Supplementary Information

On Biomineralization: Enzymes Switch On Mesocrystal Assembly

Ashit Rao, Teresa Roncal-Herrero, Elina Schmid, Markus Drechsler, Martin Scheffner, Denis Gebauer, Roland Kröger, Helmut Cölfen

Contents

Methods

Protein Expression and Purification

Preparation of constructs and protein purification is done by following previous protocols[1]. Full length SM50 and also the CTLD and GRR domains are intracellularly expressed as 6XHis-SUMO-fusion products in *Escherichia coli* BL21 CodonPlusRIL (Stratagene, Agilent Technologies). After cell lysis, the CTLD was purified from the soluble fraction by Ni^{2+}-NTA affinity chromatography. The full-length protein and GRR domains are purified from the respective urea (6 M)-solubilized cell pellet fractions. The purified proteins are profusely dialyzed against either milliQ water, HEPES buffer (10 mM) supplemented with sodium chloride (20 mM), or carbonate buffer (10 mM). Final protein contents are estimated by using a Bradford assay in a microtitre plate format. All protein additives present an N-terminal SUMO tag to aid protein expression. In absence of the tag, protein toxicity lead to poor protein yields.

Mineralization/Demineralization experiments

Potentiometric titrations and corresponding data analyses are performed by following established protocols[2]. By using a software-controlled titration system (Metrohm GmbH), calcium chloride (10 mM) is dosed at a constant rate of 0.01 mL/min to a carbonate buffer (10 mM) containing an additive (at either 0.1 or 1 mg/ml). During the measurements, the pH and free Ca^{2+} contents are monitored by utilizing a flat-membrane glass electrode and an ion-selective electrode, respectively. A constant pH is maintained by the automatic counter-titration of sodium hydroxide (10 mM). Corresponding reference and calibration experiments are also performed.

Spines of adult *Strongylocentrotus purpuratus* specimens (MBL, University of Chicago) are cleaned with cold hypochlorite (2%), followed by several rinses with DI water. The biomineral is demineralized for 1 h in acetic acid (10%) solution. The reaction is quenched using Tris buffer (500 mM, pH 7.8) followed by prolonged dialysis of the mixture against Tris buffer. The resultant mineral structures are investigated (Fig. 1).

Biochemical Studies

CD spectra are measured by using a spectropolarimeter (Jasco 715) with a Peltier temperature control unit (Jasco Hachioji, Tokyo, Japan) and quartz cuvettes. The spectra are recorded at 0.5 nm intervals between 190 and 240 nm at 20°C at a 20 nm/s scan rate. Secondary structure

compositions were predicted from the spectra by using K2D3. Western blots utilize rabbit derived primary antibodies in form of either an anti-SM50 antiserum[3] or a CTL-specific serum fraction produced by reverse immunoaffinity chromatography. Commercial Type VI collagenase (Worthington Biochemicals) is used for biochemical assays.

Solution and Structural Studies

Sedimentation velocity experiments utilize a XL-I ultracentrifuge (Beckman Coulter Inc.) operated in absorbance mode. Collected at 25°C, the data are analysed using the program SEDFIT[4] (version 9.4) to generate sedimentation coefficient and molar mass distributions. Double sector titanium centrepieces were filled with protein solutions (absorbance of 1 OD at 280 nm) and corresponding buffer references. For determining frictional coefficients (f/f_o), 2-dimensional spectrum analysis (2-DSA, Monte Carlo) are performed with 20 iterations and a grid resolution of 3600 by using Ultrascan III[5]. The axial ratios for given frictional ratios are also simulated. Density gradient experiments are performed in sucrose gradients. The nucleated mineral particles are dispersed in sucrose solutions (63%, density ϱ= 1.304 g/cc) and then fractionated in an AUC operated for 16 h until equilibrium was reached, with absorption scans at 280 nm.

Mineral products are analyzed using a Zeiss Axio Imager-M2m polarization microscope in transmission mode. Quantitative birefringence microscopy is also performed (Abrio CRi). Cryo-TEM is performed by blotting a drop onto a lacey carbon film covered Cu grid and vitrified by plunging into a liquid ethane bath in a temperature controlled cryo-box (Zeiss Microscopy, Jena, Germany). The vitrified specimens are examined in a Zeiss/LEO EM922Omega EFTEM instrument (Zeiss Microscopy, Jena, Germany). Images are recorded with a slow scan CCD camera (Ultrascan 1000, Gatan, Munich, Germany) using a software package (Digital Micrograph, Gatan, Munich, Germany). TEM imaging utilizes a Libra 120 (Zeiss SMT GmbH) instrument operated at 120 kV with a beam current of 4 μA. An atomic force microscope (AFM, Nanowizard, JPK Instruments) with silicon nitride cantilevers is operated in scanning mode to retrieve height and phase images. Liquid cell scanning transmission electron microscopy (STEM) experiments utilize a Poseidon 200 liquid flow cell holder (Protochips Inc.) and an aberration corrected Jeol 2200 FS. The microscope is operated at 200kV using a high-angle annular dark-field scanning (HAADF-STEM) detector. Pictures were recorded using a K2camera of a physical resolution of 1024 x 1024 pixels^2 at dwell time of 18 μs per pixel resulting in a scan time of 1.5 s per frame giving an electron dose of 8.8×10^5 e/nm^2. Videos were acquired using the Camtasia software allowing for the capture of live-

view images using Digital Micrograph (JEM 2200). Capture were down-sampled to 512 x512 pixels^2 and stored at 2 fps. Mineralization conditions are described in section S9C. Thermogravimetric analyses (Netzsch, Selb, Germany) are performed by heating from 293 to 1273 K at a heating rate of 5 K/min under a constant oxygen flow. Thickness of protein films are estimated by using an EP^{3}SW ellipsometer (Nanofilm Technologies, Gottingen, Germany). Films are prepared on Si100 substrates by drop casting, followed by water rinses. SiO_2 thickness is measured using blank Si100 substrates. Ψ and Δ values are estimated from the change in intensity and phase of the incident and reflected light at different angles of incidence.

Liquid cell STEM of mineralization in the presence of biomineralization proteins.

For the protein-controlled mineralization reactions, the sample observation was carried out through two 500 nm thick silicon nitride membranes with a total accessible area of 550x50 μm with a nominal thickness 500 μm. 10 nm axial length gold nanorods were added to the top membrane to have a marker. The two solutions used are: (1) a mixture of sodium bicarbonate (9 mM), sodium carbonate (1 mM) and the protein additive (1 mg/ml) and (2) calcium chloride (10 mM). A droplet (3-5 μL) of solution 1 was sandwiched onto the surface of the top silicon nitride membrane and the holder was assembled. Then, the two different solutions were continuously pumped through the cell by a syringe pump (Harvard Apparatus Inc.) at a flow rate varying from 10 to 300 μL/h. In presence of the proteins, the dynamics of mineralization are retarded possibly due confinement effects[6] and representative snapshots of the mineralization reactions are shown (Fig. 2e). Vesicles formed during CTLD-regulated mineralization are depicted in Supplementary Video 1. As considered in Supplementary Section 6, the vesicles present homogenous core compositions. For SM50-controlled mineralization, the two solutions used are: (1) a mixture of sodium bicarbonate (9 mM), sodium carbonate (1 mM) and the protein additive (1 mg/ml) and (2) calcium chloride (10 mM) and the protein additive the calcium chloride solutions are supplemented with Type VI collagenase (10 μg/ml).

Supplementary Sections

1. Sequence properties and characterization of SM50

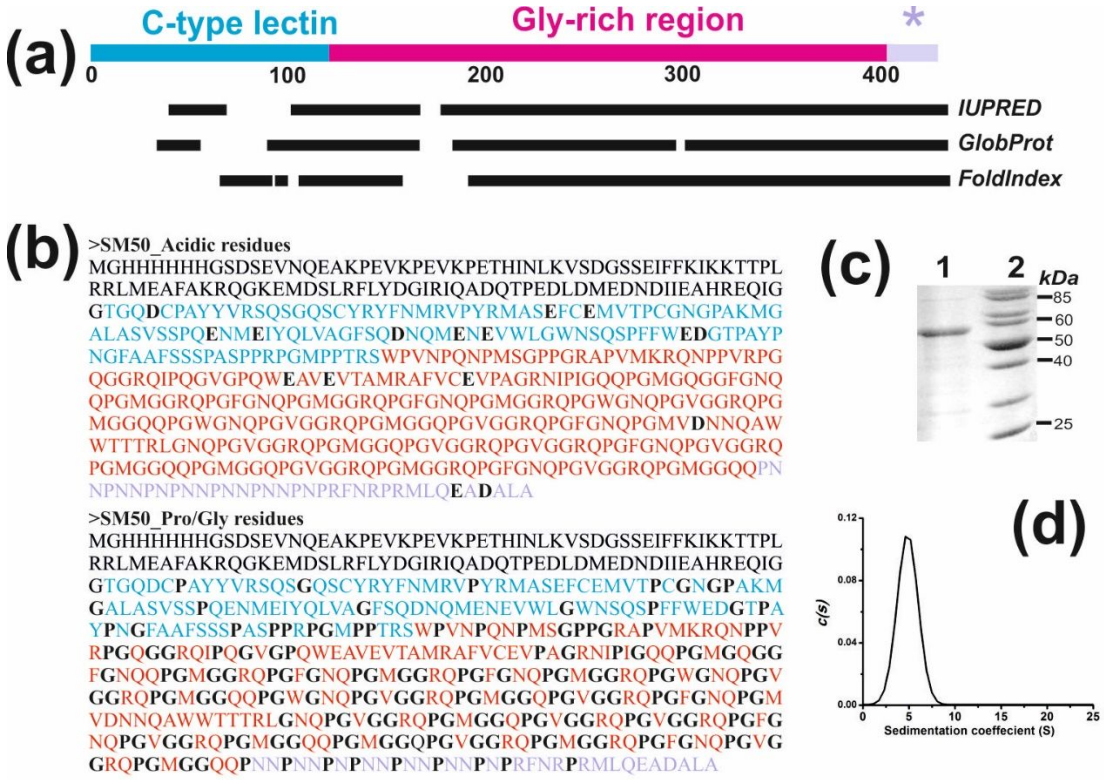


Figure S1. (a) Schematic representation of SM50 domains (without the secretion signal) consisting of the C-type lectin domain (CTLD, 1–130, blue), the glycine rich region (GRR, 131–402, pink) and a short proline rich region (PRR, 403–432, purple)[1]. Regions with high structural disorder propensities are highlighted (black bars), as predicted using IUPRED, GlobProt and FoldIndex algorithms[7-9]. (b) Shown with the N-terminal SUMO tag (normal black font), the primary sequence of SM50 highlights the scarcity of acidic residues (aspartate and glutamate) and a compositional enrichment of glycine and proline residues represented in bold font. (c) Representative SDS–PAGE profile of purified SM50 (*lane 1*) and standard molecular weight marker (*lane 2*). (d) SEDFIT[4]-derived continuous sedimentation coefficient distribution from velocity experiments of the recombinant SUMO tagged full-length SM50 presenting a 4.9±0.8 S value (molar mass of 70±20 kDa) for the monomeric species.

2. Structural characterization of the SUMO-CTLD

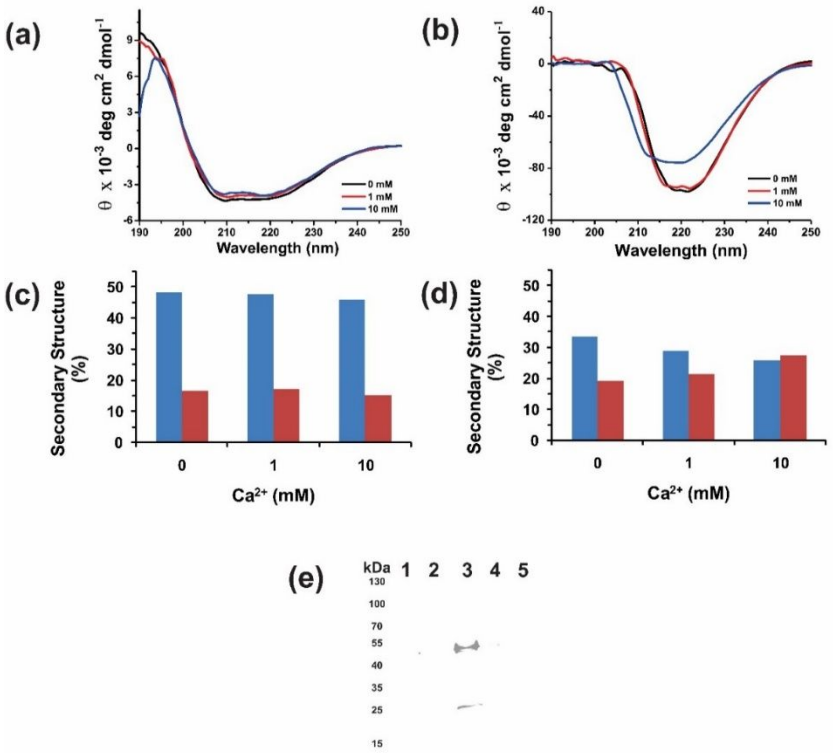


Figure S2. Representative CD spectra of the CTLD fusion in carbonate buffer at pH 9.0 in the (a) absence and (b) presence of β-mercaptoethanol with different contents of Ca^{2+} ions. (c, d) Corresponding secondary structure contents for the respective CD spectra (a) and (b) estimated by using K2D3[10], indicating α helix (blue) and β strand (red) compositions. Since the SUMO tag is devoid of cysteine residues, the distinct secondary structures identify the roles of disulfide bridges in conferring conformational stability to the CTLD. (e) Identification of the sugar ligand of CTLD by immunoblots of elutes from sugar affinity matrices charged with (1) Ara, (2) Glc, (3) GlcNAc, (4) Gal and (5) Man, for protein purified under non-reducing conditions. Chromatography matrices are prepared using vinylsulfone based chemistry[11] and the Western blots are developed by using a primary anti-SM50 serum fraction[3]. Note the stabilization of a dimeric form of the CTLD, after interactions with its sugar ligand, GlcNAc.

3. Calcium carbonate nucleation regulated by CTLD

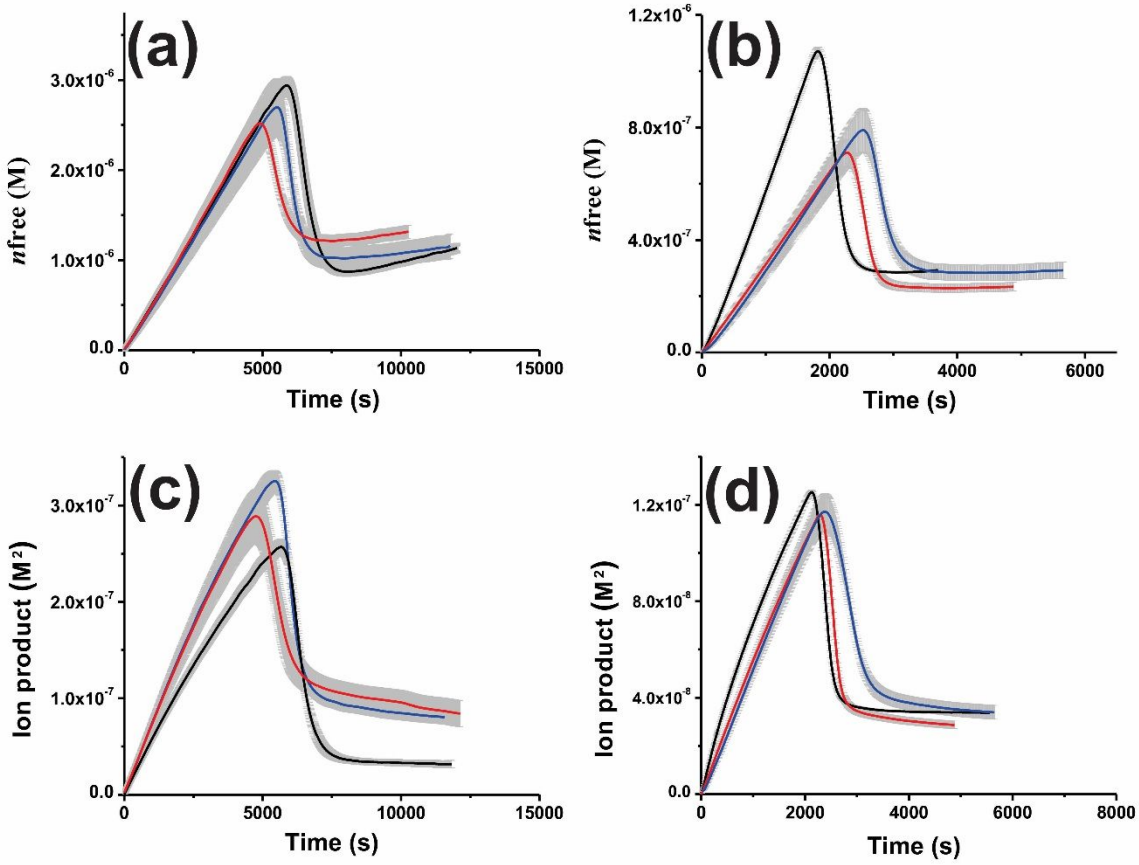


Figure S3A. Time developments of (a, b) free Ca^{2+} ion contents and (c, d) ion products in presence of the CTLD applied at 0.1 (red) and 1 (blue) mg/ml and corresponding reference experiments (black). The pH is maintained constant at either (a, c) 9.0 or (b, d) 9.75 by the automatic titration of base solutions. The line shadows (grey) represent ±1 standard deviation for data collected at intervals of 10 s. Depicted in Figure S3B, the development of free Ca^{2+} ions and base addition are used for quantitatively assessing mineral nucleation. At pH 9.0, note the solubility of the mineral product (1×10^{-7} M^2) is higher than the solubility of amorphous $CaCO_3$ (3.0×10^{-8} M^2). This reflects the pH-related contributions of the CTLD in transiently stabilizing highly hydrated mineral precursors formed via liquid-liquid demixing[12], significantly prior to particle nucleation.

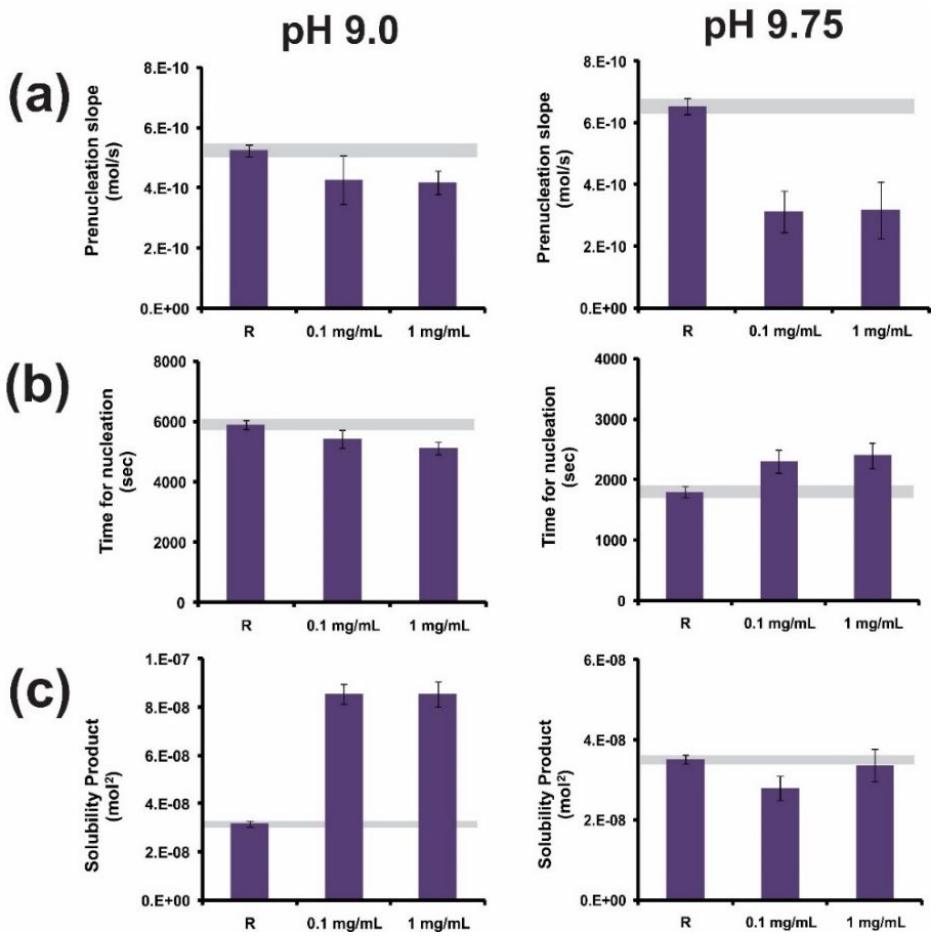


Figure S3B. Bar plots representing the effects of the CTLD on the nucleation of $CaCO_3$ at pH 9.0 and 9.75 in terms of the (a) pre-nucleation slope, (b) time of nucleation, and (c) solubility of the initially precipitated mineral phase. Error bars depict ± one standard deviation based on the data depicted in Figure S3A for reference (R) and additive-containing experiments. Note the pH dependent effects viz. relative destabilization of ion-clusters, a minor promotion of nucleation and the precipitation of highly soluble mineral phases specifically at low pH. A comparison with previously described nucleation trends with β-mercaptoethanol-treated CTLD reveals that the conformational and redox states of macromolecular additives are important determinants of mineral formation and growth[1]. Hence the cysteine residues of the CTLD partake in disulphide bond formation, appear essential for native protein structure and function in mineralization.

4. Solution studies of mineralization-induced CTLD self-assembly

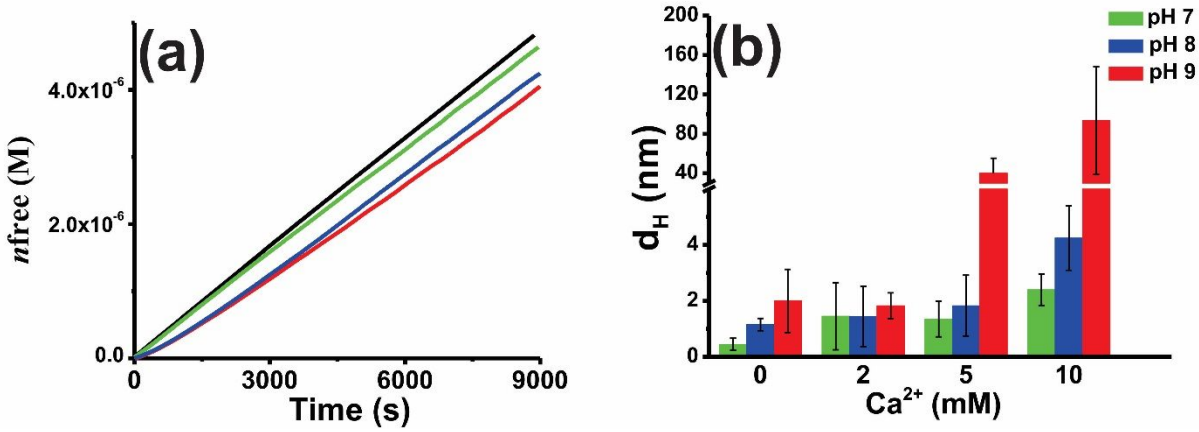


Figure S4A. (a) Time development of free Ca^{2+} ions in absence of protein (black) and in presence of the CTLD (1 mg/ml) at different pH values in carbonate-free, dilute HEPES buffer (1 mM). At pH 7.0 (green), there is negligible Ca^{2+} association by the additive. The offset becomes pronounced at higher pH, 8.0 (blue) and 9.0 (red), suggesting conducible protein conformations or speciation of side chain residues for Ca^{2+} binding. From the X-axis intercept, the complexation stoichiometry is quantified. About 1.2 and 2.6 Ca^{2+} ions are bound to a protein molecule at pH 8.0 and 9.0, respectively. In view of the primary sequence, this corresponds to less than 1 Ca^{2+} ion for each carboxylate group. (b) From dynamic light scattering, particle size distributions are measured for CTLD (1 mg/ml) with varying Ca^{2+} contents at different pH values in the absence of mineralization conditions. Without Ca^{2+} ions, particle sizes are about 0.5±0.2, 1.2±0.2 and 2.3±1.1 nm at pH 7.0, 8.0 and 9.0, respectively. The particle sizes increase in proportion to the Ca^{2+} ion contents and this effect is most pronounced at pH 9.0.

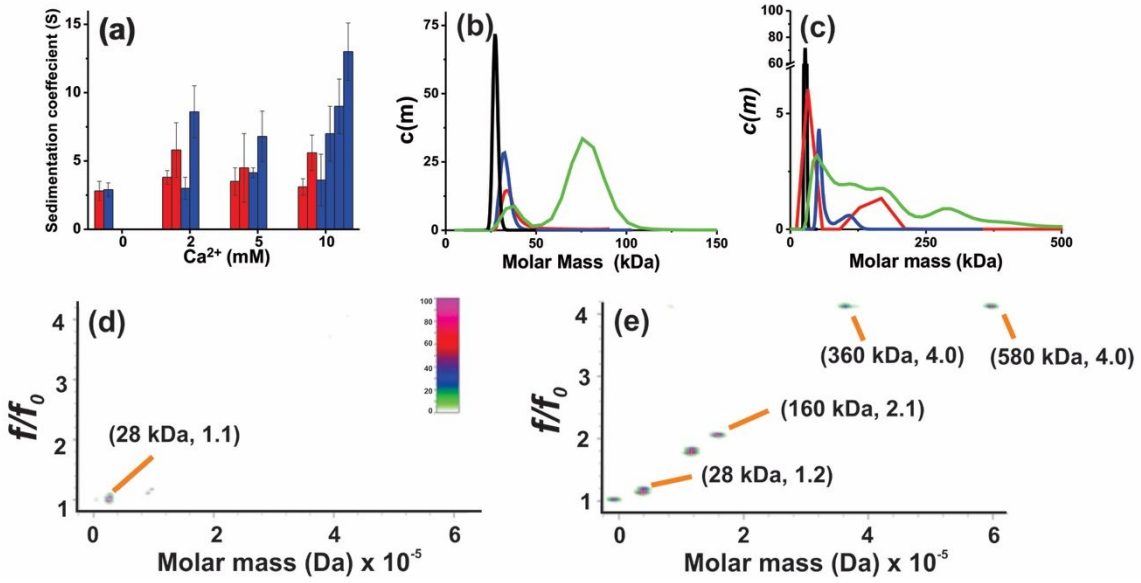


Figure S4B. From AUC velocity experiments, (a) Bar plots representing the distribution of sedimentation coefficients of the self-associated states of CTLD species in HEPES (red, pH 7.4) and carbonate (blue, pH 9.0) buffer with varying Ca^{2+} contents. Each bar represents a single state of protein self-association, with tabulated species abundance and error values (±1 standard deviations) corresponding to peak integrals and the FWHMs of Gaussian fits to the experimental distributions respectively (Table S1). (b, c) Continuous molar mass and sedimentation coefficient distributions of CTLD generated by fitting the experimental data to the Lamm equation model using SEDFIT[13], assuming a partial specific volume of 0.718 ml/g. The distributions represent (b) HEPES (pH 7.4) and (c) carbonate (pH 9.0) buffer with 0 (black), 2 (red), 5 (blue) and 10 (green) mM Ca^{2+} ion contents. In the absence of Ca^{2+} ions, single peaks corresponding to 2.8 S are observed in both buffer systems. These correspond to monomeric species with molar masses of about 26 kDa. Hence in the absence of Ca^{2+} ions, the buffer constituents have negligible effects on protein self-association. In HEPES buffer, 2 mM Ca^{2+} ions lead to species corresponding to 3.8±0.5 and 5.8±2.0 S. These species present molar masses of 32±5 and 70±39 kDa, respectively. Molar masses between monomeric (27 kDa) and dimeric (54 kDa) indicate reversible interactions on a time scale faster than the duration of the ultracentrifuge experiment. Protein species with a molar mass of 70±39 kDa correspond to a trimeric form. In carbonate buffer, protein self-association is proportionate to Ca^{2+} ion concentrations, however, is also more pronounced. At 10 mM Ca^{2+} ion concentration, four species are observed corresponding to 3.6±1.9, 7±1.9, 9±1.9 and 13±2 S. These indicate molar masses of 43±35, 111±55, 175±65 and 292±65 kDa, respectively. Therefore, in relation to the self-association states observed in HEPES buffer, the oligomerization of CTLD is enhanced

under conditions of calcium carbonate nucleation. Pseudo 3-dimensional plots are depicted from 2-DSA analyses of sedimentation profiles of CTLD samples supplemented with (d) 0 and (e) 10 mM Ca^{2+} ions in carbonate buffer (10 mM, pH 9.0). Values of molar mass and frictional coefficients are indicated in brackets. Without Ca^{2+} ions, the CTLD sample presents a peak at about 27 kDa corresponding to the monomer species. This species has a f/f_o value of 1.1 indicating a spherical hydrodynamic shape and globular structure with a = 1.94 nm. This is in reasonable agreement with the corresponding size ascertained by DLS (Figure S4A). In presence of 10 mM Ca^{2+} ions, multiple species are observed. Along with the monomeric species, larger assemblies (360 and 600 kDa) with *f/fo* values of 4.0 indicate a highly asymmetric structure (providing dimensions of 100 nm (a) × 1 nm (b) for the 360 kDa species and of 110 nm (a) × 1.2 nm (b) for the 600 kDa species based on prolate models[5]). In light of cryo-TEM observations (Figure 1g-i), these species might indicate sheet-like structures. Additionally, intermediate protein assemblies with molar masses of about 120 and 160 kDa are observed with 1.8 and 2.1 *f/fo* values, respectively. By approximating the hydrodynamic shape to a prolate ellipsoid of revolution, the axial ratios a/b are estimated. The 120 kDa protein species has axial dimensions of 44 and 0.8 nm whereas the species at 160 kDa exhibits axial dimensions of 22 and 1.4 nm (a and b, respectively). This indicates an extended conformation of the intermediate assemblies which is consistent with the cryo-TEM data (Figure 1e, f). Thus, conditions of mineralization drive the self-association of CTLD towards discrete self-associates that subsequently produce the organic scaffold.

Table S1. Summary of molar mass and sedimentation coefficient distributions of CTLD from AUC velocity experiments.

Ca^{2+} (mM)	Sedimentation coefficient (S)	Molar mass (kDa)	Relative concentration
HEPES buffer			
0	2.8±0.7	27±3	100%
2	3.8±0.5	32±5	64%
	5.8±2	70±39	26%
5	3.5±1.0	33±8	53%
	4.5±2.5	75±20	47%
10	3.1±0.6	36±10	13%
	5.6±1.3	75±20	87%
Carbonate buffer			
0	2.9±0.5	26±2	100%
2	3±0.8;	32±15	71%
	8.6±1.9	107±50	29%
5	4.14±0.36	32±10	83%
	6.8±1.86	169±25	17%
10	3.6±1.9	43±35	27%
	7±1.9	111±55	27%
	9±1.9	175±65	27%
	13±2	292±65	19%

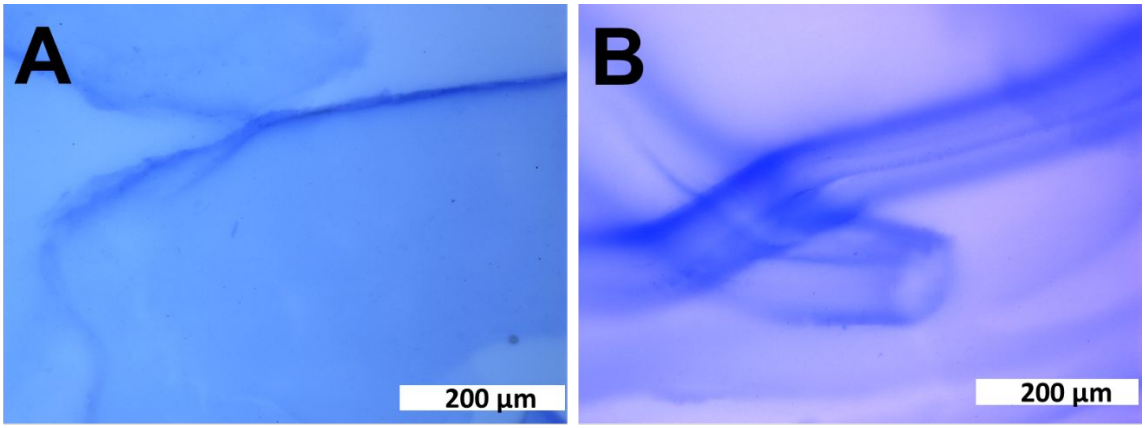


Figure S4C. Coomassie Brilliant Blue stained CTL assemblies representing (A) sheet-like and (B) the rolling of these sheets to tube-like micro-structures. Self-association occurs in mixtures of SUMO-CTL (1 mg/ml) and $CaCl_2$ (10 mM), upon aging at 4°C for 1 week.

5. Structural Characterization of CTLD-Calcite mesocrystals

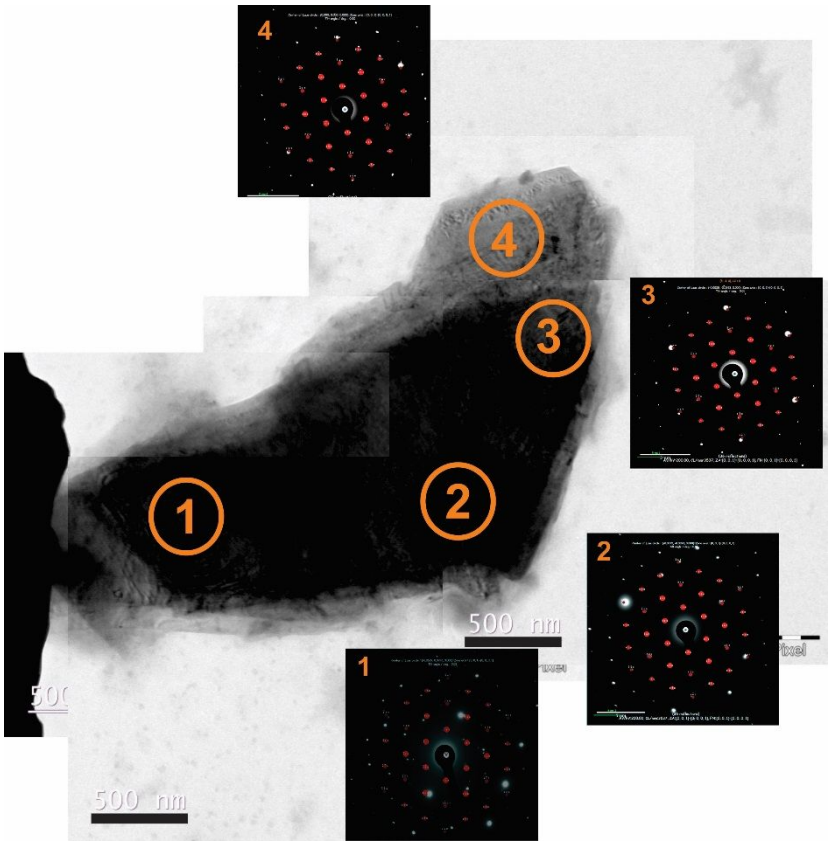


Figure S5A. Combined TEM micrographs of a crystalline product obtained from mineral nucleation in the presence of CTLD at pH 9.0. Corresponding single-crystal-like ED patterns are indexed to the calcite [001] zone axis by using JEMS-SAAS. Usually the (001) face is not stable and undergoes restructuring in aqueous solutions because of a net dipole moment from hexagonal arrangements of planar carbonate and Ca^{2+} ions. In view of mesocrystal formation, the CTLD effectively stabilizes this crystal face against restructuration, with possible mechanistic factors including Ca^{2+} binding ability (Figure S5A) and self-assembly towards sheet-like structures that provide surface coverage and stabilization of the mineral nanoparticles (Figure S5B).

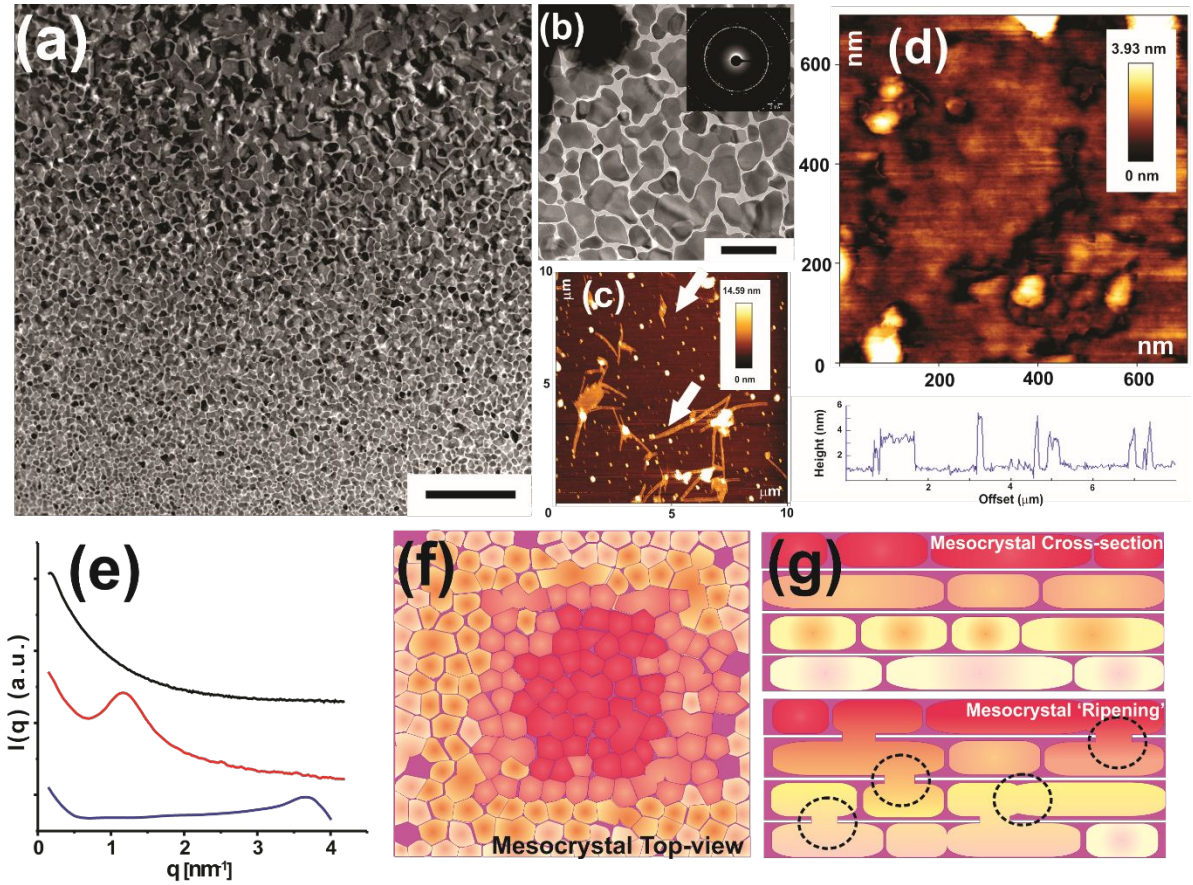


Figure S5B. (a, b) Representative TEM images of mineral composites after CTLD elimination by exposure to proteinase K. After removal of the organic matrix, the residual inorganic particles are irregular, calcite nanoplates that present an overall jigsaw-like appearance. These represent a structural organization at the nanoscale, a level of structural hierarchy in addition to the laminar arrangement. (c, d) AFM-derived height maps of structures after protein cross-linking by glutaraldehyde, following acid etching for demineralization and corresponding signals for a single line scan. Arrows (c) indicate collapse of the organic sheets due to removal of the inorganic phase. The AFM data indicate a porous nature of the organic laminae, corresponding to a height distance of about 3-6 nm and a variable width in the range of 0.1 to 2 μm. This size dimension reflects the plate-like morphology of mineral particles (a, b) constituting the superstructure. Thus, the selective removal of either the organic or inorganic phases produce complementary structural architectures and thus reveal an integration of the protein and mineral constituents. The size distributions of the mineral particles conform to previous descriptions of the spine superstructure[14-15]. (e) SAXS measurements are conducted for the products of CTLD-controlled mineralization at 1 h (blue), 2 h (red) and 5 h (black) after nucleation. This indicates an increase in the average thickness of the mineralized laminae with

time from 1.7±0.1 nm to 6.9±1.9 nm to above 100 nm (upper limit of instrument resolution). Thus, a process of mesocrystal ripening possibly involving partial grain coalescence and growth is identified, driven under the conditions of supersaturation. Schematic depiction of the layered mesocrystal structure from (f) top- (protease treatment derived, Figure 2e) and (g) side- (ultramicrotomy produced, Figure 2d) views indicating the proteinaceous matrix (purple) and space-filling arrangement of mineral particles in multilayers. (g) The process of mesocrystal ripening is illustrated by the growth and coalescence of mineral building units (dotted circles) forming mineral bridges, as reported previously[15].

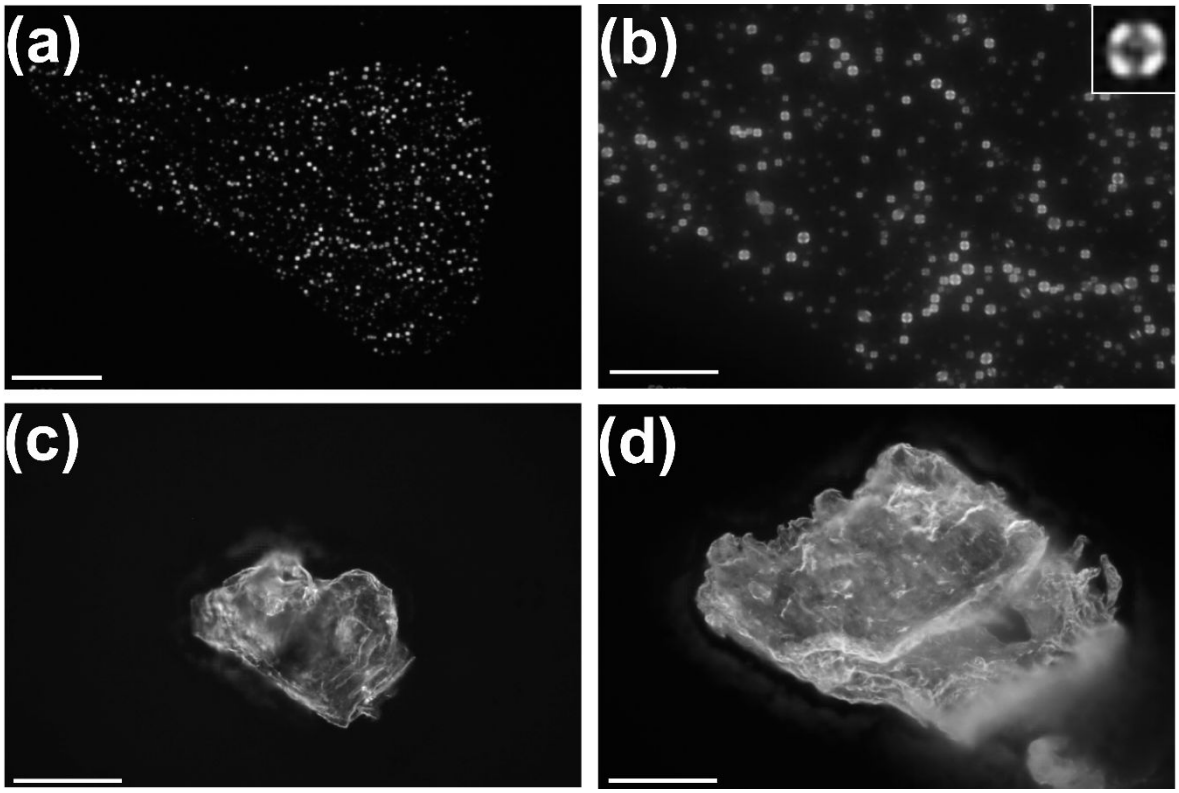


Figure S5C. Representative dark field polarization microscopy images for structures sampled (a, b) immediately and (c, d) 1 h after nucleation at pH 9.0. Scale bars represent (a) 100, (b) 50 and (c, d) 500 μm. Observations of immediate products of mineral nucleation shows co-localized birefringent particles in certain areas, suggesting an association with protein sheets (a, b). The Maltese cross consisting of two isogyrses in the inset (b) is a conoscopic pattern suggestive of an optically inactive material. Similar observations are reported for biological vesicles[16] as well as mineralization via PILPs[17]. For samples corresponding to 1 h post-nucleation, birefringent sheets are observed (c, d), indicative of mesocrystalline structures.

6. Derivation of core-shell parameters of vesicular composites

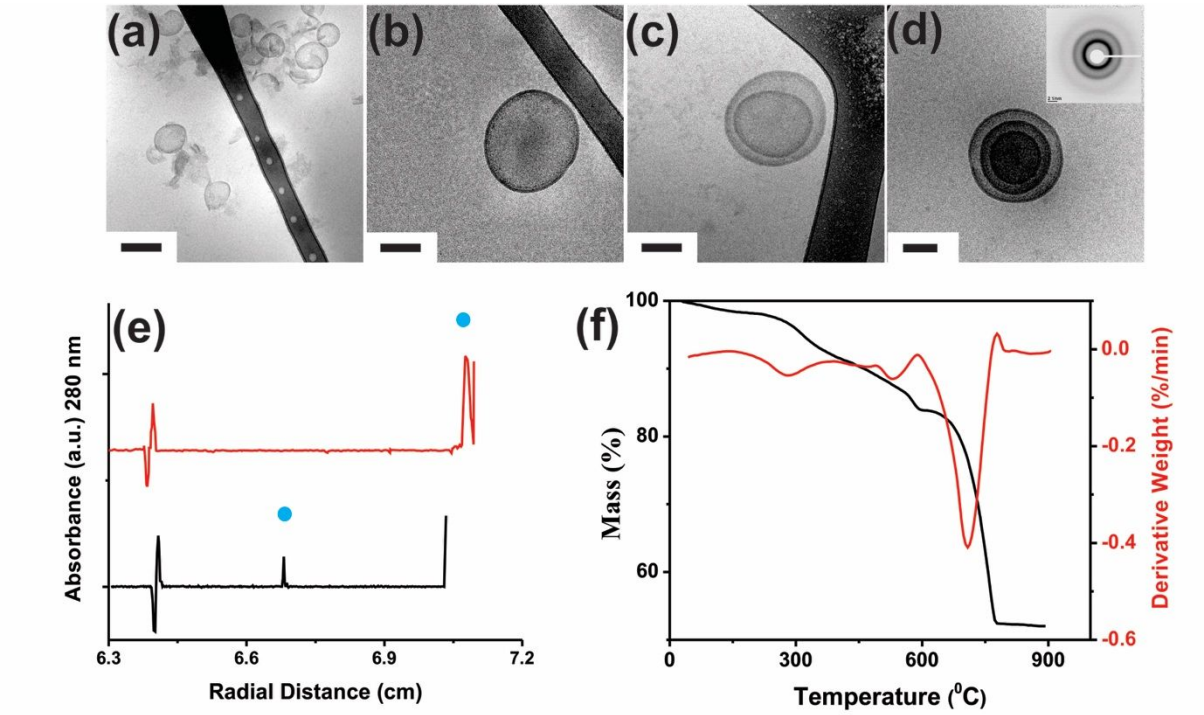


Figure S6A. (a-d) Amorphous vesicles formed in the presence of SUMO-GRR[18] during $CaCO_3$ nucleation present radii in the approximate range of 150 to 250 nm. Scale bars represent (a) 200 and (b-d) 100 nm. (e) Density gradient profiles of polystyrene particles (black) and vesicles (red) in sucrose gradients. (f) TGA analyses for nucleated mineral products particles in the presence of SUMO-GRR.

Equation (1) is used to determine the buoyant density of particles

$$\varrho_x = \varrho_i + \frac{\omega^2 r_i}{\beta}(r_x - r_i) \quad (1)$$

where the density (ρ_x) corresponds to a radial distance r_x, ω is the angular velocity and r_i is the radial distance corresponding to the hinge point. The β value for 63% sucrose is determined using calibration measurements and the calculated iso-concentration/hinge point, described below.

For calibration measurements, the determined radial distance for polystyrene nanoparticles in a 63% sucrose gradient is 6.739 cm (Fig. S6A (e)) and this distance corresponds to a density of 1.05 g/cc.

The iso-concentration/hinge point is determined to be 6.88 cm by using the equation,

$$r_s = \sqrt{\frac{r_m^2 + r_b^2}{2}} \quad (2)$$

where r_m and r_b are radial distances of the meniscus and cell bottom respectively. This radial distance corresponds to the initial density of sucrose (63%), $\rho_s = 1.3043\ gm/cc$

Thus, by using equation (1), the β value is calculated to be 4.22×10^7 rad^2 cm^5 sec^{-2} g^{-1}. The observed radial distance (7.08 cm) for the SM50GRR vesicles thus corresponds to a buoyant density of 1.4±0.085 g/cc.

From figure (f), three phases of weight loss are observed during thermogravimetry indicating water, protein and mineral constituents. The relative contents of water, protein and calcium carbonate are determined as 2.3, 18.6 and 79.1%. This corresponds to a stoichiometry of $(CaCO_3)_1(protein)_{0.14}(H_2O)_{63.9}$ for the vesicles. The net water fraction is representative of the amorphous mineral phase as well as protein molecules.

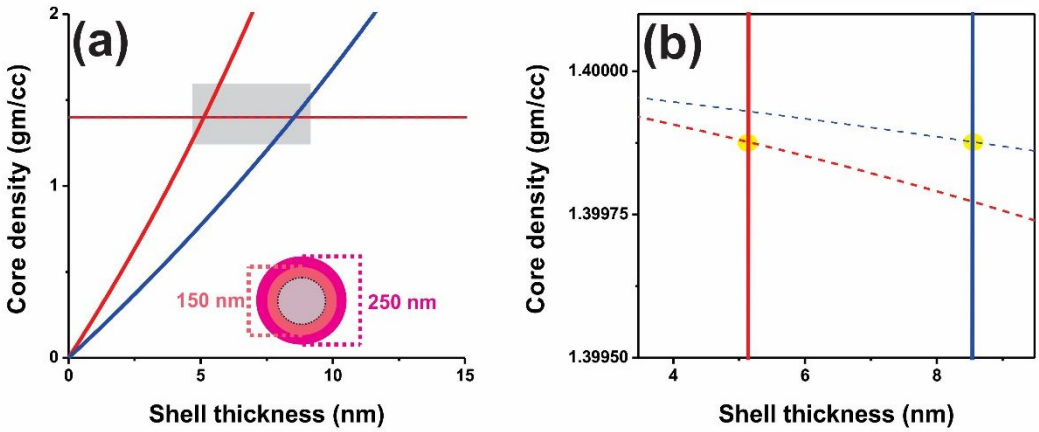


Figure S6B. (a) Plots of equations (4) and (6) estimating the density of the amorphous core and the protein shell thickness for vesicles of radii 75 (red curves) and 125 nm (blue curves) confining a mineral phase. (b) Enlarged version of the area shaded grey in (a) with yellow dots representing the solutions.

Using information from Figure S6A, the density of the amorphous core and the protein shell thickness are determined. For these estimations, we assume that the vesicles are perfect spheres with an organic shell and a mineral core of respective homogenous compositions. As an example, for vesicles 100 nm in radius and shells of x nm thickness, volume of core (V_c) and shell (V_s) can be expressed as:

$$V_c = \frac{4}{3}\pi(100 - x)^3$$

$$V_s = \frac{4}{3}\pi[(100)^3 - (100 - x)^3]$$

Secondly, the ratio of the core and shell masses can be expressed as:

$$\frac{mass_{core}}{mass_{shell}} = \frac{V_c \varrho_c}{V_s \varrho_s} \tag{3}$$

where ϱ_c is the density of the core. The shell density, $\varrho_s = 1.40\ gm/cm^3$ is determined from protein sequence. By using the mass ratio deduced from TGA analysis and the volume expressions, equation (3) is represented as

$$\varrho_c = 5.92 \times 10^{-21} \times \frac{(100)^3 - (100 - x)^3}{(100 - x)^3} \quad gm/nm^3 \tag{4}$$

We also apply a mass balance for an individual core-shell particle,

$$mass_{core} + mass_{shell} = mass_{total}$$

or

$$\varrho_c = \frac{mass_{total} - (V_s \times \varrho_s)}{V_c} \tag{5}$$

Since the total density of a particle is $1.40 \pm 0.09\ gm/cm^3$, $mass_{total}$ can be determined for a vesicle of known size (100 nm). Hence equation (5) can be expressed as:

$$\varrho_c = \frac{5.86 \times 10^{-15} - \frac{4\pi}{3} \cdot 1.40 \times 10^{-21} \cdot [(100)^3 - (100 - x)^3]}{\frac{4\pi}{3}(100 - x)^3} \quad (6)$$

In view of the vesicle radii in the range of 75 to 125 nm (Figure S6A), the corresponding solutions for equations (5) and (6) are plotted in Figure S6B.

The thickness of GRR monolayers is determined to be 2.8 nm from ellipsometry measurements. The data fit corresponds to a three layer model of silicon, silicon dioxide and an uppermost protein monolayer, considering a refractive index of the protein as 1.3[19]. Thus, the protein shell initially presents a bilayer architecture, transitioning to a multilayer structure in course of mineralization. This might reflect an amphiphilic nature of the protein in view of the N-terminal SUMO domain being globular and the disordered GRR region serving as a hydrophobic chain. This model can be extended to the full-length SM50.

In view of the low density of the amorphous core (1.39 g/cc) for both vesicles of radii 75 and 125 nm, the phase enclosed by the protein shell appears to be either highly hydrated form of amorphous calcium carbonate or a dense liquid-like phase. The literature value for the density of amorphous calcium carbonate is 1.49 g/cc as determined by SAXS and AUC[20-21]. Hence the mass fractions of the ACC (a) and water ($1 - a$) are estimated by using the following equation,

$$\varrho_c = a \cdot \varrho_{ACC} + (1 - a) \cdot \varrho_{H_2O}$$

where $\varrho_c = 1.399\ g/cm^3$, $\varrho_{ACC} = 1.49\ g/cm^3$ and $\varrho_{H_2O} = 1\ g/cm^3$. This provides respective values of 0.81 (a) and 0.19 ($1 - a$), revealing a significantly higher hydration content of the confined mineral phase relative to ACC, thus suggesting the confinement-mediated stabilization of a fluid-like mineral precursor. Considering the SAXS derived[22] density of ACC as $\varrho_{ACC} = 1.62\ g/cm^3$, even higher hydration contents of the core mineral phase are indicated as 0.64 (a) and 0.36 ($1 - a$).

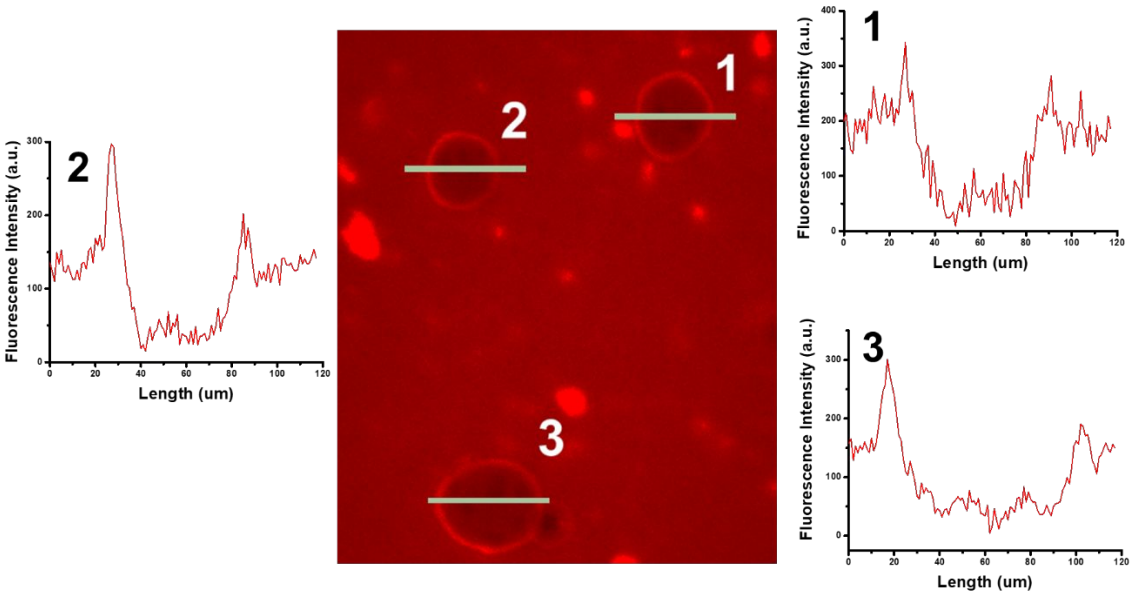


Figure S6C. Fluorescence microscopy images of vesicles composed of rhodamine-labeled SUMO fusion SpSM50 in water:THF mixtures[18]. Line plots of the normalized fluorescence signals depict depletion of the protein molecules from vesicle cores, suggesting THF rich core compositions as well as the enrichment of protein at liquid boundaries. On the other hand, the extravesicular aqueous phase contains soluble protein and aggregate particles.

7. Mass spectrometry derived sequence coverage for protein fragments in the range of 40 to 25 kDa (Figure 3a, c) derived from Type VI collagenase treated SM50

Sequence	# PSMs	Modifications	IonScore	Exp Value	Charge	MH+ [Da]	ΔM [ppm]
VWLGWNSQSPFF	3		89	2,535E-08	2	1467,69546	-6,92
GRQPGMGGQPGVGGR	1		56	4,864E-05	3	1410,69154	-7,41
GRQPGMGGQPGVG	1		51	0,0001698	2	1197,55214	-22,76
ASVSSPQENME	2		49	0,0002786	2	1178,51078	9,62
LGWNSQSPFF	1		44	0,000853	2	1182,61824	51,04
NMRVPYRMASE	2		41	0,0014926	3	1353,70279	46,18
VWLGWNSQSPF	1		39	0,0022854	2	1320,65110	10,53
MRVPYRMASE	2		34	0,0081838	3	1239,57844	-15,25
VTAmRAF	1	M4(Oxidation)	33	0,009593	2	811,40764	-6,77
GRQPGMGGRQPG	1		30	0,0189215	2	1197,55214	-32,14
MRVPYRmASE	3	M7(Oxidation)	28	0,0298508	2	1255,57808	-11,29
YVRSQSGQScYRYFN	1	C10(Carbamidomethyl)	28	0,0331098	3	1914,87067	8,08
YVRSQSGQScYRYF	1	C10(Carbamidomethyl)	27	0,0420685	3	1800,78361	-15,91
ASVSSPQENM	1		26	0,0535743	2	1049,45646	-0,37
VEVTAMRA	1		24	0,0717723	2	876,48712	30,01
MRVPYRMASEF	2		23	0,0933161	2	1386,65664	-6,58
MRVPYRM	2		23	0,1068948	3	952,39270	-97,54
KRQNPPVRPGQGGRQIPQGVGPQW	1		19	0,2365683	3	2637,25768	-62,85
GNQPGVGGRQPGmGGQPGVGGR	1	M13(Oxidation)	19	0,2477174	3	2035,84834	-66,65
GRQIPQGVGPQW	1		19	0,2534875	2	1322,72304	20,10
RVPYRMASE	1		19	0,2716168	2	1108,52240	-31,07
RVPYRM	1		17	0,4017506	2	821,45614	13,45
GRQPGMGGQPGVGGRQPGVG	1		17	0,4187517	3	1848,89323	-17,02
MRVPYRm	2	M7(Oxidation)	16	0,4931245	2	968,48050	0,00
RYFNMRVPYRM	1		15	0,6698176	3	1532,80177	26,35
SSSPASPPRPGmPPTR	1	M12(Oxidation)	15	0,6933565	3	1637,79373	-7,80
MRVPYR	1		14	0,7567572	2	821,45614	13,45
MRVPYRmASEF	1	M7(Oxidation)	14	0,8240557	2	1402,62888	-22,67

8. Mineral nucleation in the presence of SM50 and a matrix metalloprotease

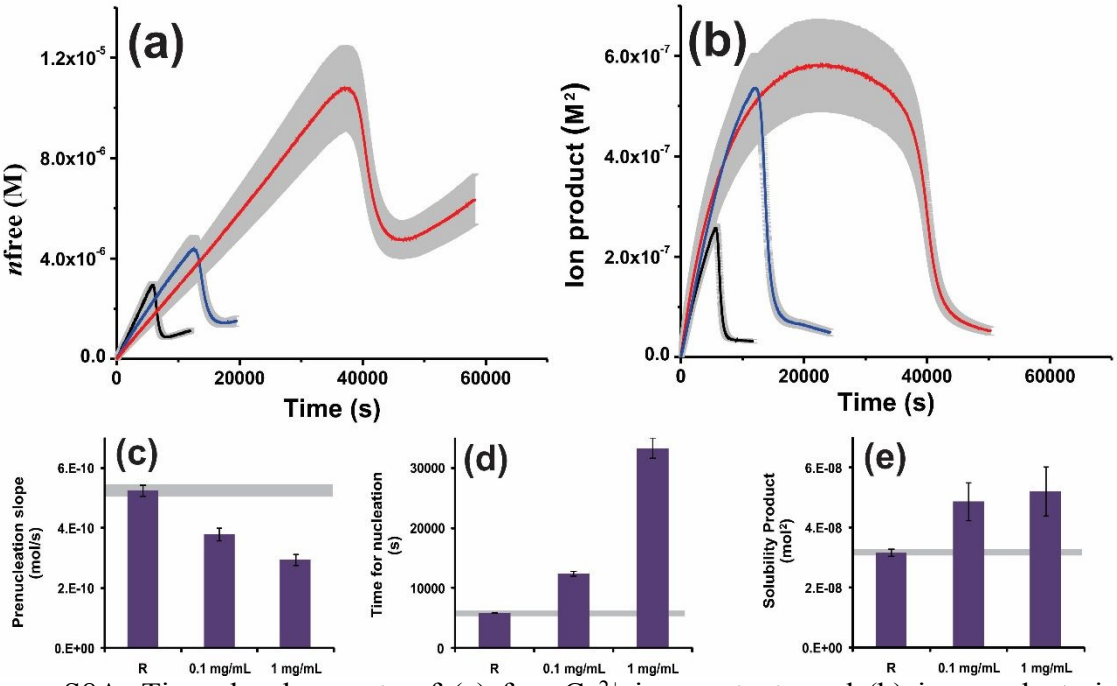
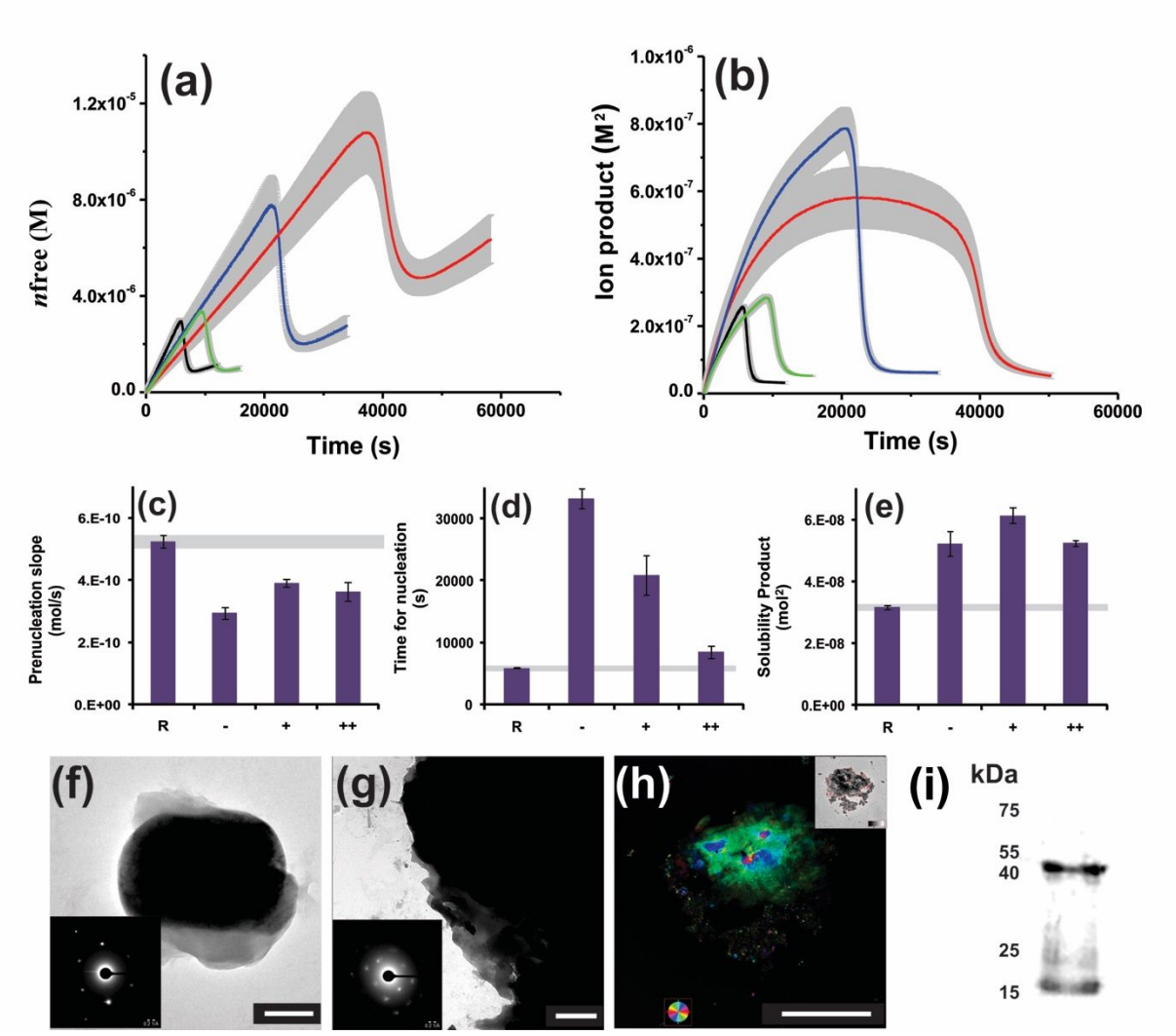


Figure S8A. Time developments of (a) free Ca^{2+} ion contents and (b) ion products in the presence of 0.1 (red) and 1 (blue) mg/ml SM50 and corresponding reference experiments (black) at pH 9.0. The line shadows (grey) represent ±1 standard deviation for data collected at intervals of 10s. The bar plots represent the effects of SM50 on the nucleation of $CaCO_3$ at pH 9.0 in view of the (c) pre-nucleation slope, (d) time for mineral nucleation, and (e) solubility of the initially precipitated mineral phase. Error bars represent ± one standard deviation. The time required for nucleation as well as the solubility of the initially formed mineral phase are accentuated by higher additive contents, while the prenucleation slope decreases indicating an increasing stabilization of prenucleation clusters. The inhibition of mineral nucleation is not solely due to ion-complexation, considering the scarcity of acidic amino acid residues (Figure S1). The values of solubility products suggest either the transient formation and/or stabilization of a highly soluble, intermediate phase of amorphous calcium carbonate. Similar inhibitory effects towards mineral nucleation are also observed during mineral nucleation in the presence of GRR[18]. Also considering that the mineral-containing vesicles are transiently stabilized in the presence of full length SM50 as well as its GRR domain, the influences of the intrinsically disordered domain on the solution state organic-inorganic interactions and their structural consequences towards mineral nucleation supersede those of the N-terminal CTLD.



S8B. Time developments of (a) free Ca^{2+} ion contents and (b) ion products in the presence of SM50 (1 mg/ml) supplemented with either 0 (red/-), 10 (blue/+) or 100 (green/++) µg/ml Type VI collagenase and in reference experiments (black/R) at pH 9.0. The line shadows (grey) represent ±1 standard deviation. Bar plots represent the titration data in terms of the (c) pre-nucleation slope, (d) time for mineral nucleation, and (e) solubility of the initially precipitated mineral phase. Error bars represent ± one standard deviation. The inhibitory effect of SM50 towards mineral nucleation is clearly diminished in the presence of type IV collagenase. (f) In the presence of SM50 alone, TEM reveals the formation of vesicles that crystallize to calcite after drying. Similar observations are noted for GRR as the nucleation additive[18]. However, with the combined application of SM50 and Type VI collagenase, laminar calcitic structures are formed that present single crystal-like (g) diffraction and (h) Abrio™ -derived polarization patterns (a method to measure spatial distribution of birefringence). Scale bars represent (f) 200 nm, (g) 1 µm and (h) 20 µm. (i) Western blots of arginine-solubilized fraction of the sea urchin spine extract developed with CTLD specific primary antibodies. In addition to the full-

length SpSM50 (40 kDa), a second band corresponding to 15 kDa suggest an intact CTLD fragment. This indicates a targeted *in vivo* proteolysis of SM50, possibly by matrix metalloproteases (MMP). This is also supported by an inhibited biomineral growth and accumulation of vesicles in mesenchymal cells of embryonic sea urchin treated with MMP inhibitors[23] as well as a general enrichment of glycine and proline in the proteolysis sites of MMPs[24]. Similar immunoblot profiles have been recently reported for SM50 from the sea urchin teeth[25].

References

1. Rao, A.; Seto, J.; Berg, J. K.; Kreft, S. G.; Scheffner, M.; Cölfen, H., Roles of larval sea urchin spicule SM50 domains in organic matrix self-assembly and calcium carbonate mineralization. *Journal of structural biology* **2013,** *183* (2), 205-215.
2. Gebauer, D.; Völkel, A.; Cölfen, H., Stable prenucleation calcium carbonate clusters. *Science* **2008,** *322* (5909), 1819-1822.
3. Killian, C. E.; Wilt, F. H., Characterization of the proteins comprising the integral matrix of Strongylocentrotus purpuratus embryonic spicules. *Journal of Biological Chemistry* **1996,** *271* (15), 9150-9159.
4. Schuck, P., Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical journal* **2000,** *78* (3), 1606-1619.
5. Brookes, E.; Cao, W.; Demeler, B., A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. *European Biophysics Journal* **2010,** *39* (3), 405-414.
6. Kröger, R.; Verch, A., Liquid Cell Transmission Electron Microscopy and the Impact of Confinement on the Precipitation from Supersaturated Solutions. *Minerals* **2018,** *8* (1), 21.
7. Dosztányi, Z.; Csizmok, V.; Tompa, P.; Simon, I., IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005,** *21* (16), 3433-3434.
8. Linding, R.; Russell, R. B.; Neduva, V.; Gibson, T. J., GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic acids research* **2003,** *31* (13), 3701-3708.
9. Prilusky, J.; Felder, C. E.; Zeev-Ben-Mordehai, T.; Rydberg, E. H.; Man, O.; Beckmann, J. S.; Silman, I.; Sussman, J. L., FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **2005,** *21* (16), 3435-3438.
10. Louis-Jeune, C.; Andrade-Navarro, M. A.; Perez-Iratxeta, C., Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins: Structure, Function, and Bioinformatics* **2012,** *80* (2), 374-381.
11. Lopez-Jaramillo, F. J.; Ortega-Muñoz, M.; Megia-Fernandez, A.; Hernandez-Mateo, F.; Santoyo-Gonzalez, F., Vinyl sulfone functionalization: A feasible approach for the study of the lectin–carbohydrate interactions. *Bioconjugate chemistry* **2012,** *23* (4), 846-855.
12. Sebastiani, F.; Wolf, S. L.; Born, B.; Luong, T. Q.; Cölfen, H.; Gebauer, D.; Havenith, M., Water Dynamics from THz Spectroscopy Reveal the Locus of a Liquid–Liquid Binodal Limit in Aqueous CaCO3 Solutions. *Angewandte Chemie International Edition* **2017,** *56* (2), 490-495.
13. Lebowitz, J.; Lewis, M. S.; Schuck, P., Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein science* **2002,** *11* (9), 2067-2079.
14. Seto, J.; Ma, Y.; Davis, S. A.; Meldrum, F.; Gourrier, A.; Kim, Y.-Y.; Schilde, U.; Sztucki, M.; Burghammer, M.; Maltsev, S., Structure-property relationships of a biological mesocrystal in the adult sea urchin spine. *Proceedings of the National Academy of Sciences* **2012,** *109* (10), 3699-3704.
15. Oaki, Y.; Imai, H., Nanoengineering in echinoderms: the emergence of morphology from nanobricks. *Small* **2006,** *2* (1), 66-70.
16. Regev, O.; Guillemet, F., Various bilayer organizations in a single-tail nonionic surfactant: unilamellar vesicles, multilamellar vesicles, and flat-stacked lamellae. *Langmuir* **1999,** *15* (13), 4357-4364.
17. Cheng, X.; Varona, P. L.; Olszta, M. J.; Gower, L. B., Biomimetic synthesis of calcite films by a polymer-induced liquid-precursor (PILP) process: 1. Influence and incorporation of magnesium. *Journal of Crystal Growth* **2007,** *307* (2), 395-404.
18. Rao, A.; Drechsler, M.; Schiller, S.; Scheffner, M.; Gebauer, D.; Cölfen, H., Stabilization of Mineral Precursors by Intrinsically Disordered Proteins. *Advanced Functional Materials* **2018**, 1802063.
19. Vörös, J., The density and refractive index of adsorbing protein layers. *Biophysical journal* **2004,** *87* (1), 553-561.

20. Bolze, J.; Peng, B.; Dingenouts, N.; Panine, P.; Narayanan, T.; Ballauff, M., Formation and growth of amorphous colloidal CaCO3 precursor particles as detected by time-resolved SAXS. *Langmuir* **2002,** *18* (22), 8364-8369.
21. Cölfen, H.; Völkel, A., Application of the Density Variation Methodon Calciumcarbonate Nanoparticles. In *Analytical Ultracentrifugation VIII*, Springer: 2006; pp 126-128.
22. Liu, J.; Pancera, S.; Boyko, V.; Shukla, A.; Narayanan, T.; Huber, K., Evaluation of the particle growth of amorphous calcium carbonate in water by means of the porod invariant from SAXS. *Langmuir* **2010,** *26* (22), 17405-17412.
23. Ingersoll, E. P.; Wilt, F. H., Matrix metalloproteinase inhibitors disrupt spicule formation by primary mesenchyme cells in the sea urchin embryo. *Developmental biology* **1998,** *196* (1), 95-106.
24. Parks, W. C.; Wilson, C. L.; López-Boado, Y. S., Matrix metalloproteinases as modulators of inflammation and innate immunity. *Nature Reviews Immunology* **2004,** *4* (8), 617-629.
25. Mao, Y.; Satchell, P. G.; Luan, X.; Diekwisch, T. G., SM50 repeat-polypeptides self-assemble into discrete matrix subunits and promote appositional calcium carbonate crystal growth during sea urchin tooth biomineralization. *Annals of Anatomy-Anatomischer Anzeiger* **2016,** *203*, 38-46.