The Structure of CaSO4 Nanorods - the Precursor of Gypsum

Journal:	*The Journal of Physical Chemistry*
Manuscript ID	jp-2019-04268w.R1
Manuscript Type:	Article
Date Submitted by the Author:	01-Jun-2019
Complete List of Authors:	Stawski, Tomasz; Deutsches Geoforschungszentrum Potsdam, Van Driessche, Alexander; Centre National de la Recherche Scientifique, ISTerre Besselink, Rogier; Deutsches Geoforschungszentrum Potsdam, Interface Geochemistry Section 4.4 Byrne, Emily; Curtin University, Department of Chemistry Raiteri, Paolo; Curtin University, Department of Chemistry Gale, Julian; Curtin University, Nanochemistry Research Institute, Department of Chemistry Benning, Liane; Deutsches Geoforschungszentrum Potsdam,

The Structure of $CaSO_4$ Nanorods - the Precursor of Gypsum

Tomasz M. Stawski[1], Alexander E.S. Van Driessche[2]**, Rogier Besselink[1,2], Emily H. Byrne[3], Paolo Raiteri[3]***, Julian D. Gale[3], Liane G. Benning[1,4,5]*

[1]*German Research Centre for Geosciences, GFZ, Interface Geochemistry, Telegrafenberg, 14473, Potsdam, Germany;*

[2]*Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTerre, 38000, Grenoble, France;*

[3]*Curtin Institute for Computation, Curtin University, PO Box U1987, Perth, WA 6845, Australia*

[4]*Department of Earth Sciences, Free University of Berlin, Malteserstr. 74-100 / Building A, 12249, Berlin, Germany;*

[5]*School of Earth and Environment, University of Leeds, Woodhouse Lane, LS2 9JT, Leeds, UK.*

Corresponding authors:

** tomasz.stawski@gmail.com; www.researchgate.net/profile/Tomasz_Stawski*

** *alexander.van-driessche@univ-grenoble-alpes.fr*

*** *P.Raiteri@curtin.edu.au*

Abstract

The understanding of the gypsum ($CaSO_4 \cdot 2H_2O$) formation pathway from aqueous solutions has been the subject of intensive research in the last couple of years. This interest stems from the fact that gypsum appears to fall into a broader category of crystalline materials whose formation does not follow classical nucleation and growth theories. The pathways involve transitory precursor cluster species, yet the actual structural properties of such clusters are not very well understood. Here, we show how in situ high-energy X-ray diffraction (HEXD) experiments and molecular dynamics (MD) simulations can be combined to derive the structure of small $CaSO_4$ clusters, which are precursors to crystalline gypsum. We fitted several plausible structures to the derived pair distribution functions (PDFs), and explored their dynamic properties using unbiased MD based on both rigid-ion and polarizable force-fields. Determination of the structure and (meta)stability of the primary species is important both from a fundamental and applied perspective; for example, this will allow for improved design of additives for greater control of the nucleation pathway.

Introduction

In recent years, we have come to appreciate the astounding complexity of the processes leading to the formation of mineral phases from ions in aqueous solutions. The original, and rather naive, 'textbook' image of these phenomena, stemming from the adaptation of classical nucleation and growth theories, has increased in complexity due to the discovery of a variety of precursor and intermediate species. These include solute clusters (e.g. prenucleation clusters, PNCs), liquid(-like) phases, as well as amorphous and nanocrystalline solids, among others[1–9]. Such precursor or intermediate species constitute different points along the pathway from dissolved ions to the final solids (e.g. crystals)[10]. Despite a plethora of indirect experimental[4] and modeling-based[11] evidence for the existence of such precursor species there has yet to be any direct structural information derived from experimental observations. Recently, we reported[12,13] that elongated sub-3 nm primary species are the fundamental building units for dihydrated calcium sulfate crystals, i.e. gypsum ($CaSO_4 \cdot 2H_2O$). Based on in situ and time-resolved small- and wide-angle X-ray scattering data (SAXS/WAXS), we showed that the formation, aggregation, rearrangement and the partial coalescence of these structural units constitutes the actual mechanism behind gypsum crystallization. Nevertheless, from those experiments we were unable to elucidate the nature or the molecular

structure of the primary species. To address this issue, we carried out in situ high-energy X-ray diffraction experiments and combined these with molecular dynamics simulations. From both in situ and in silico data we have derived the molecular structure of the precursor species. Determination of the actual structure and (meta)stability of the primary species is important both from a fundamental and applied perspective; for example, this will allow for improved design of additives for greater control of the nucleation pathway.

Experimental

Synthesis and scattering experiments

$CaSO_4$ was synthesized by reacting equimolar aqueous solutions (ultrapure deionized water, MilliQ, resistivity >18 MΩ·cm) of $CaCl_2 \cdot 2H_2O$ (pure, Sigma) and Na_2SO_4 (> 99%, Sigma) at final concentrations of 50 and 100 mmol/L at T = 10 °C. Prior to mixing, all solutions were equilibrated in a fridge at the desired reaction temperatures. The compositions of solutions (Table 1), under the assumption that gypsum was the final solid phase, were calculated with PHREEQC[14] based on the following reaction:

$$CaCl_2 + 2H_2O + Na_2SO_4 \rightarrow CaSO_4 \cdot 2H_2O\downarrow + 2NaCl$$

$CaSO_4$ formation reactions were performed in a 200 mL temperature-stabilized glass reactor. The reaction temperature in the reactor was maintained by means of an active cooling circulator (Huber Ministat 125), and the feedback thermocouple was placed in the reactor.

The reacting solutions were continuously stirred at 350 rpm, and circulated through a custom-built PEEK flow-through cell with Kapton capillary (ID 1.5 mm) using a peristaltic pump (Gilson MiniPuls 3, flow ~10 mL/second). Typically, an experiment started with 40 mL of a temperature-stabilized $CaCl_2$ aqueous solution inside the reactor. This solution was circulated through the capillary cell while scattering patterns were collected continuously as described below. $CaSO_4$ formation reactions were initiated remotely through the injection of 40 mL of a temperature-stabilized Na_2SO_4 aqueous solution. Injection and mixing (within 30 s) of the two solutions was achieved with the use of a secondary peristaltic pump (Gilson MiniPuls 3) as a remote fast injection system. Depending on the supersaturation, reactions were followed for up to 4 hours. The experimental approach used for these in situ HEXD experiments is identical to that used in our previous in situ SAXS experiments.

Table 1. Composition of $CaSO_4$ solutions from PHREEQC[14] modelling.

	50 mM $CaSO_4$	**100 mM $CaSO_4$**
Initial composition	$[Ca^{2+}] = [SO_4^{2-}] = 35$ mM	$[Ca^{2+}] = [SO_4^{2-}] = 69$ mM
	$[CaSO_4]^* = 15$ mM	$[CaSO_4]^* = 31$ mM
Final composition	$[Ca^{2+}] = [SO_4^{2-}] = 16.3$ mM	$[Ca^{2+}] = [SO_4^{2-}] = 20$ mM
	$[CaSO_4]^* = 3.9$ mM	$[CaSO_4]^* = 3.9$ mM
	[solid equivalent] = 29.8 mM	[solid equivalent] = 76.1 mM

* - dissolved species. The model considers here an ion pair.

In order to follow the formation processes in situ, and in a time-resolved manner, the flow-through cell was mounted on a high-energy X-ray diffraction (HEXD) system and the measurements were carried out at beamline I15 of the Diamond Light Source (UK). The

atomic pair distribution functions (PDFs) were obtained by using the PDFgetX3 software package[15]. Details regarding the data collection and analysis are provided in the ESI.

Atomistic simulations

Classical molecular dynamics simulations were run with the LAMMPS program[16] at 300 K and 1 atm in a cubic simulation cell of approximately 50 Å in side, which contained ~4,130 water molecules and the relevant $CaSO_4$ cluster. The long-range electrostatics were computed with the Particle Mesh Ewald algorithm with an accuracy of 10^{-5} while the bonded and van der Waals interactions were described by our recently developed force field[17] (FF1). The real space cut-off was set to 9 Å with a tapering function applied between 6 and 9 Å. The water was described with the SPC/Fw model[18], and no tapering was applied to the water-water interactions. The equations of motions were integrated with a 1 fs time-step using the velocity Verlet algorithm, while the temperature was controlled with the Canonical Sampling through Velocity Rescaling algorithm[19]. The cell was constrained to remain cubic throughout the simulations.

All simulations with the AMOEBA polarizable force field[20] were run with the openMM7.1 code[21]. The equations of motion were integrated with the Langevin algorithm using a 0.5 fs time-step, a friction coefficient of 1 ps^{-1} and a temperature of 300 K. The pressure was maintained at 1 atm using an isotropic Monte Carlo barostat[22,23] with a collision frequency of 25 steps. The real space cut-off was set to 10 Å for all non-bonded interactions, while the long-range electrostatic interactions were computed with the Particle Mesh Ewald algorithm,

again with an accuracy of 10^{-5}. Induced dipoles were solved for iteratively at each time-step until an accuracy of 10^{-5} was reached. The initial configurations were the same as used in the rigid-ion MD simulations.

The trajectories for cluster I calculated for both force fields are included as a part of the ESI: Supporting Files.

Results and Discussion

Background-corrected and time-resolved in situ diffraction patterns from a 50 mM $CaSO_4$ solution at 10 °C are presented in Fig 1A. From the very first frame (265 s) and during the initial ~2,000 s, the measured scattering data contains only an invariant broad diffuse pattern i.e. diffraction peaks were absent. After this induction time, the diffraction pattern of a crystalline phase gradually started to develop. The crystalline phase coexisted throughout the duration of the measurements with the diffuse pattern (Fig. 1B). Importantly, the as-formed diffuse pattern is significantly different from the scattering patterns obtained from the solvent, i.e. H_2O, and the aqueous solutions containing the various components of the system under consideration (Fig. 1C). Therefore, the diffuse pattern can be attributed to the presence of a disordered phase of $CaSO_4$ as is demonstrated below. Selected diffraction patterns were converted to pair distribution functions (PDFs, $G(r)$s) using a numerical inverse Fourier transform, in order to analyse, identify and model the structural changes in real-space. Throughout the reaction, the crystalline material was fitted with a gypsum model structure (Fig. 1D) without any trace of other crystalline $CaSO_4$ phases (i.e. bassanite or anhydrite).

The $G(r)$ resulting from the diffuse scattering shows several inter-atomic distances common to all calcium sulfate phases (Fig. 1E): S-O at 1.49 Å; Ca-O at ~2.6 Å; shortest (1st neighbour) Ca-Ca and S-S at ~4 Å. With respect to the crystalline forms this disordered phase lacks the longer, i.e. 2nd neighbour, Ca-Ca/Ca-S distance found at ~5-6 Å, although other longer distances are still present. Finally, and most importantly, the PDF plot shows that the structural coherence of the species constituting the diffuse phase does not exceed ~1.5 nm, and thus we can refer to it as a nanophase. Overall, the PDF plots reveal that the species from which the nanophase is built are structurally more complex than simple ions or ion pairs in solution. They have the form of clusters containing at least one single Ca-Ca distance.

The diffuse diffraction patterns shown in Fig. 1A-C are lacking detail, and only show very broad diffraction maxima, which is characteristic of small and thermally labile species in solution. Consequently, one expects that for the primary species a whole family of possible structures can fit to the $G(r)$, rather than one unique solution. Based on our previous small-angle X-ray scattering results, we can narrow down the number of possibilities by constraining the form of such clusters to rod-shaped species with a cross-sectional diameter of ~0.5 nm and a length of several nm[12,13]. Taking into account these constraints, we constructed template clusters, whose structure was further optimized to fit the experimental $G(r)$. Our template clusters were constructed by choosing a primary motif, which contained at least one single Ca-Ca bond distance, stacking a number of such motifs to obtain a rod, and then optimizing the structure using the randomized simplex method so that the simulated PDF of the structure would match as closely as possible the experimental one. In Fig. 2 we present several possible structures of the primary species, based on the best fits to the $G(r)$ of

the nanophase obtained by optimizing the initial template clusters using a DiffPy library[24].

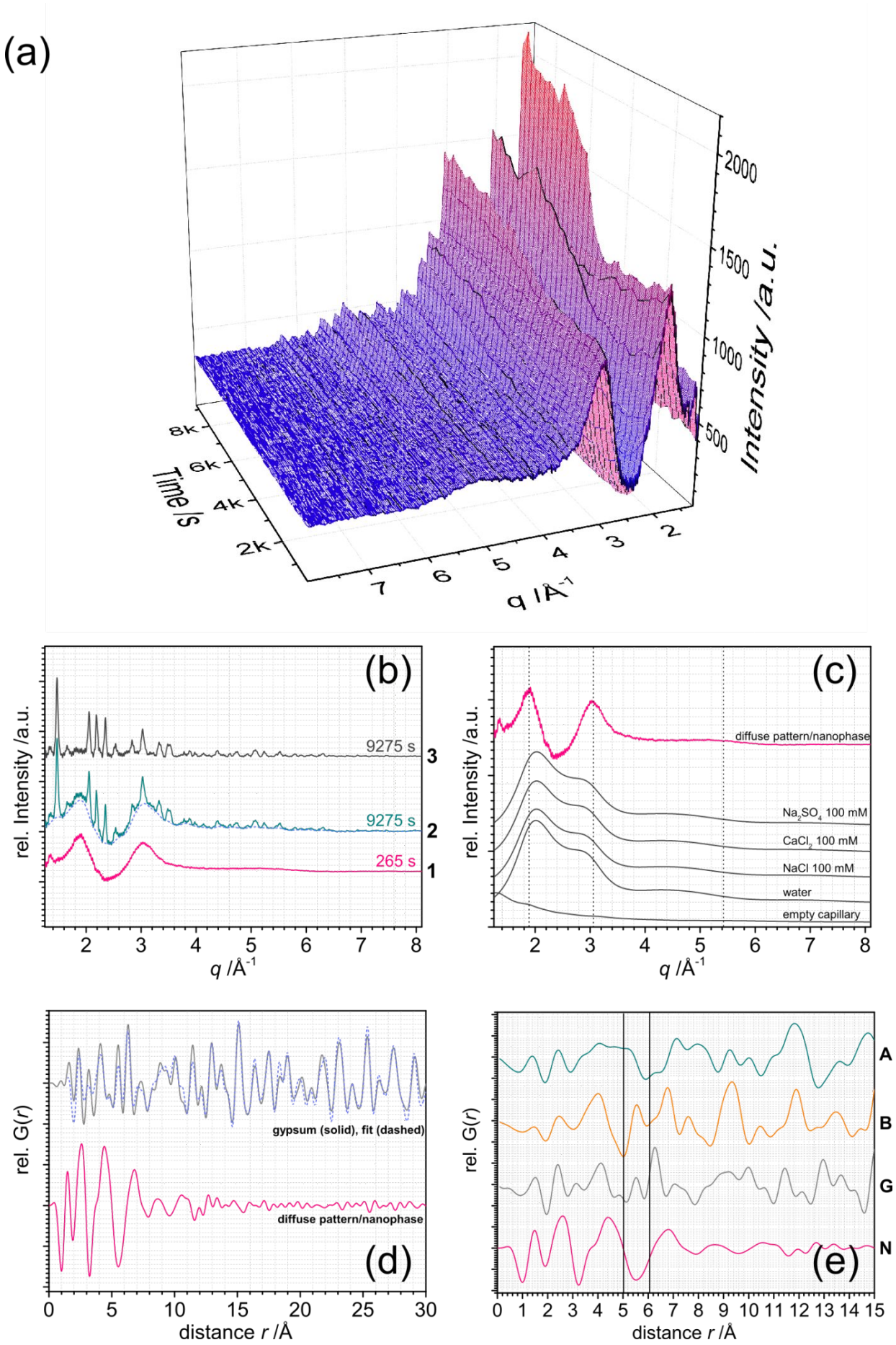
(a)
Intensity /a.u.
2000
1500
1000
500
Time /s
8k
6k
4k
2k
q /Å-1
7
6
5
4
3
2
(b)
rel. Intensity /a.u.
9275 s
3
9275 s
2
265 s
1
q /Å-1
2
3
4
5
6
7
8
(c)
rel. Intensity /a.u.
diffuse pattern/nanophase
Na2SO4 100 mM
CaCl2 100 mM
NaCl 100 mM
water
empty capillary
q /Å-1
2
3
4
5
6
7
8
rel. G(r)
gypsum (solid), fit (dashed)
diffuse pattern/nanophase
(d)
0
5
10
15
20
25
30
distance r /Å
rel. G(r)
A
B
G
N
(e)
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
distance r /Å

Fig. 1. Experimental data that follow the formation of solids in a solution with an initial concentration of 50 mM $CaSO_4$ at 10 °C. (a) time-resolved in situ HEXD pattern in which the first frame corresponds to 265 s and the last to 9,275 s; (b) comparison between the diffraction patterns of the initial nanophase 1 (solid, pink) and the final crystalline product 2 (solid, cyan). This final diffraction pattern is a juxtaposition of the diffraction pattern of the nanophase (dashed, blue) and the second purely crystalline phase 3 (solid, black); (c) comparison between the diffraction patterns of the nanophase and other components of the investigated system; (d) selected PDFs from (a) highlighting the structural difference between the nanophase and the crystalline phase clearly identified as gypsum; (e) comparison of PDFs from the nanophase - N, and other $CaSO_4$ phases: gypsum - G, bassanite - B, and anhydrite - A. The difference in the $G(r)$s for the distance r ~5-6 Å is highlighted with two black lines. For (b)-(e) offset is introduced for clarity.

The four plausible clusters (Fig. 2; see ESI for the structural files) were optimized based on the following cross-sectional motifs: I – a simple 2-member square-shaped Ca-SO_4-Ca-SO_4 unit, where 6 stacks are arranged in an AB pattern and each Ca is coordinated by two in-plane water molecules; II – 3 member triangular Ca-SO_4-Ca-SO_4-Ca-SO_4 units, where 6 such units are stacked into a helix with a torsion of 70.4°; III - analogous to II but with a torsion of 70.7° and each Ca is coordinated by two in-plane water molecules; IV – analogous to I, but with a repetitively stacked $(Ca\text{-}SO_4)_2$ unit for which the initial positions were taken from a gypsum fragment (which results in a longer length of the cluster). We used a Pearson correlation coefficient (R) as a proxy to benchmark the structures based on their simulated $G(r)$s with respect to the experimental one. All the proposed models reproduce almost ideally the experimental $G(r)$ up to r = 6 Å (R > 0.99), with minor deviations for higher r < 12 Å. Our analysis shows that the proposed family of solutions should fulfill three conditions: (1) the

clusters are rod-shaped; (2) a single structural unit in the cross-section of the rod does not expand beyond a single Ca-Ca distance, with either of 2 or 3 Ca ions per unit. The diameter of such a cross-sectional unit is ~0.5 nm; (3) the rods are "internally" an anhydrous $CaSO_4$ structure, and bound water is only present at their "surface"; the presence of the coordinated water improves measurably the quality of the fits for r > 6 Å (i.e. compare otherwise identical clusters II and III with *R* of 0.868 *vs*. 0.960).

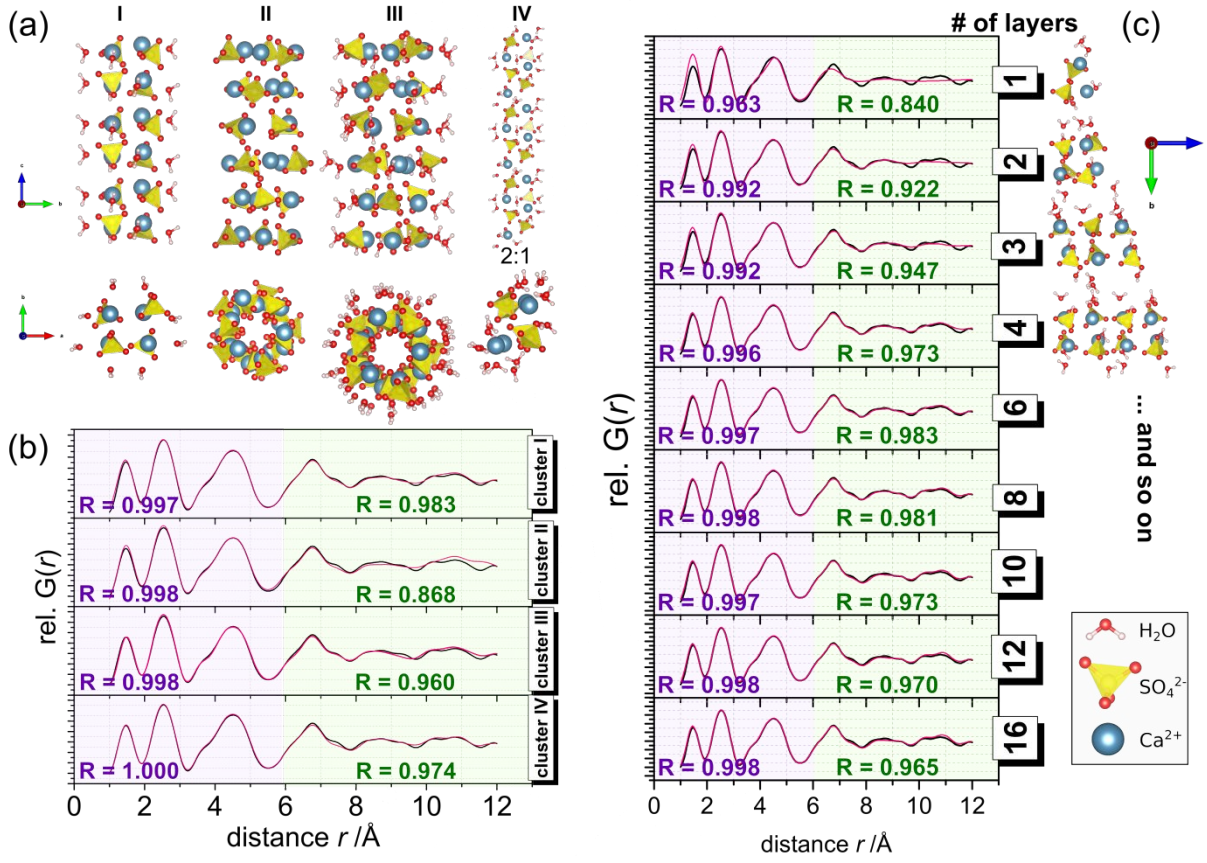

Fig. 2. (a) Selected optimized structural models of $CaSO_4$ rod-shaped clusters for the observed nanophase. Upper row: longitudinal view, Lower row: cross-sectional view. Cluster: I – Ca-Ca AB stack hydrated with 6 layers, II – Ca-Ca-Ca helix stack 70.4° anhydrous with 6 layers, III Ca-Ca-Ca helix stack 70.7° hydrated with 6 layers, IV – Ca-Ca AA stack hydrated with 6 layers. See ESI for the files containing these models; (b) Experimental PDF from the nanophase compared with the PDFs simulated from the structural models of clusters

I-IV; (c) Experimental PDF vs. simulated PDF for structures based on cluster III, in which the number of layers was varied from 2 to 30.

In Fig. 2C we compare the simulated PDFs of clusters similar to type I (with 6 layers) with the experimental data varying the number of layers from 1 to 16. All the clusters considered reproduce the experimental $G(r)$ up to r = ~6 Å (R > 0.99). However, the profile of the PDF is reproduced better for longer clusters with 4 to 16 layers, which is also reflected by the higher R-values in comparison with clusters containing 1, 2 or 3-layers for r > ~6 Å. As the number of layers is further increased the R-values for r > ~6 Å start to marginally decrease again, but they remain relatively high at R > 0.9. The values of R suggest that the optimum structure has 4 to 10 layers (length between 1.1 and 3.1 nm), but the clusters could be longer. Nonetheless, taking into account the overall coherence length, which was limited to ~1.5 nm (Fig. 1E), we can calculate the maximum physical length of a cluster. For anisotropic rod-shaped species, we can use their radius of gyration, R_g, as a proxy for the measured isotropic coherence length by using the dimensions of a cylinder of length L_c and radius R_c, and the following equation[25] $R_g^2 = R_c^2/2 + L_c^2/12$. The clusters considered have a cross-sectional diameter[12] of ~0.50 nm, which results in a length that does not exceed 5.2 nm. This in turn is equivalent to a cluster composed of 16 layers at most. Similar trends were found for structures I, III and IV, where the optimal fits are obtained for rods composed of ~4-12 stacks of structural units, yielding a total length for the rod in the range of ~1.5 - 3.5 nm depending on the packing of the structural units. The estimate of the length of the clusters obtained from the PDF analysis is in accordance with the earlier SAXS measurements of these species[12,13].

Not only do diffraction data allow us to resolve the possible structures of the observed

clusters. They are also useful to evaluate the population of such species with respect to the initial ion concentrations and the final expected gypsum content by calculating the scattering invariants of the crystalline and disordered phases. We performed such analysis using diffraction data of gypsum precipitating at 10 ºC from solution containing initially 50 and 100 mM $CaSO_4$. For both conditions, diffraction patterns of an identical nanophase profile were collected, but as expected, the crystallization process progressed faster for the higher initial $CaSO_4$ concentration[13]. As shown in Fig. 3, the nanophase constituted ~100% of the initial scattering invariant for the 50 mM solution, and >90% for the 100 mM $CaSO_4$. In both cases, by the end of the experiment the nanophase was still persistent and constituted in fact the majority contribution to the invariant at ~82% and ~75% for 50 mM and 100 mM $CaSO_4$ solutions, respectively. Based solely on the HEXD data, it is not possible to determine whether the diffuse pattern originated from solute clusters or from clusters aggregated into larger morphologies. Using thermodynamic modelling approach, we evaluated the expected initial speciation, the final composition of the solution at the equilibrium, and the amount of the precipitated solid for the $CaSO_4$ solution under the as-defined physicochemical conditions (Table 1). Such an approach allowed us to calculate the maximum yield of gypsum (the precipitated solid) that one could obtain from a solution of a given starting concentration of ions (but only including the phases which are in the database). As we compare the final contribution of the amount of the nanophase with respect to the expected amount of gypsum from both solutions, it becomes apparent that the total population of the clusters considerably exceeds what one would expect to find within a crystalline solid, even if we assume that some the clusters could “dwell” in solution (and being equivalent to what the calculation

considers to be the free ions or the ion pairs). This in turn translates into the fact that for the both considered conditions the "final" amount of crystalline gypsum is lower than anticipated at the equilibrium, and hence that the processes is not at the equilibrium by the end of the experiments (although the diffraction patterns were not evolving any more). Such an observation can be rationalized if we assume that gypsum single crystals are in fact mesocrystal-like[26] in nature, composed of crystalline domains "glued" together by the disorganized cluster ("bricks-in-the-wall")[27], and thus a large proportion of the clusters may co-exist inside the solid together with crystalline gypsum. Such imperfect crystals would mature over time converting clusters into the crystalline lattice approaching the equilibrium.

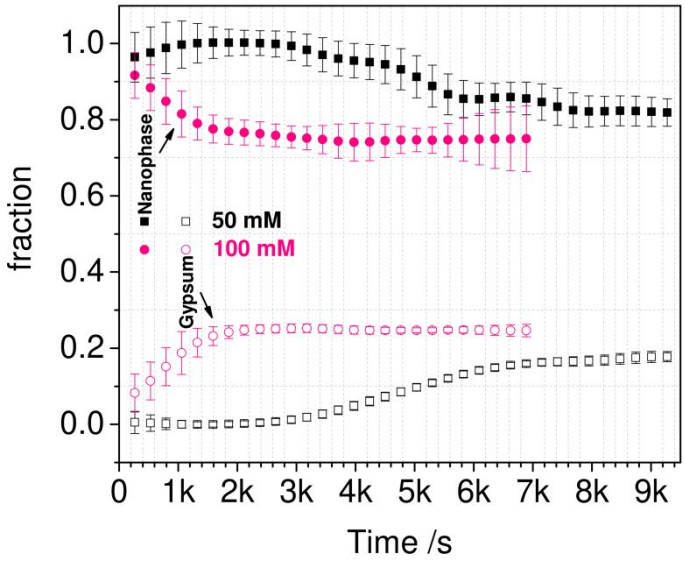

Fig. 3. Contribution as a function of time of the nanophase (filled symbols) and gypsum (open symbols) to the total scattering invariant with an initial concentration of 50 mM (black) and 100 mM (pink) $CaSO_4$ at 10 °C.

To complement the diffraction data, we also attempted to quantify the overall thermodynamic stability of the described above $CaSO_4$ clusters by using metadynamics[28]

simulations with our recently developed[17] force field model and a suite of different collective variables. Although these simulations did not produce unequivocal values for the formation free energy of small clusters (up to 4 formula units), the aggregates did consistently show rod-like shapes (please consult our additional discussion in ESI for the further details on the issues in obtaining the reliable thermodynamic properties of the clusters from the simulations). We therefore decided to use unbiased Molecular Dynamics (MD) simulations, with rigid-ion (*i.e.* classical) and polarizable (*i.e.* AMOEBA) force fields, to assess the stability of the clusters that gave the best fit to the diffraction data (Fig. 2). The scattering-derived clusters were annealed for 5 ns at 300 K. Independent of the force field used and for all the clusters considered, only cluster I maintained its structure throughout the whole MD simulation (Fig. 4), while the remaining models significantly rearranged. The AMOEBA polarizable force field showed a high degree of instability for the remaining clusters, which tended toward elongated highly distorted structures (e.g. Fig. 4A for cluster II), while on the other hand, the rigid-ion force field seemed to favour more compact clusters that resembled two cluster I rods side-by-side. This different behaviour can be explained by considering that the rigid-ion force field gives a dissolution free energy for the anhydrous $CaSO_4$ phase[17] that is too endothermic, and therefore the ions tend to form more compact structures where they are in direct contact. A structural comparison between the relaxed geometry of cluster I at the end of the MD simulations shows that there are significant differences between the rigid-ion and polarizable force fields (Fig. 4B, upper; ESI - the trajectory files). In particular, the former model shows a much more rigid and ordered arrangement of the ions, which form a stack of squares rotated 90 degrees with respect to each other, while the AMOEBA

simulations show a more mobile structure where the ions form a stack of distorted squares. Nevertheless, if we consider the overall trajectories of all the atoms in the cluster over a period of 5 ns, then despite this apparent "disordered" structure resulting from the use of the polarizable AMOEBA forcefield, this cluster structure is more consistent with the experimental data than the one from the non-polarizable model (SI: the trajectory files). These differences are evident in the Ca-Ca, Ca-S and S-S radial distribution functions, $g(r)$s (not PDFs), form trajectories obtained from the MD simulations (Fig. 4B, and ESI: trajectory files). In particular, the rigid-ion MD simulations show strong S-S peaks in the 4–5.5 Å region and a Ca – Ca peak in the 5–6 Å region (Fig. 4B, lower) that are neither in the AMOEBA MD simulations nor in the experimentally determined cluster I structure. On the other hand, the Ca-S $g(r)$ from both the rigid-ion and AMOEBA forcefield are consistent with the experimental data. Based on the above results, the AMOEBA forcefield seems to outperform the non-polarizable rigid ion forcefield and cluster I is the most likely candidate for the model structure of the $CaSO_4$ clusters as it reproduces the recorded scattering pattern. From the MD simulations, it is evident that the terminal parts of the cluster are more mobile, due to having only half of the ionic interactions, which suggests that the core part of the rod is contributing the most to the scattering signal, and therefore the clusters might be slightly longer than the 4-12 stack length that best fits the scattering data. Furthermore, a highly dynamic character is intuitively expected for such small species in solution, yet is practically impossible to implement in our fitting routines used to obtain the rigid structural models in Fig. 2 from the diffraction data. The imposition of constraints results from the fact that we had to limit the number of degrees of freedom to achieve convergence of the template

models. Typically, such constraints are valid assumptions for crystalline structures and are based on symmetry restrictions. Nonetheless, for small clusters such constraints yield merely a valid transitory averaged model (since it fits the experimental PDF), but not necessarily the minimum energy model or any equivalent low-symmetry counterpart structure. Hence, the MD simulations are essential to probe the configuration space in which our cluster I, as defined in Fig. 2, is actually just one of many possibilities. From the MD simulations it is not possible to gauge the actual thermodynamic stability of these clusters, but dynamic models using the AMOEBA force-field strongly support the HEXD and previous SAXS data[12], which show that $CaSO_4$ forms rod-shaped structures, upon creation of a supersaturated solution, that are thermodynamically stable/kinetically persistent in solution for sufficiently long time to be observed experimentally. It is very important to also note that at present the solubility of the $CaSO_4$ phases with the AMOEBA force field is unknown as it was parametrized for the solution phase, and as such it is possible that it may underestimate the stability of the solids. Estimations of the relative stability of a 4 formula structural unit based on cluster I relative to that of a $CaSO_4$ ion pair using *ab initio* quantum mechanical methods suggest that both species are likely to be similar in free energy (see ESI).

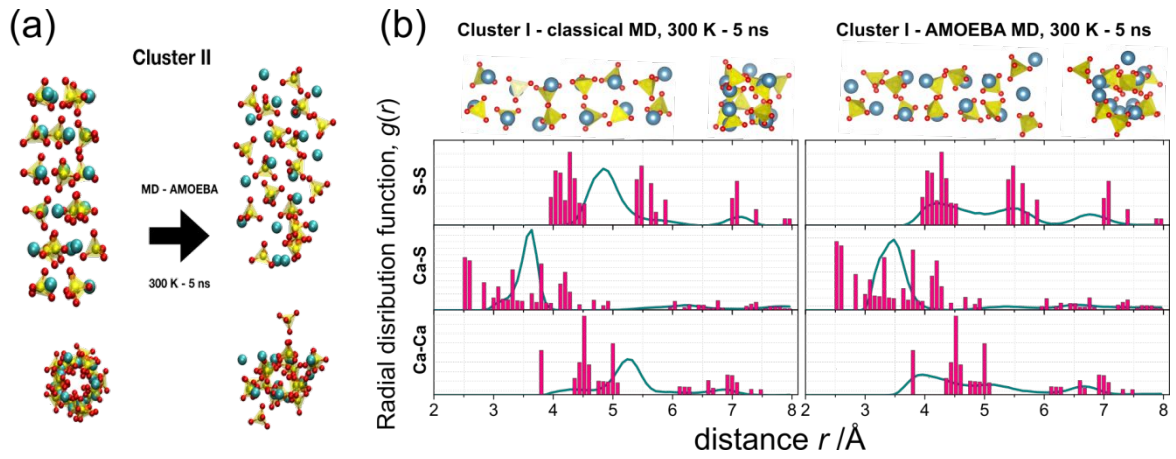

Fig. 4. a) Unstable cluster II after 5 ns of an MD simulation using the AMOEBA force field; the cluster shows significant disorder, but it essentially preserves its' original elongated shape; b) Comparison between the structures of cluster I as predicted from the rigid-ion classical (left) and AMOEBA polarizable (right) MD simulations. The models show the structures after 5 ns. The graphs show the radial distribution functions ($g(r)$s) computed from the whole MD trajectories (see ESI for the files) over the course of 5 ns (cyan lines) overlaid with those computed from the static structure of cluster I obtained from the experiments (magenta bars/histograms). The magenta $g(r)$ is generated from the optimized cluster I structure (see Fig. 2), hence the discrete histogram form.

Implications

Although our current diffraction and earlier scattering data[12,13], as well as the current MD simulations, clearly point to the presence of small $CaSO_4$ clusters as a precursor to gypsum and possibly other calcium sulfate crystalline phases[29], it is not clear what their role exactly is. It appears that they might be akin to prenucleation clusters reported for the calcium carbonate system[4,11], though our experimental approach considers in fact supersaturation

conditions. Regardless of the classification of the clusters (e.g. PNCs), their presence and evolution imply that the early stages of gypsum formation from solution are non-classical[1]. Important is also the fact that these primary clusters are numerous and persistent both in aqueous solution as well as later in the solid. Their presence along the crystalline gypsum is consistent with the surface structure factor theory of surface fractal aggregates that we developed recently[27], and the associated SAXS results which showed that the large aggregates of clusters had formed before any crystalline phases were present and persisted after gypsum crystallization[12]. Most significantly, we recently confirmed using dark field transmission electron microscopy the actual "brick-in-the-wall" (mesocrystalline) character of various natural and synthetic single $CaSO_4$ crystals[30]. These earlier results, together with the current data, strongly support our conclusion that gypsum crystallizes from the inside of large aggregates of structurally distinct, but flexible, precursor clusters.

Conclusions

In this work we used in situ high-energy X-ray diffraction experiments and molecular dynamics simulations to derive the structure of the precursor clusters of crystalline gypsum. We fitted several plausible cluster structures to the derived pair distribution functions (PDFs), and show that the proposed family of solutions has to fulfill three conditions: (1) the clusters are rod-shaped; (2) a single structural unit in the cross-section of the rod does not expand beyond a single Ca-Ca distance, with either of 2 or 3 Ca ions per unit; (3) the cluster structure are "internally" anhydrous and bound water is only present at their "surface". The dynamic

properties of the plausible structures derived from PDF analysis were tested using unbiased MD based on both rigid-ion and polarizable force-fields. Of the suite of clusters tested, the one with a simple 2-member square-shaped Ca-SO_4-Ca-SO_4 unit, where 6 stacks are arranged in an AB pattern and each Ca is coordinated by two in-plane water molecules**,** resulted to be the only structure stable throughout the whole MD simulation (5 ns). Although it was not possible to gauge the actual thermodynamic stability of these clusters, the dynamic models using the AMOEBA force-field strongly support the HEXD and previous SAXS data[12], showing that $CaSO_4$ forms rod-shaped structures, upon creation of a supersaturated solution, are thermodynamically stable/kinetically persistent in solution and play a crucial role in the formation pathway of gypsum.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research was made possible by a Marie Curie grant from the European Commission: the NanoSiAl Individual Fellowship, Project No. 703015 to TMS. We also acknowledge the financial support of the Helmholtz Recruiting Initiative grant No. I-044-16-01 to LGB. The authors would like to thank Diamond Light Source for beamtime (proposal EE11320), and the staff of beamlines I15, in particular Dr. Philip Chater for assistance with HEXD data collection. AESVD acknowledges financial supported by the French national programme EC2CO - Biohefect, SULFCYCLE. P.R. and J.D.G. thank the Australian Research Council

for funding through FT130100463 and FL180100087. The Australian National Computational Infrastructure (NCI) and the Pawsey Supercomputing Centre are acknowledged for the provision of computing resources through the NCMAS and Partners allocation schemes. E.H.B thanks the Department of Chemistry at Curtin for her PhD scholarship.

Electronic Supporting Information description

1. High-energy X-ray diffraction measurements and data analysis

2. Ab initio calculation of cluster stability.

3. Structural files of Ca-SO_4 clusters from Fig. 2 in the main text are included as *cluster_I.cif*, *cluster_II.cif*, *cluster_III.cif* and *cluster_IV.cif*.

4. Trajectory files from the MD simulations for cluster I from which Fig. 4B in the main text was derived are included as *cluster_I_classical_trj.xyz* and *cluster_I_amoeba_trj.xyz.*

References

(1) De Yoreo, J. J.; Gilbert, P. U. P. A.; Sommerdijk, N. A. J. M.; Penn, R. L.; Whitelam, S.; Joester, D.; Zhang, H.; Rimer, J. D.; Navrotsky, A.; Banfield, J. F.; et al. Crystallization by Particle Attachment in Synthetic, Biogenic, and Geologic Environments. *Science* **2015**, *349* (6247), aaa6760.

(2) Penn, R. L.; Li, D.; Soltis, J. A. A Perspective on the Particle-Based Crystal Growth of Ferric Oxides, Oxyhydroxides, and Hydrous Oxides. In *New Perspectives on Mineral Nucleation and Growth*; Springer International Publishing: Cham, 2017; pp 257–273.

(3) Karthika, S.; Radhakrishnan, T. K.; Kalaichelvi, P. A Review of Classical and Nonclassical Nucleation Theories. *Cryst. Growth Des.* **2016**, *16* (11), 6663–6681.

(4) Gebauer, D.; Völkel, A.; Cölfen, H. Stable Prenucleation Calcium Carbonate Clusters. *Science (80-.).* **2008**, *322* (5909), 1819–1822.

(5) Kellermeier, M.; Raiteri, P.; Berg, J.; Kempter, A.; Gale, J.; Gebauer, D. Entropy Drives Calcium Carbonate Ion Association. *ChemPhysChem* **2016**, *17* (21), 3535–3541.

(6) Raiteri, P.; Demichelis, R.; Gale, J. D.; Kellermeier, M.; Gebauer, D.; Quigley, D.; Wright, L. B.; Walsh, T. R. Exploring the Influence of Organic Species on Pre- and Post-Nucleation Calcium Carbonate. *Faraday Discuss.* **2012**, *159* (1), 61.

(7) Gebauer, D.; Kellermeier, M.; Gale, J. D.; Bergström, L.; Cölfen, H. Pre-Nucleation Clusters as Solute Precursors in Crystallisation. *Chem. Soc. Rev.* **2014**, *43* (7), 2348–2371.

(8) Smeets, P. J. M.; Finney, A. R.; Habraken, W. J. E. M.; Nudelman, F.; Friedrich, H.; Laven, J.; De Yoreo, J. J.; Rodger, P. M.; Sommerdijk, N. A. J. M. A Classical View on Nonclassical Nucleation. *Proc. Natl. Acad. Sci.* **2017**.

(9) *New Perspectives on Mineral Nucleation and Growth*; Van Driessche, A. E. S.,

Kellermeier, M., Benning, L. G., Gebauer, D., Eds.; Springer International Publishing: Cham, 2016.

(10) Stawski, T. M.; Benning, L. G. Chapter Five - SAXS in Inorganic and Bioinspired Research. In *Methods in Enzymology*; James, J. D. Y., Ed.; Academic Press, 2013; Vol. Volume 532, pp 95–127.

(11) Demichelis, R.; Raiteri, P.; Gale, J. D.; Quigley, D.; Gebauer, D. Stable Prenucleation Mineral Clusters Are Liquid-like Ionic Polymers. *Nat. Commun.* **2011**, *2* (590), 1–8.

(12) Stawski, T. M.; van Driessche, A. E. S.; Ossorio, M.; Diego Rodriguez-Blanco, J.; Besselink, R.; Benning, L. G. Formation of Calcium Sulfate through the Aggregation of Sub-3 Nanometre Primary Species. *Nat. Commun.* **2016**, *7*, 11177.

(13) Ossorio, M.; Stawski, T.; Rodríguez-Blanco, J.; Sleutel, M.; García-Ruiz, J.; Benning, L.; Van Driessche, A. Physicochemical and Additive Controls on the Multistep Precipitation Pathway of Gypsum. *Minerals* **2017**, *7* (8), 140.

(14) Parkhurst, D. L.; Appelo, C. A. J. *Description of Input and Examples for PHREEQC Version 3: A Computer Program for Speciation, Batch-Reaction, One-Dimensional Transport, and Inverse Geochemical Calculations*; US Geological Survey, 2013.

(15) Juhás, P.; Davis, T.; Farrow, C. L.; Billinge, S. J. L. PDFgetX3: A Rapid and Highly Automatable Program for Processing Powder Diffraction Data into Total Scattering Pair Distribution Functions. *J. Appl. Crystallogr.* **2013**, *46* (2), 560–566.

(16) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J.*

Comput. Phys. **1995**, *117* (1), 1–19.

(17) Byrne, E. H.; Raiteri, P.; Gale, J. D. Computational Insight into Calcium–Sulfate Ion Pair Formation. *J. Phys. Chem. C* **2017**, *121* (46), 25956–25966.

(18) Wu, Y.; Tepper, H. L.; Voth, G. A. Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties. *J. Chem. Phys.* **2006**, *124* (2), 024503.

(19) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.

(20) Lambrecht, D. S.; Clark, G. N. I.; Head-Gordon, T.; Head-Gordon, M. Exploring the Rich Energy Landscape of Sulfate-Water Clusters SO 42- (H 2O) N=3-7: An Electronic Structure Approach. *J. Phys. Chem. A* **2011**, *115* (41), 11438–11454.

(21) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLOS Comput. Biol.* **2017**, *13* (7), e1005659.

(22) Chow, K.-H.; Ferguson, D. M. Isothermal-Isobaric Molecular Dynamics Simulations with Monte Carlo Volume Sampling. *Comput. Phys. Commun.* **1995**, *91* (1–3), 283–289.

(23) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular Dynamics Simulations of Water and Biomolecules with a Monte Carlo Constant Pressure Algorithm. *Chem. Phys. Lett.* **2004**, *384* (4–6), 288–294.

(24) Juhás, P.; Farrow, C. L.; Yang, X.; Knox, K. R.; Billinge S. J. L.and IUCr, *Acta Crystallogr. Sect. A Found. Adv.*, 2015, **71**, 562–568.

(25) Guinier, A.; Fournet G. *Small angle scattering of X-rays*, 1955.

(26) Niederberger, M.; Cölfen, H. Oriented Attachment and Mesocrystals: Non-Classical Crystallization Mechanisms Based on Nanoparticle Assembly. *Phys. Chem. Chem. Phys.* **2006**, *8* (28), 3271–3287.

(27) Besselink, R.; Stawski, T. M.; Van Driessche, A. E. S.; Benning, L. G. Not Just Fractal Surfaces, but Surface Fractal Aggregates: Derivation of the Expression for the Structure Factor and Its Applications. *J. Chem. Phys.* **2016**, *145* (21), 211908.

(28) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562–12566.

(29) Van Driessche, A. E. S.; Stawski, T. M.; Benning, L. G.; Kellermeier, M. Calcium Sulfate Precipitation Throughout Its Phase Diagram. In *New Perspectives on Mineral Nucleation and Growth*; Springer International Publishing: Cham, 2017; pp 227–256.

(30) Stawski, T. M.; Freeman, H. M.; Van Driessche, A. E. S.; Hövelmann, J.; Besselink, R.; Wirth, R.; Benning, L. G. Particle-mediated nucleation pathways are imprinted in the internal structure of calcium sulfate single crystals http://arxiv.org/abs/1812.03022 (accessed Mar 25, 2019).

TOC Graphic