**Authors for correspondence:**
Manjula P. Mummadisetti
e-mail: mmpdsetti@gmail.com
Paul G. Falkowski
e-mail: falko@marine.rutgers.edu

## THE ROYAL SOCIETY
### PUBLISHING

# The spatial network of skeletal proteins in a stony coral

Manjula P. Mummadisetti[1], Jeana L. Drake[2,3] and Paul G. Falkowski[1,4]

[1]Environmental Biophysics and Molecular Biology Program, Department of Marine and Coastal Sciences, Rutgers, The State University of New Jersey, 71 Dudley Rd, New Brunswick, NJ 08901, USA
[2]Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, 595 Charles E. Young Drive East, Los Angeles, CA 90095, USA
[3]Department of Marine Biology, University of Haifa, 199 Aba Khoushy Avenue, Mount Carmel, Haifa 2498838, Israel
[4]Department of Earth and Planetary Sciences, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
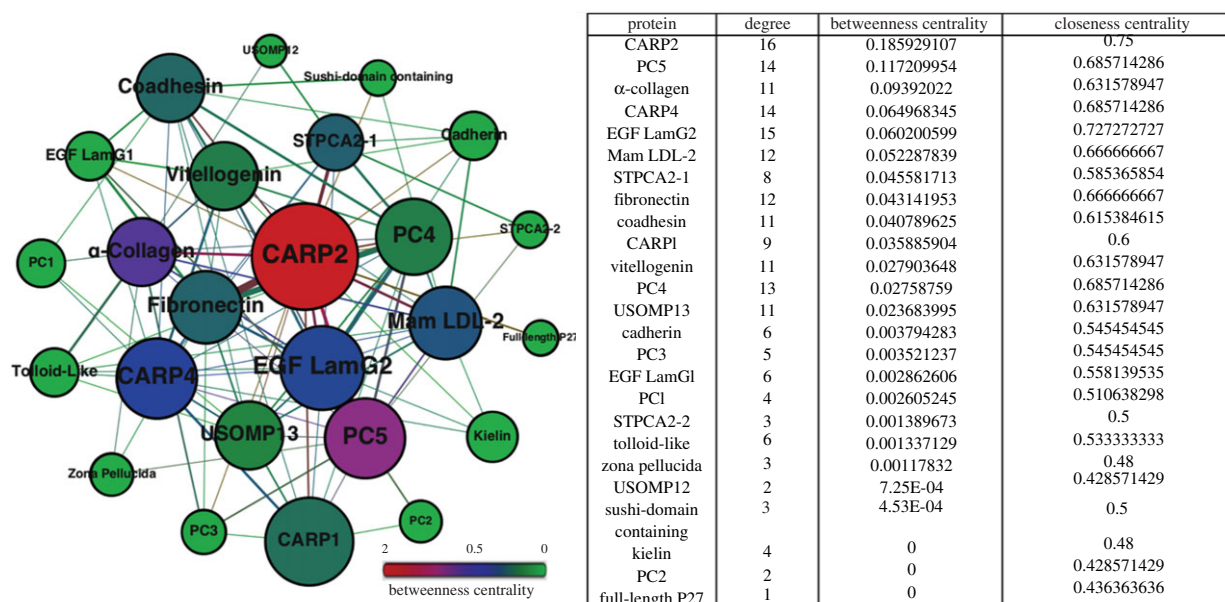
MPM, 0000-0001-8547-8123; JLD, 0000-0002-9851-379X; PGF, 0000-0002-2353-1969

Coral skeletons are materials composed of inorganic aragonitic fibres and organic molecules including proteins, sugars and lipids that are highly organized to form a solid biomaterial upon which the animals live. The skeleton contains tens of proteins, all of which are encoded in the animal genome and secreted during the biomineralization process. While recent advances are revealing the functions and evolutionary history of some of these proteins, how they are spatially arranged in the skeleton is unknown. Using a combination of chemical cross-linking and high-resolution tandem mass spectrometry, we identify, for the first time, the spatial interactions of the proteins embedded within the skeleton of the stony coral *Stylophora pistillata*. Our subsequent network analysis revealed that several coral acid-rich proteins are invariably associated with carbonic anhydrase(s), alpha-collagen, cadherins and other calcium-binding proteins. These spatial arrangements clearly show that protein–protein interactions in coral skeletons are highly coordinated and are key to understanding the formation and persistence of coral skeletons through time.

## 1. Introduction

Stony corals (phylum Cnidaria, class Anthozoa) evolved over 485 Mya [1,2], and over geological time formed massive reefs in tropical and subtropical seas. Despite their geological and ecological importance, the biomineralization process in these organisms is poorly understood. Further, a complete understanding of how they will respond to climate change requires a deeper knowledge of the biomineralization process than currently exists [3]. Over the past decade, a core set of more than 100 proteins has been identified in coral skeletons [4–7]. This 'skeletome' [8] appears to be highly conserved across coral taxa. At the nanoscale level, these proteins and other biomolecules, termed the extracellular matrix (ECM) of the skeleton or more specifically the skeletal organic matrix (SOM), are hypothesized to control the biomineral deposition process, shape, size and three-dimensional organization [2,9,10], as well as the mechanical properties and plasticity of the biomineral material [11].

Proteomic analysis of the skeletome has identified acid-rich, metal-binding, framework and adhesion proteins [2,4–7]. Of these, acid-rich proteins and carbonic anhydrases are the most highly characterized [2,12,13]. Coral acid-rich proteins (CARPs; alternatively called skeletal aspartic acid-rich proteins (SAARPs) [5]) can precipitate calcium carbonate from unamended seawater [12] and modify the mineral polymorph, or differing crystal structures for minerals with the same chemical composition [14]; in the case of stony corals, this is predominantly aragonite [15]. In addition, the coral calcification process uses

| protein | degree | betweenness centrality | closeness centrality |
|---|---|---|---|
| CARP2 | 16 | 0.185929107 | 0.75 |
| PC5 | 14 | 0.117209954 | 0.685714286 |
| α-collagen | 11 | 0.09392022 | 0.631578947 |
| CARP4 | 14 | 0.064968345 | 0.685714286 |
| EGF LamG2 | 15 | 0.060200599 | 0.727272727 |
| Mam LDL-2 | 12 | 0.052287839 | 0.666666667 |
| STPCA2-1 | 8 | 0.045581713 | 0.585365854 |
| fibronectin | 12 | 0.043141953 | 0.666666667 |
| coadhesin | 11 | 0.040789625 | 0.615384615 |
| CARPl | 9 | 0.035885904 | 0.6 |
| vitellogenin | 11 | 0.027903648 | 0.631578947 |
| PC4 | 13 | 0.02758759 | 0.685714286 |
| USOMP13 | 11 | 0.023683995 | 0.631578947 |
| cadherin | 6 | 0.003794283 | 0.545454545 |
| PC3 | 5 | 0.003521237 | 0.545454545 |
| EGF LamGl | 6 | 0.002862606 | 0.558139535 |
| PCl | 4 | 0.002605245 | 0.510638298 |
| STPCA2-2 | 3 | 0.001389673 | 0.5 |
| tolloid-like | 6 | 0.001337129 | 0.533333333 |
| zona pellucida | 3 | 0.00117832 | 0.48 |
| USOMP12 | 2 | 7.25E-04 | 0.428571429 |
| sushi-domain containing | 3 | 4.53E-04 | 0.5 |
| kielin | 4 | 0 | 0.48 |
| PC2 | 2 | 0 | 0.428571429 |
| full-length P27 | 1 | 0 | 0.436363636 |

**Figure 1.** Interaction network of insoluble matrix proteins in *S. pistillata* skeleton. The network map shows individual proteins as nodes (circles) and their relations/interactions as edges (lines). These nodes and edges are scaled based on betweeness centrality, which measures the shortest paths among every pair of nodes in a network, which influences the colour of the node from green to red (green with the lowest interactions and red with the highest). The size of the node and the edge thickness depend on the number of connections between proteins.

soluble bicarbonate ions as the primary source of carbonate [12,16]. In the biomineralization process, carbon dioxide is produced. Hence carbonic anhydrases—both membrane-bound *Stylophora pistillata* carbonic anhydrase (STPCA) and skeletal STPCA2 [13,17]—allow bicarbonate replenishment through the rapid conversion of carbon dioxide to bicarbonate [18,19]. Once formed, the aragonite crystals must be cemented to each other and to the underlying skeleton. Proteins responsible for the attachment processes include collagens, cadherins and other calcium-binding and adhesion proteins [4,20]; however, the spatial organization of these molecules within the skeleton is completely unknown.

Chemical cross-linking combined with mass spectrometry sequencing is a well-established approach to understanding the spatial organization of proteins. This approach does not suffer from some of the problems encountered when using immunological methods and can be used to study spatially close skeletal proteins [21,22]. Here, we probed the spatial interactions in the Indo-Pacific stony coral *S. pistillata*, using a classical, well-characterized cross-linker, bis(sulfosuccinimi-dyl)suberate (BS3), which tethers primary amines up to 11.4 Å apart in the intact skeleton [23]. Our results, across both experimental and biological replicates, are highly reproducible at the protein level and reveal that, while many different proteins are required to work together to create the optimal conditions for mineralization to occur, they are not located randomly but are spatially highly organized. Here, we begin to clarify the spatial patterning of the calcification space as a new mineral is formed between the living tissue of the animal and the substrate or older skeleton.

## 2. Results and Discussion

### 2.1. The interactome—a topological model

The cross-link networks, termed 'interactomes' [24], reveal generalized patterns among skeleton building proteins (figure 1 and electronic supplementary material, SI appendix,

figures S1–S4 and tables S1 and S2). These spatial interactions among the skeletal proteins are inferred upon treatment with BS3 to the cleaned, powdered coral skeleton. *S. pistillata* skeleton contains EDTA-soluble and -insoluble proteins in its aragonitic skeleton [4], of which individual coral skeletal proteins have the ability to precipitate calcium carbonate *in vitro* [14,25]. BS3 cross-linking impeded their ability to precipitate calcium carbonate (see below; electronic supplementary material, SI appendix, figure S10). For spatial interactions, the skeletal material is cross-linked and decalcified, and the cross-linked peptides, and hence proteins, are identified by high-resolution mass spectrometry. The separate interactomes for the soluble and insoluble proteins from the decalcified coral skeleton are presented in figure 1 and electronic supplementary material, SI appendix, figure S1, respectively.

The interaction networks are presented in the form of nodes, where each node represents a specific protein. The mathematically derived interactions based on the cross-linked proteins are inferred to represent true biochemical interactions responsible for skeletal formation. The networks are scaled based on betweenness centrality, which refers to the frequency at which each node (protein) occurs in the shortest paths between two other proteins (e.g. if a far-reaching node, protocadherin 2 (PC2), has direct connections only with CARP1 and PC5, it is now connected to many other proteins via CARP1 and PC5, and hence CARP1 and PC5 have greater betweenness centrality than PC2). Higher betweenness suggests the bridging property of a protein among clusters in a network. We also identify the centrality of proteins with another factor: closeness centrality, which refers to the frequency of interaction among two nodes as well as the interactions of the node to connect with the entire network. These topological measures are used to identify the key components in a network of multiple proteins.

The interactomes suggest that CARPs are spatially close to the two isoforms of the skeletal carbonic anhydrase STPCA2 and other CARPs, as well as adhesion and framework proteins including cadherins, vitellogenin, thrombospondin, epidermal

growth factor (EGF) and laminin-G domain-containing (EGF LamG) proteins, and coadhesin. We also detected two uncharacterized proteins. The first, a von Willebrand factor type A (vWFA) domain-containing adhesion glycoprotein (SpiUSOMP 12 or USOMP12), was identified as a phosphopentothenoylcysteine decarboxylase [XP_022783323.1]; however, it does not contain any conserved domains found in other similar decarboxylases. It also has 28% sequence similarity to a protein in the stony coral *Acropora millepora*, annotated as collagen alpha-1(XIV) chain-like (XP_029187150.1) [6]. A second uncharacterized protein detected in our sequencing (SpiUSOMP 13 or USOMP13) appears to be a coral-specific protein [XP_022780049.1] [26]. In addition to the previously known skeletal carbonic anhydrase (STPCA2), we also identified an isoform and name it STPCA2-2. STPCA2-2 has an acidic insert region (electronic supplementary material, SI appendix, figure S5) comparable to STPCA2 (here, STPCA2-1).

## 2.2. The coral acid-rich proteins

Six CARPs/SAARPs have been characterized to date [4,6,12], and a seventh was sequenced in the present study. Most of these proteins contain more than 100 amino acids with approximately 30% or more of the sequence consisting of aspartic and/or glutamic acids, or share an orthologous relationship to such highly acidic proteins [27,28]. Owing to the ultra-high abundance of acidic residues, these proteins have isoelectric points between 3 and 4.5 and can precipitate aragonite in unamended seawater [12]. In this study, we found five of the seven previously known acid-rich proteins: CARP1, CARP2, CARP4/SAARP1, CARP5/SAARP2, an aspartic acid-rich protein identified here with high sequence similarity to partial-P27/acidic SOMP/SAARP3 [AGG36350.1] [4,5] and that we call full-length P27 (electronic supplementary material, SI appendix, figure S6), and a novel glutamic acid-rich protein that we have named CARP6 (electronic supplementary material, SI appendix, figure S7). This is also the first report of CARP1 and CARP2 being identified by mass spectrometry, although previous studies identified them by immunolocalization [29].

Our cross-linking analysis suggests that CARPs interact with each other in a variety of ways (electronic supplementary material, SI appendix, figures S3 and S4). CARP1 is the only CARP in close proximity to the E-rich CARP2 and D-rich CARP4. CARP1 has previously been shown, using immunolocalization and calcein staining of cell cultures of *S. pistillata*, to be a member of the ECM proteins and specifically adjacent to freshly formed $CaCO_3$ particles [30]. Micro-Raman and nuclear magnetic resonance spectroscopy suggest that it and CARP3 bind to the aragonite phase in newly settled planulae [31]. CARP1 has seven EF-hand domains, which bind $Ca^{2+}$, and has an 83% sequence similarity to calumenin [12], which is suggested for its role in bone morphogenetic protein (BMP)-2-dependent calcification and interactions with hydroxyapatite or calcite crystals [32].

Another highly acidic coral protein found to promote aragonite formation *in vitro* by binding to calcium is CARP2 [12]. Its glutamic acid residues have been detected in centres of calcification of *S. pistillata* spat [31]. It is the only known CARP which is upregulated in the pre-settled stage of planulae of the related pocilloporid coral *Pocillopora damicornis* [33]. This protein is suggested to precipitate a Mg-rich amorphous calcium carbonate (ACC), which is the first biomineral formed in the centres of calcification [15]. Our analysis found close spatial interactions

of CARP2 with CARP1 and full-length P27, along with two carbonic anhydrases, STPCA2-1 and STPCA2-2 (electronic supplementary material, SI appendix, figure S4).

The glutamic acid-rich proteins CARP2 and the novel CARP6, both detected here for the first time by mass spectrometry, are the only CARPs to interact with an α-collagen, a framework protein that plays an important role in biomineralization. A similar glutamic acid-rich protein in vertebrate bones, bone sialoprotein (BSP), interacts with type 1 collagen [34] and promotes biomineralization in bone tissue by binding collagen and integrin [35]. BSP interacts with integrin at its arginine–glycine–aspartic acid sequence, and such an arginine–glycine–glutamic acid sequence is also seen in CARP2 (but not in other CARPs). BSP is a glycosylated phosphoprotein expressed exclusively in new mineralization zones [36], similar to CARP2 in aragonitic coral skeleton [31], and is responsible for hydroxyapatite nucleation and binding [37]. While the abundant negative charges from BSP's poly-E-rich regions are necessary for calcium-binding and initial interactions with collagen, the high-affinity binding with collagen is found in the protein's non-acidic regions [35,38], and BSP exhibits high-affinity binding to $Ca^{2+}$ both *in vivo* and *in vitro* [34]. Based on the functional similarity, we performed sequence alignments of the N-terminus of BSP and CARP2 (electronic supplementary material, SI appendix, figure S8). Based on these alignments, the previously studied [38] BSP–collagen interaction sites are in close proximity to the cross-linked lysines of CARP2 and α-collagen (electronic supplementary material, SI appendix, figure S8). Another glutamic acid-rich protein which contains extended poly-E residues, as also seen in the novel CARP6 (electronic supplementary material, SI appendix, table S3), is known from vertebrate rod photoreceptors and has low-affinity binding to calcium but at a higher capacity [39]. Taken together, glutamic acid-rich regions probably play a major role to act as calcium buffers/concentrator at the centres of calcification.

The interaction of CARP2 with acidic carbonic anhydrase STPCA2-2 may have a specific role in centres of calcification, with the acidic domain of STPCA2-2 (electronic supplementary material, SI appendix, figure S5) contributing to the initial mineralization and binding to calcium ions upon its interaction with CARP2, while CARP2 (although acid rich) assists in protein interactions and increased ACC nanoparticle accumulation.

The interactions of CARP4 with CARP5 and CARP5 with CARP6 are especially interesting. These three CARPs have stretches of glutamic acid or aspartic acids (electronic supplementary material, SI appendix, figure S7), suggesting a similar role of CARP4/5 and CARP6 in calcium binding/concentrating during the biomineralization process. Based on their sequences (electronic supplementary material, SI appendix, figure S7), these three CARPs are intrinsically disordered [40,41]. The presence of such highly disordered proteins in mineralization is seen widely across the tree of life and yet their role is not well understood [4,6,34].

## 2.3. Carbonic anhydrases—STPCA2 and STPCA2-2

In addition to CARPs, another well-studied protein found in the coral skeleton is carbonic anhydrase. Previous immunolocalization studies have identified the widespread distribution of the carbonic anhydrase STPCA2 in the skeleton [4,29]. The *S. pistillata* genome encodes 16 carbonic anhydrases,

of which three (STPCA, STPCA2 and STPCA3) are characterized for their chemical properties [42]. Proteomic studies of the *S. pistillata* skeleton, however, have only found STPCA2 to date [4,7] and a different carbonic anhydrase was detected in *Acropora digitifera* skeleton [6] but not in *A. millepora* [5]. We found a second isoform of STPCA2, here named STPCA2-2. These STPCA2 isoforms have a high sequence identity of about 74% except at their N-termini and at the insert region of STPCA2-2 (electronic supplementary material, SI appendix, figure S5). This region contains an abundance of acidic residues and appears to be specific to the two STPCA2 isoforms identified so far compared with the rest of the α-carbonic anhydrase superfamily as it is absent in mammalian carbonic anhydrases and the other cellular coral carbonic anhydrases (electronic supplementary material, SI appendix, figure S5).

Since STPCA2 isoforms contain evolutionarily conserved domains compared with the rest of the skeletal proteins, we modelled these two isoforms (STPCA2-1 and STPCA2-2) using I-TASSER [43]. All lysines involved in inter-protein interactions, based on our cross-linking work, from both STPCA2 isoforms appear at the surface of the proteins' proposed tertiary structures (electronic supplementary material, SI appendix, figure S9), suggesting that the STPCA2 structure may evolutionarily support a highly positively charged surface (electronic supplementary material, SI appendix, figure S9); this would allow its interactions with negatively charged domains (e.g. CARPs).

Protein cross-linking results in the reduction/loss of protein flexibility and therefore function [44]. This is consistent with our results from *in vitro* calcium carbonate precipitation assays using a calcite-promoting seawater medium to which we added cross-linked SOM, un-cross-linked SOM or no protein (electronic supplementary material, SI appendix, figure S10). The pair of reacting residues within a protein (lysines in this case for BS3) must be within the reacting distance and at the same time accessible to the cross-linker for the proteins to become cross-linked. This non-cleavable linkage makes the local region of the protein static with subsequent loss of flexibility, dynamic properties and protein function [45,46]. We observed that, when EDTA-solubilized SOM proteins were incubated in seawater, precipitates containing calcium and magnesium were formed (electronic supplementary material, SI appendix, figure S10); however, when cross-linked- and then EDTA-solubilized SOM proteins were incubated in seawater, such precipitates were not formed. This confirms that cross-linking restrains SOM proteins and the dynamic behaviour required for their function of aragonite precipitation is subsequently lost.

We then mapped the STPCA2 cross-link pairs onto their modelled structures and found that all the intra-STPCA2 cross-link pairs are located near each other and are on the same side of the folded protein (figure 2). Three of the four cross-linked pairs in the models are approximately 11.4 Å apart, strongly suggesting that BS3 reacts with native proteins in the skeletal environment and does not produce random cross-linked products. The fourth cross-link pair ($^{63}$K::$^{150}$K) has residues approximately 20 Å apart in the models (figure 2d), which could be for two reasons: these residues lie in the highly flexible coiled regions of STPCA2-2 (which is evident from the I-TASSER structure; figure 2) and/or these residues are derived from the STPCA2 isoforms. We tested this second hypothesis through molecular docking of the two identified STPCA2 isoforms (figure 2e,f). Molecular docking is a method to model two or more proteins (protein–ligand) at their interaction sites, either using predictive algorithms or from experimental data. In our study, we used experimentally identified cross-linked residues $^{63}$K::$^{150}$K (since Lys-63 is present in both isoforms) as the interaction site of the two STPCA2 isoforms modelled in the interactive docking prediction program ZDOCK, which applies unison of shape complementarity and electrostatics to statistically score and generate models with high structural stability and predictive accuracy. As shown in figure 2e,f, once the STPCA2 isoforms are docked at STPCA2-1 $^{63}$K::STPCA2-2 $^{150}$K, these residues are within the cross-linking distance of BS3. Additionally, this interaction also brings the interacting residues of STPCA2 within the cross-linking distance of CARP2 and CARP4, which would otherwise be far apart (electronic supplementary material, SI appendix, figure S11).

Although we could not capture a direct interaction between CARP2 and CARP4, it would seem necessary for the transition from ACC to aragonitic fine needles in the coral skeleton. While most CARP genes are upregulated after the settlement of coral planulae, CARP2 is the only CARP gene that is highly expressed in the pre-settled stage and is downregulated post-settlement in *Pocillopora damicornis* [31,47]. CARP2 has also recently been found associated with the ACC phase and was suggested to be involved in the formation of initial mineral precursors [31]; this is consistent with our cross-linking interaction between CARP2 and the framework protein α-collagen.

## 2.4. Structural and calcium-binding domain-containing proteins

Several transmembrane domain-containing proteins that become part of the skeleton include cadherin-like and other calcium-binding proteins. One such class of proteins is von Willebrand factor (vWF) proteins. The vWFA domain is seen in a large family of adhesion glycoproteins, whose structures are stabilized upon calcium binding [48]. These proteins enhance mineralization of bones *in vitro* [49], suggesting their role in the initial mineralization process during skeleton development. Genes for these proteins have previously been studied in pre- and post-settled larvae from *P. damicornis* [33] and *A. millepora* [50] and are upregulated in the pre-settled stage of coral larvae before mineralization commences. The vWFA proteins identified in our work include thrombospondin, USOMP12 and collagen. Thrombospondin contains a highly conserved protein-binding domain (TSP-1) and has previously been suggested for its role in biomineral remodelling in higher organisms [51]. The α-collagen probably plays an important role in controlling mineralization. Previous *in vitro* studies suggest its role in amorphous-phase mineral formation, which requires infiltration of the mineral into fibrils followed by their organization into oriented crystals [52]. α-Collagens are highly abundant, fibrous, insoluble and ECM proteins in eukaryotes and are known for their structural role. Several studies show *in vitro* nucleation and mineral formation in the presence of collagen, with the interaction of positive and negative charges on the proteins being suggested as a mechanism for mineral growth in solution [53]. We observed spatial interactions from vWFA domain-containing proteins such as thrombospondin, USOMP12 and collagen with CARPs and the STPCA2 isoforms.
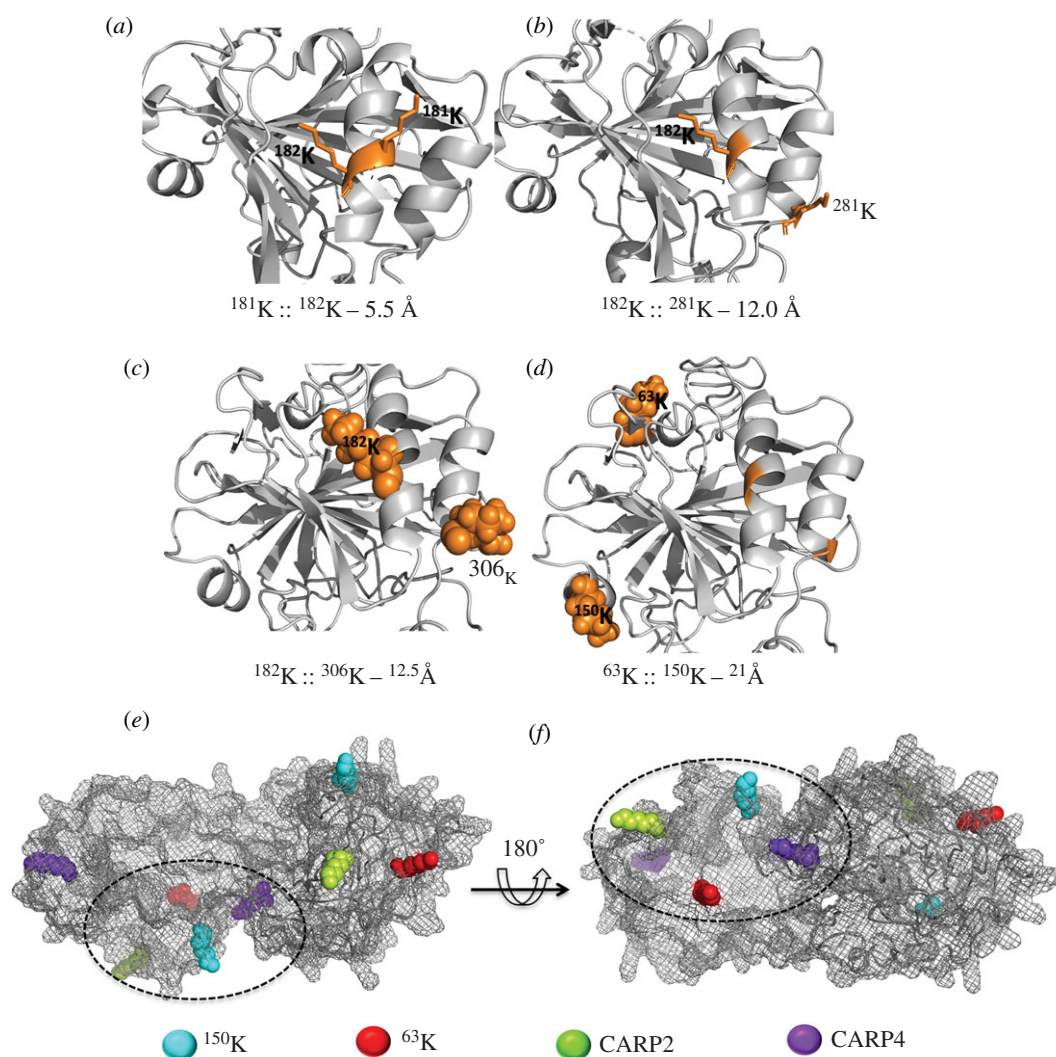
**Figure 2.** Interactions within STPCA2 isoforms. Carbonic anhydrase (STPCA2-1 and STPCA2-2) protein models were generated by I-TASSER. (*a,b*) The intra-molecular cross-links identified for STPCA2-1 by mass spectrometry are mapped onto these models, shown in orange sticks. (*c,d*) STPCA2-2; sphere representation of cross-linked residues in orange. Distances (below *a–d*) between cross-linked residues were measured in PyMOL. (*e,f*) Molecular docking of $^{63}$K and $^{150}$K from two STPCA2 isoforms. The lysines from STPCA2 interacting with CARP2 and CARP4, when mapped onto the STPCA2 docking model, show that all four residues are spatially close.

The von Willebrand factor type D (vWFD) domain-containing proteins are another set of proteins that bind calcium and are an important structural component of the skeletal ECM. The skeletal proteins seen in our study with this domain include sushi domain-containing protein [AGG36340] and vitellogenin [XP_022779720.1]. Vitellogenin is notable for its interaction with lipopolysaccharides, peptidoglycans and polysaccharides such as glucan and laminarin by pattern recognition, and belongs to a group of multivalent pattern recognition receptor binding proteins that have been suggested for their role in innate immune defence [54]. Vitellogenin is also a precursor of egg yolk protein, helping to provide nutrients to developing embryos by its properties of recognizing proteins and lipids. Such pleiotropic roles of vitellogenin in fighting bacterial infections and its protein and lipid interactions may be important for coral skeleton formation. We found interactions of sushi domain-containing protein with coadhesin (which has vWFA- and FA58C-carbohydrate-binding domains), MAM-and-LDL domain-containing protein and CARP2 (figure 1, electronic supplementary material, SI appendix, figures S1 and S2), whereas vitellogenin interacts with several adhesion proteins (coadhesin, fibronectin,

collagen, EGF LamG, protocadherin) and CARP4. Based on our observations of interactions of vWF domain-containing proteins with other adhesion proteins, and their expression in early stages of calcifying spat [33], it is clear that they play an important role in framework building, cell adhesion, and protein–protein, protein–lipid and protein–crystal interactions during the skeleton development process in corals.

We observed several other calcium-binding proteins, including the following. (i) Kielin-like and EGF LamG domain-containing proteins which contain one or more LamG domains that, in addition to binding calcium, perform a variety of functions, including adhesion and migration. (ii) MAM-and-LDL domain-containing proteins, which, as extracellular proteins, impart properties of homo-oligomerization and adhesion. (iii) The five sequences PC1–PC5 were also detected by mass spectrometry (electronic supplementary material, SI appendix, table S3). We compared these sequences with the skeletal protocadherin-like protein from *A. millepora* [JT JT011093.1] (electronic supplementary material, SI appendix, figure S13), which is predicted as a 450 kDa protein. Based on our alignment analysis, four out of five of the *S. pistillata* protocadherin sequences (PC1–PC4) have a very
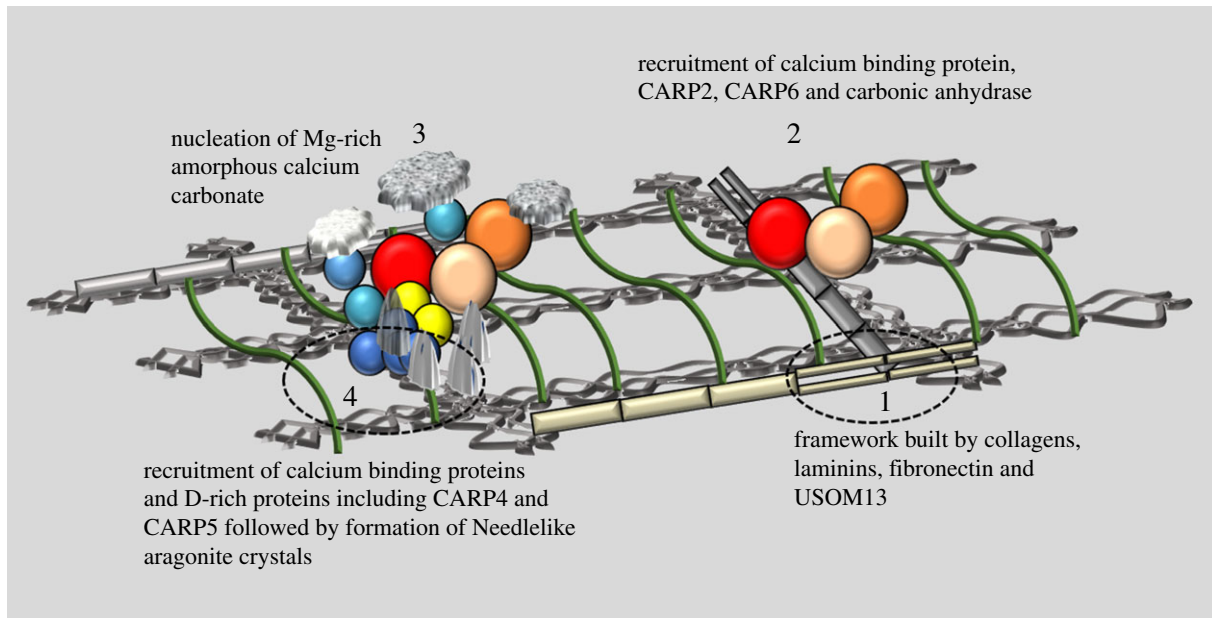
**Figure 3.** Working model of biomineralization. Based on previous research and our interactome analyses presented here, we suggest the following working model of stony coral biomineralization. 1. Skeletal vWF domain-containing proteins collagen, USOMP13 and laminins form a structural framework. 2. Glutamic acid-rich proteins CARP2 and CARP6 bind to collagen in the presence of STPCA2-2 (carbonic anhydrase). 3. Initiation of ACC phase formation, binding of other adhesion and calcium-binding proteins along with aspartic acid-rich proteins CARP4 and CARP5. 4. Formation of needle-like aragonite crystals. Confirmation of the timing of each step remains to be resolved.

high sequence similarity to *A. millepora* protocadherin and they are each other's reciprocal best blast hit between the two species. We therefore propose that the sequences PC1–PC4 are actually part of a larger cadherin orthologous to the *A. millepora* skeletal protocadherin-like protein that was incompletely predicted in *S. pistillata*. We further note that, with their large number of extracellular cadherin repeats, EGF-like and LamG domains and cytoplasmic catenin-binding domain, these orthologous cadherins observed in *S. pistillata* [4], *A. millepora* [6] and *A. digitifera* [5] skeleton are likely to be classical cadherins [55]. (iv) A CUB (complement C1r/C1s, Uegf, Bmp1) domain-containing tolloid-like protein was also observed in our analysis. Tolloid-like proteins are known for their role in orchestrating the formation of ECM, patterning the biomineral and BMP signalling in morphogenetic processes [56]. Binding to these structural proteins as described above would allow the highly acidic proteins to be arrayed in an ordered fashion in the calcifying space for tighter control by corals over biomineralization.

To the extent that we can generalize patterns of protein interactions across biomineralizing taxa outside of Cnidaria, our results suggest that highly acidic mineralization-related proteins such as aspein, dentins and the SIBLING family of proteins [57] would also be associated with other acidic proteins [4,6,58] and carbonic anhydrases. The cementing proteins will depend on the biomineral that is precipitated. However, based on the limited number of proteomic analyses available at this time, we suggest that collagens and cadherin-like proteins would be broadly distributed in metazoan biominerals.

## 2.5. Concluding working model

A working model for the spatial organization of SOM proteins in the stony coral skeleton is shown in figure 3. We propose that the major ECM-forming skeletal proteins that contribute to the formation of the organic framework of the calcifying space (vWFA, vWFD family of proteins) first structure the microenvironment of the calcifying space. This is followed

by their interaction with glutamic acid-rich protein CARP2 (and potentially CARP6) and carbonic anhydrases. Our study identified interactions of ECM-forming framework proteins with CARP2 and carbonic anhydrases. Previous expression studies indicate that vWFA, vWFD domain-containing proteins along with CARP2 are highly expressed during the initial, newly released stage of larval development [33]. The known characteristics of these proteins suggest that they lead to the formation of an initial ACC phase, followed by the recruitment of other adhesion and $Ca^{2+}$-binding proteins (such as cadherins), a matrix-patterning tolloid-like protein, a BMP signalling protein and aspartic acid-rich proteins (e.g. CARP4 and CARP5). This is consistent with the late-stage expression of CARP4 and CARP5 in the larval development as settled calcifying spat [33]. Our interactome analysis supports that the calcium-binding proteins interact directly with acid-rich proteins or indirectly via cadherins and other adhesion proteins. The presence of such skeletal ECM-forming proteins (framework and adhesion proteins) along with matrix-patterning proteins would be necessary to provide polarity, organizing the CARPs and other acid-rich (i.e. disordered) proteins in the calcifying space during the biomineralization process [59].

## 3. Material and Methods

Fully detailed methods can be found in electronic supplementary material, Methods.

### 3.1. Study organism

Fragments of *S. pistillata* were grown in an in-house 800 litre flow-through system.

### 3.2. Cleaning coral skeleton

Skeletons were soaked overnight in 3% (wt/vol) sodium hypochlorite, copiously rinsed in deionized water and dried

overnight at 37°C. Dried skeletons were ground to less than 150 µm diameter and then bleached, rinsed and dried again. The dried powder was further ground to less than 60 µm.

### 3.3 Skeletal organic matrix extraction

Cleaned skeletal powder for replicate samples (5 g each) was decalcified in 0.5 M EDTA at room temperature while shaking. Insoluble material was pelleted and washed in water followed by two washes in 10 mM phosphate-buffered saline (PBS), pH 8.0, 10 µM $CaCl_2$; the final pellets were resuspended in 100 mM PBS, pH 8.0, 10 µM $CaCl_2$. The EDTA-soluble fractions were concentrated and washed twice in 10 mM PBS, pH 8.0, 10 µM $CaCl_2$ and then once in 100 mM PBS, pH 8.0, 10 µM $CaCl_2$ in 10 kDa Amicon centrifugal filter units (Millipore Sigma). At this stage, samples were concentrated to 200 µl and stored at −20°C. Data from three biological replicates and experimental replicates were pooled together to generate our final results.

### 3.4. Protein cross-linking

Cross-linking reactions using BS3 (Proteochem) dissolved in mass spectrometry-grade water were performed on the cleaned skeletal powders at 5 mM and 10 mM BS3 in PBS pH 7.4 for 1 h at room temperature. The reaction was stopped using 40 mM ammonium bicarbonate for 20 min at room temperature. Pelleted skeleton containing cross-linked proteins was rinsed to remove any unreacted cross-linker before dissolution in EDTA as described above.

### 3.5. Electrophoresis and protein digestion

Cross-linked proteins were stacked into gels by sodium dodecyl sulfate polyacrylamide gel electrophoresis; the entire band was excised from the gel, fixed in 50% ethanol and 5% acetic acid for 30 min, washed in water several times, washed in a solution of 25 mM ammonium bicarbonate and 50% acetonitrile several times [60], dried in a SpeedVac and stored frozen until enzymatic digestion. Immediately prior to mass spectrometric analysis, proteins were treated with dithiothreitol and iodo-acetic acid followed by trypsin digestion [60].

### 3.6. Mass spectrometry

Tryptic digests were resolved on a C18 reversed-phase column. Spectrometry was performed on a QE Exactive Orbitrap instrument. Proteins were identified by both the MassMatrix search engine and X!Tandem using a FASTA search library created from the *S. pistillata* transcriptome [61]. Identified proteins with at least two unique peptides and a $p$-value $\leq 10E^{-10}$ were retained to create a custom cross-link search database. Cross-linked peptides with a $p$-value $\leq 10E^{-3}$ were considered for further analysis.

### 3.7. Cross-link peptide search

Mass spectrometry generated files were first analysed by MassMatrix version 2.4.2 using a ≤1% false discovery rate on a decoy search. Then, peptide matches from the first step were searched using MassMatrix's search algorithm to identify the cross-linked peptides based on the BS3 range constraint (11.4 Å). In the third step, the quality of each peptide match was measured by a single $p$-value calculated from a probabilistic score obtained from the number of matched peaks, a probabilistic score obtained from the ion intensity distribution of matched peaks and a probabilistic score obtained from consecutiveness of matched peaks.

### 3.8. Protein modelling and network maps

Carbonic anhydrase models were generated using I-TASSER by homology modelling [43]. The PDB files of protein structure generated by I-TASSER were visualized in PyMOL. These PDB files were also used for docking analysis performed using ZDOCK. The results from BS3 cross-linking are presented as network maps. These are created with the network analysis and visualization program Gephi [62].

### 3.9. *In vitro* calcium carbonate precipitation assay

This assay was performed to identify if the ECM proteins were functional after application of protein cross-linker and their extraction from the coral skeleton (electronic supplementary material, SI appendix, figure S11). Briefly, 10 µg of total proteins from the EDTA-solubilized skeletal powder (with and without BS3 treatment) was incubated in artificial seawater in sterile six-well plates. Precipitates were collected from the surface of the wells, diluted and imaged under scanning electron microscopy. A detailed method for this assay is included in the supplementary methods.

## References

1. Scrutton C. 1999 Palaeozoic corals: their evolution and palaeoecology. *Geol. Today* **15**, 184–193. (doi:10.1046/j.1365-2451.1999.1505005.x)

2. Drake JL, Mass T, Stolarski J, Von Euw S, van de Schootbrugge B, Falkowski PG. 2020 How corals made rocks through the ages. *Glob. Chang. Biol.* **26**, 31–53. (doi:10.1111/gcb.14912)

3. Clark MS. 2020 Molecular mechanisms of biomineralization in marine invertebrates. *J. Exp. Biol.* **223**, jeb206961. (doi:10.1242/jeb.206961)

4. Drake JL, Mass T, Haramaty L, Zelzion E, Bhattacharya D, Falkowski PG. 2013 Proteomic analysis of skeletal organic matrix from the stony coral *Stylophora pistillata*. *Proc. Natl. Acad. Sci.* **110**, 3788–3793. (doi:10.1073/pnas.1301419110)

5. Takeuchi T, Yamada L, Shinzato C, Sawada H, Satoh N. 2016 Stepwise evolution of coral biomineralization revealed with genome-wide proteomics and transcriptomics. *PLoS ONE* **11**, e0156424. (doi:10.1371/journal.pone.0156424)

6. Ramos-Silva P, Kaandorp J, Huisman L, Marie B, Zanella-Cléon I, Guichard N, Miller DJ, Marin F. 2013 The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by co-option and domain shuffling. *Mol. Biol. Evol.* **30**, 2099–2112. (doi:10.1093/molbev/mst109)

7. Peled Y, Drake JL, Malik A, Almuly R, Lalzar M, Morgenstern D, Mass T. 2020 Optimization of skeletal protein preparation for LC–MS/MS sequencing yields additional coral skeletal proteins in *Stylophora pistillata*. *BMC Mater.* **2**, 1–5. (doi:10.1186/s42833-020-00014-x)

8. Marin F, Bundeleva I, Takeuchi T, Immel F, Medakovic D. 2016 Organic matrices in metazoan calcium carbonate skeletons: composition, functions, evolution. *J. Struct. Biol.* **196**, 98–106. (doi:10.1016/j.jsb.2016.04.006)

9. Gilbert P, Porter SM, Sun CY, Xiao S, Gibson BM, Shenkar N, Knoll AH. 2019 Biomineralization by particle attachment in early animals. *Proc. Natl Acad. Sci. USA* **116**, 17 659–17 665. (doi:10.1073/pnas.1902273116)

10. Chen Y, Feng Y, Deveaux JG, Masoud MA, Chandra FS, Chen H, Zhang D, Feng L. 2019 Biomineralization forming process and bio-inspired nanomaterials for biomedical application: a review. *Minerals* **9**, 68. (doi:10.3390/min9020068)

11. Kim YY *et al*. 2016 Tuning hardness in calcite by incorporation of amino acids. *Nat. Mater.* **15**, 903–910. (doi:10.1038/nmat4631)

12. Mass T, Drake JL, Haramaty L, Kim JD, Zelzion E, Bhattacharya D, Falkowski PG. 2013 Cloning and characterization of four novel coral acid-rich proteins that precipitate carbonates in vitro. *Curr. Biol.* **23**, 1126–1131. (doi:10.1016/j.cub.2013.05.007)

13. Moya A, Tambutté S, Bertucci A, Tambutté E, Lotto S, Vullo D, Supuran CT, Allemand D, Zoccola D. 2008 Carbonic anhydrase in the scleractinian coral *Stylophora pistillata*: characterization, localization, and role in biomineralization. *J. Biol. Chem.* **283**, 25 475–25 484. (doi:10.1074/jbc.M804726200)

14. Laipnik R, Bissi V, Sun CY, Falini G, Gilbert PUPA, Mass T. 2019 Coral acid rich protein selects vaterite polymorph in vitro. *J. Struct. Biol.* **209**, 107431. (doi:10.1016/j.jsb.2019.107431)

15. Von Euw S *et al*. 2017 Biological control of aragonite formation in stony corals. *Science* **356**, 933–938. (doi:10.1126/science.aam6371)

16. Comeau S, Tambutté E, Carpenter RC, Edmunds PJ, Evensen NR, Allemand D, Ferrier-Pagès C, Tambutté S, Venn AA. 2017 Coral calcifying fluid pH is modulated by seawater carbonate chemistry not solely seawater pH. *Proc. R. Soc. B* **284**, 20161669. (doi:10.1098/rspb.2016.1669)

17. Bertucci A, Tambutté S, Supuran CT, Allemand D, Zoccola D. 2011 A new coral carbonic anhydrase in *Stylophora pistillata*. *Mar. Biotechnol.* **13**, 992–1002. (doi:10.1007/s10126-011-9363-x)

18. Bertucci A, Moya A, Tambutté S, Allemand D, Supuran CT, Zoccola D. 2013 Carbonic anhydrases in anthozoan corals—a review. *Bioorganic Med. Chem.* **21**, 1437–1450. (doi:10.1016/j.bmc.2012.10.024)

19. Zoccola D, Innocenti A, Bertucci A, Tambutté E, Supuran CT, Tambutté S. 2016 Coral carbonic anhydrases: regulation by ocean acidification. *Mar. Drugs* **14**, 109. (doi:10.3390/md14060109)

20. Mass T *et al*. 2017 Amorphous calcium carbonate particles form coral skeletons. *Proc. Natl Acad. Sci. USA* **114**, E7670–E7678. (doi:10.1073/pnas.1707890114)

21. Petrotchenko EV, Borchers CH. 2010 Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom. Rev.* **29**, 862–876. (doi:10.1002/mas.20293)

22. Zhao A, Hao G, Gu J. 2013 Chemical crosslinking and mass spectrometric identification of interaction sites within soluble aggregate of protein therapeutics. *J. Pharm. Biomed. Anal.* **73**, 99–102. (doi:10.1016/j.jpba.2012.05.006)

23. Rappsilber J. 2011 The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* **173**, 530–540. (doi:10.1016/j.jsb.2010.10.014)

24. Sanchez C, Lachaize C, Janody F, Bellon B, Röder L, Euzenat J, Rechenmann F, Jacq B. 1999 Grasping at molecular interactions and genetic networks in *Drosophila melanogaster* using FlyNets, an Internet database. *Nucleic Acids Res.* **27**, 89–94. (doi:10.1093/nar/27.1.89)

25. Mass T, Drake JL, Haramaty L, Rosenthal Y, Schofield OME, Sherrell RM, Falkowski PG. 2012 Aragonite precipitation by 'proto-polyps' in coral cell cultures. *PLoS ONE* **7**, e35049. (doi:10.1371/journal.pone.0035049)

26. Voolstra CR, Li Y, Liew YJ, Baumgarten S, Zoccola D, Flot JF, Tambutté S, Allemand D, Aranda M. 2017 Comparative analysis of the genomes of *Stylophora pistillata* and Acropora digitifera provides evidence for extensive differences between species of corals. *Sci. Rep.* **7**, 1–4. (doi:10.1038/s41598-017-17484-x)

27. Drake JL, Whitelegge JP, Jacobs DK. 2020 First sequencing of ancient coral skeletal proteins. *Sci. Rep.* **10**, 1–11. (doi:10.1038/s41598-020-75846-4)

28. Zaquin T, Malik A, Drake JL, Putnam HM, Mass T. 2021 Evolution of protein-mediated biomineralization in scleractinian corals. *Front. Genet.* **12**, 52.

29. Mass T, Drake JL, Peters EC, Jiang W, Falkowski PG. 2014 Immunolocalization of skeletal matrix proteins in tissue and mineral of the coral *Stylophora pistillata*. *Proc. Natl Acad. Sci.* **111**, 12 728–12 733. (doi:10.1073/pnas.1408621111)

30. Mass T, Drake JL, Heddleston JM, Falkowski PG. 2017 Nanoscale visualization of biomineral formation in coral proto-polyps. *Curr. Biol.* **27**, 3191–3196.e3 (doi:10.1016/j.cub.2017.09.012)

31. Akiva A, Neder M, Kahil K, Gavriel R, Pinkas I, Goobes G, Mass T. 2018 Minerals in the pre-settled coral *Stylophora pistillata* crystallize via protein and ion changes. *Nat. Commun.* **9**, 1–9. (doi:10.1038/s41467-018-04285-7)

32. Rose-Martel M, Smiley S, Hincke MT. 2015 Novel identification of matrix proteins involved in calcitic biomineralization. *J. Proteomics* **116**, 81–96. (doi:10.1016/j.jprot.2015.01.002)

33. Mass T, Putnam HM, Drake JL, Zelzion E, Gates RD, Bhattacharya D, Falkowski PG. 2016 Temporal and spatial expression patterns of biomineralization proteins during early development in the stony coral *Pocillopora damicornis*. *Proc. R. Soc. B* **283**, 20160322. (doi:10.1098/rspb.2016.0322)

34. Fujisawa R. 2012 Acidic bone matrix proteins and their roles in calcification. *Front. Biosci.* **17**, 1891. (doi:10.2741/4026)

35. Kruger TE, Miller AH, Wang J. 2013 Collagen scaffolds in bone sialoprotein-mediated bone regeneration. *Sci. World J.* **2013**, 812718. (doi:10.1155/2013/812718)

36. Ganss B, Kim RH, Sodek J. 1999 Bone sialoprotein. *Crit. Rev. Oral Biol. Med.* **10**, 79–98. (doi:10.1177/10454411990100010401)

37. Wuttke M, Müller S, Patric Nitsche D, Paulsson M, Hanisch FG, Maurer P. 2001 Structural characterization of human recombinant and bone-derived bone sialoprotein. Functional implications for cell attachment and hydroxyapatite binding. *J. Biol. Chem.* **276**, 36 839–36 848. (doi:10.1074/jbc.M105689200)

38. Tye CE, Hunter GK, Goldberg HA. 2005 Identification of the type I collagen-binding domain of bone sialoprotein and characterization of the mechanism of interaction. *J. Biol. Chem.* **280**, 13 487–13 492. (doi:10.1074/jbc.M408923200)

39. Haber-Pohlmeier S, Abarca-Heidemann K, Körschen HG, Dhiman HK, Heberle J, Schwalbe H, Klein-Seetharaman J, Kaupp UB, Pohlmeier A. 2007 Binding of $Ca^{2+}$ to glutamic acid-rich polypeptides from the rod outer segment. *Biophys. J.* **92**, 3207–3214. (doi:10.1529/biophysj.106.094847)

40. Bürgi J, Xue B, Uversky VN, Van Der Goot FG. 2016 Intrinsic disorder in transmembrane proteins: roles in signaling and topology prediction. *PLoS ONE* **11**, e0158594. (doi:10.1371/journal.pone.0158594)

41. Van Der Lee R *et al*. 2014 Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631. (doi:10.1021/cr400525m)

42. Del Prete S *et al*. 2018 Comparison of the anion inhibition profiles of the $\alpha$-CA isoforms (SpiCA1, SpiCA2 and SpiCA3) from the scleractinian coral *Stylophora pistillata*. *Int. J. Mol. Sci.* **19**, 2128. (doi:10.3390/ijms19072128)

43. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2014 The I-TASSER suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8. (doi:10.1038/nmeth.3213)

44. Patramani I, Katsiri K, Pistevou E, Kalogerakos T, Pavlatos M, Evangelopoulos AE. 1969 Glutamic-aspartic transaminase—antitransaminase interaction: a method for antienzyme purification. *Eur. J. Biochem.* **11**, 28–36. (doi:10.1111/j.1432-1033.1969.tb00734.x)

45. Chen ZA, Rappsilber J. 2018 Protein dynamics in solution by quantitative crosslinking/mass spectrometry. *Trends Biochem. Sci.* **43**, 908–920. (doi:10.1016/j.tibs.2018.09.003)

46. Schlaich NL, Häner M, Lustig A, Aebi U, Hurt EC. 1997 In vitro reconstitution of a heterotrimeric nucleoporin complex consisting of recombinant Nsp1p, Nup49p, and Nup57p. *Mol. Biol. Cell* **8**, 33–46. (doi:10.1091/mbc.8.1.33)

47. Neder M, Laissue PP, Akiva A, Akkaynak D, Albéric M, Spaeker O, Politi Y, Pinkas I, Mass T. 2019 Mineral formation in the primary polyps of pocilloporoid corals. *Acta Biomater.* **96**, 631–645. (doi:10.1016/j.actbio.2019.07.016)

48. Jakobi AJ, Mashaghi A, Tans SJ, Huizinga EG. 2011 Calcium modulates force sensing by the von Willebrand factor A2 domain. *Nat. Commun.* **2**, 1–9. (doi:10.1038/ncomms1385)

49. Ohyama Y, Katafuchi M, Almehmadi A, Venkitapathi S, Jaha H, Ehrenman J, Morcos J, Aljamaan R, Mochida Y. 2012 Modulation of matrix mineralization by Vwc2-like protein and its novel splicing isoforms. *Biochem. Biophys. Res. Commun.* **418**, 12–16. (doi:10.1016/j.bbrc.2011.12.075)

50. Hayward DC, Hetherington S, Behm CA, Grasso LC, Forêt S, Miller DJ, Ball EE. 2011 Differential gene expression at coral settlement and metamorphosis—a subtractive hybridization study. *PLoS ONE* **6**, e26411. (doi:10.1371/journal.pone.0026411)

51. Gutner-Hoch E, Ben-Asher HW, Yam R, Shemesh A, Levy O. 2017 Identifying genes and regulatory pathways associated with the scleractinian coral calcification process. *PeerJ* **5**, e3590. (doi:10.7717/peerj.3590)

52. Nudelman F, Pieterse K, George A, Bomans PHH, Friedrich H, Brylka LJ, Hilbers PAJ, De With G, Sommerdijk NAJM. 2010 The role of collagen in bone apatite formation in the presence of hydroxyapatite nucleation inhibitors. *Nat. Mater.* **9**, 1004–1009. (doi:10.1038/nmat2875)

53. Nudelman F, Lausch AJ, Sommerdijk NAJM, Sone ED. 2013 In vitro models of collagen biomineralization. *J. Struct. Biol.* **183**, 258–269. (doi:10.1016/j.jsb.2013.04.003)

54. Li Z, Zhang S, Zhang J, Liu M, Liu Z. 2009 Vitellogenin is a cidal factor capable of killing bacteria via interaction with lipopolysaccharide and lipoteichoic acid. *Mol. Immunol.* **46**, 3232–3239. (doi:10.1016/j.molimm.2009.08.006)

55. Hulpiau P, Van Roy F. 2011 New insights into the evolution of metazoan cadherins. *Mol. Biol. Evol.* **28**, 647–657. (doi:10.1093/molbev/msq233)

56. Hopkins DR, Keles S, Greenspan DS. 2007 The bone morphogenetic protein 1/Tolloid-like metalloproteinases. *Matrix Biol.* **26**, 508–523. (doi:10.1016/j.matbio.2007.05.004)

57. Boskey AL, Villarreal-Ramirez E. 2016 Intrinsically disordered proteins and biomineralization. *Matrix Biol.* **52-54**, 43–59. (doi:10.1016/j.matbio.2016.01.007)

58. Ramos-Silva P *et al.* 2014 The skeleton of the staghorn coral *Acropora millepora*: molecular and structural characterization. *PLoS ONE* **9**, e97454. (doi:10.1371/journal.pone.0097454)

59. Addadi L, Weiner S. 1985 Interactions between acidic proteins and crystals: stereochemical requirements in biomineralization. *Proc. Natl Acad. Sci.* **82**, 4110–4114. (doi:10.1073/pnas.82.12.4110)

60. Mummadisetti MP, Frankel LK, Bellamy HD, Sallans L, Goettert JS, Brylinski M, Limbach PA, Bricker TM. 2014 Use of protein cross-linking and radiolytic footprinting to elucidate PsbP and PsbQ interactions within higher plant Photosystem II. *Proc. Natl Acad. Sci.* **111**, 16 178–16 183. (doi:10.1073/pnas.1415165111)

61. Liew YJ, Aranda M, Voolstra CR. 2016 Reefgenomics.Org—a repository for marine genomics data. *Database* **2016**, baw152. (doi:10.1093/database/baw152)

62. Bastian M, Heymann S, Jacomy M. 2009 Gephi: an open source software for exploring and manipulating networks. See https://gephi.org/publications/gephi-bastian-feb09.pdf.

63. Perez-Riverol Y *et al.* 2019 The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450. (doi:10.1093/nar/gky1106)