White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/215138/

Version: Accepted Version

Article:
Sandei, I., Gaule, T., Batchelor, M. et al. (2024) Phage Display Identifies Affimer Proteins that Direct Calcium Carbonate Polymorph Formation. Biomaterials Science, 12 (20). pp. 5215-5224. ISSN: 2047-4830

https://doi.org/10.1039/D4BM00165F

This is an author produced version of a journal article accepted for publication in Biomaterials Science, made available under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Phage Display Identifies Affimer Proteins that Direct Calcium Carbonate Polymorph Formation

Ilaria Sandei[a], Thembaninkosi Gaule[a,b], Matthew Batchelor[b], Emanuele Paci[b], Yi-Yeoun Kim[a], Alexander N. Kulak[a], Darren Tomlinson[b] and Fiona C. Meldrum[a]*

A key factor in biomineralization is the use of organic molecules to direct the formation of inorganic materials. However, identification of molecules that can selectively produce the calcium carbonate polymorphs calcite or aragonite has proven extremely challenging. Here, we use a phage display approach to identify proteins – rather than the short peptides typically identified using this method – that can direct calcium carbonate formation. A 1.3×10^{10} library of Affimer proteins was displayed on modified M13 phage, where an Affimer is a $\approx$ 13 kDa protein scaffold that displays two variable regions of 9–13 residues. The phage displaying the Affimer library were then screened in binding assays against calcite and aragonite at pH 7.4, and four different strongly-binding proteins were identified. The two aragonite-binding proteins generated aragonite when calcium and magnesium ions were present at a 1:1 ratio, while the calcite-binding proteins produce magnesium-calcite under the same conditions. Calcite alone formed in the presence of all four proteins in the absence of magnesium ions. In combination with molecular dynamics simulations to evaluate the conformations of the proteins in solution, this work demonstrates the importance of conformation in polymorph control, and highlights the importance of magnesium ions, which are abundant in seawater, to reduce the energetic barriers associated with aragonite formation.

Introduction

Evolution has enabled organisms to achieve astounding control over mineralization processes. This is seen in the formation of biominerals such as bones, seashells and fish-scales that are optimized for a wide range of functions including mechanical support, protection and even optical properties.[1] In forming these structures, nature achieves a degree of control over crystallization that surpasses that achieved in synthetic systems, such as the generation of crystals with complex morphologies or specific polymorphs.[2] This is principally achieved using organic molecules that can adopt roles ranging from soluble organic matrices to soluble

additives. There has therefore been significant interest in translating these strategies to synthetic systems.

The calcium carbonate system, which precipitates as three anhydrous polymorphs under ambient conditions, provides one of the best examples of control over crystal polymorph in nature. Of these phases, calcite is the thermodynamically stable form at room temperature, aragonite is only slightly less stable, and vaterite is a kinetic polymorph.[3, 4] However, while both calcite and vaterite readily precipitate under ambient conditions, few synthetic organic additives have been identified that promote aragonite, and it is notable that all contain basic moieties.[5-7] Yet, aragonite is a common biomineral, and some organisms such as mollusks can even switch between calcite and aragonite with perfect fidelity.[8]

Significant efforts have therefore been made to identify biomolecules that can select for calcite or aragonite, where biomolecules are extracted from aragonite and calcite biominerals and then employed as crystallization additives.[8-11] This strategy has provided significant insight into the mechanisms by which organisms control calcium carbonate biomineralization, where the macromolecules are often highly acidic. However, there are few reports of the production of aragonite using proteins or protein fragments extracted from aragonite biominerals[12] where most only generate aragonite when combined with more complex environments such as chitin substrates,[13, 14] an insoluble β-chitin/silk-fibroin matrix,[8] or magnesium ions.[15] Many questions therefore remain about the mechanisms by which organisms employ organic molecules to control polymorph formation.

One of the most effective approaches for the discovery of organic molecules that can direct the formation of inorganic materials is screening libraries of biomolecules. In most cases, this is conducted by expressing libraries of short peptides on phage or cells, and screening for their binding affinity to the target material. Application of this strategy to calcium carbonate has identified peptides that bind strongly to calcite, aragonite and vaterite, but which have only generated calcite or vaterite when they were employed in crystallization assays.[16-20] Due to their small size, peptides lack the conformational rigidity of proteins that can lead to enhanced specificity in many scenarios.

Here, we have used a phage display approach to screen libraries of proteins for their ability to direct the formation of calcite or aragonite, and thus explore the role of composition and conformational freedom on their activities. Affimer proteins were selected for this study, where these are small, stable recombinant proteins that possess two variable loops that can incorporate between 12 and 36 amino acids (Figure 1). The protein scaffold constrains the possible conformations that the peptides in the loops can adopt, thereby distinguishing our approach from screens of phage-display libraries of short peptides. Screening of phage-display libraries of Affimer proteins has demonstrated strong binding to multiple targets.[21-23]

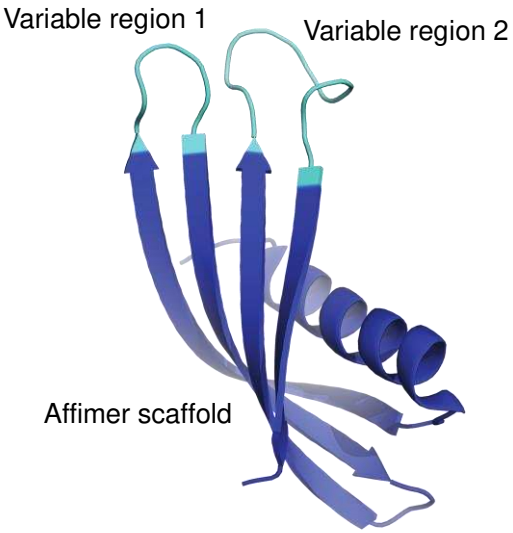

Figure 1. General representation of an Affimer protein. The two variable regions (or loops) are highlighted in cyan while the Affimer scaffold is highlighted in blue.

This was achieved by displaying Affimer proteins on the pIII minor coat protein of M13 phage, and screening for their ability to bind to calcite or aragonite. The most strongly-binding proteins are then expressed and employed in crystallization assays to determine their influence of calcium carbonate polymorph. Our experiments show that while polymorph selectivity is not achieved with the Affimer proteins alone, co-precipitation with low concentrations of magnesium ions enable the aragonite- and calcite-binding Affimer proteins to generate aragonite and calcite respectively. Magnesium ions are abundant in seawater[24] and are expected to act in combination with biomolecules to select polymorph in calcium carbonate biomineralization. Molecular dynamics (MD) simulations were used to evaluate the stability of the Affimer scaffold and the ensemble of conformations populated by the likely

flexible variable loops. These data provide important insights into the role of the conformations of organic additives in achieving control over calcite/aragonite polymorphism.

Results

Design of Experiments

Affimers are 12–15 kDa non-antibody synthetically engineered proteins that are based on the consensus sequence of cystatins (cysteine protease inhibitors). They exhibit monomeric structures and stability over a range of pH and temperature and do not possess glycosylation sites or disulfide bonds.[23, 25] Affimer proteins (originally called 'Adhirons') consist of two main regions: a scaffold comprising a 4-stranded antiparallel β-sheet and one α-helix, and two randomized peptide regions (often called loops or variable regions 1 and 2) of variable amino acid composition.[22, 23, 25] (Figure 1). The two variable regions are 9–13 amino acids in length and can bind to a target of interest, where their conformations are restricted by the protein scaffold. Randomization of the loop regions enables the construction of a large library (10^{10}) of Affimer variants that can be employed in a phage display process to select high-affinity binder sequences towards small targets.[21-23]

A library of M13 phage encoded Affimer proteins can be generated by inserting the gene for the Affimer into the pIII gene of the minor coat protein of M13 phage. Subsequent transformation of the recombinant phagemid into *E. coli* cells followed by infection with a helper phage enables the production of phage particles displaying a specific Affimer variant on the pIII coat protein, with the two variable regions free to interact with small particles or other targets. The library used in our work was estimated to comprise 1.3×10^{10} individuals that vary in the amino acid composition of the variable loop regions, and are displayed on the pIII coat protein of M13 bacteriophage.[21, 23, 25] Each phage therefore displays Affimer proteins that possess the same scaffold but differ in the compositions of the variable loops. The phagemid carries the genetic information to express that protein. Affimer proteins that bind strongly to calcite or aragonite are identified using a bio-panning process in which the Affimer phage library is incubated with calcite or aragonite crystals respectively (Figure 2). These "winning" proteins are sequenced to determine the composition of the variable regions. They

are also expressed and used in crystallization assays to determine their influence on the polymorph and morphology of calcium carbonate.

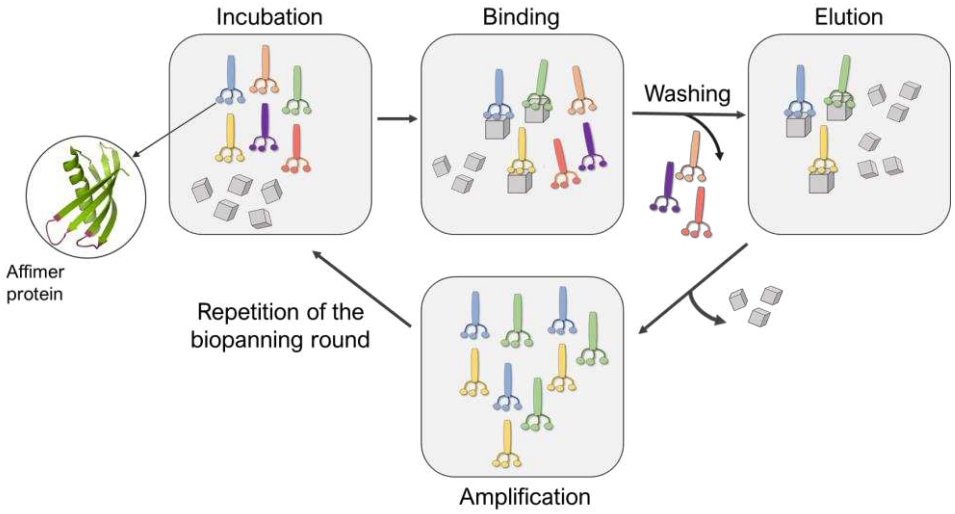

Figure 2. Schematic representation of the panning rounds performed against calcite and aragonite crystals.

Selection of Affimer Proteins

Affimer proteins that bind strongly to calcite or aragonite were identified by incubating the Affimer phage library with 70–160 nm calcite crystals synthesized using a carbonation method[26, 27] (Figure S1) or 50–100 µm long aragonite crystals synthesized at elevated temperatures[28] (Figure S2). Briefly, calcite or aragonite crystals were (i) incubated with the Affimer phage library in a buffer solution at pH 7.4 at room temperature for 1 hour, then (ii) washed to remove unbound proteins, and finally (iii) dissolved to release the strongly-bound proteins. A total of four bio-panning rounds were performed, in each case starting with the individuals eluted from the previous round. The proteins isolated from the fourth round were then amplified and analyzed. As the phage DNA encodes for the specific Affimer protein displayed on the virion surface, the Affimer proteins can be identified by sequencing the corresponding phagemid. The sequencing was achieved by Genewiz® through the Sanger method.

Sequences of Affimer Proteins that Bind to Calcium Carbonate

A total of 20 clones were selected after the fourth bio-panning rounds against both calcite and aragonite. The four most abundant clones were then selected for further investigation, where they were recombinantly expressed and tested in crystallization assays. DNA sequencing revealed that out of the 20 clones, two unique clones were selected that bound strongly to calcite (CBA-1 and CBA-2) and two that bound strongly to aragonite (ABA-1 and ABA-2). Although the Affimer phage library was designed to have two variable regions that are 9 amino acids long, CBA-2 has a 12 amino acid long variable region.

Affimer	Loop 1[a]	Loop 2[a]	Affimer pI[b]	Charge[c]	Hydrophilicity[d]
CBA-1	GKHKKYFMK	PFYSKLKTM	9.44	+6	+1.8
CBA-2	AHHRARIRFKVLK	QMKYMNSKI	9.63	+7	+5.4
ABA-1	KKAYMMLDR	ARFWKRYHR	9.52	+6	+7.6
ABA-2	ALFHRYRWV	KKSRMWHEW	9.30	+4	-2.8
Control Affimer	AAAA	AAE	6.75	-1	-3.9

[a] Amino acids are color coded on the basis of charge at pH 7.4. Positively charged residues shown in blue; negatively charged ones in red; hydrophobic residues in black and polar residues shown in green.[54]

[b] Theoretical pI of Affimer protein obtained using the ExPASy ProtParam tool.

[c] Charge was calculated by subtracting the number of basic residues from the number of acidic residues, taking into account protonation at pH 7.4.

[d] Hydrophilicity was calculated using the Hopp and Woods index,[29] where a positive value corresponds to hydrophilic and a negative value to hydrophobic.

Table 1. Table summarizing the characteristics of the selected calcite-binding Affimer proteins (CBA-1 and CBA-2) and aragonite-binding Affimer proteins (ABA-1 and ABA-2).

The distribution of amino acids across the variable regions of CBA-1 and CBA-2 were similar. Around 30% of the variable regions 1 and 2 comprise positively charged residues (R and K) which results in overall isoelectric points (pI) of 9.44 and 9.63 (Table 1) and an overall positive charge at pH 7.4. Whilst polar residues and hydrophobic residues make up 20% and 50% respectively of the variable regions, the Affimer proteins are expected to have high hydrophilicity (Table 1). Interestingly no negatively charged residues (D and E) were observed. The variable regions of the aragonite-binding Affimer proteins showed similar amino acid distributions to the calcium-binding Affimer proteins. However, the variable regions of both ABA-1 and ABA-2 do contain an acidic amino acid, namely D or E (Table 1). Although both Affimer proteins have a 50% distribution of hydrophobic residues within the variable regions and have an overall positive charge, ABA-2 is the most hydrophobic (Table 1) as it has more tryptophan residues.

Probing the Interactions of the Selected Affimer Proteins with Calcite and Aragonite

The specificity of the identified calcite- and aragonite-binding Affimer proteins for these crystal polymorphs was then investigated using a protein ELISA. A control Affimer protein was also generated to determine whether the scaffold itself binds to the calcium carbonate crystals. This has very short variable regions comprising four alanines in the first loop and two alanine and one glutamic acid in the second loop (Table 1). This test was conducted by exposing the proteins to equal amounts of the individual polymorphs for 1 hour in casein blocking buffer, and then washing the crystals to remove unbound proteins. After the addition of Anti-His antibodies that bind to the His-tag on the Affimer proteins, and subsequent addition of TMB (3,3',5,5'-Tetramethylbenzidine), the color of the solution changes to light blue if Affimer proteins are bound to the crystals. This was quantified by measuring the absorbance of the solution at 620 nm. No color change was observed after incubating calcite or aragonite with the Affimer control proteins, which demonstrates that binding of the proteins to the crystals is mediated through the variable loop regions.

The ELISAs of the ABA and CBA proteins confirmed that all of these proteins bind to the polymorphs used in the bio-panning rounds. However, while ABA-1 and ABA-2 only bind to aragonite and have no affinity for calcite, CBA-1 and CBA-2 bind to both polymorphs, albeit

with a slightly higher affinity for calcite. This was indicated by the lack of color in the solution of ABA-1 and ABA-2 incubated with calcite, and an intense blue color when ABA-1, ABA-2, CBA-1 or CBA-2 were incubated with calcite or aragonite targets (Figure 3).

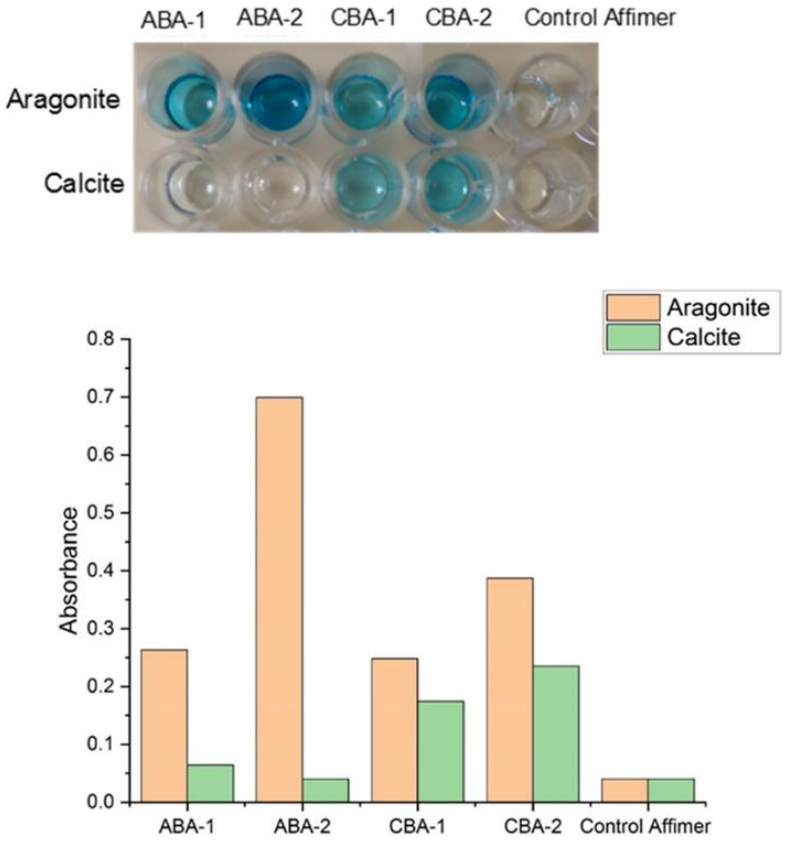

Figure 3. ELISA of the aragonite and calcite-binding Affimer protein and the control Affimer protein against calcite and aragonite, and absorption measurements taken at 620 nm. The blue color is indicative of binding. The ABA-1 and ABA-2 proteins only bind to aragonite, while the CBA-1 and CBA-2 bind to both aragonite and calcite, with a slightly higher affinity for calcite. The scaffold, as represented by the control Affimer protein, binds to neither polymorph, confirming that the binding of the proteins to the crystals is mediated through the loop regions.

Calcium Carbonate Precipitation in the Presence of the Selected Affimer Proteins

Crystallization experiments were then conducted to investigate the influence of the identified CBA and ABA proteins on calcium carbonate crystallization. The Affimer proteins were expressed as described previously.[23] They were then purified using Ni-NTA resin chromatography and buffer exchanged into H_2O to remove phosphate ions or buffers that

could affect the crystallization. SDS-PAGE analyses demonstrated that the selected Affimer proteins were expressed with good yields and purities (Figure S3). Calcium carbonate was precipitated in the presence of these Affimer proteins using the ammonia diffusion method (ADM),[30] in which solutions comprising [$CaCl_2$] = 2 mM and [Affimer] = 10 µM were exposed to ammonium carbonate vapor in a sealed chamber. Crystals formed on glass slides placed at the base of the reaction solution and were subsequently removed for characterization. Calcite formed in the presence of all four ABA and CBA proteins and the control Affimer protein (Figures 4a–f).

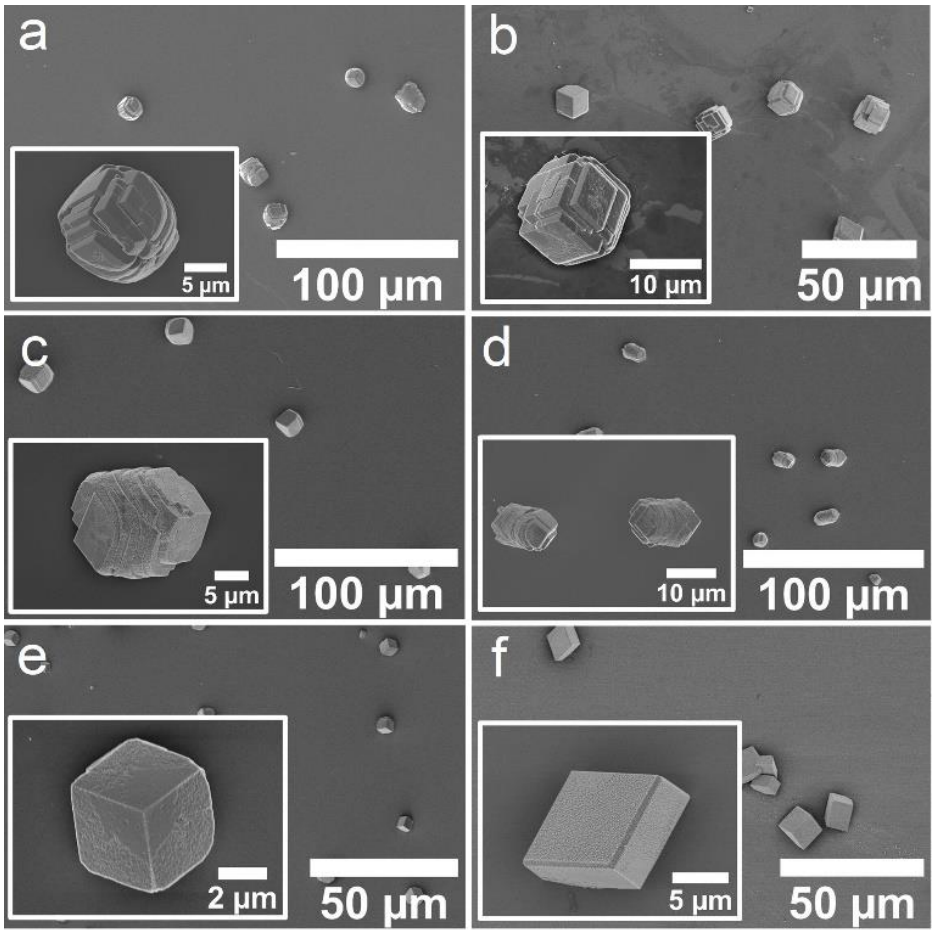

Figure 4. SEM micrographs of crystals obtained using the ADM method in the presence of (a) ABA-1, (b) ABA-2, (c) CBA-1, (d) CBA-2, (e) the control Affimer protein, and (f) in the absence of any protein.

Characterization of the crystals using scanning electron microscopy (SEM) showed some morphological differences; crystals formed with ABA-1 and ABA-2 were rhombohedral in form with blocky steps on the {104} faces (Figures 4a and 4b), while those grown in the presence

of CBA-1 and CBA-2 were elongated along the *c*-axis (Figure 4c and 4d), which is indicative of preferential binding of the proteins to the acute over the obtuse steps on the {104} faces.[26, 31] CBA-2 induced slightly greater elongation, and in both cases the equatorial zone featured macroscopic, blocky steps that derive from step bunching. Control experiments performed with the control Affimer protein yielded rhombohedral calcite crystals with slightly roughened surfaces (Figure 4e), while those conducted in the absence of protein yielded 10–20 μm rhombohedral calcite crystals with smooth faces (Figure 4f).

The influence of the Affimer proteins on calcium carbonate precipitation was also explored in the presence of magnesium ions, since magnesium ions are abundant in seawater[24] and are expected to play an important role in calcium carbonate biomineralization. Notably, magnesium ions are recognised to promote the formation of aragonite at room temperature in certain concentration regimes.[32-34] Experiments were conducted at $[CaCl_2] = [MgCl_2] = 2$ mM, where calcite forms in the absence of Affimer protein (Figure 5f).

Near identical magnesium-calcite crystals with peanut-like morphologies formed in the presence of both CBA proteins (Figures 5c and 5d), and the polymorph was confirmed by Raman spectroscopy (Figure 6). In contrast, the ABA proteins both yielded aragonite as the principal phase (Figures 5a and 5b), where the crystals were spherical (ABA-1) or resembled overgrown dumbbells and comprised bundles of needles (ABA-2). 60% aragonite and 40% Mg-calcite formed in the presence of ABA-1, while ABA-2 generated 90% aragonite and 10% Mg-calcite crystals. The Mg-calcite crystals formed in these experiments were lozenge-shaped.

Control experiments performed with the control Affimer protein yielded calcite crystals with blocky steps (Figure 5e), while those conducted in the absence of protein yielded near-spherical calcite crystals with a roughened surface (Figure 5f). Additional transmission electron microscopy (TEM) analyses were performed on the Mg-calcite and aragonite crystals after preparing thin sections using focused ion beam (FIB) milling. This demonstrated that the Mg-calcite were single crystals, and the aragonite was polycrystalline; polymorphs were confirmed using selected area electron diffraction (SAED) (Figure S4 and Figure S5).

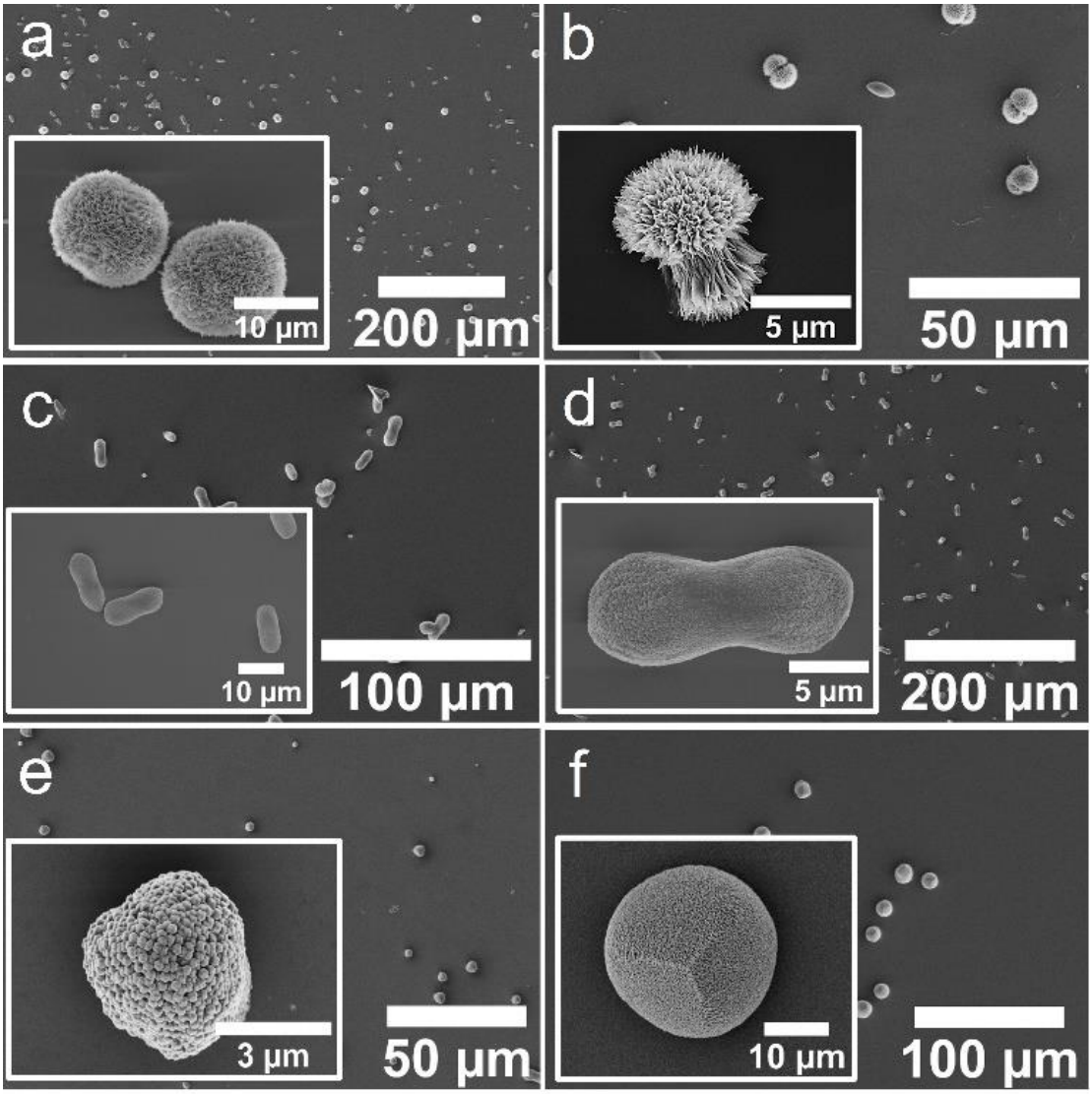

Figure 5. SEM micrographs of crystals obtained using the ADM method in the presence of Mg^{2+} ions at a Ca:Mg ratio 1:1 and (a) ABA-1, (b) ABA-2, (c) CBA-1, (d) CBA-2, (e) the control Affimer protein, and (f) in the absence of any protein. The images shown in the insets are (a and b) aragonite crystals and (c-f) calcite crystals.

Calcium Carbonate Precipitation in the Presence of Peptides

The Affimer proteins provide a scaffold that limits the conformational freedom of the variable loops. The role of the conformation of the loops in their ability to interact with calcium carbonate surfaces was explored by precipitating calcium carbonate in the presence of 10 µM peptides with the same compositions as the variable loops. The peptides investigated are summarized in Table S1. Calcite alone was obtained in the presence of all of the peptides tested, both in the presence and absence of Mg^{2+} ions, where representative data are given

in Figures S6, S7, S8 and S9 for peptides with the same sequences as loops 1 and 2 from CBA-1. Addition of Mg^{2+} ions resulted in somewhat less regular morphologies, and increased surface roughness (Figures S7 and S8). In the absence of Mg^{2+} ions, perfect calcite rhombohedra formed in the presence of loop 1 peptides, loop 2 peptides or both loop 1 and 2 peptides (Figure S8). These results suggest that the constrained conformation adopted by the peptides within the Affimer scaffold is crucial in their interaction with the crystals.

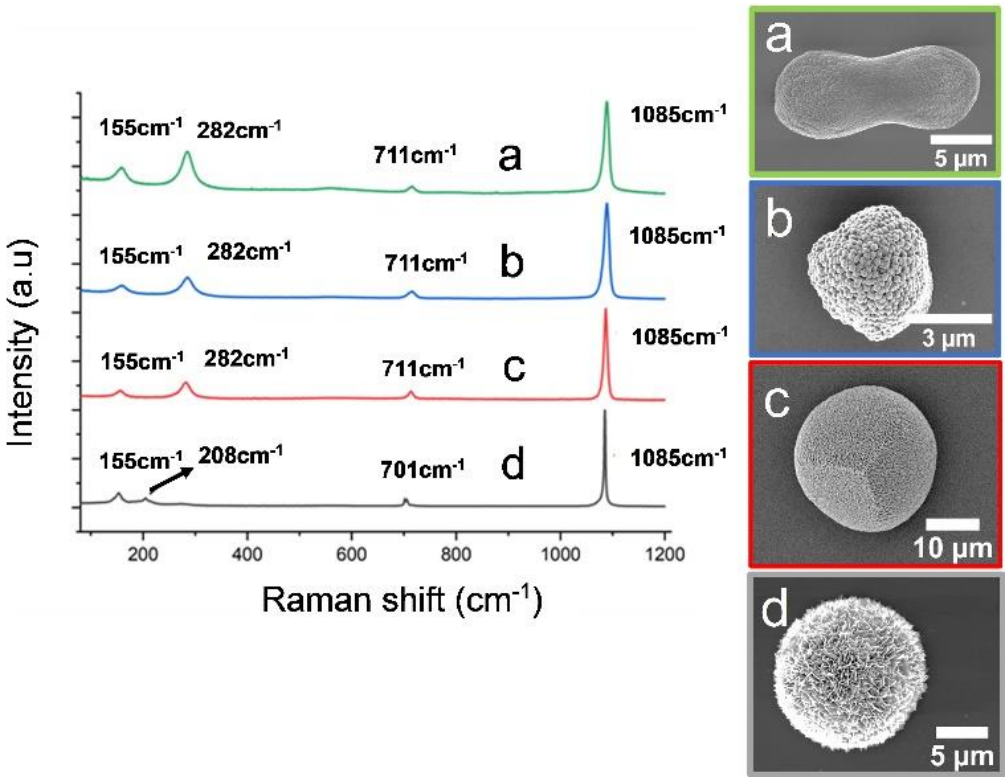

Figure 6. Raman spectra of crystals obtained from the different experiments conducted in the presence and absence of proteins: (a) calcite crystals obtained using ADM in the presence of Mg^{2+} ions at a Ca:Mg ratio 1:1 and CBA proteins, (b) calcite crystals obtained using ADM in the presence of Mg^{2+} ions at a Ca:Mg ratio 1:1 and the control Affimer protein, (c) calcite crystals obtained using ADM in the presence of Mg^{2+} ions at a Ca:Mg ratio 1:1 and in the absence of any protein, (d) aragonite crystals obtained using ADM in the presence of Mg^{2+} ions at a Ca:Mg ratio 1:1 and ABA proteins. All of them show the characteristic peak at 1085 cm^{-1} related to the symmetric stretching vibration of the carbonate group while the peak of the in-plane bending of the CO_3^{2-} group shifts from (a–c) 711 cm^{-1}, to (d) 701 cm^{-1}. For (d) the characteristic peak of calcite at 282 cm^{-1} is absent while we see the peaks at 155 cm^{-1} and 208 cm^{-1} related to the translation and libration modes of the carbonate ions in aragonite.

Molecular Dynamics (MD) Simulations

MD simulations were used to model the conformations adopted by the Affimer scaffold and the variable loops of CBA-1, ABA-1 and ABA-2 in solution environments. These three Affimer proteins have the same scaffold sequence and loops of the same length as an Affimer for which there is a published X-ray crystal structure (PDB code: 6YXW).[35] CBA-2 has longer loops and so was not modelled. Initial structures for simulation of CBA-1, ABA-1 and ABA-2 were built by mutating the loop residues of the 6YXW model to match the required sequence (Table 1). The simulations show that the scaffold is structurally stable in all cases. The secondary structure motifs are an α-helix (residues ≈ 8–22) and four β-strands (residues $\beta1$ ≈ 28–35, $\beta2$ ≈ 49–57, $\beta3$ ≈ 63–71 and $\beta4$ ≈ 84–91) and all were highly conserved in the simulations (Figure 7a). There were some differences in the observed length of the β strands, however, where loop 1 contributed to a longer pair of strands ($\beta1$ and $\beta2$) in ABA-2, while loop 2 contributed to a longer pair ($\beta3$ and $\beta4$) in ABA-1 and CBA-1.

The root mean square deviation (RMSD) describes the variation in the structure of the protein over time.[36] For the scaffold—*i.e*. all residues except the two variable loops (residues 39–47 and 73–81)—the RMSD values are low at 1–2 Å. This shows that the Affimer scaffolds have similar structures in solution to the 6YXW crystal structure (Figure 7b). As expected, the RMSD values for the variable loop regions (Figure 7c and 7d) are generally higher than for the scaffold, demonstrating some deviation from the loop structures of 6YXW (which has different loop sequences). RMSD values are particularly high for loop 1 in the CBA-1 Affimer. This reflects a change in structure as compared with the initial structure, and could result from the adoption of an alternative static structure and/or an increase in the dynamics or flexibility of the protein or protein region.

Root mean square fluctuation (RMSF) values describe, for each residue in the sequence, the average deviation of that residue away from its simulation average position. The RMSF therefore presents a simple metric for comparing more stable regions of a protein with areas that are more flexible or dynamic. As can be seen from the plots of RMSF for the three Affimer proteins (Figure 7e), the protein regions with secondary structure have low flexibility whereas the N-terminus and loops—and particularly the two variable loops—show higher variability.

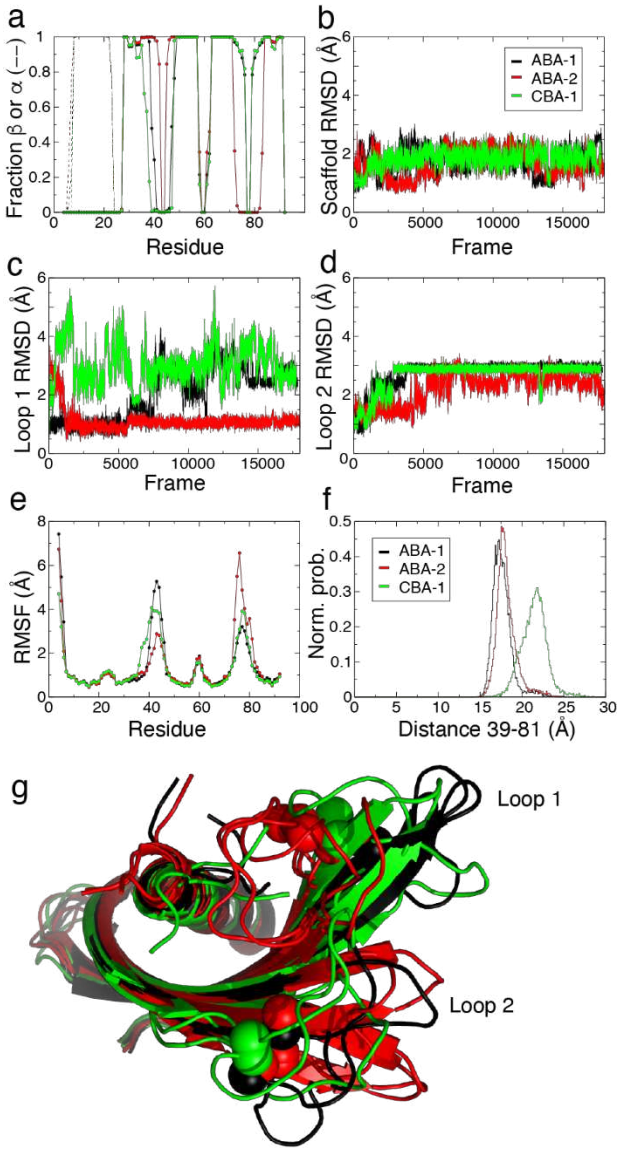

Figure 7. Molecular dynamics simulations of Affimers: ABA-1 (black), ABA-2 (red) and CBA-1 (green). (a) Secondary structure analysis on a per residue basis averaged over the simulation; dotted lines show the position of the α helix, solid lines and circles show the positions of the β strands. (b) RMSD of Cα atoms with respect to the initial structure of Affimer scaffolds (residues 4–38, 48–72, 82–92) over the course of the simulation. (c) Cα RMSD of loop 1 (residues 39–47). (d) Cα RMSD of loop 2 (residues 73–81). (e) RMSF values for the three simulations. (f) The distance between the outer edges of the two loops as gauged by the Cα to Cα distance for residues 39 and 81. (g) Representative structures from major clusters (>10% occupancy, accounting for >70% of the simulation in total) observed in the three simulations. Three major clusters were observed for ABA-1 (28, 26, 20%) and CBA-1 (49, 12, 11%), four major clusters were observed for ABA-2 (31, 26, 20, 14%).

Interestingly, ABA-1 and ABA-2 both have one very flexible loop and one more constrained loop, while CBA-1 has two intermediately flexible loops. The flexibility of loop 1 for CBA-1 also extends further towards the N-terminus than the aragonite-binding Affimers, reducing the effective length of the first β strand (see Figure 7a). As a result, the gap between the outer edges of the four strands making up the β sheet (or the edges or 'reach' of the two loops, as described by the distance between residues 39 and 81) is larger in the case of CBA-1 compared to the two aragonite-binding Affimers (Figure 7f).

Further insight can be gained by identifying similar conformations seen in the simulations (based on relative RMSD values) and comparing representative structures for each of these clusters for each of the different Affimer proteins. Equivalent cluster analyses gave a total of nine clusters for ABA-1 and ten clusters for both ABA-2 and CBA-1, and reversible interconversion between conformations belonging to different clusters was observed (Figure S10). Representative structures for the most highly populated clusters (observed for >10% of the simulation) are shown with the scaffolds overlaid in Figure 7g. The structures overlay very well away from the loop regions, although two different positions are observed for the β3/β4 hairpin for all Affimer proteins. This is seen most clearly for ABA-2 where the hairpin is longest and loop 2 shortest (Figure 7g). The representative cluster structures clearly show a range of preferred loop conformations, where loop 1 conformations are folded closer to the helix in ABA-2, and CBA-1 has loop 2 conformations closer to the helix. The simulations therefore reveal a range of conformational preferences around the variable loop regions for the three different Affimers, which although difficult to deconvolute, could go some way to explain preferred binding to the different calcium carbonate polymorphs.

Discussion

Phage display techniques have been widely used to identify peptides that bind strongly to biological and inorganic substrates.[17, 20, 21, 37-42] At the heart of this strategy is the ability to display, and then subsequently screen vast combinatorial libraries of peptides against the chosen target. Multiple screening rounds identify the peptide sequences that form the strongest interactions with the target, and the inherent link with the genotype provided by the phage enables the individual "winning" peptides to be sequenced and then expressed.

The effect of the specific sequences on mineral formation can then be studied in scaled-up experiments.

A number of previous studies have used phage display to identify peptides that bind strongly to the individual calcium carbonate polymorphs.[16, 17, 19, 20, 43] The protein sequences identified have many similarities to those found in the current study, in which basic and hydrophobic amino acids were prevalent. Schüler et al. identified two heptamers following three bio-panning rounds against vaterite crystals that only contained hydroxylic and hydrophobic residues, of which proline was the most abundant.[19] Vaterite formed in crystallization assays using one of these heptamers, while calcite formed when glycine had been substituted for proline in the same peptide. Gaskin et al. identified 10-mer peptides that bound strongly to calcite or aragonite and again showed that these were not rich in acidic amino acids, where they contained just one Glu or Asp each.[17] Li et al. screened a 15-mer peptide library against calcite crystals, and identified nine strongly-binding peptides that were typically rich in hydroxylated (Ser and Thr) and hydrophobic residues.[16] Gebauer et al. studied the influence of 12-mer peptides – identified as strong binders to geological aragonite – on calcium carbonate crystallization.[43] The three peptides studied were basic, and all inhibited calcite formation but did not generate aragonite. Finally, Völkle et al compared the effects of mixtures of proteins extracted from sea urchin skeletal plates with 12-mer peptides identified in binding assays of phage libraries to calcite and aragonite at pH 7.5.[20] Both sets of peptides exhibited a broad range of isoelectric points and were depleted in negatively-charged residues, while the aragonite-binders were enriched in lysine and histidine. However, calcite was the only polymorph produced in all crystallization experiments.

It is valuable to compare the sequences of these peptides with those of proteins found associated with calcite and aragonite biominerals. Proteins associated with the aragonite nacreous layer of mollusks can be acidic or basic. The acidic matrix protein Pif has been widely investigated,[9, 13, 14, 44] and the C-terminal (Pif80) and N-terminal regions (Pif97) appear to have different roles in aragonite formation.[9, 13, 44] Both regions are highly acidic with pI values of 4.99 and 4.65 respectively, but Pif80 has a higher percentage of charged residues and exhibits a highly repetitive four-residue motif (Asp-Asp-Arg(Lys)-Lys(Arg)) and a cluster of acidic amino acids (Asp-Glu-Asp).[13, 14] It has been suggested that Pif97 may bind to the chitin layer *in vivo*,

while Pif80 may be involved in controlling the nucleation and orientation of aragonite crystals.[44] Proteins that are rich in basic residues are also believed to contribute to aragonite formation *in vivo*: KRMP-3,[45] N40,[46] PfN23,[47] Prisilkin-39[48] and PNU9.[49] Proteins of the KRMP family include a lysine-rich basic region and a Gly/Tyr-rich region,[45] which is comparable to the Affimer proteins selected here for binding to aragonite.

Proteins found in the prismatic layer of mollusk shells are often rich in glycine and aspartic acid.[45, 50, 51] MSP-1,[51] aspein[50] and asprich[52] are examples of proteins with high aspartic acid contents, while MSI31,[53] MSP-1[51] and prismalin-14[54] are glycine-rich, featuring GGY domains. MSI31 has 10 poly(glycine) repeats of 3-5 residues and six acidic motifs principally comprising Glu residues.[55] The glycine-rich region may participate in the formation of β-sheet structures, while the acidic domain would enable binding of Ca^{2+} ions.[55] MSP-1, in contrast, is an acidic glycoprotein with -SG-, D and K rich domains.[51] Calprismin and Caspartin, two proteins extracted from the prismatic layer of *Pinna nobilis*, are mainly acidic, where Ala, Asx, Thr, and Pro dominate the amino acid composition of calprismin, and Asp is the main residue in Caspartin.[56] In contrast, the shell matrix proteins (shematrins) identified in the pearl oyster *Pinctada fucata* and associated with the formation of the calcitic prismatic layer possess repetitive basic RKKKY, RRKKY or RRRKY domains at their C-termini.[57]

These data highlight the challenge and complexity of exploring the link between protein/ peptide composition and calcium carbonate polymorph, where comparison of all of these studies suggests that there is no stand-out distinction between calcite- and aragonite-binding molecules. In an attempt to identify patterns, a recent study[58] used machine-learning to analyse the compositions of calcite and aragonite-binding 12-mers identified in the study by Völkle et al.[20] This revealed some differences in the residues located at positions 1–12 in the peptides, and suggested that most difference occurred in tiny, aliphatic, aromatic, acidic and basic residues.

This body of work, together with our observations that the compositions of the Affimer proteins selected in the calcite and aragonite binding assays are similar (where all principally comprise basic and hydrophobic residues) strongly suggests that the amino acid compositions of the peptides and proteins are not the sole determinant in biological control over calcium

carbonate polymorph. The main differences are the presence of a single acidic residue in the ABAs, and the principal amino acid in the ABAs is arginine rather than lysine in the CBAs. Modelling studies of the interactions between a calcite surface and the primary (-NH_2) amine in the Lys side chain and the secondary amine of the Arg side chain have shown that they interact strongly with the carbonate anions at all adsorption sites.[5] The amines are positively charged under the experimental conditions, and Lys binds more strongly than Arg. No comparable data exist for aragonite.

This indicates that the conformations and conformational freedom of the proteins are also crucial in determining their interactions with the crystal surface, and their ability to direct calcium carbonate crystallization. In this way, the hydrophobic residues, which do not bind directly to the crystal surface but do dictate the conformation of the proteins, play important roles in determining the protein behavior. Changes in the conformations of Affimers close to the variable loop regions were clearly shown here using MD simulations, and these differences will likely account for how the proteins interact with the crystal surface. This is also supported by the studies conducted with peptides of compositions identical to those of the loop regions, in which calcite formed rather than aragonite with ABA proteins with identical loop sequences. The constraints imposed on the peptides by the Affimer scaffold is therefore important in defining the interactions between the loops and the crystal surface.

Finally, our study highlights the likelihood that magnesium ions may act in combination with biomolecules to direct calcium carbonate crystallization. Few proteins or protein fragments[12] direct aragonite formation without combining them with more complex environments such as chitin substrates,[13, 14] an insoluble β-chitin/silk-fibroin matrix,[8] or magnesium ions.[15] The addition of magnesium ions is one of the most robust methods of generating aragonite under ambient conditions, where they inhibit the formation of calcite, allowing magnesian calcite and ultimately aragonite to form as the Mg/Ca ratio in solution increases.[33, 59, 60] Given the high concentration of magnesium ions in seawater, their combination with biomacromolecules could offer a straightforward means of tuning crystal polymorph *in vivo*.[5, 61]

Conclusions

In summary, this article demonstrates the use of phage display to screen a combinatorial library of Affimer proteins and identify individual proteins that bind strongly to calcite and aragonite. As compared to randomized peptides that have previously been selected as crystallization additives using phage display,[16, 17, 19] proteins offer the possibility of constraining the conformations of the variable loops, and provide easier genetic manipulation of the phagemid for protein production.[21, 23] Affimer proteins are particularly well suited to this approach, since they are stable under a wide range of conditions.[22, 23, 25] While polymorph selectivity was not achieved in the presence of the proteins alone, the aragonite-binding Affimer proteins generated aragonite when low concentrations of magnesium ions were present, and the calcite-binding proteins generate calcite. Notably, peptides with sequences corresponding to the variable loops of the aragonite-binding proteins only formed calcite. The conformations of Affimer proteins were shown to differ close to the variable loop regions in dynamic molecular models of the aragonite- and calcite-binding proteins. These results demonstrate that the conformation of the organic additives can play a significant role in determining their interaction with growing crystals, and also highlight the value of additives acting in combination with magnesium ions to direct calcium carbonate formation *in vivo*. Indeed, lower concentrations of magnesium ions are required to generate aragonite when these organic molecules are present, giving the advantage of faster crystallization rates than are achieved at high Mg/Ca ratios. Further work will use simulations to study the interactions of the Affimer proteins direct with calcite and aragonite surfaces.

Author Contributions

IS carried out the majority of the experimental work and characterisation. TG assisted and provided guidance with the molecular biology studies, while YYK provided guidance with the crystallisation work. AK assisted with SEM. MB and EP led the simulations studies, with assistance from IS. FCM and DT supervised the project. The manuscript was written through contributions of all authors and all authors have given approval to the final version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the European Research Council (ERC) for funding an Advanced Grant for the DYNAMIN project, grant agreement No. 788968. We are grateful for access to Raman spectroscopy through the Flow-Xl facility (EPSRC grant EP/T006331/1). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising.

References

1. H. A. Lowenstam and S. Weiner, *On Biomineralization*, Oxford University Press, New York, 1989.
2. F. C. Meldrum and H. Colfen, *Chem. Rev.*, 2008, **108**, 4332-4432.
3. F. Lippmann, *Sedimentary carbonate materials* Springer-Verlag, Berlin, 1973.
4. J. W. Morse, R. S. Arvidson and A. Luttge, *Chem. Rev.*, 2007, **107**, 342-381.
5. O. Nahi, A. N. Kulak, S. Zhang, X. He, Z. Aslam, M. A. Ilett, I. J. Ford, R. Darkins and F. C. Meldrum, *Adv. Sci.*, 2022, e2203759.
6. N. Nassif, N. Gehrke, N. Pinna, N. Shirshova, K. Tauer, M. Antonietti and H. Colfen, *Angew. Chem. Int. Ed.*, 2005, **44**, 6004-6009.
7. A. W. Xu, W. F. Dong, M. Antonietti and H. Colfen, *Adv. Func. Mater.*, 2008, **18**, 1307-1313.
8. G. Falini, S. Albeck, S. Weiner and L. Addadi, *Science*, 1996, **271**, 67-69.
9. M. Suzuki, K. Saruwatari, T. Kogure, Y. Yamamoto, T. Nishimura, T. Kato and H. Nagasawa, *Science*, 2009, **325**, 1388-1390.
10. T. Takeuchi, I. Sarashina, M. Iijima and K. Endo, *FEBS Lett.*, 2008, **582**, 591-596.
11. A. M. Belcher, X. H. Wu, R. J. Christensen, P. K. Hansma, G. D. Stucky and D. E. Morse, *Nature*, 1996, **381**, 56-58.
12. R. A. Metzler, J. S. Evans, C. E. Killian, D. Zhou, T. H. Churchill, N. P. Appathurai, S. N. Coppersmith and P. Gilbert, *J. Am. Chem. Soc.*, 2010, **132**, 6329-6334.
13. S. Y. Bahn, B. H. Jo, Y. S. Choi and H. J. Cha, *Sci. Adv.*, 2017, **3**, e1700765.
14. Y. P. Du, H. H. Chang, S. Y. Yang, S. J. Huang, Y. J. Tsai, J. J. T. Huang and J. C. C. Chan, *Sci. Rep.*, 2016, **6**, 30883.
15. C. Rivera-Perez, I. A. Flores-Sanchez, J. de Areyano, D. I. R. Posadas and N. Y. Hernandez-Saavedra, *Sci. Rep.*, 2020, **10**, 20201.
16. C. M. Li, G. D. Botsaris and D. L. Kaplan, *Cryst. Growth Des.*, 2002, **2**, 387-393.
17. D. J. H. Gaskin, K. Starck and E. N. Vulfson, *Biotechnol. Letts.*, 2000, **22**, 1211-1216.
18. M. L. Lemloh, K. Altintoprak, C. Wege, I. M. Weiss and D. Rothenstein, *Materials*, 2017, **10**, 119.
19. T. Schuler, J. Renkel, S. Hobe, M. Susewind, D. E. Jacob, M. Panthofer, A. Hoffmann-Roder, H. Paulsen and W. Tremel, *J. Mater. Chem. B*, 2014, **2**, 3511-3518.

20. E. Völkle, F. Schulz, J. M. Kanold, M. Michaelis, K. Wissel, F. Brümmer, A. S. Schenk, S. Ludwigs, J. Bill and D. Rothenstein, *J. Mater. Chem. B*, 2023, **11**, 10174-10188.
21. A. E. Rawlings, J. P. Bramble, A. A. S. Tang, L. A. Somner, A. E. Monnington, D. J. Cooke, M. J. McPherson, D. C. Tomlinson and S. S. Staniland, *Chem. Sci.*, 2015, **6**, 5586-5594.
22. C. Xie, C. Tiede, X. Zhang, C. Wang, Z. Li, X. Xu, M. J. McPherson, D. C. Tomlinson and W. Xu, *Sci. Rep.*, 2017, **7**, 9608.
23. C. Tiede, A. A. S. Tang, S. E. Deacon, U. Mandal, J. E. Nettleship, R. L. Owen, S. E. George, D. J. Harrison, R. J. Owens, D. C. Tomlinson and M. J. McPherson, *Protein Eng. Des. Sel.*, 2014, **27**, 145-155.
24. J. B. Ries, *Biogeosci. Discuss.*, 2009, **6**, 7325–7452.
25. C. Tiede, R. Bedford, S. J. Heseltine, G. Smith, I. Wijetunga, R. Ross, D. Al Qallaf, A. P. E. Roberts, A. Balls, A. Curd, R. E. Hughes, H. Martin, S. R. Needham, L. C. Zanetti-Domingues, Y. Sadigh, T. P. Peacock, A. A. Tang, N. Gibson, H. Kyle, G. W. Platt, N. Ingram, T. Taylor, L. P. Coletta, I. Manfield, M. Knowles, S. Bell, F. Esteves, A. Maqbool, R. K. Prasad, M. Drinkhill, R. S. Bon, V. Patel, S. A. Goodchild, M. Martin-Fernandez, R. J. Owens, J. E. Nettleship, M. E. Webb, M. Harrison, J. D. Lippiat, S. Ponnambalam, M. Peckham, A. Smith, P. K. Ferrigno, M. Johnson, M. J. McPherson and D. C. Tomlinson, *Elife*, 2017, **6**.
26. D. C. Green, J. Ihli, P. D. Thornton, M. A. Holden, B. Marzec, Y. Y. Kim, A. N. Kulak, M. A. Levenstein, C. Tang, C. Lynch, S. E. D. Webb, C. J. Tynan and F. C. Meldrum, *Nat. Commun.*, 2016, **7**, 13524.
27. J. Ihli, W. C. Wong, E. H. Noel, Y. Y. Kim, A. N. Kulak, H. K. Christenson, M. J. Duer and F. C. Meldrum, *Nat. Commun.*, 2014, **5**, 3169.
28. L. F. Wang, I. Sondi and E. Matijevic, *J. Coll. Int. Sci.*, 1999, **218**, 545-553.
29. T. P. Hopp and K. R. Woods, *Proc. Natl. Acad. Sci. USA*, 1981, **78**, 3824-3828.
30. J. Ihli, P. Bots, A. Kulak, L. G. Benning and F. C. Meldrum, *Adv. Func. Mater.*, 2013, **23**, 1965-1973.
31. C. A. Orme, A. Noy, A. Wierzbicki, M. T. McBride, M. Grantham, H. H. Teng, P. M. Dove and J. J. DeYoreo, *Nature*, 2001, **411**, 775-779.
32. J. Lee and J. W. Morse, *Geol.*, 2010, **38**, 115-118.
33. J. W. Morse, Q. W. Wang and M. Y. Tsio, *Geol.*, 1997, **25**, 85-87.

34. D. Wang, L. M. Hamm, A. J. Giuffre, T. Echigo, J. D. Rimstidt, J. J. De Yoreo, J. Grotzinger and P. M. Dove, *Faraday Disc.*, 2012, **159**, 371-386.
35. K. Z. Haza, H. L. Martin, A. Rao, A. L. Turner, S. E. Saunders, B. Petersen, C. Tiede, K. Tipping, A. A. Tang, M. Ajayi, T. Taylor, M. Harvey, K. M. Fishwick, T. L. Adams, T. G. Gaule, C. H. Trinh, M. Johnson, A. L. Breeze, T. A. Edwards, M. J. McPherson and D. C. Tomlinson, *Nat. Commun.*, 2021, **12**, 4045.
36. L. Martinez, *Plos One*, 2015, **10**, e0119264.
37. G. P. Smith and V. A. Petrenko, *Chem. Rev.*, 1997, **97**, 391-410.
38. J. X. Huang, S. L. Bishop-Hurley and M. A. Cooper, *Antimicrob. Agents Chemother.*, 2012, **56**, 4569-4582.
39. A. Artzy-Schnirman, E. Abu-Shah, M. Dishon, H. Soifer, Y. Sivan, Y. Reiter, I. Benhar and U. Sivan, *J. Peptide Sci.*, 2014, **20**, 446-450.
40. R. H. Hoess, *Chem. Rev.*, 2001, **101**, 3205-3218.
41. G. P. Smith, *Science*, 1985, **228**, 1315-1317.
42. P. Golec, J. Karczewska-Golec, M. Los and G. Wegrzyn, *J. Nanopart. Res.*, 2012, **14**, 1218.
43. D. Gebauer, A. Verch, H. G. Borner and H. Colfen, *Cryst. Growth Des.*, 2009, **9**, 2398-2403.
44. S. Y. Bahn, B. H. Jo, B. H. Hwang, Y. S. Choi and H. J. Cha, *Cryst. Growth Des.*, 2015, **15**, 3666-3673.
45. J. Liang, G. R. Xu, J. Xie, I. Lee, L. Xiang, H. Z. Wang, G. Y. Zhang, L. P. Xie and R. Q. Zhang, *Plos One*, 2015, **10**, e0131868.
46. Z. G. Yan, G. Jing, N. P. Gong, C. Z. Li, Y. J. Zhou, L. P. Xie and R. Q. Zhang, *Biomacromol.*, 2007, **8**, 3597-3601.
47. D. Fang, C. Pan, H. J. Lin, Y. Lin, G. Y. Zhang, H. Z. Wang, M. X. He, L. P. Xie and R. Q. Zhang, *J. Biol. Chem.*, 2012, **287**, 15776-15785.
48. Y. W. Kong, G. Jing, Z. G. Yan, C. Z. Li, N. P. Gong, F. J. Zhu, D. X. Li, Y. R. Zhang, G. L. Zheng, H. Z. Wang, L. P. Xie and R. Q. Zhang, *J. Biol. Chem.*, 2009, **284**, 10841-10854.
49. J. J. Kong, C. Liu, D. Yang, Y. Yan, Y. Chen, Y. J. Liu, G. L. Zheng, L. P. Xie and R. Q. Zhang, *Crystengcomm*, 2019, **21**, 1250-1261.
50. D. Tsukamoto, I. Sarashina and K. Endo, *Biochem. Biophys. Res. Commun.*, 2004, **320**, 1175-1180.

51. I. Sarashina and K. Endo, *Marine Biotechnol.*, 2001, **3**, 362-369.
52. B. A. Gotliv, N. Kessler, J. L. Sumerel, D. E. Morse, N. Tuross, L. Addadi and S. Weiner, *Chembiochem*, 2005, **6**, 304-314.
53. Y. Zhang, L. P. Xie, Q. X. Meng, T. M. Jiang, R. L. Pu, L. Chen and R. Q. Zhang, *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.*, 2003, **135**, 565-573.
54. M. Suzuki, E. Murayama, H. Inoue, N. Ozaki, H. Tohse, T. Kogure and H. Nagasawa, *Biochem. J.*, 2004, **382**, 205-213.
55. S. Sudo, T. Fujikawa, T. Nagakura, T. Ohkubo, K. Sakaguchi, M. Tanaka, K. Nakashima and T. Takahashi, *Nature*, 1997, **387**, 563-564.
56. C. Zhang and R. Q. Zhang, *Marine Biotechnol.*, 2006, **8**, 572-586.
57. M. Yano, K. Nagai, K. Morimoto and H. Miyamoto, *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.*, 2006, **144**, 254-262.
58. A. L. S. Nidoy and J. I. B. Janairo, *Chem. Lett.*, 2024, **53**.
59. M. M. Reddy and K. K. Wang, *J. Cryst. Growth*, 1980, **50**, 470-480.
60. S. H. Zhang, O. Nahi, L. Chen, Z. Aslam, N. Kapur, Y. Y. Kim and F. C. Meldrum, *Adv. Func. Mater.*, 2022, **32**, 2201394.
61. S. M. Stanley, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 2006, **232**, 214-236.