Title
Supersaturated calcium carbonate solutions are classical

Journal
Science Advances, 4(1)

ISSN
2375-2548

Authors
Henzler, Katja
Fetisov, Evgenii O
Galib, Mirza
et al.

Publication Date
2018-01-05

DOI
10.1126/sciadv.aao6283

Peer reviewed

CHEMICAL PHYSICS

Supersaturated calcium carbonate solutions are classical

Katja Henzler,[1]* Evgenii O. Fetisov,[2]* Mirza Galib,[3]* Marcel D. Baer,[3]* Benjamin A. Legg,[3,4]* Camelia Borca,[1] Jacinta M. Xto,[1,5] Sonia Pin,[1] John L. Fulton,[3] Gregory K. Schenter,[3] Niranjan Govind,[6] J. Ilja Siepmann,[2,7] Christopher J. Mundy,[3,8]† Thomas Huthwelker,[1]† James J. De Yoreo[3,4]†

Mechanisms of $CaCO_3$ nucleation from solutions that depend on multistage pathways and the existence of species far more complex than simple ions or ion pairs have recently been proposed. Herein, we provide a tightly coupled theoretical and experimental study on the pathways that precede the initial stages of $CaCO_3$ nucleation. Starting from molecular simulations, we succeed in correctly predicting bulk thermodynamic quantities and experimental data, including equilibrium constants, titration curves, and detailed x-ray absorption spectra taken from the supersaturated $CaCO_3$ solutions. The picture that emerges is in complete agreement with classical views of cluster populations in which ions and ion pairs dominate, with the concomitant free energy landscapes following classical nucleation theory.

INTRODUCTION

For more than a century, nucleation from electrolyte solutions has been viewed through the lens of classical nucleation theory (CNT), which is based on two key principles (*1*). First, there is a free energy barrier to nucleation due to the surface tension associated with creation of the solid-liquid interface; this leads to a critical size that clusters must exceed before the chemical potential driving force allows them to grow spontaneously. Second, clusters are able to overcome this barrier due to the inherent thermal fluctuations of the system, which allow for the addition of individual ions or molecules despite the uphill trend in free energy. These two principles exhibit a self-consistency because work must be expended to increase the size of a subcritical cluster, and the number density of clusters falls off exponentially with size. Consequently, small species (that is, ions and ion pairs) dominate cluster populations (Fig. 1). This picture of nucleation has been recently called into question based on experimental observations and classical molecular mechanics (MM) simulations of cluster formation in supersaturated solutions of $CaCO_3$, which cast doubt on one principle or on both principles (*2*–*7*).

These studies suggest that these solutions contain populations of oligomeric species referred to as prenucleation clusters in numbers far in excess of what would be expected from classical solution models. Titration experiments demonstrate that, depending on pH, approximately 30 to 75% of Ca^{2+} ions in $CaCO_3$ solutions exist in complexes (*2*). The models of nucleation based on prenucleation clusters hold that the majority of this high degree of complexation reflects the dominance of large, oligomeric clusters (*2*–*4*). In contrast, classical descriptions predict that species beyond ions and ion pairs should exist at concentrations that are lower by orders of magnitude (*1*) and thus attribute the bulk of this complexation to the formation of ion pairs, such as $CaCO_3$ (aq) or $CaHCO_3^+$ (aq), at concentrations that can be predicted on the basis of equilibrium constants for ion pairing (*8*). Moreover, nucleation of the solid in prenucleation cluster models is proposed to proceed by aggregation of these oligomeric clusters with only a small free energy barrier, perhaps via the formation of a dense but disordered liquid precursor; free ions and ion pairs play little or no role in nucleation (Fig. 1).

The essential discrepancy between CNT and prenucleation cluster models lies, then, in the free energy of cluster formation, which is directly reflected in cluster populations. However, although cryogenic transmission electron microscopy (cryo-TEM) data appeared to initially provide evidence for the existence of oligomeric clusters in $CaCO_3$ solutions (*7*), later experiments and reevaluation of the previous data led to the conclusion that they could not be detected by this method (*9*). Thus, no oligomeric clusters of any size in saturated $CaCO_3$ solutions have been seen directly by microscopy techniques, and to date, they have evaded detection by scattering methods. Their existence was inferred from analytical ultracentrifugation (AUC) data (*2*, *3*, *6*), but no direct connection between the oligomeric species detected in AUC and the degree of complexation evident in titration experiments has been established.

Molecular simulation of solution structure is equally challenging because of the difficulties of developing pairwise additive classical MM potentials that correctly describe ion-pairing interactions and simulating sufficient periods of time to adequately sample cluster configurations. Thus, the use of standard molecular dynamics simulations of slightly supersaturated solutions to achieve equilibrium and correctly predict cluster stability, size distribution, and free energy of formation remains a formidable challenge (*10*) and can lead to contrasting results. Although initial applications of MM to the $CaCO_3$ system supported the conclusion that oligomeric species far larger than simple ions or ion pairs were abundant and stable relative to the latter (*4*), more recent work using the same classical potentials extending the simulations to longer times in dilute solutions concluded that the clusters decay into ions and ion pairs (*9*). A complete $CaCO_3$ solution model that correctly describes ion-pairing interactions, gives the distribution of species as a function of solution concentration, and correctly predicts solution equilibrium

[1]Swiss Light Source, Paul Scherrer Institut, 5232 Villigen, Switzerland. [2]Department of Chemistry and Chemical Theory Center, University of Minnesota, 207 Pleasant Street SE, Minneapolis, MN 55455, USA. [3]Physical Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA. [4]Department of Materials Science and Engineering, University of Washington, Seattle, WA 98195, USA. [5]Department of Chemistry, ETH Zürich, 8093 Zürich, Switzerland. [6]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99354, USA. [7]Department of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Avenue SE, Minneapolis, MN 55455, USA. [8]Department of Chemical Engineering, University of Washington, Seattle, WA 98195, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: chris.mundy@pnnl.gov (C.J.M.); thomas.huthwelker@psi.ch (T.H.); james.deyoreo@pnnl.gov (J.J.D.Y.).

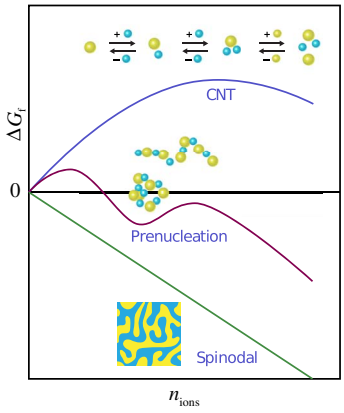

Fig. 1. A schematic representation of the different pathways for the formation of solid polymorphs of $CaCO_3$. In CNT, nucleation proceeds uphill in free energy through the addition of Ca^{2+} (cyan), CO_3^{2-} (yellow), or ion pairs where large clusters are considered rare (but necessary) and the metastable solution phase is dominated by monomers and ion pairs. Prenucleation is characterized by a metastable solution phase dominated by populations of larger clusters that undergo further association to form the solid polymorph $CaCO_3$ (*2*). Spinodal decomposition occurs spontaneously, creating the solid polymorph and ions at a high degree of supersaturation.

constants is lacking, as is a direct connection between the predicted species and those seen experimentally. Here, we combine sophisticated experiments, simulations using advanced sampling techniques and complex interaction potentials, and electronic structure theory to overcome many of these limitations and thereby achieve a detailed molecular picture of ion pairing, cluster size distribution, and the initial stages of nucleation in supersaturated $CaCO_3$ solutions. This methodology can be used to study other chemical systems as well.

RESULTS

Constructing a theoretical solution model

To create a theoretical solution model, we first construct potentials of mean force (PMFs) between Ca^{2+} and either CO_3^{2-} or HCO_3^- using first-principles molecular dynamics simulations with the interactions described by Kohn-Sham density functional theory (DFT) (*11*, *12*). The advantage of DFT over more common MM approaches is that DFT provides an excellent description of aqueous response to both small and large interfaces (*13*) and better representation of the local complex structure around ions, such as Ca^{2+}, as differentiated by x-ray absorption fine structure measurements (*14*). Because precipitation in the $CaCO_3$ system occurs already at concentrations $<10^{-3}$ M, accurate information on the solvent-mediated ion-ion interactions can be gleaned from simulations for systems containing a single ion pair and 96 water molecules (see the Supplementary Materials). The comparison shown in Fig. 2B between the PMFs of $CaCO_3$ ion pairing generated from DFT and MM representations (*15*, *16*) illustrates the following important differences: (i) DFT yields a 50% smaller free energy gain for ion pairing (at 3 to 4 Å); (ii) DFT slightly favors bidentate versus monodentate ion pairs, whereas the opposite holds for the MM description; and (iii) DFT yields a significantly larger barrier separating contact ion pairs from solvent shared ion pairs (Fig. 2B). The key consequence of these differences is a significantly reduced propensity to form ion pairs in the DFT model as compared to the MM model.

Knowledge of the DFT-based PMFs allows us to construct a reduced molecular representation consisting of Ca^{2+} and CO_3^{2-} ions in a solvent treated in an implicit manner (Fig. 2A). A key assumption is that the PMFs that describe the short-range interaction are independent of concentration. For both the MM and the DFT methods, PMFs are augmented with continuum electrostatics (CE) represented by the Coulomb potential modified by the experimentally determined dielectric constant for water to describe the long-range interactions (Fig. 2B). Thus, we refer to the underlying potential for the reduced model for the MM as MM/CE. Because of the relatively small systems used to perform the DFT calculations, additional assumptions need to be made to construct the reduced interaction potential. Specifically, the DFT PMF requires an additional matching to the MM/CE in the solvent-shared region. This is based on the observation that the coordination number of the Ca^{2+} ion with respect to water is seven at Ca^{2+}–CO_3^{2-} distances >4 Å reflecting the isolated cation (*17*). The coordination number at this separation is also consistent with a recent two-dimensional PMF using the MM interaction potential (*18*). Thus, we refer to the reduced potential based on the DFT short-range PMF as DFT + MM/CE, which is shown in Fig. 2B. It is of interest to note that although we performed a one-dimensional PMF, we have access to Ca^{2+} coordination numbers with respect to water in our biased ensemble. For the DFT, the coordination number of Ca^{2+} with respect to water is roughly 4.5 at Ca^{2+}–CO_3^{2-} separations between 2.5 and 3 Å. For separations between ~3 and ~3.8 Å, the coordination number rises steeply from 4.5 to 5.5. For separations between ~3.5 and ~4 Å, the coordination number is 6. Beyond the distance of ~4 Å, the coordination number is 7. Last, the like-charge repulsive interactions for both the DFT + MM/CM and the MM/CM models were taken from the PMFs using the MM model.

We use this reduced model to predict cluster size and shape distributions and the equilibrium constants between the interconversion of species as a function of concentration. Recent work suggests that cluster configurational entropy can significantly influence the size and shape distributions of small nucleation clusters (*19*), making efficient statistical sampling of the molecular configurations for concentrated $CaCO_3$ a necessity. In keeping with the assumption of a dilute limit, we neglect many-body correlations between the ions but retain both two-body correlation effects and the important water-ion correlations that are encoded in the PMFs (Fig. 2B), as done previously in predicting the collective phenomena of electrolytes at low concentrations (*17*, *20*). Thus, our reduced model comprises two different attractive point particles to represent Ca^{2+} and CO_3^{2-} (Fig. 2A) that interact according to the corresponding PMFs (Fig. 2B), where the corresponding terms for Ca^{2+}–Ca^{2+} and CO_3^{2-}–CO_3^{2-} are strictly repulsive (see the Supplementary Materials for details).

This reduced model in conjunction with the aggregation-volume-bias Monte Carlo (AVBMC) method (see the Supplementary Materials for details) (*21*) allows for the fully converged statistical sampling required to obtain the cluster concentrations, n_{ij}

$$n_{ij} = exp(-\beta G_{ij})(n_{10})^i(n_{01})^j \quad (1)$$

The subscript *ij* denotes stoichiometric Ca^{2+} and CO_3^{2-}, respectively, $\beta = 1/k_BT$ (where k_BT is Boltzmann's constant multiplied by absolute temperature), and G_{ij} is the free energy of formation of the *ij* cluster relative to the corresponding nonassociated cations and anions. The resulting cluster populations for the DFT + MM/CE description are dominated by free ions (n_{10} and n_{01}) and ion pairs (n_{11}), and the concentrations of all larger species are orders of magnitude smaller (Fig. 2C, bottom), for example, by >100 and >1000 for n_{21} and n_{22}, respectively.

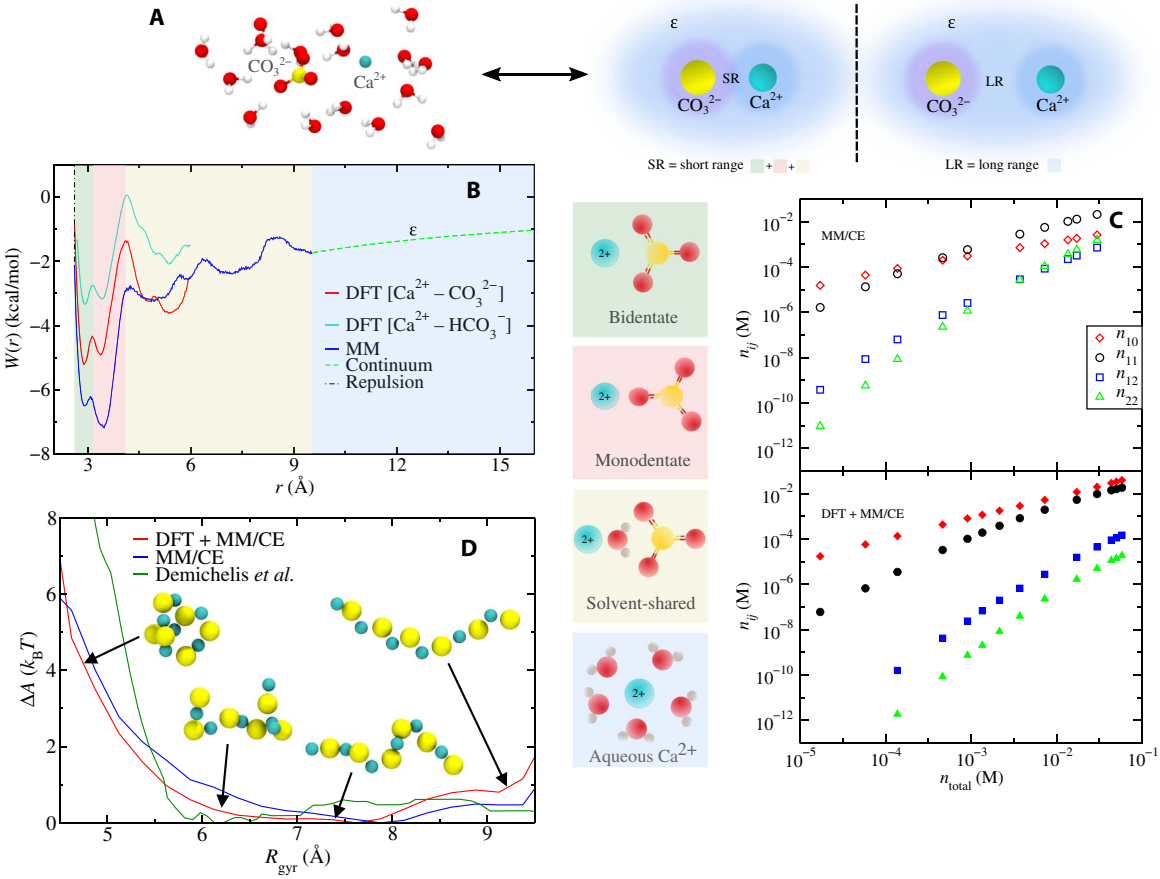

Fig. 2. Theoretical basis of the molecular base $CaCO_3$ solution model. (**A**) Schematic of the mapping of the explicit molecular model to the reduced molecular model. The short-range interaction of the reduced model is computed through a PMF using a molecular model, and the long-range interaction is determined by a CE description. (**B**) The computed PMFs. The color-coded maps correspond to the short-range structural moieties that are depicted in the schematics. (**C**) Equilibrium populations n_{ij} of small $CaCO_3$ clusters versus the total concentration of $CaCO_3$ as determined by the reduced model [shown in (A), right]. SEM is smaller than a symbol size. (**D**) Free energy of $(CaCO_3)_6$ clusters versus the radius of gyration (with some representative structures shown) comparing the present work to the explicit molecular model of Demichelis *et al.* (*4*).

Further connection of our MM/CE version of our solution model to previous MM simulations that used explicit water is achieved by directly comparing the predictions of absolute G_{ij} to those reported by Wallace *et al.* (*5*) (see fig. S3A) and to shape distributions for $(CaCO_3)_6$ clusters to that of Demichelis *et al.* (*4*) (Fig. 2D). These comparisons yield good agreement, demonstrating the efficacy and dominance of the PMF in describing ion pairing under dilute solution conditions. An important prediction of our simulations is that, despite the agreement in cluster shape distribution obtained here with MM/CE, DFT + MM/CE, and past simulation studies (*4*), neither our MM/CE nor our DFT + MM/CE models find larger species, such as $(CaCO_3)_6$, to be present in solution at any relevant concentrations. From the fact that we are able to converge the sampling of the cluster populations at the correct experimental concentrations, we conclude that these larger species are insignificant in slightly and moderately supersaturated solutions.

From PMFs to bulk thermodynamic properties

Taking the species populations from the solution model, we calculate the equilibrium constants $K_{ij} = n_{ij}/(n_{10}^i n_{01}^j)$ for interconversion between species (Fig. 3B). We use these DFT-based constants to model the free Ca^{2+} concentration in bulk solutions during titration of Ca^{2+} into either carbonate or bicarbonate buffers and obtain good agreement between the predictions and previously reported experimental titration data (Fig. 3A and table S2) (*2*). Although this calculation ignores the secondary effects of the background electrolyte, Na^+ and Cl^-, inclusion of these species in classical solution models demonstrates that these species have negligible contribution to the total speciation (table S3). Moreover, classical solution models based on accepted equilibrium constants (*8*) predict amounts of ion pairs (tables S2 and S3) and titration curves (fig. S7A) similar to those obtained in the DFT + MM/CE model. In contrast, because of the stronger ion pairing seen in the MM model that leads to 10- to 100-fold increases in predicted ion complexation (Figs. 2C and 3B), the degree of ion complexation occurring during Ca^{2+} titration is strongly overestimated (Fig. 3A). Although both the DFT + MM/CE and MM/CE models produce the same qualitative conclusions regarding speciation, the bulk titration data support the picture put forth by the DFT-based PMFs where free Ca^{2+} followed by ion pairs, with almost negligible contribution from larger clusters, dominate at experimentally relevant concentrations.

Using umbrella sampling (*22*) in conjunction with AVBMC, we also calculated the free energy of cluster formation ΔG_f as a function of cluster size (Fig. 3C). In keeping with the predicted dominance of ions and ion pairs, we find that ΔG_f exhibits the size dependence expected

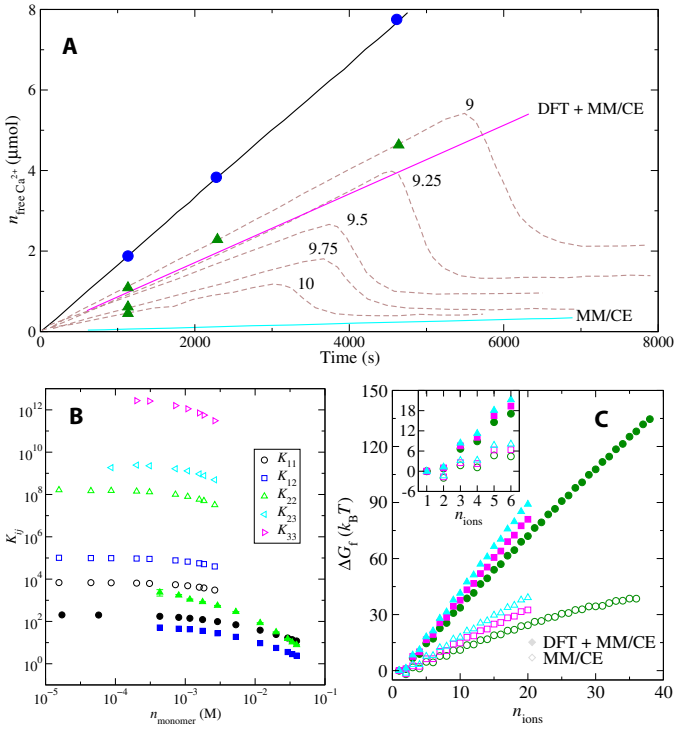

Fig. 3. Macroscopic outcomes of the solution model. (**A**) Computed (magenta and cyan) and experimental (*2*) titration curves of $CaCl_2$ into a carbonate buffer solution at different pH values: The black solid line shows the total amount of Ca^{2+} titrated into solution, and brown dashed lines show the amount of unbound Ca^{2+}. The blue circles correspond to the total amount of calcium added to solutions and the green triangles refer to the amount of free calcium in the solution under conditions where the XANES experiments are conducted (see the Supplementary Materials, including tables S2 and S3, for details). (**B**) Computed equilibrium constants for the formation of small $CaCO_3$ clusters versus equilibrium concentration of monomers. Filled symbols denote DFT + MM/CE results, and open symbols indicate MM/CE results. SEM is smaller than a symbol size if not shown. (**C**) Free energy of formation of $CaCO_3$ clusters versus cluster size. Dark green, magenta, and cyan symbols correspond to total $CaCO_3$ concentrations of 17.22, 7.264, and 3.719 mM, respectively. These results represent the free energy landscape for nucleation and are distinct from those of Wallace *et al.* (*5*), which only consider the free energy of individual clusters. SEM is smaller than a symbol size.

from CNT using either the DFT + MM/CE or MM/CE solution models, but the greater tendency toward ion pairing in the MM/CE model leads to a weaker ΔG_f dependence and thus smaller nucleation barriers. Using the ΔG_f size dependence of the DFT + MM/CE model, we deduce that the interfacial free energy for the formation of amorphous $CaCO_3$ (ACC) is between 14 and 25 mJ/m^2, depending on the number density (*23*). This value implies that at a supersaturation of between 1.5 and 2.0, the barrier to ACC formation falls below k_BT and the solution spontaneously phase-separates. Assuming solubility on the order of 10 to 15 mM for ACC, this estimate compares favorably to the recent experimental determinations of the point at which $CaCO_3$ solutions undergo spinodal decomposition (*24*).

In situ molecular-level measurements of solution speciation

We experimentally determined the Ca speciation in $CaCO_3$ solutions in situ by using synchrotron-based Ca K-edge x-ray absorption near-edge structure (XANES) spectroscopy (*25*, *26*) to directly probe the local chemical environment of aqueous Ca^{2+} ions. Spectra are taken using a liquid microjet (see the Supplementary Materials) (*27*, *28*) at specific Ca^{2+} concentrations and pH values lying along published titration curves (Fig. 3A, blue circles and green triangles) (*2*). The use of state-of-the-art undulator beamlines (*29*) enabled measurements on dilute supersaturated solutions extending down to submillimolar Ca^{2+} concentrations that are close to the solubility limit of the most stable phase (calcite) (see the Supplementary Materials for details). XANES spectra for various solid carbonates are also acquired for comparison.

The obtained XANES spectra of both the Ca^{2+} (for example, $CaCl_2$) and $CaCO_3$ solutions (Fig. 4A) are distinct from one another and from those of solid ACC, aragonite, and calcite in the regions of the pre-edge, near-edge, and white line (fig. S6). The normalized difference spectra (Fig. 4B), that is, the difference between the $CaCO_3$ solution spectra and the spectrum of fully hydrated Ca^{2+} (in a $CaCl_2$ solution), highlight the spectral changes that are due to the presence of associating ions whose structure deviates from that of normal sevenfold coordinated aqueous Ca^{2+} ions (*30*). We observe that, for constant solution pH, the difference spectra show little change with Ca^{2+} concentration (fig. S2A), whereas with increasing pH (9 to 10), the magnitude increases (fig. S2B). We also consider the possibility of compact nanoclusters consisting of >10 cation/anion assemblies that are structural analogs of ACC, calcite, or aragonite that will have XANES spectra that are similar to those of the bulk solid (fig. S6) (*31*). A linear combination of any of these solid phases with a representative fraction of Ca^{2+} (see the Supplementary Materials) yields a unique concentration-dependent XANES spectrum that does not reproduce the experimental $CaCO_3$ solution spectra. Thus, we can exclude the presence of nanoscale solid $CaCO_3$ phases (*10*) even at supersaturations substantially above the phase boundary seen in the titration data (Fig. 3A; see Comparison of the theoretical spectra with experimental references in the Supplementary Materials).

To identify the ion-pair species that are responsible for the observed difference spectra, we directly compute the time-dependent DFT (TDDFT) XANES spectra (*32*–*34*) using the species and distributions harvested from the DFT-based PMFs and the AVBMC populations in Fig. 2 (B and C, respectively). When the TDDFT method was applied to the solid-phase standards, the calculated and experimental spectra were in reasonable agreement (fig. S6). For the aqueous phases, the DFT methods were used to calculate the spectra for mixtures of monodentate and bidentate contact ion pairs, solvent-shared ion pairs, and free Ca^{2+} in the two limiting conditions of pure carbonate (pH 9.75) and pure bicarbonate (pH 7 in Fig. 4 and fig. S6). These two theory conditions bracket the experimental pH values, providing an approximate amplitude estimate for the experimental spectra acquired at pH 9 and 9.75. Intermediate pH values would be represented as a linear combination of these carbonate, bicarbonate, and free Ca^{2+} theory spectra. The perturbation of the aqueous Ca^{2+} structure due to the presence of the anion in the first or second solvation shell alters the XANES spectra in different ways (fig. S4B). For the experimental condition of pH 9.75 in Fig. 4 (C and D), the theoretical distribution of Ca species (Ca^{2+} with CO_3^{2-}) is in agreement with experimental models based on available thermodynamic data (table S2). The predicted difference spectrum (Fig. 4D) shows exceptional agreement with the experimental spectrum in both the position of the features and their amplitudes that lie just outside of the experimental uncertainties. Moreover, an additional observation verifying internal consistency is that the experimental data, the classical solution thermodynamics, which is based on formation energies (*8*) as implemented in MINTEQ (*35*) and

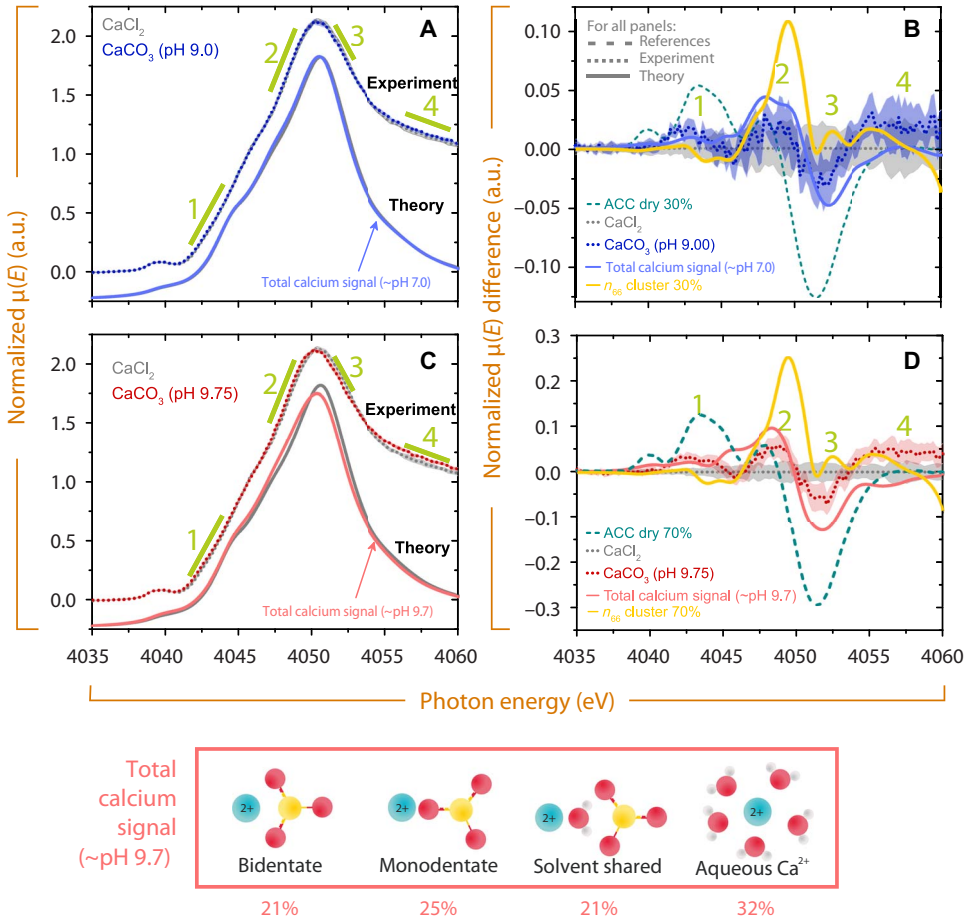

Fig. 4. Comparison of experimental and theoretical XANES spectra of carbonate solutions at pH 9 and 9.75. (**A**) Experimental XANES Ca K-edge spectra [in arbitrary units (a.u.)] for a pure $CaCl_2$ solution (gray dots) and for a freshly mixed 1:1 solution of $CaCl_2$ (150 μM; pH 9) and carbonate buffer (20 mM; pH 9) (blue dots), compared to the simulated spectra for a $CaCl_2$ (as Ca^{2+}) solution (gray line) and for a mixture of $CaHCO_3^+$ ion pairs with speciation as calculated from DFT theory (blue line). The regions marked as 1, 2, 3, and 4 indicate the regions where the spectrum of a Ca^{2+} solution differs from a solution with ion pairs. The speciation of the ions pairs for pH 7 is summarized in tables S2 and S3. (**B**) Difference spectrum between a solution with ion pairs and the $CaCl_2$ standard solution for the cases shown in (A). Gray shaded region centered about the *x* axis represents the experimental error bars for the measurement of the pure $CaCl_2$ solution. The blue dotted line with a light blue error region represents the difference spectrum between the $CaCO_3$ and $CaCl_2$ solutions of the experimental data shown in (A). The solid blue line represents the difference spectrum for the theoretical data shown in (A). For comparison, the difference spectra of the experimental ACC (dry) and the theoretical $(CaCO_3)_6$ clusters are shown. For these two standards, a mixing ratio was applied to match the amount of bound species measured by Gebauer *et al.* (*2*), in which 30% is the amount of bound species in a supersaturated $CaCO_3$ solution at pH 9. (**C** and **D**) Same set of analysis for a solution at pH 9.75 (below the four-panel figure), where the speciation percentages are determined from the R^2-weighted integral of the DFT-MM/CE potential in Fig. 2B. In this case, the amount of bound species is about 70%.

GEMS (Gibbs Energy Minimization Selektor) (*36*), and the TDDFT correctly predict the increasing fraction of ion pairs with rising pH that is observed in titration experiments by Gebauer *et al.* (*2*).

Although our MM- and DFT-based solution models predict only minute concentrations of multi-ion clusters [for example, $(CaCO_3)_6$], it is instructive to ascertain whether their XANES features can be used to rule them out. We compute TDDFT XANES spectra for a representative mixture of the four $(CaCO_3)_6$ clusters ranging from globular to linear (Fig. 2D) that were predicted by the MM model at a concentration needed to account for the bound species measured experimentally (*2*). The resulting spectrum exhibits a line shape that is not consistent with the experimental solution spectra (fig. S6C). Thus, the ability to quantitatively calculate the XANES spectrum of candidate Ca^{2+} cluster species provides the opportunity to test their existence as components in experimental spectra, and the results obtained here imply that they are not present in significant numbers.

DISCUSSION

The agreement between the experimental and calculated XANES spectra demonstrates the exquisite sensitivity of the local electronic response to changes in the structural motifs in the first and second solvation shells of Ca^{2+} ions. The agreement between the XANES difference spectra under varying pH conditions strongly supports the conclusion that the solution structure is dominated by ion pairing, even near the solution–solid-phase boundary. The observed trends in both the experimental and theoretical XANES spectra are also consistent with speciation calculations based on the previously determined and widely used experimental equilibrium constants (see the Supplementary Materials) (*8*). These equilibrium constants corroborate the dominance of ions and ion pairs and show that the overall fraction of "bound" n_{11} $CaCO_3$ clusters increases with pH at constant concentrations, whereas the ratio of free $CaCO_3$/bound (Ca) remains constant during titration at constant pH. Both observations

are consistent with the titration experiments (Fig. 3B; see also the Supplementary Materials).

Although the calculated free energy of cluster formation (Fig. 3C) predicts that large clusters are necessary for nucleation, the speciation analysis, both theoretical and experimental, implies that these clusters are exceedingly rare except at supersaturations for which the free energy barrier approaches k_BT. Thus, our findings rule out the presence of large prenucleation clusters as a dominant species, producing a clear picture of $CaCO_3$ nucleation proceeding by monomer addition from a solution rich in isolated ions and their pairs, although our findings are agnostic with regard to whether the first condensed phase is solid ACC or a dense liquid state forming in a binodal region as concluded in previous computational (*5*) and experimental (*9*) studies. This picture is further corroborated by XANES experiments demonstrating that the speciation is unchanged as a function of concentration at a fixed pH (fig. S2A) and ion pairing becoming less prominent at lower pH (fig. S2B), both in accordance with a linear titration curve (Fig. 3A), where the ratio of bound/unbound Ca depends solely on pH. Moreover, it is in agreement with the conclusions of Smeets *et al.* (*9*), in which cryo-TEM was used to rule out the existence of clusters larger than 9 Å in diameter and MM simulations were extended to adequate time scales to predict that clusters placed in dilute solutions decay into ions and ion pairs and solutions containing free ions generated few additional species beyond ion pairs (*9*).

It is worthwhile to briefly mention two equally important criteria that must be fulfilled to meet the most stringent requirements of a particular speciation model. First, the experimental titration curve is linear, reflecting equilibrium with simple ion pairs such as $CaCO_3$. In contrast, a model invoking equilibrium between free ions with primarily a dominance of larger, oligomeric clusters would be strongly nonlinear in accordance with Eq. 1 (fig. S7). Second, an examination of the titration data by Gebauer *et al.* (*2*) reveals that, despite significant changes in total calcium concentration, the total amount of bound calcium at the moment when the nucleation starts is a constant of ~3.9 μM for all pH values in line with observations of Kellermeier *et al.* (*37*), adding additional credence to the validity of the solution model derived herein that ion pairs are the species responsible for cluster formation and eventual precipitation. Hence, it is their concentration that determines the onset of nucleation. Our conclusion that ion pairs are the fundamental unit leading to nucleation of the first $CaCO_3$ condensed phase parallels recent findings that this same species is the fundamental unit leading to growth of calcite (*38*–*40*).

There is close agreement between the predicted and published equilibrium constants, titration curves, XANES spectra, and pH dependence of bound Ca^{2+}. There is agreement with the classical dependence of the free energy of cluster formation on size and between the predicted and measured concentrations at which spontaneous phase separation occurs. In total, $CaCO_3$ solutions contain negligible concentrations of species beyond simple ions and ion pairs and form clusters at a rate, in relative numbers, and with energies of formation that are expected from CNT. Hence, all these factors point to a picture of supersaturated $CaCO_3$ solutions defined by classical views of solution speciation, cluster populations, and the initial stages of nucleation.

MATERIALS AND METHODS

Chemicals

Analytical grade calcium chloride dihydrate, sodium carbonate, sodium bicarbonate, $CaCO_3$, and dimethyl carbonate were purchased from Sigma-Aldrich and used without any further purification. Sodium hydroxide solutions (0.1 and 1 M) and urea were purchased from Merck KGaA. Deionized water (electrical resistance, >18.2 megohms) was obtained from a Milli-Q water purification system (Millipore).

Synthesis of references

Aragonite

The synthesis protocol described by Wang *et al.* (*41*) was used for the preparation of aragonite reference. Briefly, 20 ml of deionized water, 0.735 g of $CaCl_2{\cdot}2H_2O$, and 0.9006 g of urea were sealed in a 50-ml Pyrex bottle. The mixture was held at 90°C in a convection oven for 4 hours. The precipitate was filtered using a glass fiber membrane (Whatman, 1820-025) and then washed several times with deionized water and, afterward, with pure ethanol. The precipitate was then dried at 90°C in a convection oven for 16 hours.

Dry ACC

The dry ACC reference material was prepared using the procedure described by Faatz *et al.* (*42*). A mixture of 0.001 mol of $CaCl_2{\cdot}2H_2O$, 0.0049 mol of dimethyl carbonate, and 80 ml of deionized water was combined in a 250-ml polypropylene beaker and stirred at a rate of 675 rpm (Teflon-coated magnetic stir bar). Twenty milliliters of 0.5 M sodium hydroxide solution was then added, leading to the hydrolysis of the dimethyl carbonate, followed by the precipitation of ACC. After 2.5 min, the products were divided into two portions, transferred to 50-ml centrifugal tubes, and centrifuged for 7 min (T = 20°C) at an acceleration of 200*g* (Eppendorf centrifuge 5804R). After removal of the supernatant with a pipette, the precipitate was dispersed in 30 ml of ethanol using an ultrasonic sound bath and a VWR vortex mixer. This was followed by another centrifugation round of 5 min at 200*g* and at a temperature of 20°C, followed by removal of the supernatant with a pipette. Subsequently, two additional washing steps with acetone followed by centrifugation for 2 min at 80*g* were performed. After the last washing, the wet precipitate was transferred to a 50-ml round-bottom flask, and the residual solvent was removed by freeze-drying. The material was stored in a desiccator under vacuum until measurement.

In situ *ACC*

The in situ ACC reference was freshly prepared in solution and measured in an in situ XANES apparatus. Briefly, pure CO_2 was bubbled (440 ml/min) through 100 ml of freshly prepared 15 mM Na_2CO_3 solution while stirring with a magnetic stirrer at 300 rpm to set the pH at 6.3. One hundred milliliters of 10 mM $CaCl_2$ solution (pH ~5.5) was then added, yielding an equilibration pH of 6 of the resulting solution mixture. The pH of the solution was then slowly increased by changing the flow rate ratio of a gas mixture of pure CO_2 and 99% N_2/1% CO_2 at a total flow rate of 20 ml/min through the solution. The increase in pH reduced the $CaCO_3$ solubility in the supersaturated solution and eventually triggered precipitation, as marked by a spontaneous pH decrease. Continuously, a small amount of the reacting mixture was sucked out (0.05 ml/min) of the reaction vessel with a syringe pump and fed into an in situ flow-through cell [tube inner diameter (ID), 1 mm]. On the basis of a particle-size versus sedimentation-velocity estimation, only agglomerates smaller than 1 μm will reach the cell, where XANES spectra were taken. Consequently, a mixture of nonsedimented ACC nanoparticles in solution, ion pairs, and free Ca^{2+} ions arrives at the probing position of the x-ray beam. Although the relative composition of these components in the reacting solution is not known at the moment, it is clear that any spectra taken from this solution after the spontaneous decrease in pH represent spectra from a nucleated solution, in contrast to spectra

taken from the liquid microjet, where supersaturated solutions are considered (see below). Here, we adopt the name "in situ ACC" for spectra taken from this solution, because ACC is the well-accepted first building block for carbonate formation. We note that any attempts to analyze the XANES spectra by a linear combination fitting of the aqueous Ca^{2+} and dry ACC failed to reproduce the measured spectrum, indicating that the in situ ACC spectra resemble the signature of the first nucleated species in a solution with a different structural configuration from the dry ACC.

XANES spectroscopy

XANES spectroscopy measurements are presented for both solid calcium carbonate samples and aqueous supersaturated calcium carbonate solutions. In XANES spectroscopy, x-ray photons excite Ca core 1s electrons to valence states (*25*, *26*), providing a spectral signature that reflects the structure of the first, second, and, to a lesser extent, third solvation shell around the Ca^{2+} ion.

PHOENIX beamline

All experiments were conducted at the PHOENIX I (Photons for the Exploration of Nature by Imaging and XAFS) beamline at the Swiss Light Source (Paul Scherrer Institut, Villigen, Switzerland), which provides a high flux of photons in the tender x-ray region (0.8 to 8 keV). The source of the beamline was an elliptical undulator (APPLE II). The optical design followed that of the LUCIA (line for ultimate characterization by imaging and absorption) beamline (*29*), which is now located at the synchrotron Soleil. Monochromatic light was generated by a double-crystal monochromator using a pair of silicon (111) crystals. The x-ray beam can be focused down to ~3 μm × 3 μm using a pair of Kirkpatrick-Baez mirrors, providing micrometer-scale resolution for x-ray spectroscopic measurements [μ-XAS (micro–x-ray absorption spectroscopy) and μ-XRF (micro–x-ray fluorescence)]. The photon flux at 3 keV is ~1×10^{11} photons/s. The end station can accommodate measurements performed under different vacuum conditions (from 10^{-5} to 800 mbar) as required for the specific user experiments. The flux of the incoming photons, I_0, was measured by the total electron yield (TEY) from a nickel-coated polyester foil that is contained in a separate vacuum chamber having a pressure of about 5×10^{-8} mbar. This I_0 chamber was located about 1 m upstream from the experimental chamber. The end station of the beamline has a modular construction, which provides the possibility to perform in situ studies, such as a liquid microjet or other user-supplied setups.

Sample environment for solid samples

Solid powder samples were dispersed onto a conducting carbon tape that was affixed directly to an electrically isolated copper plate. XAS spectra were acquired simultaneously in both TEY and fluorescence modes (beam size, 2 mm × 2 mm). For the TEY mode, the drain current into the powder sample was measured. For the work presented here, all solid references were taken in the TEY mode to avoid self-absorption distortions.

Liquid microjet for in situ studies of carbonate solutions

Ca K-edge spectra of freshly mixed supersaturated $CaCO_3$ solutions were obtained in situ using a liquid microjet, which was mounted in the end station and connected to the liquid flow system, as depicted in Fig. 5. Separate solutions of aqueous $CaCl_2$ and aqueous carbonate buffer ($NaHCO_3/Na_2CO_3$) were continuously pumped through the liquid flow system (flow rate, 1 ml/min; pump manufacturer, Micro-

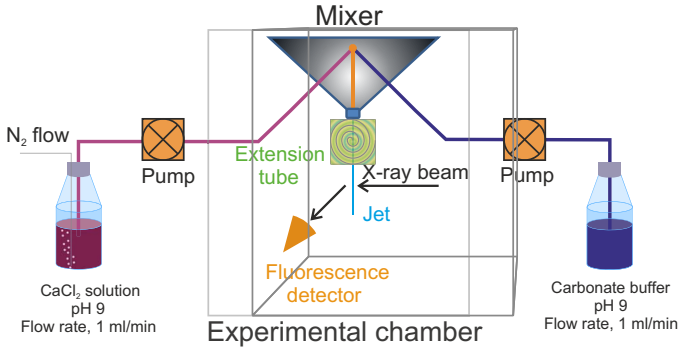

Fig. 5. Schematic of the liquid jet setup. Two solutions were mixed in a mixing tee and passed through a delay tube (residence time, ca. 43 s) and then into a perfluoroalkoxy (PFA) nozzle (100 μm ID), forming a stable liquid jet in the experimental chamber that was held at a pressure of about 20 mbar. The delay tube ensures dissipation of flow turbulence and complete mixing of the liquid.

liquids GmbH) and were then mixed using a commercially available static mixing tee (U-466, IDEX Health & Science; mixing time, 66 ms). The freshly mixed solution was then pumped through a short polyetheretherketone tube [ID, 0.35 mm; outer diameter (OD), 1.5875 mm; length, 50.93 mm] into a Teflon "delay" tube (ID, 1 mm; OD, 3 mm; length, 991 mm) and then into the jet nozzle. The jet nozzle consisted of a section of PFA tube (ID, 100 μm; OD, 1.5875 mm; length, 25 mm). The delay tube (residence time, ~43 s) served several purposes. First, it helped to ensure complete mixing from the mixing tee. Second, the flow turbulence from the mixer must be dissipated before entering the nozzle to assure jet stability and complete solution mixing. Finally, the delay tube aided in indirectly detecting nanoparticulate amorphous or crystalline agglomerates because the presence of these would likely stick to walls, serving as sites for crystal growth, and, hence, would cause a decrease or instability in the measured absolute Ca Kα fluorescence signal. Because no such fluctuations were observed under any conditions used in the presented spectra, we are confident that no nucleation occurred in the solution. The entire apparatus (mixer, extension tube, and microjet) was mounted into the PHOENIX I end station and operated in a He/H_2O atmosphere with a total pressure of about 20 mbar, thus effectively operating the jet under pressure conditions close to the water vapor pressure at 25°C. The ejected mixture from the jet was collected in a Teflon catcher, which was maintained at a temperature of approximately 15°C using copper cooling tubes that were wound around the catcher. The cooling tubes shown in fig. S1D prevented undesired condensation of water vapor on the fluorescence detector surface and, in principle, kept the microjet liquid interface region at a slight degree of supersaturation. However, because the microjet velocity was high (4.2 m/s), the diffusive transport limitations in the gas phase near the liquid jet surface established a de facto equilibrium of the liquid surface with its own water vapor. Moreover, it has been shown that for microjet systems that shoot liquids into vacuum, the liquid interface is unperturbed in terms of vapor pressure and temperature (*43*). Hence, the induction of nucleation by strong disequilibria on the jet surface can be excluded, and the XAS spectra shown here represent conditions of the bulk solution.

The jet setup offers a windowless method of studying supersaturated $CaCO_3$ solutions. Its specific advantages include the minimization of radiation damage effects by continuous sample renewal and the elimination of stationary surfaces (such as x-ray windows) where possible nucleation products could accumulate and disturb the XAS spectra acquisition.

Fluorescent x-ray photons were detected by an energy-dispersive single-element silicon drift diode (Ketek), which was mounted in 90° geometry with respect to the incoming beam. The x-ray beam was focused to about 50 μm × 20 μm (H × W). The detection limit for calcium atoms in aqueous solutions was mainly caused by the elastic scattering from water that added a background component to the Ca Kα fluorescence. To minimize this effect, the position and distance of the jet relative to the incoming beam and the detector were optimized to keep the dead time below 20% and to minimize the elastic scattering from the bulk water. This was achieved by a two-dimensional scan of the liquid jet position in the horizontal plane at a fixed photon energy of 4300 eV. This procedure takes advantage of the linear polarization of synchrotron radiation, which has no elastic scattering component in the polarization direction. Figure S1 (A and B) illustrates this procedure by showing the measured Ca Kα fluorescence intensity and the elastic scattering as function of the y position (along the beam axis) of the liquid jet. For the shown example, the final position was chosen at the minimum intensity of the elastic scattering at around $y = -9.25$ mm, marked as a dotted line.

The horizontal position of the microjet was optimized by taking line scans as shown in fig. S1C. Because of inevitable mechanical imperfections of the monochromator, the beam moved horizontally by a few tens of micrometers when scanning the full energy range from 3900 to 4300 eV. In principle, this shift can be compensated by automated corrections to the monochromator mechanics. We avoided this approach because the additional motorization can cause small mechanical instabilities, which may affect the high position accuracy needed in the local XANES region of about 10 eV around the main features after the K-edge. For scans over this small energy range, the beam movement can be neglected. To ensure that the whole scan region can be used for proper normalization, two line scans across the liquid jet were taken, one at 4050 eV (around the white line maximum) and a second one at 4300 eV in the post-edge region of the spectrum (fig. S1C). The liquid jet was positioned in such a way that the beam hit the jet at its center, that is, at the fluorescence intensity maximum over the whole energy range of the spectral scan, as indicated by the black vertical line. By using this overall procedure, the current detection limit for Ca atom solutions was about 75 μM.

Solution preparation for liquid jet experiments

Before mixing, the pH of the two reacting solutions (20 mM carbonate buffer and $CaCl_2$ solutions at a concentration of 150, 300, or 600 μM) was adjusted to ensure equivalent pH of both solutions (that is, pH 9, 9.75, or 10). The desired pH of the carbonate buffer was set by adjusting the $NaHCO_3/Na_2CO_3$ ratio, whereas the pH of the $CaCl_2$ solution was achieved by titrating with 1 M NaOH solution. This adjustment, along with the excess of carbonate in the system, helped to minimize pH changes after mixing. The pH of the product solutions typically did not differ by more than 0.1 pH units from that of the starting solutions. To remove any traces of dissolved CO_2 from the water used to prepare the $CaCl_2$ solutions, the water was degassed by bubbling N_2 gas for several hours before usage. Measurements were made on the same day in which the pH-adjusted $CaCl_2$ solution was prepared. To prevent $CaCO_3$ precipitation from absorption of atmospheric CO_2 into the pH-adjusted $CaCl_2$ stock solution, the $CaCl_2$ solution was kept under a continuous trickle of nitrogen flow during the experiment. The carbonate buffer was set up in contact with the atmosphere to minimize degassing of the CO_2 from the pH-sensitive CO_2/bicarbonate/carbonate solution equilibrium during the experiment. Because the liquid was taken from the lower part of the solution in the bottle, the composition used in the experiments should be unaffected (no stirring or shaking of the solution during measurement) from contact with air.

Analysis of the experimental XANES data

To ensure high-quality data, the stability of energy calibration, and the reproducibility of the data taken on different beamtimes, spectra of pure dilute $CaCl_2$ solutions were taken during each beamtime as a reference standard. Collection of up to 12 individual spectra for each averaged spectrum ensured a total of about $N = 100{,}000$ counts at the Ca white line (~4050 eV). The statistical error at each energy of the averaged spectra can be expressed as the SD (about 1.5% in the XANES region), which is slightly larger than the statistical error from total counts N of about 0.3% (that is, $\Delta N = N^{1/2}$ of total counts per energy point). In all figures, we show the SD as error. The error of the difference spectra was calculated by Gaussian error propagation. The obtained XAS spectra were processed using the XAS analysis software package Athena (*44*) to remove glitches and to normalize each individual spectra. Afterward, the data were exported from Athena, and all further data treatment, such as averaging and calculation of the statistical error, was done in IGOR Pro Software (WaveMetrics Inc.). Some illustrations of the data in the paper are created using the Origin software (OriginLab).

Construction of interaction potentials

Born-Oppenheimer first-principles molecular dynamics simulations in the canonical ensemble NVT (at $T = 300$ K) using periodic boundary conditions were performed using the CP2K simulation suite (www.cp2k.org) containing the Quickstep module for the DFT calculations (*45*, *46*). We followed a protocol similar to that of Martyna *et al.* (*47*) using a double-zeta basis set that has been optimized for the condensed phase (*48*) in conjunction with Goedecker-Teter-Hutter pseudopotentials (*48*) using a 400-rydberg cutoff for the auxiliary plane wave basis. A Nose-Hoover chain thermostat was attached to every degree of freedom (translations of nuclei) to control the temperature (*47*). The Becke exchange (*49*) and correlation due to LYP (Lee-Yang-Parr) (*50*) was used in addition to the dispersion correction (D2) put forth by Grimme (*51*) with a 40 Å cutoff (that is, beyond the minimum image convention).

The carbonate/bicarbonate simulation cell consisted of one carbonate anion in 95 water molecules and a single Ca^{2+}, in a cubic box with an edge length of 14.40 Å. The PMF was constructed by restraining the carbon of the CO_3^{2-} or HCO_3^- species and the Ca^{2+}, where sampling windows over the distance ranging from 2.4 to 5.6 Å were equally spaced by 0.2 Å using harmonic umbrella potentials of the form $V_{\mathrm{umbrella}} = k(r - r_0)^2$ with a force constant k of 4000 kJ mol^{-1} nm^{-2}. To ensure sufficient sampling in the barrier region, additional windows with a stiffer force constant were added, ranging from 3.6 to 4.6 Å equally spaced apart by 0.1 Å with a force constant of 14,000 kJ mol^{-1} nm^{-2}. In each umbrella window, a trajectory of at least 40 ps was collected after 5 ps of equilibration. The weighted histogram analysis method was used to extract a free energy profile from these histograms (*52*). Simulations for the MM potential were performed using the interaction potential of Raiteri and co-workers (*15*, *16*) under system size and simulation protocols identical with those used for the DFT calculation, but longer simulation times (1.5 ns per window) were accessible for the MM simulations. We also calculated the PMF for the MM potential using a larger simulation cell with linear dimensions of 28 Å to ensure that there were no significant finite size effects.

In principle, the above protocol should converge to a proper PMF that is correctly weighted with respect to orientation. To directly test

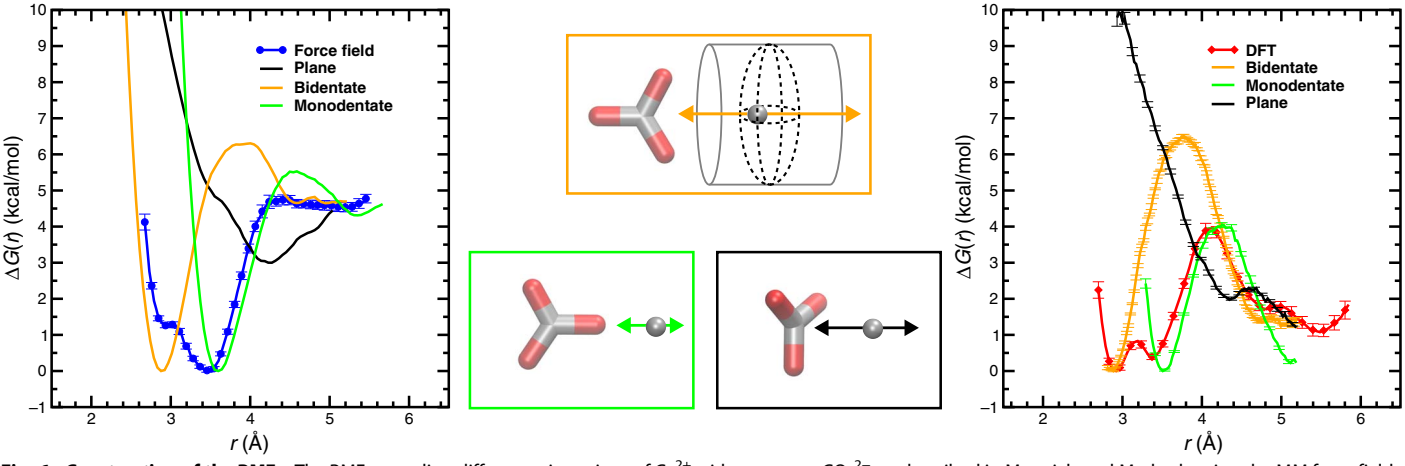

Fig. 6. Construction of the PMFs. The PMFs sampling different orientations of Ca^{2+} with respect to CO_3^{2-}, as described in Materials and Methods using the MM force field (left) (*15*, *16*) and using DFT (right). The blue PMF (left) and red PMF (right) denote the PMF used in Fig. 2B and are shown for reference. The middle panels denote the different orientations considered, corresponding to color codes in the legends of the PMFs. The cylinder in the top middle panel schematically denotes the restraining volume and representative trajectories (black dashed curves) used for all orientations.

our understanding of how the PMF can be affected by the nonspherical nature of the carbonate anion, we performed additional calculations. To this end, we construct PMFs (both DFT and MM) using the same protocol as outlined above by restraining the Ca^{2+} (i) along the bisector of a single O–C–O bend of the carbonate (bidentate), (ii) along the C–O of the carbonate, and (iii) perpendicular to the plane of the carbonate molecule as shown in Fig. 6. The results corroborate our understanding that much of the attractive interaction is in the plane of the carbonate anion. Unfortunately, it is not a straightforward task to simply add the three individual PMFs obtained in Fig. 6 to construct the ones used in Fig. 2B. This would require the proper weighting of the volume of phase space (configuration space) for each of the PMFs or an interpolation of the PMFs for other sets of orientations and, in addition, the setting of an absolute scale. Nevertheless, using the three individual PMFs in Fig. 6 as a basis, it is possible to associate features of the PMFs shown in Fig. 2B used in this study with specific structural moieties, such as monodentate and bidentate. Moreover, the overall agreement with the $(CaCO_3)_6$ structures and populations with the work of Demichelis *et al.* (*4*) demonstrates that the potentials in Fig. 2B are general enough to reproduce subtle structural features.

Monte Carlo simulations of the supersaturated $CaCO_3$ solutions

Monte Carlo (MC) simulations were performed in the canonical ensemble at $T = 298$ K and for simulation volumes of $8 \times 10^3\ nm^3 \leq V \leq 27 \times 10^6\ nm^3$ (*53*). All systems consisted of 300 Ca^{2+} cations and 300 CO_3^{2-} anions, where each ion is represented as a single interaction site. The potential energy of a configuration was calculated as the pairwise-additive sum of distance-dependent Ca^{2+}/CO_3^{2-}, Ca^{2+}/Ca^{2+}, and CO_3^{2-}/CO_3^{2-} PMFs that are truncated at 20 Å. To efficiently sample ion addition or removal to and from clusters, AVBMC moves (*21*) were used in addition to standard translational moves; the ratio between attempted translational and AVBMC moves was set to 4:1. Maximum displacements of translational moves were adjusted during equilibration to yield an acceptance ratio of 40%; AVBMC moves used 2.7 Å $\leq r \leq$ 3.9 Å to define the bonding region and explored 32 trial locations, which resulted in an acceptance rate of about 9%. In the current simulations, a distance-based criterion is used to define the ion-ion aggregate size distribution. To this end, an ion belongs to a given aggregate if its distance to at least one other unlike ion in the aggregate is less than 4.0 Å. Two different sets of simulations were carried out: On one hand, simulations without additional umbrella potentials were used to determine the size (numbers of cations and anions) and shape distributions of small aggregates. On the other hand, simulations with an additional umbrella potential acting on only one specific aggregate were used to probe the size and shape distributions of larger aggregates along the nucleation pathway. For the former set, 16 independent runs were carried out for each concentration. At least 200,000 MC cycles (where a cycle consists of $N = 600$ randomly selected MC moves) were used for equilibration, and at least 1,000,000 MC cycles were used for the production periods. In the other case, one specific ion was tagged to define the aggregate, and self-adaptive umbrella sampling (*22*) was used to flatten the size distribution for this particular cluster, thereby providing access to reliable statistics for much larger aggregates. Eight independent runs were carried out consisting of at least 200,000 MC cycles for the refinement of the umbrella potential, followed by at least 700,000 MC cycles for production.

Calculation of the XANES spectra at the Ca K-edge

XANES calculations were performed at the Ca K-edge using the restricted excitation window TDDFT approach (*32*) as implemented in the NWChem quantum chemistry program (*54*). This approach involves defining a restricted subspace of single excitations from the relevant core orbitals and no restrictions on the target unoccupied states. This is valid because excitations from the deep core levels are well separated from pure valence-valence excitations. For the solvated complexes, the Sapporo-TZP-2012 (*55*) all-electron basis set was used for the single absorbing Ca center, and the 6-311G** basis set (*56*) was used for the O and H atoms. For the solid-state systems, calcite and aragonite, the Sapporo-TZP-2012 (*55*) all-electron basis set was used for the central Ca absorbing center, whereas the remaining Ca centers were represented with the Stuttgart RLC ECPs (relativistic large-core effective core potentials) (*57*) and the C and O atoms were represented with the 6-311G** basis set. The BHLYP (*50*, *58*) exchange-correlation functional was used for all calculations. This approach has been successfully used in several studies including the K-edges of oxygen, carbon, and fluorine in a number of molecular systems (*32*); ruthenium L_3-edge in $[Ru(NH_3)_6]^{3+}$ (*32*); ruthenium L_3-edge in a series of model Ru(II) and Ru(III) complexes and mixed-valence metal (Ru/Fe) dimers (*59*); K-edge spectra of oxygen, nitrogen, and sulfur in cysteine (*60*); dissolved lithium polysulfide species

in Li–S batteries (*61*); Al K-edge studies of the aluminum distribution in zeolites (*62*); alpha-alumina, sodium aluminate, and aqueous Al^{3+} species (*33*); Cl K-edge spectra in actinide hexachloride complexes (*63*); Na K-edge studies of the hydration structure of Na^+ (*13*); and the valence-to-core x-ray emission in transition metal complexes (*64*).

For the aqueous Ca^{2+} and ion-pair species, an ensemble of 50 snapshots was harvested from a 50-ps DFT molecular dynamics trajectory. Each configuration in an ensemble consisted of either one Ca^{2+} ion or a set of ion pairs (Ca^{2+} and CO_3^{2-}) solvated by 40 water molecules. The inclusion of 40 water molecules, which provided explicit solvation about the ions, was required to achieve convergence for the calculated XANES spectrum. For the solid-state systems, the initial structure was taken from the experimental crystallographic data (*65*, *66*). A supercell having 24 $CaCO_3$ units for aragonite and 32 units for calcite was constructed, which was then relaxed in a short 10-ps DFT molecular dynamics run under periodic boundary conditions. A subset ensemble of five clusters, each having 21 Ca^{2+} and 22 CO_3^{2-}, was then carved from this simulated supercell for the XANES calculation. Because the solid structure has fewer degrees of freedom compared to the solvated systems, a larger ensemble was not required. However, for the $CaCO_3$ solids, a larger cluster was required for full convergence, and the outer shell of the cluster must be carefully terminated. These cluster sizes started to reach the practical limit for TDDFT calculations. For this reason, the aqueous ions were >90% converged, whereas the crystalline solid spectra were approximately 70% converged. For the calculations involving dynamically ordered liquid-like oxyanion polymers (DOLLOPs), a single configuration having one dollop unit solvated by 180 classical water molecules was used. Each DOLLOP contained six Ca atoms, for which six separate calculations were carried out for each of those six absorbing Ca centers. An average XANES spectrum from an extracted cluster about these Ca centers for each snapshot was used to represent each dollop. For the overall XANES spectrum representing the DOLLOP equilibrium, the individual spectra were weighted against the equilibrium population of all four types of DOLLOPs (*4*). Figure S4A shows all four DOLLOPs and the weighted spectrum using the equilibrium population. Figure S4B shows the simulated XANES spectra of each ion-pair species for both carbonate and bicarbonate systems. These spectra have later been used to calculate the difference spectra, which can be compared directly to the experiment.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/4/1/eaao6283/DC1

Supplementary Text

fig. S1. Adjustment of the liquid jet position.

fig. S2. Influence of concentration, pH, and extension line on the experimental difference spectra.

fig. S3. Summary of the calculated free energies.

fig. S4. Calculation of the XANES spectra at the Ca K-edge.

fig. S5. Normalized $\mu(E)$ spectra for crystalline and aqueous samples from the experiment compared to those from TDDFT calculations.

fig. S6. Comparison of reference materials to experimental difference spectra.

fig. S7. MINTEQ simulations of calcium titration experiments showing predicted free calcium ion concentrations versus total calcium concentration in solutions before nucleation.

table S1. Summary of the numerical values from the AVBMC simulations.

table S2. Ion-pair speciation.

table S3. Predicted ion-pair distribution.

REFERENCES AND NOTES

1. D. Kashchiev, Forms and applications of the nucleation theorem. *J. Chem. Phys.* **125**, 014502 (2006).
2. D. Gebauer, A. Völkel, H. Cölfen, Stable prenucleation calcium carbonate clusters. *Science* **322**, 1819–1822 (2008).
3. D. Gebauer, H. Cölfen, Prenucleation clusters and non-classical nucleation. *Nano Today* **6**, 564–584 (2011).
4. R. Demichelis, P. Raiteri, J. D. Gale, D. Quigley, D. Gebauer, Stable prenucleation mineral clusters are liquid-like ionic polymers. *Nat. Commun.* **2**, 590 (2011).
5. A. F. Wallace, L. O. Hedges, A. Fernandez-Martinez, P. Raiteri, J. D. Gale, G. A. Waychunas, S. Whitelam, J. F. Banfield, J. J. De Yoreo, Microscopic evidence for liquid-liquid separation in supersaturated $CaCO_3$ solutions. *Science* **341**, 885–889 (2013).
6. M. Kellermeier, P. Raiteri, J. K. Berg, A. Kempter, J. D. Gale, D. Gebauer, Entropy drives calcium carbonate ion association. *ChemPhysChem* **17**, 3535–3541 (2016).
7. E. M. Pouget, P. H. H. Bomans, J. A. C. M. Goos, P. M. Frederik, G. de With, N. A. J. M. Sommerdijk, The initial stages of template-controlled $CaCO_3$ formation revealed by cryo-TEM. *Science* **323**, 1455–1458 (2009).
8. L. N. Plummer, E. Busenberg, The solubilities of calcite, aragonite and vaterite in CO_2-H_2O solutions between 0 and 90°C, and an evaluation of the aqueous model for the system $CaCO_3$-CO_2-H_2O. *Geochim. Cosmochim. Acta* **46**, 1011–1040 (1982).
9. P. J. M. Smeets, A. R. Finney, W. J. E. M. Habraken, F. Nudelman, H. Friedrich, J. Laven, J. J. De Yoreo, P. M. Rodger, N. A. J. M. Sommerdijk, A classical view on nonclassical nucleation. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7882–E7890 (2017).
10. G. A. Tribello, F. Bruneval, C. Liew, M. Parrinello, A molecular dynamics study of the early stages of calcium carbonate growth. *J. Phys. Chem. B* **113**, 11680–11687 (2009).
11. W. Kohn, L. J. Sham, Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
12. G. Lippert, J. Hutter, M. Parrinello, A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* **92**, 477–488 (1997).
13. M. Galib, M. D. Baer, L. B. Skinner, C. J. Mundy, T. Huthwelker, G. K. Schenter, C. J. Benmore, N. Govind, J. L. Fulton, Revisiting the hydration structure of aqueous Na^+. *J. Chem. Phys.* **146**, 084504 (2017).
14. S. Bogatko, E. Cauët, E. Bylaska, G. Schenter, J. N. Fulton, J. Weare, The aqueous Ca^{2+} system, in comparison with Zn^{2+}, Fe^{3+}, and Al^{3+}: An ab initio molecular dynamics study. *Chemistry* **19**, 3047–3060 (2013).
15. P. Raiteri, J. D. Gale, D. Quigley, P. M. Rodger, Derivation of an accurate force-field for simulating the growth of calcium carbonate from aqueous solution: A new model for the calcite–water interface. *J. Phys. Chem. C* **114**, 5997–6010 (2010).
16. P. Raiteri, J. D. Gale, Water is the key to nonclassical nucleation of amorphous calcium carbonate. *J. Am. Chem. Soc.* **132**, 17623–17634 (2010).
17. M. D. Baer, C. J. Mundy, Local aqueous solvation structure around Ca^{2+} during Ca^{2+}···Cl^- pair formation. *J. Phys. Chem. B* **120**, 1885–1893 (2016).
18. P. Raiteri, R. Demichelis, J. D. Gale, Thermodynamically consistent force field for molecular dynamics simulations of alkaline-earth carbonates and their aqueous speciation. *J. Phys. Chem. C* **119**, 24447–24458 (2015).
19. B. A. Legg, J. J. De Yoreo, The energetics of prenucleation clusters in lattice solutions. *J. Chem. Phys.* **145**, 211921 (2016).
20. T. T. Duignan, M. D. Baer, C. J. Mundy, Ions interacting in solution: Moving from intrinsic to collective properties. *Curr. Opin. Colloid Interface Sci.* **23**, 58–65 (2016).
21. B. Chen, J. I. Siepmann, Improving the efficiency of the aggregation–volume–bias Monte Carlo algorithm. *J. Phys. Chem. B* **105**, 11275–11282 (2001).
22. B. Chen, J. I. Siepmann, K. J. Oh, M. L. Klein, Aggregation-volume-bias Monte Carlo simulations of vapor-liquid nucleation barriers for Lennard-Jonesium. *J. Chem. Phys.* **115**, 10903–10913 (2001).
23. B. N. Hale, Monte Carlo calculations of effective surface tension for small clusters. *Aust. J. Phys.* **49**, 425–434 (1996).
24. Z. Zou, W. J. E. M. Habraken, L. Bertinetti, Y. Politi, A. Gal, S. Weiner, L. Addadi, P. Fratzl, On the phase diagram of calcium carbonate solutions. *Adv. Mater. Interfaces* **4**, 1600076 (2017).
25. J. Stöhr, *NEXAFS Spectroscopy* (Springer, 1992).
26. J. J. Rehr, A. L. Ankudinov, Progress in the theory and interpretation of XANES. *Coord. Chem. Rev.* **249**, 131–140 (2005).
27. M. A. Brown, T. Huthwelker, A. B. Redondo, M. Janousch, M. Faubel, C. A. Arrell, M. Scarongella, M. Chergui, J. A. van Bokhoven, Changes in the silanol protonation state measured in situ at the silica–aqueous interface. *J. Phys. Chem. Lett.* **3**, 231–235 (2012).
28. S. Pin, T. Huthwelker, M. A. Brown, F. Vogel, Combined sulfur K-edge XANES–EXAFS study of the effect of protonation on the sulfate tetrahedron in solids and solutions. *J. Phys. Chem. A* **117**, 8368–8376 (2013).
29. A.-M. Flank, G. Cauchon, P. Lagarde, S. Bac, M. Janousch, R. Wetter, J.-M. Dubuisson, M. Idir, F. Langlois, T. Moreno, D. Vantelon, LUCIA, a microfocus soft XAS beamline. *Nucl. Instrum. Methods Phys. Res., Sect. B* **246**, 269–274 (2006).
30. J. L. Fulton, S. M. Heald, Y. S. Badyal, J. M. Simonson, Understanding the effects of concentration on the solvation structure of Ca^{2+} in aqueous solution. I: The perspective on local structure from EXAFS and XANES. *J. Phys. Chem. A* **107**, 4688–4696 (2003).
31. F. M. Michel, J. MacDonald, J. Feng, B. L. Phillips, L. Ehm, C. Tarabrella, J. B. Parise, R. J. Reeder, Structural characteristics of synthetic amorphous calcium carbonate. *Chem. Mater.* **20**, 4720–4728 (2008).

32. K. Lopata, B. E. Van Kuiken, M. Khalil, N. Govind, Linear-response and real-time time-dependent density functional theory studies of core-level near-edge x-ray absorption. *J. Chem. Theory Comput.* **8**, 3284–3292 (2012).
33. J. L. Fulton, N. Govind, T. Huthwelker, E. J. Bylaska, A. Vjunov, S. Pin, T. D. Smurthwaite, Electronic and chemical state of aluminum from the single- (K) and double-electron excitation ($KL_{II\&III}$, KL_I) x-ray absorption near-edge spectra of α-alumina, sodium aluminate, aqueous $Al^{3+}\cdot(H_2O)_6$, and aqueous $Al(OH)_4^-$. *J. Phys. Chem. B* **119**, 8380–8388 (2015).
34. V. Martin-Diaconescu, M. Gennari, B. Gerey, E. Tsui, J. Kanady, R. Tran, J. Pécaut, D. Maganas, V. Krewald, E. Gouré, C. Duboc, J. Yano, T. Agapie, M.-N. Collomb, S. DeBeer, Ca K-edge XAS as a probe of calcium centers in complex systems. *Inorg. Chem.* **54**, 1283–1292 (2015).
35. J. P. Gustafsson, Visual MINTEQ ver. 3.0 (Department of Land and Water Resources Engineering, KTH, 2011).
36. D. A. Kulik, T. Wagner, S. V. Dmytrieva, G. Kosakowski, F. F. Hingerl, K. V. Chudnenko, U. R. Berner, GEM-Selektor geochemical modeling package: Revised algorithm and GEMS3K numerical kernel for coupled simulation codes. *Comput. Geosci.* **17**, 1–24 (2013).
37. M. Kellermeier, A. Picker, A. Kempter, H. Cölfen, D. Gebauer, A straightforward treatment of activity in aqueous $CaCO_3$ solutions and the consequences for nucleation theory. *Adv. Mater.* **26**, 752–757 (2014).
38. D. Genovese, M. Montalti, F. Otálora, J. Gómez-Morales, M. Sancho-Tomás, G. Falini, J. M. Garcia-Ruiz, Role of $CaCO_3^\circ$ neutral pair in calcium carbonate crystallization. *Cryst. Growth Des.* **16**, 4173–4177 (2016).
39. M. P. Andersson, S. Dobberschütz, K. K. Sand, D. J. Tobler, J. J. De Yoreo, S. L. S. Stipp, A microkinetic model of calcite step growth. *Angew. Chem. Int. Ed.* **55**, 11086–11090 (2016).
40. A. Carino, A. Testino, M. R. Andalibi, F. Pilger, P. Bowen, C. Ludwig, Thermodynamic-kinetic precipitation modeling. A case study: The amorphous calcium carbonate (ACC) precipitation pathway unravelled. *Cryst. Growth Des.* **17**, 2006–2015 (2017).
41. L. F. Wang, I. Sondi, E. Matijević, Preparation of uniform needle-like aragonite particles by homogeneous precipitation. *J. Colloid Interface Sci.* **218**, 545–553 (1999).
42. M. Faatz, F. Gröhn, G. Wegner, Amorphous calcium carbonate: Synthesis and potential intermediate in biomineralization. *Adv. Mater.* **16**, 996–1000 (2004).
43. K. R. Wilson, B. S. Rude, J. Smith, C. Cappa, D. T. Co, R. D. Schaller, M. Larsson, T. Catalano, R. J. Saykally, Investigation of volatile liquid surfaces by synchrotron x-ray spectroscopy of liquid microjets. *Rev. Sci. Instrum.* **75**, 725–736 (2004).
44. B. Ravel, M. Newville, *ATHENA*, *ARTEMIS*, *HEPHAESTUS*: Data analysis for x-ray absorption spectroscopy using *IFEFFIT*. *J. Synchrotron Radiat.* **12**, 537–541 (2005).
45. J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, J. Hutter, QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **167**, 103–128 (2005).
46. J. VandeVondele, J. Hutter, Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **127**, 114105 (2007).
47. G. J. Martyna, M. L. Klein, M. Tuckerman, Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635–2643 (1992).
48. S. Goedecker, M. Teter, J. Hutter, Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **54**, 1703–1710 (1996).
49. A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic-behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
50. C. Lee, W. Yang, R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B* **37**, 785–789 (1988).
51. S. Grimme, Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **25**, 1463–1473 (2004).
52. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenberg, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
53. W. W. Wood, F. R. Parker, Monte Carlo equation of state of molecules interacting with the Lennard-Jones potential. I. A supercritical isotherm at about twice the critical temperature. *J. Chem. Phys.* **27**, 720–733 (1957).
54. M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. A. de Jong, NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **181**, 1477–1489 (2010).
55. T. Noro, M. Sekiya, T. Koga, Segmented contracted basis sets for atoms H through Xe: Sapporo-(DK)-*n*ZP sets (*n* = D, T, Q). *Theor. Chem. Acc.* **131**, 1124 (2012).
56. R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, Self-consistent molecular-orbital methods. XX. A basis set for correlated wave-functions. *J. Chem. Phys.* **72**, 650–654 (1980).
57. G. Igelmann, H. Stoll, H. Preuss, Pseudopotentials for main group elements (IIIa through VIIa). *Mol. Phys.* **65**, 1321–1328 (1988).
58. A. D. Becke, A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372–1377 (1993).
59. B. E. Van Kuiken, M. Valiev, S. L. Daifuku, C. Bannan, M. L. Strader, H. Cho, N. Huse, R. W. Schoenlein, N. Govind, M. Khalil, Simulating Ru L_3-edge x-ray absorption spectroscopy with time-dependent density functional theory: Model complexes and electron localization in mixed-valence metal dimers. *J. Phys. Chem. A* **117**, 4444–4454 (2013).
60. Y. Zhang, J. D. Biggs, D. Healion, N. Govind, S. Mukamel, Core and valence excitations in resonant x-ray spectroscopy using restricted excitation window time-dependent density functional theory. *J. Chem. Phys.* **137**, 194306 (2012).
61. M. Vijayakumar, N. Govind, E. Walter, S. D. Burton, A. Shukla, A. Devaraj, J. Xiao, J. Liu, C. Wang, A. Karim, S. Thevuthasan, Molecular structure and stability of dissolved lithium polysulfide species. *Phys. Chem. Chem. Phys.* **16**, 10923–10932 (2014).
62. A. Vjunov, J. L. Fulton, T. Huthwelker, S. Pin, D. Mei, G. K. Schenter, N. Govind, D. M. Camaioni, J. Z. Hu, J. A. Lercher, Quantitatively probing the Al distribution in zeolites. *J. Am. Chem. Soc.* **136**, 8296–8306 (2014).
63. N. Govind, W. A. de Jong, Simulating Cl K-edge x-ray absorption spectroscopy in MCl_6^{2-} (M = U, Np, Pu) complexes and $UOCl_5^-$ using time-dependent density functional theory. *Theor. Chem. Acc.* **133**, 1463 (2014).
64. Y. Zhang, S. Mukamel, M. Khalil, N. Govind, Simulating valence-to-core x-ray emission spectroscopy of transition metal complexes with time-dependent density functional theory. *J. Chem. Theory Comput.* **11**, 5804–5809 (2015).
65. H. Sitepu, B. H. O'Connor, D. Li, Comparative evaluation of the March and generalized spherical harmonic preferred orientation models using x-ray diffraction data for molybdite and calcite powders. *J. Appl. Crystallogr.* **38**, 158–167 (2005).
66. S. M. Antao, I. Hassan, The orthorhombic structure of $CaCO_3$, $SrCO_3$, $PbCO_3$ and $BaCO_3$: Linear structural trends. *Can. Mineral.* **47**, 1245–1255 (2009).

Acknowledgments: We acknowledge D. Gebauer for providing us with titration data, J. Gale for technical discussions about our theoretical approach, S. Kathmann for useful discussions on nucleation, and J. A. van Bokhoven for helpful discussions. We also thank R. Wetter and C. Frieh for providing technical support during the beamtime and for the construction of all in situ setups. **Funding:** PMF and MM simulations were performed at the Pacific Northwest National Laboratory (PNNL) with support from the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (BES), Division of Material Sciences and Engineering. The solution model and theoretical XANES calculations were supported by the DOE, Office of Science, BES, Division of Chemical Sciences, Geosciences, and Biosciences. AVBMC simulations were supported through an award from the NSF (CHE-1265849). All measurements were performed at the PHOENIX beamline at the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland. S.P. acknowledges funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 290605 (PSIFELLOW/COFUND). J.M.X. acknowledges funding from the Swiss National Science Foundation (grant no. 200021_157148). DFT simulations were performed within the Materials Synthesis and Simulation Across Scales (MS3) Initiative through the Laboratory Research and Development Program at PNNL. E.O.F. acknowledges the Alternate Sponsored Fellow program at PNNL where he spent 3 months during this project and a University of Minnesota Doctoral Dissertation Fellowship. The PMF calculations used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the DOE, Office of Science under contract no. DE-AC02-05CH11231. A portion of the research (XANES calculations) was performed at the Environmental Molecular Sciences Laboratory, which is a DOE Office of Science User Facility; all other calculations were performed using PNNL's Institutional Computing resources. The AVBMC simulations used resources of the Minnesota Supercomputing Institute. PNNL is a multiprogram national laboratory operated for the DOE by Battelle under contract no. DE-AC05-76RL01830. **Author contributions:** K.H., C.B., J.M.X., S.P., J.L.F., and T.H. performed and analyzed the XANES experimental data. M.G., N.G., and J.L.F. performed and analyzed the theoretical XANES spectra. M.D.B. and C.J.M. performed and analyzed DFT and MM calculations. E.O.F. and J.I.S. performed and analyzed MC calculations. E.O.F., J.I.S., G.K.S., B.A.L., and C.J.M. constructed and analyzed the theoretical solution model. B.A.L. and K.H. validated the solution model with known thermodynamic models. J.J.D.Y., K.H., E.O.F., T.H., J.L.F., and C.J.M. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 10 August 2017
Accepted 15 December 2017
Published 26 January 2018
10.1126/sciadv.aao6283

Citation: K. Henzler, E. O. Fetisov, M. Galib, M. D. Baer, B. A. Legg, C. Borca, J. M. Xto, S. Pin, J. L. Fulton, G. K. Schenter, N. Govind, J. I. Siepmann, C. J. Mundy, T. Huthwelker, J. J. De Yoreo, Supersaturated calcium carbonate solutions are classical. *Sci. Adv.* **4**, eaao6283 (2018).