Soft Matter

PAPER

View Article Online
View Journal | View Issue

Cite this: *Soft Matter*, 2024, **20**, 4886

Implications of intracrystalline OC17 on the protection of lattice incorporated proteins†

Huseyin Burak Caliskan ‡*[ab] and Fatma Isik Ustok[c]

Biogenic $CaCO_3$ formation is regulated by crystallization proteins during crystal growth. Interactions of proteins with nascent mineral surfaces trigger proteins to be incorporated into the crystal lattice. As a result of incorporation, these intracrystalline proteins are protected in the lattice, an example of which is ancient eggshell proteins that have persisted in $CaCO_3$ for thousands of years even under harsh environmental conditions. OC17 is an eggshell protein known to interact with $CaCO_3$ during eggshell formation during which OC17 becomes incorporated into the lattice. Understanding protein incorporation into $CaCO_3$ could offer insights into protein stability inside crystals. Here, we study the protection of OC17 in the $CaCO_3$ lattice. Using thermogravimetric analysis we show that the effect of temperature on intracrystalline proteins of eggshells is negligible below 250 °C. Next, we show that lattice incorporation protects the OC17 structure despite a heat-treatment step that is shown to denature the protein. Because incorporated proteins need to be released from crystals, we verify metal chelation as a safe crystal dissolution method to avoid protein denaturation during reconstitution. Finally, we optimize the recombinant expression of OC17 which could allow engineering OC17 for engineered intracrystalline entrapment studies.

Received 29th March 2024,
Accepted 29th May 2024

DOI: 10.1039/d4sm00371c

rsc.li/soft-matter-journal

Introduction

Ovocleidin-17 (OC17) is an eggshell protein whose role is thought to regulate $CaCO_3$ crystal growth during shell formation.[1] Simulations suggest that during the formation of an eggshell, OC17 acts like an enzyme to catalyse the amorphous-to-calcite transition.[1–3] The existence of an initial amorphous phase in eggshells has previously been verified.[4–6] The interaction of OC17 with nascent $CaCO_3$ crystals is in line with the fact that it is a member of calcium-dependent type (C-type) lectin-like protein family.[7–11] However, unlike other C-type lectins, which interact with a variety of carbohydrates,[12,13] OC17 interacts with inorganic $CaCO_3$, and as a result of interaction, OC17 becomes incorporated into the growing $CaCO_3$ lattice.[14,15] Incorporation into the lattice isolates the protein from the environment. A potential result of incorporation is the protection of the incorporated protein against thermal damage at temperatures that would degrade the protein if it were in solution (Fig. 1A and B). Because $CaCO_3$ is a low-cost, non-hazardous material which is synthesized at room temperature (RT), in-lattice protection of proteins in $CaCO_3$ crystals could address the temperature-sensitivity problem of protein-based therapeutics. The stability of proteins at high temperatures is particularly problematic in resource-limited settings where keeping a reliable cold-chain is often challenging. Incorporation of biomolecules in biogenic $CaCO_3$ has been known for decades in the context of wild-type proteins in fossil samples.[16–19] Peptide fragments from incorporated proteins are protected for eons in the $CaCO_3$ lattice for which remarkable examples have been documented lately.[20,21] Incorporated proteins store valuable information which is utilized for amino acid racemization geochronology, a method for dating archaeological samples.[22,23] In recent years, there is growing interest in incorporation of molecules into $CaCO_3$ for other practical purposes, for instance, for organic–inorganic materials[24] or for targeted release of drugs on demand.[25] Incorporation of a variety of molecules in $CaCO_3$ for bioinspired materials has also been reported.[26–28] Importantly, the effects of incorporation on lattice parameters have been thoroughly studied.[29–37] There is interest in loading small molecule drugs in crystals for protection under harsh conditions.[38] Nanoparticles[29,30,38–40] and polymers[41–45] have been incorporated, while amino acids have been used to assist in the incorporation of targets[46] or to investigate the effect of incorporation on lattice strain.[29,32–37] GFP-perlucin fusion[47] and cowpea mosaic virus[48] have been incorporated in $CaCO_3$. A recent citizen science project

[a] *University of Cambridge, Department of Engineering, Trumpington Street, CB2 1PZ Cambridge, UK. E-mail: huseyinburakcaliskan@etu.edu.tr*
[b] *University of Cambridge, The Nanoscience Centre, 11 JJ Thomson Avenue, CB3 0FF Cambridge, UK*
[c] *University of Cambridge, Cambridge Institute for Medical Research, Department of Haematology, The Keith Peters Building, Hills Road, CB2 0XY Cambridge, UK*
† Electronic supplementary information (ESI) available: Dot blot image of the optimization of OC17 expression. See DOI: https://doi.org/10.1039/d4sm00371c
‡ Present address: TOBB University of Economics and Technology, Department of Biomedical Engineering, Söğütözü Caddesi No. 43, 06560, Ankara, Turkey.

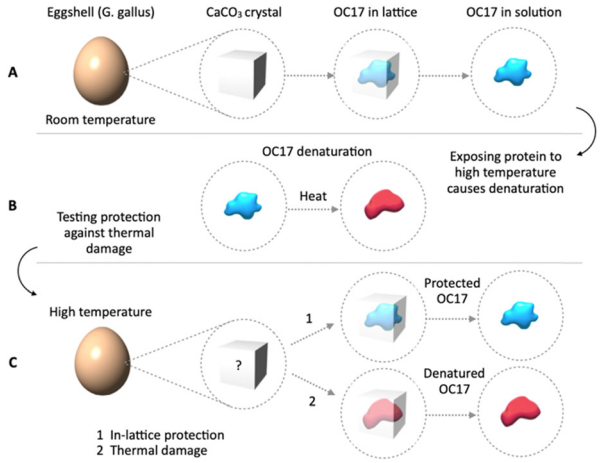

Fig. 1 (A) OC17 is an intracrystalline protein incorporated into the $CaCO_3$ lattice. OC17 could be released from the lattice into solution by dissolving the eggshell. Incorporation into the lattice could protect OC17 against harsh environmental conditions such as high ambient temperature. (B) OC17 denatures if it is exposed to high temperatures for extended periods. As a natural intracrystalline protein, OC17 could offer insights into the protection of proteins incorporated into $CaCO_3$ crystals. (C) High temperature treatment of an eggshell provides an experimental system to test whether OC17 undergoes a structural change in the lattice during heat-treatment. OC17 extracted from a heat-treated eggshell is shown to resist heat-imposed denaturation of intracrystalline OC17 (path 1, see the text for details).

also discussed the influence of amino acids and related additives on calcite and vaterite.[49] Nonetheless, research on protein incorporation in $CaCO_3$ specifically for high-temperature protection has not attracted enough attention. There is also a gap in our knowledge about whether incorporated proteins in $CaCO_3$ are structurally intact after exposure to high temperatures or after reconstitution from lattice using dissolution agents. It could be insightful to probe intracrystalline proteins for temperature stability. Studying natural intracrystalline proteins may provide a method to address temperature-sensitivity of therapeutics.

Here, we probe the effect of heat on intracrystalline proteins of eggshells, particularly focusing on OC17 to investigate whether lattice incorporation protects the protein against thermal damage (Fig. 1C). In order to avoid using harsh reconstitution solutions, we compare acid dissolution and metal chelation to optimize protein release from crystals. Then, we probe secondary and tertiary structures of a model protein (BSA) after incubation in a metal chelating agent which we show is convenient for reconstitution. Although wild-type OC17 could be extracted from eggshells, it is also desirable to recombinantly express it which would enable engineering OC17 not only for expressing fusion proteins (*e.g.* for assistive incorporation, such as GFP-OC17) but also for probing the function of OC17, which is yet to be demonstrated. For these reasons, we optimize conditions for recombinant expression of OC17. We also show an alternative method of wild-type OC17 extraction which is beset with the existence of intracrystalline lysozymes.[50] Studying natural intracrystalline proteins such as OC17, may help address the temperature-sensitivity problem of protein-based therapeutics. Structural studies probing intracrystalline proteins that are exposed to high ambient temperatures might be beneficial for developing nature-inspired methods to protect, store and deliver proteins outside the cold chain.

Results and discussion

Thermogravimetric analysis (TGA) of eggshell intracrystalline fraction

Thermal degradation of intracrystalline fraction can be probed using TGA which allows quantifying the amount of incorporated material. Because $CaCO_3$ melts at significantly higher temperatures compared to the degradation of biogenic intracrystalline fraction, the percentage of total incorporated content in $CaCO_3$ could be determined. In addition, the temperature at which a substantial weight change occurs is thought to be correlated with the effect of heat on intracrystalline proteins. We studied eggshells to investigate if protein degradation could be inferred during a heat-treatment step in a typical TGA experiment. As a negative control, we used commercial $CaCO_3$ powder. The results are shown in Fig. 2. The weight of commercial $CaCO_3$ did not change from 100 °C up to 500 °C which shows that the difference between pure $CaCO_3$ and eggshell arises from incorporated organic fraction in eggshells. From 100 °C to 300 °C, there is a slight decrease in the weight of the eggshell samples which is thought to be either loosely bound H_2O in crystals[51] or amorphous $CaCO_3$[52] which may be present in varying amounts in eggshells. The TGA data suggest that up to 250 °C, intracrystalline proteins are largely unaffected by heat or protein degradation is very low to be detected through weight change. Starting roughly from 250 °C, the observed decrease in weight is most likely related to protein integrity. This change is thought to arise from CO_2

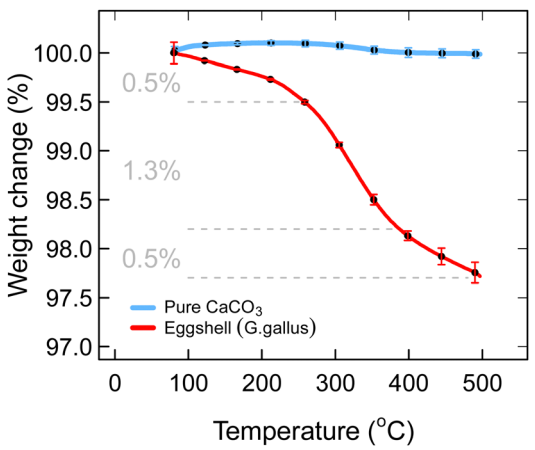

Fig. 2 Thermogravimetric analysis of eggshell crystals. Samples were heated from room temperature to 500 °C. Comparison of eggshell with pure $CaCO_3$ (blue line) shows the presence of intracrystalline materials in biogenic $CaCO_3$ (red line). The decrease in weight in the eggshell samples indicates that the effect of heat on intracrystalline material starts to a large extent only after around 250 °C. Initial change in the weight below 250 °C possibly arises from surface-bound and lattice-bound H_2O (0.5% from 100 °C to 250 °C). Note that temperatures used in the analysis are much lower than the melting point of $CaCO_3$. Note that some of the errors are too low to be distinguished.

release from the organic content.[53] It is important to note that changes in the host lattice arising from protein degradation should be taken into account on which there is a significant body of work.[29–37] After 400 °C, the rate of weight loss slightly decreases, though we did not continue analysis above 500 °C after which unrelated changes such as $CaCO_3$ transition into CaO,[54] would interfere the protein-focused analysis. The fact that purified proteins in solution degrade at much lower temperatures than 250 °C suggests that intracrystalline proteins may be protected against thermal damage as long as they are incorporated into the $CaCO_3$ lattice. Because the focus of this study is to see whether lattice incorporation might be a viable method for protein protection as described above, it might be interesting to compare the protection of intracrystalline lysozyme in eggshells with *in vitro* incorporated lysozyme. In fact, *in vitro* incorporated lysozyme has recently been shown to resist high temperature damage in crystals.[38] To avoid diverging from the main topic, we focused solely on OC17 to see if a natural intracrystalline protein is protected in the $CaCO_3$ lattice rather than comparing *in vivo versus in vitro* incorporation.

Purification and identification of OC17 for subsequent analyses

Protocols for purification of wild-type OC17 from eggshells have been reported along with the crystal structure of the protein,[9] however, the protocols[10] require organic solvents (*e.g.* acetonitrile and trifluoroacetic acid in the HPLC step) which might denature the protein. Instead of using harsh purification media, we used a batch extraction method to purify intact, properly folded OC17 for subsequent heat-treatment tests. Eggshell membranes and surface-bound chemicals (*e.g.* cuticle, natural dye molecules, *etc.*) were removed in an EDTA washing step and insoluble materials were filtered. Then, eggshells were ground into powder and dissolved in 5% acetic acid.[55,56] After an initial buffer exchange to remove acetic acid in a CM Sepharose column, step-wise pH gradient was used to elute OC17, OC23[57] (glycosylated form of OC17) and lysozyme (Fig. 3A). We did not seek further purification of contaminant proteins as their concentrations are negligible as seen in the gel image (Fig. 3B). We verified the proteins using peptide fingerprinting by LC-MS on gel-excised samples corresponding to OC17 and lysozyme. The high coverage of amino acid sequences for OC17[58] and lysozyme clearly identifies each protein (Fig. 3C). It is worth noting that removing lysozyme from OC17 is non-trivial, however, it could be removed in a number of ways. Previous protocols utilized size-exclusion steps in tandem to remove lysozyme and OC17.[10] We developed an alternative method using a protein–protein interaction between lysozyme and ivy protein. Ivy had likely evolved in certain bacteria as a defence mechanism against lysozyme.[59] Ivy binds lysozyme with high affinity which disarms lysozyme's enzymatic function and protects the integrity of the bacterial cell wall.[60] We recombinantly tagged ivy with histidine to bind lysozyme to the column through his-tagged ivy which allows eluting OC17 in a single step.

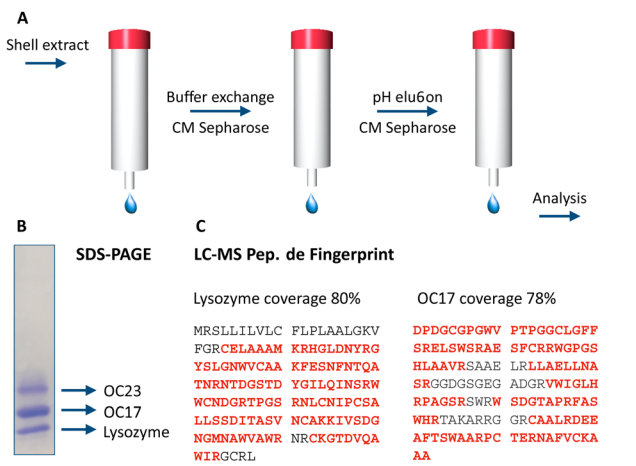

Fig. 3 (A) Extraction of wild-type OC17 from shell extract for heat-treatment experiments. After filtering insoluble fraction, the CM Sepharose medium was used in tandem for initial buffer exchange and subsequent stepwise pH gradient (see the text for details on comparison with previously published methods of OC17 purification). (B) Gel image shows the elution fraction containing lysozyme, OC23 (glycosylated OC17) and OC17. (C) Protein bands excised from the gel were analyzed with LC-MS to verify OC17 and lysozyme sequences. It is important to distinguish OC17 and lysozyme because the molecular weights of proteins are close. Red letters in the amino acid sequences show the residues that matched with sequences of each protein in peptide fingerprinting.

Probing lattice protection of OC17 at high temperatures

We investigated the effect of heat on intracrystalline OC17 by comparing the protein extracted from room temperature-stored eggshells to the protein extracted from the heat-treated eggshell. In other words, intracrystalline protein was first extracted from eggshells and then heat-treated in solution for analysis. For comparison, the eggshell was first heat-treated and then extraction of the protein was carried out from this heat-treated sample. In this way, the protective effect of lattice incorporation could be studied. For this reason, we first extracted OC17 from room temperature-stored eggshells and used circular dichroism (CD) for structural analysis (red line in Fig. 4). Then, the OC17 solution was heat-treated at 50 °C for 60 h. After treatment, the same sample was analysed again to see the effect of heat (black line in Fig. 4). Heat-treated OC17 undergoes a structural change because of heat denaturation. To probe the effect of heat on intracrystalline OC17, we first heat-treated eggshells at 50 °C for 60 h and after heat-treatment, we extracted OC17 and analysed its structure. The CD result shows that OC17 from heat-treated eggshells (blue line in Fig. 4) have almost the same secondary structure compared to OC17 extracted from room temperature-stored eggshells (red line in Fig. 4). The result indicates that the incorporation of OC17 into $CaCO_3$ lattice protects the protein against thermal damage at least at 50 °C for as long as 60 h. Because we focus on protecting proteins using an *in vitro* incorporation method, we chose experimental conditions in a way that the temperature and the time frame (50 °C, 60 h) is relevant to real-life situations, and we did not test higher temperatures for longer periods. It should be noted that the effect of heat on the lattice incorporated OC17 may be extended to other target proteins. In other

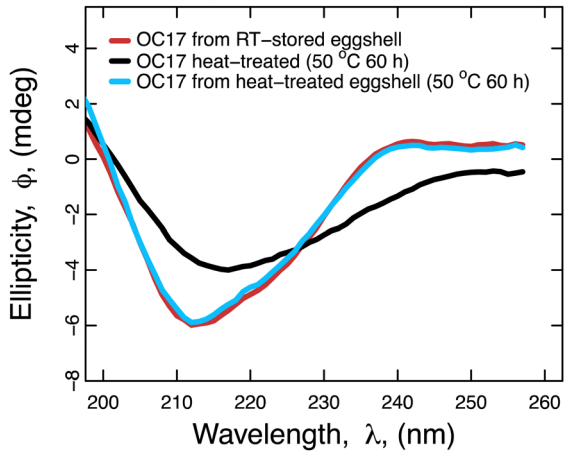

Fig. 4 Analysis of the effect of high temperature on the OC17 structure. CD spectrum of OC17 was collected after purification of the protein from room temperature-stored eggshells (red line). Then, the protein solution was heat-treated at 50 °C for 60 h. The spectrum after heat-treatment shows that at 50 °C, thermal damage causes structural change (black line). Next, eggshells were heat-treated at 50 °C for 60 h before any extraction is performed, to test whether OC17 is protected in the crystals against thermal damage under the same heat treatment (50 °C for 60 h). The superposition of OC17 spectra from room-temperature eggshells (red line) and heat-treated eggshells (blue line) suggests that incorporation into lattice provides protection to OC17 against thermal damage.

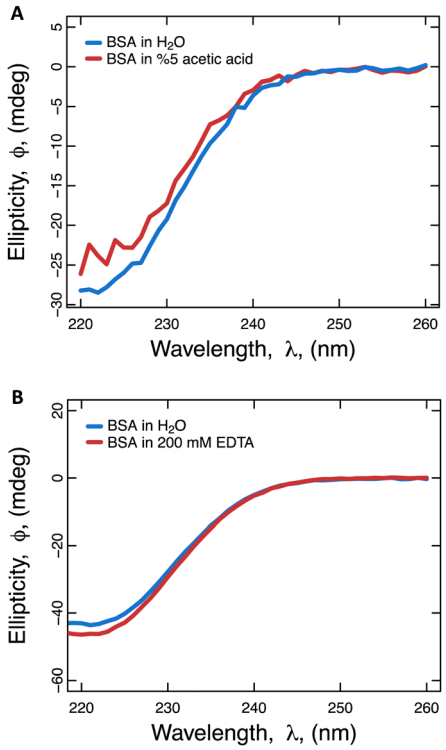

Fig. 5 CD spectra showing the effect of reconstitution solution on the model protein structure. The common method of dissolving synthetic $CaCO_3$ crystals with an incorporated target is to use an acidic solution to release as much intracrystalline content as possible in a short period of time. The caveat of using acid solution is the risk of target denaturation at low pH. The comparison of (A) 5% acetic acid (pH 5) and (B) 200 mM EDTA solutions on protein structure indicates that EDTA avoids protein denaturation. The increasing difference towards 220 nm compared to positive control (BSA in H_2O) suggests that acetic acid triggers partial unfolding. A decrease in ellipticity in acetic acid compared to EDTA may also reflect changes in protein structure. Note that acetic acid interferes strongly with CD signal below 220 nm. For extended spectra of BSA in EDTA see Fig. 6.

words a given target protein could be probed for protection efficiency in the lattice. Because the overarching aim here is to investigate the in-lattice protection of a natural intracrystalline protein, we did not extend the analysis further with different target proteins.

Probing the effect of dissolution agents on releasing intracrystalline proteins

Lattice incorporated proteins need to be released from $CaCO_3$ crystals for practical applications. If the incorporated compound is not a protein (*e.g.* a small molecule drug), then it could be taken out of the crystal using one of the $CaCO_3$ dissolution methods. It is worth noting that incorporated small molecule drugs could be taken orally because $CaCO_3$ would be dissolved in the acidic environment of the stomach. If proteins are incorporated, then releasing these targets needs to be done in a way that does not denature proteins during the reconstitution step.

There are two main ways to dissolve $CaCO_3$: acid dissolution and metal chelation. The advantage of acid dissolution is to release intracrystalline content rapidly, however, the caveat is the risk of denaturing incorporated protein at low pH. The advantage of metal chelation is that it is possible to dissolve $CaCO_3$ at neutral pH, thereby avoiding the risk of low pH denaturation. The disadvantage of metal chelation is the slow dissolution rate. We emulated intracrystalline protein releasing step by incubating a model protein, BSA, in either diluted acetic acid or in diluted EDTA.[55,56] Fig. 5A shows the CD spectra of BSA incubated in 5% acetic acid along with a positive control (BSA in H_2O) at the same concentration. The difference in the acetic acid-incubated sample from positive control particularly at small wavelengths indicates that acidic solution causes rearrangements in the protein structure at least to some extent, although the pH of the solution is adjusted to 5. In addition, a decrease in the CD signal compared to BSA in EDTA could be attributed to partial denaturation. In contrast, BSA incubated in 200 mM EDTA solution shows only negligible difference from the positive control (BSA in H_2O). Because metal chelation seems to be superior to acetic acid, we further investigated the effect of EDTA on the model protein. Fig. 6A shows extended CD spectra of EDTA-incubated BSA (*i.e.* extended data from Fig. 5B). The CD signal of the sample superposes with positive control below 210 nm and above 225 nm. The slight difference between roughly 210 and 225 nm does not indicate structural changes of the target, however, to ensure that EDTA incubation does not affect the protein, we probed the 3D structure of the sample using intrinsic tryptophan fluorescence spectroscopy. Tryptophan residues are highly sensitive to structural changes. As a complementary analysis to CD, which provides information on secondary structure, tryptophan fluorescence could reveal changes in the tertiary structure. As is

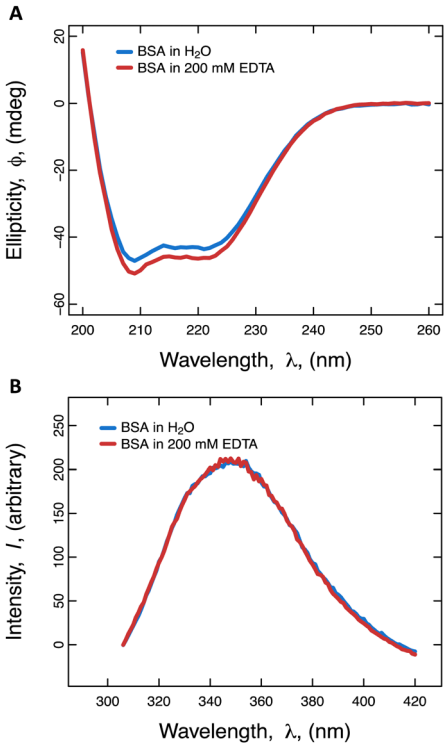

Fig. 6 Extended data for the effect of EDTA on model protein BSA. (A) CD spectra of BSA in 200 mM EDTA compared to the positive control (BSA in H_2O). Superposition of the CD signal from 200 nm to 260 nm suggests that the effect of EDTA on protein structure is negligible. (B) Intrinsic tryptophan fluorescence spectra of the same BSA samples. Intrinsic tryptophan fluorescence is highly sensitive to changes in the 3D structure and complementary to CD analysis because compared to changes in the secondary structure, tryptophan fluorescence provides information about the tertiary structure. Note that the difference in the CD data between 210 nm and 225 nm is lacking in the overall 3D structure (*i.e.* no difference between fluorescence of BSA in H_2O and in EDTA).

seen in Fig. 6B, EDTA-incubated BSA shows no difference from positive control which suggests that dissolving $CaCO_3$ in EDTA solution is suitable to release incorporated proteins at neutral pH without denaturing the protein structure. In addition, EDTA dissolution also offers a way to study the effect of incorporation itself on the protein structure (*i.e.* the structure of the same protein before and after incorporation) because any difference after releasing the target from the crystal lattice could be attributed to incorporation-related effects. It is reasonable to use EDTA dissolution despite its slower dissolution rate compared to acidic solution. An exception could be a target that is resistant to unfolding under low pH conditions.

Cloning and optimization of recombinant OC17 expression

Although OC17 could be extracted from eggshells, it would be desirable to express it recombinantly for a number of reasons. Because the function of OC17 has been proposed to regulate amorphous-to-crystal transition during shell formation, recombinant expression could allow designing experiments with engineered protein to verify molecular simulations.[1] For instance, site-directed mutagenesis could help revealing residues that were proposed to be key for interaction with the growing crystal.[2,3] Recombinant OC17 not only offers an experimental system to investigate the function of the protein but it could also advance our understanding of the amorphous-to-crystal transition in biogenic minerals. In addition, recombinant expression could allow further engineering of the protein such as insertion of small molecules into OC17 through non-canonical amino acids. Labelling OC17 might be beneficial to conclusively show that the protein binds to amorphous $CaCO_3$ and detaches once the structure transforms into calcite. Furthermore, from a practical perspective, genetic fusion of a target protein to OC17 could incorporate both proteins at once if the target cannot incorporate itself. Although an eggshell is a low-cost material, recombinant expression could yield high concentrations of protein unaffected by incorporation-related (if any) structural changes. For these reasons, we developed and optimized the expression of OC17 in *E. coli*. We designed primers to amplify the OC17 sequence with overhanging regions that target the insert site for Gibson assembly (Fig. 7A). After verifying the plasmid sequence, we tested protein expression in two different media: LB and TY supplemented with 1% glucose because of the effect of glucose on expression using pET plasmids.§ At two temperatures and different IPTG concentrations, we found that there is almost no expression in the TY medium at 25 °C (see the dot blot image in the ESI†). The pelB sequence was included in the construct for secretion of the target protein into the periplasmic space, however, low concentrations of OC17 were detected in the medium after overnight incubation at 30 °C, probably because of the small size of OC17. In subsequent analyses, both periplasmic fraction and medium were tested for the target protein. Expression in LB medium was found to be more efficient, however, at 30 °C, there is no expression above 0.1 mM IPTG. Expression markedly increases at 25 °C. We further tested lower IPTG concentrations (0.01, 0.025, 0.05, 0.1, 0.25 and 0.5 mM) to see if expression could be enhanced (Fig. 7B). His-tagged OC17 samples were analysed with western blot using anti-Histag antibodies (Fig. 7B). The highest expression was observed with 0.01 mM IPTG in the periplasmic space, however, because the protein also leaks into the medium at 0.05 mM, collection of both medium and periplasmic space fractions at this concentration would be reasonable.

As described in the previous sections, it would be illuminating to show OC17 in action, although, to our knowledge, there is no such report. The absence of experimental verification of simulations probably arises from the difficulty of designing a setup with which OC17–$CaCO_3$ interaction could be investigated directly in real-time. With recombinant OC17 at hand, it can be labelled in a way that allows monitoring the OC17–$CaCO_3$ interaction, for instance using FLIM-FRET, with which we hope to show OC17 in action.

§ R. Novy and B. Morris, Use of glucose to control basal expression in the pET system, *inNovations*, 2001, **13**, 8–10. https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/386/947/innovations-013-nvg.pdf.

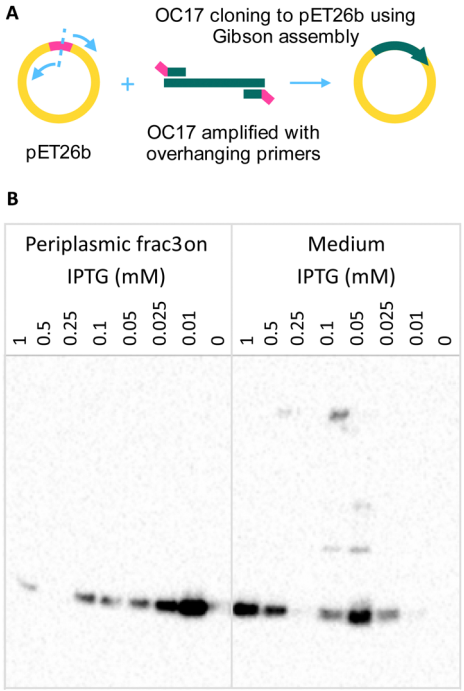

Fig. 7 (A) Cloning OC17 into pET26b using Gibson assembly for subsequent recombinant expression. OC17 DNA was amplified with overhanging primers designed for assembly into a plasmid with a 6XHisTag at the N-terminal. (B) Western blot of OC17 expression. Optimization of expression in *E. coli* BL21 cells was tested in the LB growth medium at 25 °C. Although the pelB sequence in the plasmid directs the protein into the periplasmic space, leakage into the medium was also tested. As seen in the image, OC17 leaks into the medium possibly because of the small size of the protein (15.3 kDa). A range of IPTG concentrations were used to test the higher level of expression. OC17 was detected using anti-Histag antibody.

Conclusions

Protein incorporation into $CaCO_3$ lattice might address the protection of temperature-sensitive proteins for long periods outside the cold-chain. The structural characterization of natural intracrystalline proteins for their heat-resistance might be beneficial. These proteins could be attractive and accessible models. As a step towards this goal, we focused on eggshells which are not only a low-cost, ubiquitous biomineral but also contain one of the intriguing intracrystalline proteins, OC17, reported to date. We investigated whether or not lattice incorporation provides heat-resistance to eggshell protein OC17, the study of which could inspire efficient synthetic incorporation pathways. Apart from loading target proteins in $CaCO_3$, an equally important step is to release the target protein into solution without affecting the function of the protein. For this reason, we studied the effect of $CaCO_3$ dissolution agents on a model protein. Compared to acid dissolution, we found that metal chelation using EDTA is superior because it allows dissolving $CaCO_3$ at neutral pH, though it takes longer than acid dissolution. Having recombinant OC17 in hand, it would be possible to design experiments to show the protein in action which could be interesting based on the proposed function of OC17. Looking from a practical perspective, if extended to other molecules such as nucleic acids, lattice incorporation might be applied to different biological compounds. The temperature-sensitive nature of therapeutics is a global health problem, unveiled again with challenges in vaccine delivery during the COVID-19 pandemic. Incorporation of protein-based vaccines in $CaCO_3$ may offer an alternative way to keep vaccines safe at high ambient temperatures as long as needed. Intracrystalline proteins such as OC17 may offer insights for developing biomimetic, low-cost and effective vaccine stabilization technology.

Experimental

Materials

The *E. coli* K-12 MG 1655 cells were donated by Dr Gillian Fraser from the Department of Pathology, University of Cambridge, UK. The primers and plasmid pET26b were purchased from Sigma Aldrich (Darmstadt, Germany). The XL1-Blue Competent cells were purchased from Stratagene, USA. The Q5 High-Fidelity 2X Master Mix was purchased from New England Biolabs, UK. The 2X ReddyMix PCR Master Mix and FastDigest DpnI were purchased from Thermo Scientific, UK. PCR purification, gel extraction, plasmid purification kits were purchased from the QIAGEN, UK. The OC17 gene was synthesized by Invitrogen Thermo Scientific, UK. LB growth medium, agar plates, and buffer solutions were obtained from the Department of Haematology, University of Cambridge, UK. All purification columns were purchased from GE Healthcare, UK, unless otherwise stated. The SDS-PAGE gels were purchased from Life Technologies., UK. The western blot membranes and ultrafiltration centrifugal filters were purchased from Merck Millipore, UK. All chemicals for buffer preparations, cell culture, protein purification, and proteins (chicken egg white lysozyme and lyophilized powder) and bovine serum albumin (lyophilized powder) were purchased from Sigma Aldrich (Darmstadt, Germany), unless otherwise stated. All solutions were prepared with deionized water and pH adjustment was done with concentrated HCl or NaOH if needed. The solutions were filtered either through vacuum filtration with 0.45 or 0.2 μm filter papers or using syringe-driven filters. Buffer exchange of the protein solutions was done using Amicon Ultra Centrifugal Filters (Merck, Darmstadt, Germany) along with a Sorvall ST 8 bench-top centrifuge (Thermo Scientific, Massachusetts, USA).

Methods

Circular dichroism and intrinsic tryptophan fluorescence measurements for structural analysis. CD measurements were performed in quartz cuvettes from Starna Scientific (Essex, UK). The concentration of the proteins was measured with a microplate reader from BMG Labtech (Ortenberg, Germany) at 280 nm. The pIs and molecular extinction coefficients were calculated theoretically using amino acid sequences of the proteins using ExPASy ProtParam or Innovagen PepCalc software. Intrinsic tryptophan fluorescence measurements were performed using a microplate reader BMG Labtech (Ortenberg, Germany) with excitation of the samples in the UV range.

Recording of emission spectra was performed between 300 and 420 nm.

LC/MS analysis for peptide fingerprinting. LC-MS analysis was carried out at the Cambridge Centre for Proteomics (CCP) using Thermo Scientific Lumos and Dionex 3000 RSLCnano (Massachusetts, USA) at the Department of Biochemistry. MASCOT LC-MS data analysis software (Matrix Sciences, Boston, USA) was used to match the amino acid sequences of the target proteins from the National Centre for Biotechnology Information, Basic Local Alignment Search Tool (NCBI BLAST) search (Protein IDs or OC17 and lysozyme Q9PRS8 and P00698, respectively).

Thermal analysis of biological and commercial $CaCO_3$ samples. TGA was conducted on an ExStar SII TG/DTA7300 (SII NanoTechnology, Tokyo, Japan). Eggshells were briefly cleaned with diluted EDTA solution to remove both the shell membranes and surface bound proteins and protective layers such as cuticle and commercial chemicals. Then eggshells were washed with distilled H_2O and ground into powder manually using a pestle and mortar to facilitate the dissolution step. A stainless steel test sieve (Fisher Scientific, Leicestershire, UK) with a mesh size of 90 μm was used to collect the eggshell sample. Particles bigger than 90 μm were discarded. Approximately 15 mg powdered eggshell or equivalent amount of pure commercial $CaCO_3$ powder (Sigma Aldrich, Darmstadt, Germany) were used for TGA analysis between RT to 500 °C at a heating rate of 10 °C min^{-1}.

Cloning of the OC17 gene and expression of the OC17 protein. The cDNA sequence of OC17 was obtained from the GenBank (KF835610). The primers used are given in the ESI.† The plasmid and OC17 DNA fragments were amplified in PCR using Q5 High-Fidelity 2X Master Mix according to the manufacturer's instructions. The band containing the PCR product was excised and extracted from agarose gel using a QIAquick Gel Extraction Kit. Cloning was performed using Gibson assembly and transformation was done to competent cells using the manufacturer's guidelines. To select the cells containing the plasmid, cells were grown overnight at 37 °C. The plasmid was verified by sequencing. (Genewiz, Cambridge, UK). The sequence verified plasmids were transformed into Tuner DE3 competent cells for expression and two 10 mL overnight cultures (LB and TY medium containing 1% glucose) with 50 μg mL^{-1} kanamycin were grown at 30 °C with constant shaking at 200 rpm. The following day, 50 mL media were inoculated with 0.5 mL overnight culture. The new cultures were grown at 37 °C until optical density at 600 nm (OD600) reaches 0.5–0.6. Different IPTG concentrations (0, 0.1, 0.25, 0.5 and 1 mM) were tested to find the optimum expression conditions. Following induction, cultures were incubated either at 30 °C or at 25 °C overnight. The following day, 2 mL of culture was centrifuged at 8000 rpm and the pellets of the culture were resuspended in B-PER (bacterial protein extraction reagent) containing 0.2 mg ml^{-1} lysozyme, 5UI DNase I, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$, and incubated at RT for 30 min followed by centrifugation. Because the B-PER reagent is a cell lysis medium that solubilize proteins, the whole cell protein extract was used for the dot blot analysis of OC-17. After medium selection for optimum expression, IPTG concentration was optimized for LB medium in order to find whether lower concentrations of IPTG (below 0.1 mM) would result in a higher expression level. The same procedure was repeated as above except that 1 mL of overnight culture was used for inoculation and incubated at 30 °C only until optical density at 600 nm (OD600) reaches 0.5–0.6. After IPTG addition, the temperature was decreased to 25 °C and the cultures were grown overnight. The cells were then harvested by centrifugation at 4000 × *g* for 20 min and the medium was collected. To extract the periplasmic proteins only, pellets were resuspended in 30 mM tris–HCl containing 20% sucrose, pH 8.0. The cells were kept on ice and ice-cold 500 mM EDTA was added at a final concentration of 1 mM. The solution was then centrifuged at 4000 × *g* for 20 min at 4 °C. The supernatant was discarded and pellets were resuspended in 5 mM $MgSO_4$ and incubated on ice for 10 min. The solution containing periplasmic proteins was centrifuged at 4000 × *g* for 20 min at 4 °C. Following periplasmic protein extraction, fractions were tested with SDS-PAGE and western blot immunodetection against anti-Histag antibody HRP (dilution 1:5000) to verify OC17 extraction.

Author contributions

H. B. C. and F. I. U. conceptualized the project, developed the methodology, designed and performed the experiments and wrote the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

H. B. C. would like to thank the Cambridge Commonwealth, European & International Trust, the Department of Engineering, the Nanoscience Centre, and Prof Jim Huntington from the Cambridge Institute for Medical Research, Department of Haematology, Cambridge, UK.

Notes and references

1 C. L. Freeman, J. H. Harding, D. Quigley and P. M. Rodger, *Angew. Chem., Int. Ed.*, 2010, **49**, 5135–5137.
2 C. L. Freeman, J. H. Harding, D. Quigley and P. M. Rodger, *J. Phys. Chem. C*, 2011, **115**, 8175–8183.
3 C. L. Freeman, J. H. Harding, D. Quigley and P. M. Rodger, *Phys. Chem. Chem. Phys.*, 2015, **17**, 17494–17500.
4 A. B. Rodriguez-Navarro, P. Marie, Y. Nys, M. T. Hincke and J. Gautron, *J. Struct. Biol.*, 2015, **190**, 291–303.
5 L. Stapane, N. Le Roy, J. Ezagal, A. B. Rodriguez-Navarro, V. Labas, L. Combes-Soia, M. T. Hincke and J. Gautron, *J. Biol. Chem.*, 2020, **295**, 15853–15869.
6 J. Gautron, L. Stapane, N. Le Roy, Y. Nys, A. B. Rodriguez-Navarro and M. T. Hincke, *BMC Mol. Cell Biol.*, 2021, **22**, 11.

7 M. Hincke, C. Tsang, M. Courtney, V. Hill and R. Narbaitz, *Calcif. Tissue Int.*, 1995, **56**, 578–583.
8 J. Reyes-Grajeda, D. Jauregui-Zuniga, A. Rodriguez-Romero, A. Hernandez-Santoyo, V. Bolanos-Garcia and A. Moreno, *Protein Pept. Lett.*, 2002, **9**, 253–257.
9 J. Reyes-Grajeda, A. Moreno and A. Romero, *J. Biol. Chem.*, 2004, **279**, 40876–40881.
10 K. Mann and F. Siedler, *Biochem. Mol. Biol. Int.*, 1999, **47**, 997–1007.
11 Q. Zhang, L. Liu, F. Zhu, Z. Ning, M. T. Hincke, N. Yang and Z. Hou, *PLoS One*, 2014, **9**(3), e93452.
12 K. Drickamer, *Curr. Opin. Struct. Biol.*, 1999, **9**, 585–590.
13 A. Zelensky and J. Gready, *FEBS J.*, 2005, **272**, 6179–6217.
14 Y. Nys, J. Gautron, J. Garcia-Ruiz and M. Hincke, *C. R. Palevol*, 2004, **3**, 549–562.
15 M. T. Hincke, Y. Nys, J. Gautron, K. Mann, A. B. Rodriguez-Navarro and M. D. McKee, *Front. Biosci.-Landmark*, 2012, **17**, 1266–1280.
16 G. Sykes, M. Collins and D. Walton, *Org. Geochem.*, 1995, **23**, 1059–1065.
17 S. Weiner, H. A. Lowenstam and L. Hood, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**, 2541–2545.
18 M. Salamon, N. Tuross, B. Arensburg and S. Weiner, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13783–13788.
19 C. L. Oskam, J. Haile, E. Mclay, P. Rigby, M. E. Allentoft, M. E. Olsen, C. Bengtsson, G. H. Miller, J.-L. Schwenninger, C. Jacomb, R. Walter, A. Baynes, J. Dortch, M. Parker-Pearson, M. T. P. Gilbert, R. N. Holdaway, E. Willerslev and M. Bunce, *Proc. R. Soc. B*, 2010, **277**, 1991–2000.
20 B. Demarchi, S. Hall, T. Roncal-Herrero, C. L. Freeman, J. Woolley, M. K. Crisp, J. Wilson, A. Fotakis, R. Fischer, B. M. Kessler, R. R. Jersie-Christensens, J. V. Olsen, J. Haile, J. Thomas, C. W. Marean, J. Parkington, S. Presslee, J. Lee-Thorp, P. Ditchfield, J. F. Hamilton, M. W. Ward, C. M. Wang, M. D. Shaw, T. Harrison, M. Dominguez-Rodrigo, R. D. E. MacPheel, A. Kwekason, M. Ecker, L. K. Horwitz, M. Chazan, R. Kroger, J. Thomas-Oates, J. H. Harding, E. Cappellini, K. Penkman and M. J. Collins, *eLife*, 2016, **5**, e17092.
21 B. Demarchi, M. Mackie, Z. Li, T. Deng, M. J. Collins and J. Clarke, *eLife*, 2022, **11**, e82849.
22 M. Crisp, B. Demarchi, M. Collins, M. Morgan-Williams, E. Pilgrim and K. Penkman, *Quat. Geochronol.*, 2013, **16**, 110–128.
23 G. Miller, P. Beaumont, H. Deacon, A. Brooks, P. Hare and A. Jull, *Quat. Sci. Rev.*, 1999, **18**, 1537–1548.
24 Y.-Y. Kim, J. D. Carloni, B. Demarchi, D. Sparks, D. G. Reid, M. E. Kunitake, C. C. Tang, M. J. Duer, C. L. Freeman, B. Pokroy, K. Penkman, J. H. Harding, L. A. Estroff, S. P. Baker and F. C. Meldrum, *Nat. Mater.*, 2016, **15**, 903.
25 G. Magnabosco, M. Di Giosia, I. Polishchuk, E. Weber, S. Fermani, A. Bottoni, F. Zerbetto, P. G. Pelicci, B. Pokroy, S. Rapino, G. Falini and M. Calvaresi, *Adv. Healthcare Mater.*, 2015, **4**, 1510–1516.
26 M. Di Giosia, I. Polishchuk, E. Weber, S. Fermani, L. Pasquini, N. M. Pugno, F. Zerbetto, M. Montalti, M. Calvaresi, G. Falini and B. Pokroy, *Adv. Funct. Mater.*, 2016, **26**, 5569–5575.
27 Y.-Y. Kim, L. Ribeiro, F. Maillot, O. Ward, S. J. Eichhorn and F. C. Meldrum, *Adv. Mater.*, 2010, **22**, 2082.
28 Y.-Y. Kim, A. S. Schenk, D. Walsh, A. N. Kulak, O. Cespedes and F. C. Meldrum, *Nanoscale*, 2014, **6**, 852–859.
29 B. Pokroy, A. N. Fitch, F. Marin, M. Kapon, N. Adir and E. Zolotoyabko, *J. Struct. Biol.*, 2006, **155**, 96–103.
30 E. Zolotoyabko, *Adv. Mater. Interfaces*, 2017, 4.
31 B. Pokroy, J. P. Quintana, E. N. Caspi, A. Berner and E. Zolotoyabko, *Nat. Mater.*, 2004, **3**, 900–902.
32 S. Borukhin, L. Bloch, T. Radlauer, A. H. Hill, A. N. Fitch and B. Pokroy, *Adv. Funct. Mater.*, 2012, **22**, 4216–4224.
33 I. Ben Shir, S. Kababya, I. Katz, B. Pokroy and A. Schmidt, *Chem. Mater.*, 2013, **25**, 4595–4602.
34 E. Seknazi and B. Pokroy, *Adv. Mater.*, 2018, **30**, 1707263.
35 E. Seknazi, S. Mijowska, I. Polishchuk and B. Pokroy, *Inorg. Chem. Front.*, 2019, **6**, 2696–2703.
36 M. Różycka, I. Coronado, K. Brach, J. Olesiak-Bańska, M. Samoć, M. Zarębski, J. Dobrucki, M. Ptak, E. Weber, I. Polishchuk, B. Pokroy, J. Stolarski and A. Ożyhar, *Chem. – Eur. J.*, 2019, **25**, 12740–12750.
37 S. Mijowska, I. Polishchuk, A. Lang, E. Seknazi, C. Dejoie, S. Fermani, G. Falini, N. Demitri, M. Polentarutti, A. Katsman and B. Pokroy, *Chem. Mater.*, 2020, **32**, 4205–4212.
38 O. Nahi, A. N. Kulak, T. Kress, Y.-Y. Kim, O. G. Grendal, M. J. Duer, O. J. Cayre and F. C. Meldrum, *Chem. Sci.*, 2021, **12**, 9839–9850.
39 A. N. Kulak, M. Semsarilar, Y.-Y. Kim, J. Ihli, L. A. Fielding, O. Cespedes, S. P. Armes and F. C. Meldrum, *Chem. Sci.*, 2014, **5**, 738–743.
40 Y. Ning, L. Han, M. Douverne, N. J. W. Penfold, M. J. Derry, F. C. Meldrum and S. P. Armes, *J. Am. Chem. Soc.*, 2019, **141**, 2481–2489.
41 Y. Ning, L. Han, M. J. Derry, F. C. Meldrum and S. P. Armes, *J. Am. Chem. Soc.*, 2019, **141**, 2557–2567.
42 A. N. Kulak, P. Yang, Y.-Y. Kim, S. P. Armes and F. C. Meldrum, *Chem. Commun.*, 2014, **50**, 67–69.
43 Y.-Y. Kim, K. Ganesan, P. Yang, A. N. Kulak, S. Borukhin, S. Pechook, L. Ribeiro, R. Kroeger, S. J. Eichhorn, S. P. Armes, B. Pokroy and F. C. Meldrum, *Nat. Mater.*, 2011, **10**, 890–896.
44 Y. Ning, D. J. Whitaker, C. J. Mable, M. J. Derry, N. J. W. Penfold, A. N. Kulak, D. C. Green, F. C. Meldrum and S. P. Armes, *Chem. Sci.*, 2018, **9**, 8396–8401.
45 Y. Ning, L. A. Fielding, L. P. D. Ratcliffe, Y.-W. Wang, F. C. Meldrum and S. P. Armes, *J. Am. Chem. Soc.*, 2016, **138**, 11734–11742.
46 B. Marzec, D. C. Green, M. A. Holden, A. S. Cote, J. Ihli, S. Khalid, A. Kulak, D. Walker, C. Tang, D. M. Duffy, Y.-Y. Kim and F. C. Meldrum, *Angew. Chem., Int. Ed.*, 2018, **57**, 8623–8628.
47 E. Weber, L. Bloch, C. Guth, A. N. Fitch, I. M. Weiss and B. Pokroy, *Chem. Mater.*, 2014, **26**, 4925–4932.
48 M. B. Al-Handawi, P. Commins, S. Shukla, P. Didier, M. Tanaka, G. Raj, F. A. Veliz, R. Pasricha, N. F. Steinmetz and P. Naumov, *Adv. Biosyst.*, 2018, **2**(5), e1700176.

49 C. A. Murray, P. M. Scientists, L. Holland, R. O'Brien, A. Richards, A. R. Baker, M. Basham, D. Bond, L. D. Connor, S. J. Day, J. Filik, S. Fisher, P. Holloway, K. Levik, R. Mercado, J. Potter, C. C. Tang, S. P. Thompson and J. E. Parker, *CrystEngComm*, 2024, **26**, 753–763.
50 M. Hincke, J. Gautron, M. Panheleux, J. Garcia-Ruiz, M. McKee and Y. Nys, *Matrix Biol.*, 2000, **19**, 443–453.
51 Z. Zou, X. Yang, M. Alberic, T. Heil, Q. Wang, B. Pokroy, Y. Politi and L. Bertinetti, *Adv. Funct. Mater.*, 2020, **30**, 2000003.
52 A. V. Radha, T. Z. Forbes, C. E. Killian, P. U. P. A. Gilbert and A. Navrotsky, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 16438–16443.
53 Y. Levi-Kalisman, S. Raz, S. Weiner, L. Addadi and I. Sagi, *J. Chem. Soc., Dalton Trans.*, 2000, 3977–3982.
54 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.
55 F. Marin and G. Luquet, *C. R. Palevol*, 2004, **3**, 469–492.
56 F. Marin, B. Pokroy, G. Luquet, P. Layrolle and K. De Groot, *Biomaterials*, 2007, **28**, 2368–2377.
57 K. Mann, *FEBS Lett.*, 1999, **463**, 12–14.
58 K. Mann and F. Siedler, *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.*, 2006, **143**, 160–170.
59 C. Abergel, V. Monchois, D. Byrne, S. Chenivesse, F. Lembo, J.-C. Lazzaroni and J.-M. Claverie, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 6394–6399.
60 V. Monchois, C. Abergel, J. Sturgis, S. Jeudy and J. Claverie, *J. Biol. Chem.*, 2001, **276**, 18437–18441.

