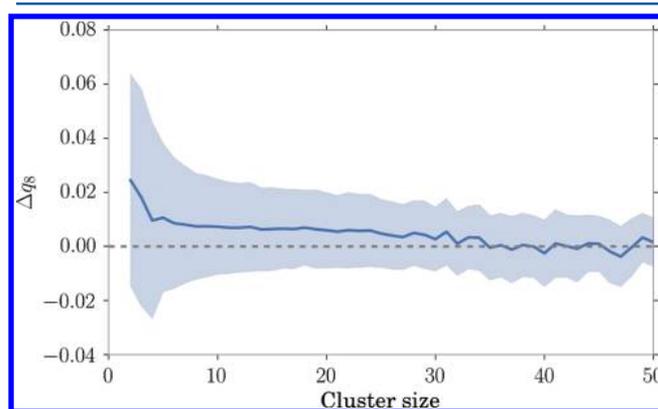


Figure 16. Δq_8 as a function of cluster size. The shaded area indicates one standard deviation. The average value is slightly positive, indicating that Na^+ ions tend to occupy more ordered environments than those occupied by Cl^- ions.

note that for small clusters Δq_8 is positive on average, indicating that Na^+ tends to lie in more ordered environments than Cl^- . As we would expect, the distinction between Na^+ and Cl^- grows smaller as clusters grow in size, and Δq_8 approaches zero for large clusters. Survival functions and nucleation probabilities were analyzed separately for positive and negative values of Δq_8 , but no significant differences were found. Thus, although there is a structural distinction for small clusters, this effect does not have a significant influence on cluster survival or the probability of nucleation.

4. SUMMARY AND CONCLUSIONS

Direct MD simulations have been employed to identify and investigate factors that influence the nucleation of NaCl crystals in model supersaturated aqueous solutions. We develop methods that allow potential nuclei to be detected as small

clusters (approximately six ions) and followed in time until nucleation is achieved, which occurs very rarely, or the cluster dissolves back into solution.

Our analysis clearly demonstrates that cluster size is not the only property that has an important influence on the expected lifetime and nucleation probability of a particular cluster. We show that the geometric arrangement of the ions in the cluster, as measured by a single parameter that we call cluster crystallinity, is also very influential. For example, for clusters of 10 ions, the median lifetime for clusters of high crystallinity is double that of those of low crystallinity, and their probability of achieving nucleation is ~ 8 times greater. Similarly, smaller and larger clusters also have significantly longer lifetimes and greater probabilities of nucleation.

Physically, it is not entirely clear why crystallinity (as measured by our parameter) has such a large influence on cluster lifetime and nucleation probability, and this is especially true for small clusters of 6 or 10 ions. One possibility is that there is a connection with the binding energy of a cluster. However, we did not find a strong correlation between binding energy and our crystallinity parameter, indicating that the crystallinity parameter is not merely a proxy for energetic stability. For small clusters, we did find that Na^+ ions had some preference for a “central” position, in accordance with Zahn’s observation,¹⁸ but such cluster arrangements had no influence on the probability of nucleation.

Given that there is no obvious energetic explanation for the increased stability of small clusters of higher crystallinity, it appears that the advantage must lie in the microscopic dynamics of cluster growth and/or disintegration. One possibility is that ions in more ordered environments are less exposed to water molecules and hence high-crystallinity clusters are less susceptible to disintegration.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.6b05291.

Comparison of additional cluster properties (average neighbor count, volume, surface area, sphericity, hydration, and radius of gyration) for clusters that nucleate and those that do not (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: gabriele.janaro@gmail.com (G.L.).

*E-mail: patey@chem.ubc.ca. Tel: (604) 822-2996 (G.N.P.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The financial support of the Natural Science and Engineering Research Council of Canada is gratefully acknowledged. This research has been enabled by the use of WestGrid and Compute/Calcul Canada computing resources, which are funded in part by the Canada Foundation for Innovation, Alberta Innovation and Science, BC Advanced Education, and the participating research institutions. WestGrid and Compute/Calcul Canada equipment is provided by IBM, Hewlett Packard, and SGI.

REFERENCES

- (1) Blatt, H.; Tracy, R.; Owens, B. *Petrology: Igneous, Sedimentary, and Metamorphic*, 3rd ed.; W. H. Freeman: New York, NY, 2005.
- (2) Webber, K. L.; Falster, A. U.; Simmons, W. B.; Foord, E. E. The Role of Diffusion-Controlled Oscillatory Nucleation in the Formation of Line Rock in Pegmatite-Plite Dikes. *J. Petrol.* **1997**, *38*, 1777–1791.
- (3) Arosio, P.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. Chemical Kinetics for Drug Discovery to Combat Protein Aggregation Diseases. *Trends Pharmacol. Sci.* **2014**, *35*, 127–135.
- (4) Rabinow, B. E. Nanosuspensions in Drug Delivery. *Nat. Rev. Drug Discovery* **2004**, *3*, 785–796.
- (5) Laaksonen, A.; Talanquer, V.; Oxtoby, D. W. Nucleation: Measurements, Theory, and Atmospheric Applications. *Annu. Rev. Phys. Chem.* **1995**, *46*, 489–524.
- (6) Kulmala, M.; Petäjä, T.; Ehn, M.; Thornton, J.; Sipilä, M.; Worsnop, D.; Kerminen, V. M. Chemistry of Atmospheric Nucleation: on the Recent Advances on Precursor Characterization and Atmospheric Cluster Composition in Connection with Atmospheric New Particle Formation. *Annu. Rev. Phys. Chem.* **2014**, *65*, 21–37.
- (7) Aastuen, D.; Clark, N.; Cotter, L.; Ackerson, B. J. Nucleation and Growth of Colloidal Crystals. *Phys. Rev. Lett.* **1986**, *57*, 1733.
- (8) Wang, Z.; Wang, F.; Peng, Y.; Han, Y. Direct Observation of Liquid Nucleus Growth in Homogeneous Melting of Colloidal Crystals. *Nat. Commun.* **2015**, *6*, No. 6942.
- (9) Desarnaud, J.; Derluyn, H.; Carmeliet, J.; Bonn, D.; Shahidzadeh, N. Metastability limit for the nucleation of NaCl crystals in confinement. *J. Phys. Chem. Lett.* **2014**, *5*, 890–895.
- (10) Sleutel, M.; Lutsko, J.; Van Driessche, A. E.; Durán-Olivencia, M. A.; Maes, D. Observing Classical Nucleation Theory at Work by Monitoring Phase Transitions with Molecular Precision. *Nat. Commun.* **2014**, *5*, No. 5598.
- (11) Pienack, N.; Bensch, W. In-Situ Monitoring of the Formation of Crystalline Solids. *Angew. Chem., Int. Ed.* **2011**, *50*, 2014–2034.
- (12) Jiménez-Ángeles, F.; Firoozabadi, A. Nucleation of Methane Hydrates at Moderate Subcooling by Molecular Dynamics Simulations. *J. Phys. Chem. C* **2014**, *118*, 11310–11318.
- (13) Chakraborty, D.; Patey, G. N. Evidence That Crystal Nucleation in Aqueous NaCl Solution Occurs by the Two-Step Mechanism. *Chem. Phys. Lett.* **2013**, *587*, 25–29.
- (14) Chakraborty, D.; Patey, G. How Crystals Nucleate and Grow in Aqueous NaCl Solution. *J. Phys. Chem. Lett.* **2013**, *4*, 573–578.
- (15) Anwar, J.; Zahn, D. Uncovering Molecular Processes in Crystal Nucleation and Growth by Using Molecular Simulation. *Angew. Chem., Int. Ed.* **2011**, *50*, 1996–2013.
- (16) Salvalaglio, M.; Perego, C.; Giberti, F.; Mazzotti, M.; Parrinello, M. Molecular-Dynamics Simulations of Urea Nucleation from Aqueous Solution. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, E6–E14.
- (17) Matsumoto, M.; Saito, S.; Ohmine, I. Molecular Dynamics Simulation of the Ice Nucleation and Growth Process Leading to Water Freezing. *Nature* **2002**, *416*, 409–413.
- (18) Zahn, D. Atomistic Mechanism of NaCl Nucleation from an Aqueous Solution. *Phys. Rev. Lett.* **2004**, *92*, No. 040801.
- (19) Zimmermann, N. E.; Vorselaars, B.; Quigley, D.; Peters, B. Nucleation of NaCl from Aqueous Solution: Critical Sizes, Ion-Attachment Kinetics, and Rates. *J. Am. Chem. Soc.* **2015**, *137*, 13352–13361.
- (20) Gebauer, D.; Kellermeier, M.; Gale, J. D.; Bergström, L.; Cölfen, H. Pre-Nucleation Clusters as Solute Precursors in Crystallisation. *Chem. Soc. Rev.* **2014**, *43*, 2348–2371.
- (21) Kellermeier, M.; Picker, A.; Kempter, A.; Cölfen, H.; Gebauer, D. A Straightforward Treatment of Activity in Aqueous CaCO₃ Solutions and the Consequences for Nucleation Theory. *Adv. Mater.* **2014**, *26*, 752–757.
- (22) Merikanto, J.; Zapadinsky, E.; Lauri, A.; Vehkamäki, H. Origin of the Failure of Classical Nucleation Theory: Incorrect Description of the Smallest Clusters. *Phys. Rev. Lett.* **2007**, *98*, No. 145702.
- (23) Gebauer, D.; Völkel, A.; Cölfen, H. Stable Prenucleation Calcium Carbonate Clusters. *Science* **2008**, *322*, 1819–1822.
- (24) Vekilov, P. G. Dense Liquid Precursor for the Nucleation of Ordered Solid Phases from Solution. *Cryst. Growth Des.* **2004**, *4*, 671–685.
- (25) Anderson, V. J.; Lekkerkerker, H. N. Insights into Phase Transition Kinetics from Colloid Science. *Nature* **2002**, *416*, 811–815.
- (26) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
- (27) Hassan, S. A. Microscopic Mechanism of Nanocrystal Formation from Solution by Cluster Aggregation and Coalescence. *J. Chem. Phys.* **2011**, *134*, No. 114508.
- (28) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (29) Benavides, A.; Aragonés, J.; Vega, C. Consensus on the Solubility of NaCl in Water from Computer Simulations Using the Chemical Potential Route. *J. Chem. Phys.* **2016**, *144*, No. 124504.
- (30) Mester, Z.; Panagiotopoulos, A. Z. Temperature-dependent Solubilities and Mean Ionic Activity Coefficients of Alkali Halides in Water from Molecular Dynamics Simulations. *J. Chem. Phys.* **2015**, *143*, No. 044505.
- (31) Moučka, F.; Nezbeda, I.; Smith, W. R. Molecular Force Fields for Aqueous Electrolytes: SPC/E-compatible Charged LJ Sphere Models and their Limitations. *J. Chem. Phys.* **2013**, *138*, No. 154102.
- (32) Manzanilla-Granados, H. M.; Saint-Martin, H.; Fuentes-Azcatl, R.; Alejandre, J. Direct Coexistence Methods to Determine the Solubility of Salts in Water from Numerical Simulations. Test case NaCl. *J. Phys. Chem. B* **2015**, *119*, 8389–8396.
- (33) Kobayashi, K.; Liang, Y.; Sakka, T.; Matsuoka, T. Molecular Dynamics Study of Salt-solution Interface: Solubility and Surface Charge of Salt in Water. *J. Chem. Phys.* **2014**, *140*, No. 144705.
- (34) Aragonés, J.; Sanz, E.; Vega, C. Solubility of NaCl in Water by Molecular Simulation Revisited. *J. Chem. Phys.* **2012**, *136*, No. 244508.
- (35) *The Merck Index*, 14th ed.; Merck Inc.: Whitehouse Station, NJ, 2006.
- (36) Hess, B.; Kutzner, C.; Van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (37) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (38) Berendsen, H. J.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (39) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N Log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (40) Hassan, S. A. Morphology of Ion Clusters in Aqueous Electrolytes. *Phys. Rev. E* **2008**, *77*, No. 031501.
- (41) Lanaro, G.; Patey, G. N. Molecular Dynamics Simulation of NaCl Dissolution. *J. Phys. Chem. B* **2015**, *119*, 4275–4283.
- (42) Espinosa, J. R.; Vega, C.; Valeriani, C.; Sanz, E. The Crystal-Fluid Interfacial Free Energy and Nucleation Rate of NaCl From Different Simulation Methods. *J. Chem. Phys.* **2015**, *142*, No. 194709.

- (43) Lechner, W.; Dellago, C. Accurate Determination of Crystal Structures Based on Averaged Local Bond Order Parameters. *J. Chem. Phys.* **2008**, *129*, No. 114707.
- (44) Radhakrishnan, R.; Trout, B. L. Nucleation of Hexagonal ice (I_h) in Liquid Water. *J. Am. Chem. Soc.* **2003**, *125*, 7743–7747.
- (45) Reinhardt, A.; Doye, J. P.; Noya, E. G.; Vega, C. Local Order Parameters for use in Driving Homogeneous ice Nucleation with all-atom Models of Water. *J. Chem. Phys.* **2012**, *137*, No. 194504.
- (46) Li, T.; Donadio, D.; Russo, G.; Galli, G. Homogeneous Ice Nucleation from Supercooled Water. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19807–19813.
- (47) Steinhardt, P. J.; Nelson, D. R.; Ronchetti, M. Bond-Orientational Order in Liquids and Glasses. *Phys. Rev. B* **1983**, *28*, 784.
- (48) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. *A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, OR, Aug 24, 1996; AAAI Press: Palo Alto, CA, 1996.
- (49) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (50) Jaccard, P. The Distribution of the Flora in the Alpine Zone. *New Phytol.* **1912**, *11*, 37–50.
- (51) Kaplan, E. L.; Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481.