Optimization of skeletal protein preparation for LC-MS/MS sequencing yields additional coral skeletal proteins in *Stylophora pistillata*

Yanai Peled[1*], Jeana Drake[1*^], Assaf Malik[1], Ricardo Almuly[1], Maya Lalzar[2], David Morgenstern[3], Tali Mass[1+]

[1]Marine Biology Department, University of Haifa, Haifa, Israel
[2]Bioinformatics Core Unit, University of Haifa, Haifa, Israel
[3]De Botton Protein Profiling Institute of the Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science, Israel
*equal contribution
[^]Corresponding author during review: jeanadrake@g.ucla.edu
[+]Corresponding author upon publication: tmass@univ.haifa.ac.il

Abstract

Stony corals generate their calcium carbonate exoskeleton in a highly controlled biomineralization process mediated by a variety of macromolecules including proteins. Fully identifying and classifying these proteins is crucial to understanding their role in exoskeleton formation, yet no optimal method to extract and isolate and characterize coral skeletal proteins has been established and their complete composition remains obscure. Here, we tested four skeletal protein extraction protocols using acetone precipitation and ultrafiltration dialysis filters to present a comprehensive scleractinian coral skeletal proteome. We identified a total of 60 proteins in the coral skeleton, 44 of which were not present in previously published stony coral skeletal proteomes. Extracted protein treatment protocols carried out in this study revealed that there is no “one optimal method” and each protocol revealed a unique set of method-exclusive proteins. To better understand the general mechanism of skeletal protein transportation, we further examined the proteins’ gene ontology, transmembrane domains, and signal peptides. We found that transmembrane domain proteins and signal peptide secretion pathways, by themselves, could not explain the transportation of proteins to the skeleton . We therefore propose that proteins are transported to the skeletal via vesicles and possibly non-traditional secretion pathways.

Keywords

Skeleton organic matrix proteins (SOMPs), biomineralization, Scleractinia

Introduction

Scleractinian corals, or stony corals, are the most prolific biomineralizers in phylum Cnidaria [1]. They are a key components of shallow-water tropical reefs, often forming massive structures that serve as the foundation of an ecosystem which hosts some of the more biodiverse communities on the planet [2, 3]. They are amongst the oldest biomineralizing metazoans, producing calcium carbonate ($CaCO_3$) exoskeletons in the form of aragonite through biologically-directed mechanisms [4, 5], which makes up ≥95% of the entire skeletal mass, the remainder of which is the skeletal organic matrix (SOM) [6-8].

Biomineralization refers to the ability of a living organism to selectively exploit elements from its surrounding environment to build a biologically functioning crystalline structure [9, 10]. Biomineralizers can be found throughout all kingdoms of life, from bacteria [e.x. 11], to algae [e.x., 12], mollusks [e.x., 13], corals [e.x., 10] and mammals [e.x., 14]. The minerals formed through this process differ in structure from their non-biological counterparts and their formation is mediated by a variety of organic molecules (the SOM), which have been intensively studied since the 1960's in diverse organisms [reviewed by 10]. The SOM consists of proteins, lipids, and polysaccharides, which are not present in the geological mineral form [reviewed by 15]. SOM proteins (SOMPs) embedded within the mineral are hypothesized to serve as a framework for crystal nucleation [reviewed by 16]. The proteins involved in the process of skeletal biomineralization have been described and studied extensively in echinoderms [e.x., 17, 18] [and reviewed by 19, 20], mollusks [e.x., 21, 22-24] and

mammals [e.x., 25, 26, 27], among others [28]. At present, the best described SOMP complex is of mammalian bone and teeth [15].

In stony corals, the most current knowledge of SOMPs is limited and based on intraskeletal protein extraction [29-33] . It has been suggested that coral SOMPs aid in the molecular processes of crystallization as well as in the development and strengthening of the minerals ([34-37] among others). The constant advancement of mass spectrometry technology has broadened our capability to identify many proteins in skeletal extracts, even those proteins in low abundance [38]. However, this technology is sensitive to contamination by organic matter remnants from soft tissue and cell debris from the study organism, and the little-addressed issue remains of contamination by researchers during the protein extraction, preparation, and sequencing steps [39-41]. Upon extraction, the SOMPs are usually divided into two fractions: acid-soluble matrix and acid-insoluble matrix proteins (ASM and AIM respectively), based on their solubility in the acid of choice [22, 31-33, 42]. While some past attention has been directed towards the soluble fraction [30, 43, 44], Pereira-Mouriès et al. [45] showed that, in the bivalve *Pinctada maxima*, the classification of AIM and ASM is misleading and that both fractions share common features. Furthermore, Goffredo et al. [42] found in the stony coral *Balanophyllia europaea* that both fractions consist of the same macromolecules; they associated the degree of solubility to the difference in cross-linking. They also showed that each solubility fraction has a different influence on calcium carbonate crystal morphology, aggregation, and polymorphism in-vitro. In contrast, Ramos-Silva et al. [32] observed a different SOMP composition between solubility fractions in the scleractinian coral *Acropora millepora*,. Out of 36 SOMPs, only two were found exclusively in the soluble fraction and twelve were exclusive to the insoluble fraction. These examples

demonstrate the attempts to attribute different properties to the two fractions but the data remain inconclusive.

To date, three major coral skeletal proteomes have been published [31-33] with each proteome consisting of 30-40 proteins. Of the 30 proteins sequenced from *A. digitifera* skeleton, 26 were also detected in *A. millepora* skeleton [32, 33]. They consist mostly of either transmembrane (TM) domain proteins or secretory proteins [33]. However, only 12 of the proteins identified in *A. millepora* skeleton matched those found in *S. pistillata* skeleton [31, 32]. In *A. millepora*, 11 TM domain-containing proteins were identified, as well as two proteases that were not detected in *S. pistillata* [32]. The authors suggested that the proteases' role is in cleaving the extracellular domain of TM proteins and incorporating them into the skeleton.

The coral skeletal proteomes published to date reveal an overlap of several detected proteins, but at least 10 proteins from each species appear to be unique. It is currently unknown if this is truly due to species-specific gene expression and protein localization or to methods in extracting, purifying, and sequencing the proteins. In this study we analyzed several methods for protein purification to increase the detection of proteins from cleaned coral skeleton powder. We show that the use of acetone precipitation versus centrifugal filter washing, and the degree to which each purification method is performed, affects the numbers and types of proteins that can be sequenced by mass spectrometry. Further, we suggest that there is no one 'best' method for coral skeletal protein preparation such that future research projects may need to utilize several preparation methods to detect the full breadth of proteins embedded in coral skeleton.

Methods

Sample collection and preparation for protein extraction

The hermatypic coral *Stylophora pistillata* (Esper, 1797) was collected under a special permit from the Israeli Natural Parks Authority in the waters in front of the H. Steinitz Marine Biology Laboratory, Eilat, Israel, Red Sea (29°30 N, 34°56 E), using SCUBA diving.

We fragmented one *S. pistillata* colony into small pieces, approximately 2x2 cm, with a diamond band saw. Coral fragments were transferred to 50-ml Falcon brand conical vials (Falcon tubes) and oxidized with 20 mL 1:1 of 30% H_2O_2: 3% NaClO solution for 1 hour, during which 1.5 mL of 3% NaClO solution were gently added to the tubes every 20 minutes and continued the incubation overnight at room temperature following modified methods of Stoll et al. [46]. Fragments were washed with double distilled water five times for one minute each time and dried at 60 C˚ overnight. We crushed the cleaned fragments to ≤63 µm diameter with a mortar and pestle. Skeleton powder was then oxidized in sterile Falcon tubes and washed as above three more times (i.e., four complete rounds of oxidative cleaning) to ensure that no organic residue remained on the skeletal grains. In each cycle, the removal of the oxidizing or washes solution was performed by centrifugation at 5,000 x g for 3 min at 4 C˚. Cleaned skeletal powder was then dried overnight at 60 C˚. We carried out all the described processes in a laminar flow biological hood (apart from oven drying) with all preparation tools and surfaces bleached to avoid contamination.

To monitor the removal of proteins from the skeletal fragments, we checked the cleaning efficiency under SEM after the fourth oxidative cleaning. Samples were vacuum-coated with 4 nm gold prior to examination under a ZEISS Sigma TM scanning electron microscope an in-lens detector (5kV, WD = 5-7mm) (SI Figure 1a,b). In addition, we sonicated the powder at 4˚C in filter-sterilized phosphate buffered saline (PBS, pH 7.4) for 30 minutes, pelleted the powder at 5,000 x *g* for 3

minutes at 4˚C, concentrated the supernatant on a 3-kDa cutoff centrifugal filter unit (Amicon) and loaded samples of supernatant on a 8-16% SDS-PAGE TGX stain free gels (Bio-Rad) (SI Figure 1c).

Extraction of skeletal proteins

We decalcified approximately 1.3 g cleaned skeleton powder per protein extraction protocol in 0.5 M acetic acid (30 ml acid/g cleaned skeleton powder) in Falcon tubes while rotating the tubes at room temperature for 3 hours. Samples were then centrifuged at 5,000 x *g* for 5 min at 4°C and supernatant was transferred to a new tube and stored at 4°C. We continued the decalcification of the undissolved pellets with a second volume of 0.5 M acetic acid and allowed decalcification to proceed to completion. We then combined both liquid fractions (70 ml total) for each sample, froze the total volumes at -80˚C, and dried them by overnight lyophilization. The dried pellets were stored in -80°C until further processing.

The lyophilized pellets were re-suspended in 12 ml MilliQ water and the proteins were concentrated on 3 kD cutoff Amicon® Ultra 15 centrifugal filter units (Merk-Millipore) 5,000 x *g* at 4˚C to reach a final volume of 0.5 ml. To continue desalting the samples we diluted them again to 12 ml in MilliQ water and repeated the concentration procedure as above. At this stage, we observed an insoluble pellet in all samples.

Because washes of the extracted pellet can reduce the representation of proteins, in this study we examined two protein concentrating and cleaning methods performed for previously published coral skeletal proteomes; (i) centrifugation ultrafiltration (CF methods [32]) and (ii) acetone precipitation (ACT methods, [31]) (Table 1; SI Table 1). Further, for each method we separately examined the ASM and AIM fractions.

Two filter-concentrated samples containing both soluble and insoluble fractions were examined for the effects of centrifugal filtration. Sample CF2 was centrifuged at 5,000 x g for 5 min at 4°C to pellet the AIM fraction. Both the ASM and AIM of CF2 were divided into separate sterile 1.5 ml microcentrifuge tubes, lyophilized, and stored at -80°C until further use. Sample CF4 was desalted as described above for a third time before separating ASM and AIM fractions.

Both ACT samples were centrifuged as above for CF2 to separate ASM and AIM fractions. To sample ACT1 ASM (concentrated to 0.5 ml by centrifugal filtration as described above) was added 2 ml 100% ice cold acetone. The sample was vortexed for 10 seconds, incubated at -20°C for 30 minutes, and centrifuged at 4,300 x g for 30 min at 4°C. Finally, the pellet was washed three more times with 2 mL of 80% ice cold acetone and stored then at -80°C until further use. The AIM fraction was similarly washed four times with 80% acetone. Both fractions of sample ACT3 were treated as in ACT1 but with one less washing step of each fraction.

LC MS\MS

S. pistillata skeletal protein samples were dissolved in 5% SDS and digested with trypsin using the S-trap method overnight at room temperature. We analyzed the resulting peptides with nanoflow ultra-performance liquid chromatograph (nanoAcquity) coupled to a high resolution, high mass accuracy mass spectrometer (Fusion Lumos). The sample was trapped on a Symmetry C18 0.18*20mm trap column (Waters, Inc) and separated on a HSS T3 0.075*250 mm column (Waters, Inc.) using a gradient of 4-28% (80% acetonitrile, 0.1% Formic acid) for 150 minutes. Spray voltage was set to +2kV. The data were acquired in the Fusion Lumos using a Top Speed Data-Dependent Acquisition method using a cycle time of 3 s. An MS1 scan was performed in the Orbitrap at 120,000 resolution with a maximum injection

time of 60 ms. The data were scanned between 300-1800 m/z. MS2 was selected using a monoisotopic precursor selection set to peptides, peptide charge states set to +2-+8 and dynamic exclusion set to 30 s. MS2 was performed using HCD fragmentation scanned in the Orbitrap, with the first mass set to 130 m/z at a resolution of 15,000. Maximum injection time was set to 60 ms with automatic gain control of $5x10^{-4}$ ions as a fill target. The resulting data were searched against the NCBI *Stylophora pistillata* protein database using the Byonic search engine (Protein Metrics Inc.) – the first search was carried out without any false discovery rate (FDR) filtering, to generate a focused database for a second search. The second search was set to 1% FDR, allowing fixed carbamidomethylation on C and variable oxidation on MW, deamidation on NQ and protein N-terminal acetylation. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the Pride [47] partner repository under the dataset identifier PXD017891; reviewer login username and password have been provided to the reviewers. Data will be made publicly available without a required login in ProteomeXchange upon publication of the manuscript.

Data sorting

We used the *S. pistillata* genome database as a reference peptide database for the mass spectrometry analysis [48] (NCBI BioProjects PRJNA281535 and PRJNA415215) appended with known NCBI *S. pistillata* skeletal proteins[e.x., 31, 49] *S. pistillata* carbonic anhydrase, ACE95141.1). We also included a common contaminants database. Despite this inclusion of contaminants databases, several proteins likely of human origin were sequenced and attributed to *S. pistillata*. To filter out these potential contaminants from our final list of coral-specific proteins, we BLASTed all sequences against the 'Primates' database in NCBI using Blast2GO.

We then examined NCBI-generated sequence alignments of coral versus *Homo sapiens* proteins with e-values lower than e^{-50} and percent mean similarity greater than 50%, all sequences with e-values lower than e^{-100}, and all sequences with percent similarity greater than 80%, and removed from our final list of coral proteins any sequences with three or more peptides each of seven or more amino acids in length that were identical between *S. pistillata* and humans.

All proteins identified by the LC MS\MS analysis were filtered to proteins with at least two significant spectra and at least one significant unique peptide. Skeletal proteins were first sorted by fractions and methods (ASM and AIM for each extraction method). Next, we sorted all skeletal proteins by their gene ontology (GO) terms [50, 51] to differentiate plasma membrane proteins and extracellular matrix (ECM) proteins. Finally, we grouped all by terms of interest: protein modification, transmembrane, and ECM; membrane processing; metal binding and vesicular.

Proteins detected by LC-MS/MS were annotated using the Trinotate pipeline which relies on both pfam and Uniprot data [52]. GO terms and PFAM annotations were assigned to *Stylophora pistillata* predicted proteins and transcripts using Trinotate 3.0.1 (https://github.com/Trinotate). Transmembrane regions were predicted using the TMHMM server v2.0 [53]. Signal peptides on the N-termini of proteins were predicted using SignalP 5.0 [54, 55]. Attachment of proteins to the exterior of the cell membrane by glycosylphosphatidylinositol (GPI) anchors was predicted using PredGPI [56]. These computational analyses used program default settings and cutoffs. Completeness of protein sequences was determined by comparing all returned coral proteins to the *Acropora digitifera* genome [57]; NCBI RefSeq assembly GCF_000222465.1). Detection of various proteins across solubility and protocol fractions was visualized in Venny 2.1 online software.

We determined orthologous biomineralization genes, known from skeletal proteomic analysis, across coral taxa. First, we estimated orthology relationships between all non-redundant genes of selected metazoa species using OrthoFinder. We included all genes of all Cnidaria species with known genome-based annotations. OrthoFinder generates orthology-groups (Orthogroups) based on normalized reciprocal best Blast hits' bit scores [58], and then estimates orthologues genes pairs within Orthoroups [59]. We then selected all pairs of *Acropora* spp. orthologs to *S. pistillata* (1:1, 1:many, many:many relationships) ([31-33], this study). From these pairs we further selected *S. pistillata* spectra-based identified proteins, or skeletal proteins known in the literature. Since not all skeletal protein annotations from the literature were included in our reference Orthofinder proteome datasets, we further found their best matches in the reference orthofinder proteome using blastp.

Results:

After extensive cleaning of the powdered skeleton and acid-extraction of embedded organic matter we identified in total 60 coral-specific proteins, with two or greater spectra or one unique peptide with at least 10 spectra and an identification score of 250 or greater, in *S. pistillata* skeleton as predicted by the species genome [48] (Table 2, SI Table 2). Trinotate [60] [52] returned GO annotations [50] for all of these proteins, although many remain uncharacterized (SI Table 3).

In order to evaluate the efficacy and improve current methods for stony coral skeletal protein extraction, we examined four different protocols; two centrifugation ultrafiltration filters (CF) and two further acetone precipitation (ACT) protocols. Proteomes of the four methodologies differed in composition and variety (Figure 1). Combining results of all CF fractions identified 52 coral-specific proteins while combined ACT protocols yielded 13 such proteins (Figure 1A). Moreover,

redundancy between methodologies was low. Only 8.3% of the proteins overlapped between methods while 78.3% and 13.3% of the proteins were exclusive to combined CF and combined ACT fraction data, respectively.

To evaluate the extracted protein treatment efficiency of each protocol, we first compared the number of proteins detected in each method (e,x,, ACT1 vs ACT 3 and CF2 vs CF4). Of the 8 proteins found only in samples from the acetone wash protocols (ACT 1 and ACT 3), one was observed only in ACT1 while five were observed only ACT3 (Figure 1, SI Table 2); both ACT1 solubility fractions had one more wash step than did those in ACT3. In contrast, of the 47 proteins observed only in CF samples, 11 were found in CF2, which went through two filter centrifugation steps, while 16 were found in CF4, which went through a third filter centrifugation step. Further, only two proteins were observed in all four purification fractions in at least one solubility form: CARP4/SAARP1 and synapsin 2-like. Two additional proteins were observed in three of the four protocols: STPCA2 and an uncharacterized proteins.

We next examined the difference in protein composition obtained from the solubility fractions ASM versus AIM. We identified different distributions of proteins in the ASM vs. AIM in both purification methods (Figures 1B & C). Notably, when combining all methods, more proteins were identified in the AIM compared to the ASM. A total of 57 coral proteins were identified in the AIM fraction compared to a total of 8 in the ASM fraction (SI Table 2), of which 5 (8.3%) were identify both in ASM and AIM.

Our data compared to other coral skeletal proteomes

Since our analysis yielded a large amount of new skeletal proteins, we also compared our results with the three previously published proteome of *S. pistillata, A. digitifera*

and *A. millepora*. Out of our entire identified skeletal proteome containing 60 proteins, using OrthoFinder and BlastP, only 16 were found to be similar to proteins identified in these studies. Yet, this proportion of overlap (16 out of 60) is significantly greater than the expected proportion by chance, since the proportion of known skeletal matrix proteins in the reference coral proteomes is extremely small (less than ~0.2%). Seven proteins were found to overlap all four proteomes: a coadhesin-like protein, an EGF and laminin G domain-containing protein, a hypothetical protein, a MAM and LDL-receptor class A domain-containing protein, a mucin, aspartic acid-rich protein 2-like, and a ZP domain-containing protein (Table 3).

Skeletal proteome characterization

We interrogated the mechanisms by which proteins may be exported from or attached to the cell (SI Table 2). Seventeen coral skeletal proteins with likely complete N-terminus predictions possess signal peptides as a potential mechanism for export from the cell. Eight proteins contain at least one transmembrane span suggesting that they are embedded in the cell membrane. Further, 10 proteins likely interact with the exterior of the cell membrane by GPI anchors. In total, 25 of the 60 sequenced coral skeletal proteins exhibit documented characteristics for localization in the ECM.

Because the majority of the proteins sequenced from the *S. pistillata* skeleton do not possess features for signaling their export from the cell, we queried the data set for further suggestions of positioning the proteins in the membrane or that the proteins may be exported by vesicles such as those that may be involved in calcium concentration. To do this, we examined the skeletal proteome annotations and GO classifications toward finding common features to allow grouping of proteins. Out of the entire skeletal proteome sequenced in this study (60 proteins), 39 genes were

returned with GO terms that allowed their classification into five groups of interest based on their cellular component, biological process, and molecular function to suggest likely cellular locations pertinent to the calcification mechanism, which may therefore be indicative of their function in this process: lipid\phosphate\glycan related proteins (i.e., membrane processing); ECM-related, transmembrane, and protein modification proteins; and metal binding proteins vesicular/secretion related proteins, (SI Table 3).

Of the proteins with GO terms, 10 proteins are suggested to be involved with processing of the cell membrane (Figure 2), A much larger number are related to vesicles/secretion as well as binding metal, with 21 and 25 assigned to each category, respectively. We combined these two categories in our proposed cellular location Figure 2 as some of the skeletal proteins proposed to be found intracellularly in vesicles are also known to bind calcium [34, 61]. Finally, 19 and 12 are potential ECM proteins or are involved in protein modification, respectively. It should be noted that many protein are assigned to multiple categories.

Discussion:

In this study we show the importance of using complementary post-extraction methods to concentrate coral skeletal proteins for sequencing the full breadth of the skeletal proteome. Our results show a clear and marked difference in detected proteins between protein purification methods, with only two proteins observed in all methods. Centrifugal Filter (CF) methods yielded a much greater abundance and diversity of proteins than did acetone precipitation (ACT) (Figure 1). Moreover, we found that protein detection is largely method-exclusive (Figure 1). We hypothesize that the protein yield differences between extraction methods is based on the different properties of each method. CF is a mechanical filtration based on size and has a bias

toward hydrophobic proteins, whereas ACT is based mainly on the chemical interactions of proteins resulting in increased precipitation of hydrophilic proteins [62]. We speculate that the CF methods yielded more total proteins in our study as acetone precipitated proteins may be difficult to resolubilize, potentially leading to the possible loss of many proteins in the pellet which is not transferred to the trypsin digestion step.

A major challenge in working with intra-skeletal proteins is isolating the true skeletal proteins from soft tissue contamination [40, 41]. In the present study, as in previous work on *S. pistillata* skeletal proteins [40] we carried out an intensive oxidative cleaning step on the skeletal powder, in addition to cleaning the skeletal fragments, to avoid contamination. We did not observe any organic residues by SEM on intact skeletons following the second cleaning step (SI Figure 1a,b) or in PBS soaked on cleaned powders and concentrated (SI Figure 1c), and we are therefore confident that all sequenced proteins are endogenous to the skeleton and are not coral cellular contamination.

In this study, we examined the protein composition of the ASM and AIM proteins separately. Although we see variations due to the different purification techniques (Figure 1), our results indicate that, broadly, the AIM is distinguished in composition from the ASM. Across all purification techniques, the AIM fractions yielded more total proteins and fraction-exclusive proteins than the ASM fraction did (Figure 1B, C). Since all extraction and precipitation experiments were carried out simultaneously, we can rule out batch effect and technical differences. This AIM versus ASM skeletome fractional composition is similar to previous studies carried out on other marine biomineralizers such as mollusks [45] as well as to the stony corals *A. millepora* [32] and *S. pistillata* [31]. However, the overlap between the two

solubility fractions reported for *A. millepora* [32] was 61%, in contrast to our study which shows an overlap across all purification treatments combined of only 10% (Figure 1). One plausible explanation for this difference is the use of improved LC-MS/MS performance over the past several years enabled us to detect more proteins and obtain a more comprehensive proteome. In previous studies of coral skeletal proteomes which detailed differences between solubility fractions, the instruments used were low resolution, low mass accuracy, which tend to result in a lower percent identification of the data. Additionally, most search engines and FDR-calculating algorithms struggle with very small datasets [63, 64]. Our use of Byonic helps to alleviate statistical limitations that can result in false negative results.

Further differences in proteomes, beyond species differences, are the differing reagents used in precipitation (compared to [31]) and our smaller centrifugal filter cutoffs (compared to [32]). This second difference is particularly important with respect to protein degradation. Even when embedded in biominerals so that amino acids and even short peptides persist, proteins may still succumb to degradation [65]. This result in small peptides fragments that may be lost from centrifugal filter units with 10 kDa pore sizes and larger. It is reasonable to assume that coral skeletal proteins go through the same process, and if so, the cutoff of the membrane directly affects the number of peptide spectra matches (PSM). Using smaller cutoff filters in this study might have allowed us to capture some of these sheered peptides and led to higher PSMs.

Traditional protein trafficking by signal peptides and transmembrane domain does not explain the full extent of protein transportation to skeleton

Corals' skeleton is external to the animal; therefore, proteins in the skeletal matrices must be transported outside the cells or span the membrane and have an extracellular

portion to reach the skeletal crystallization front. Indeed, recent studies of anthozoans reveal a significant proportion of TM domain proteins (~35%) in the SOM [32, 33], similar to that found in the better-studied Echinoidia [66] . Based on these findings, we examined the hypothesis that many of the skeletal proteins originate in the plasma membrane.

Our analysis revealed that TM domain proteins are not the major component of the SOM protein complex; in the present proteome, TMHMM prediction suggests that eight are embedded in the membrane. Hence we examined a cellular secretion option. Ramos-Silva et al. [32] reported 15 proteins (41%) with SPs that did not also possess a TM domain in the *A. millepora* proteome. In their study of the *A. digitifera* proteome, Takeuchi et al. [33] reported a similar proportion of proteins with SPs but no TM domains (40%). When examining the combined S. *pistillata* proteomes, 17 out of 60 (28%) were positive for SPs (SI Table S2, [31]), likely a combination of incomplete gene prediction and other mechanisms for exporting the proteins to the calicoblastic space.

Hydrophobic regions of TM domains are similar to those of SPs, making SP prediction difficult [53, 67]. We therefore cannot say for sure, based on these two analyses (TMHMM and SignalP) alone, which is the more dominant pathway for protein transport. While SP by itself can help predict traditional secretory pathways [67], proteins that do not possess SPs may still leave the cell via other means without classic SPs. This includes membrane pores [67], ATP-binding cassette transporters [68] and autophagosome/endosome-based secretion [69, 70]. These are but a few examples of secretory pathways which bypass the conventional TM/SP signaling pathway (reviewed in [67, 70, 71]).

A further method by which protein may be exported to the site of calcification is a vesicular pathway that differs from the conventional SP and TM pathway. Previous studies have shown Ca^{2+} rich granules in the calicoblastic epithelium (skeletogenic cells), but not in the other tissue layers, suggesting their role as a Ca^{2+} reservoirs in the cells. Vesicles were previously identified in corals [72, 73]; however, their origin and content was not detailed and they are sometimes attributed to preservation byproducts. These intracellular ion-rich vesicles may endocytose sea water [74], after which they are enriched in carbonate ions and then form hydrated ACC and anhydrous ACC precursors stabilized by acidic biomolecules including CARPs [75]. Using cell cultures, Mass et al. [61] suggest that the vesicles, which contain Asp rich proteins, then transport their contents to the ECM, releasing their content by exocytosis. The biomineral then further develops extracellularly, likely aided by other ECM proteins [30, 35, 49, 76] as well as other biomolecules [77-80]. At present the process of calcium delivery to the skeleton and the role of most of the proteins in skeletal depositing remain to be determined.

Since TM and SP analysis did not fully explain the mechanism of protein transportation, we classified the likely functions and locations of the skeletal proteins according to their GO terms as lipid\phosphate\glycan related proteins, metal binding proteins, vesicular/secretion related proteins, ECM-related and transmembrane proteins, and protein modification proteins (Figure 2, SI Table 3). Looking at the proposed locations or functions of skeletal proteins (Figure 2), we find comparable numbers of vesicular and metal binding proteins versus TM domain/ECM proteins. These results are in-line with the trend reported in previous coral skeletal proteomes [31-33]. While we do not exclude the role of TM proteins in the skeletal deposition

process, we suggest that the biological mechanism of SOM transportation is enhanced by vesicular secretion and other possible, non-traditional, secretion pathways.

It is difficult to map and characterize proteins in non-model organisms such *S. pistillata* [81-83]. High quality proteomic mapping requires knowledge of phosphorylation, glycosylation, proteolytic cite activities and other modifications [31, 84-86], in order to create a more thorough database [81-83]. Further, many of the coral skeletal proteins reported to date remain uncharacterized. Uncharacterized proteins were reported in *Acropora* skeletons at a rate of approximately 25% [32, 33] while, they were previously reported at less than 10% in *Stylophora* skeleton [31]. Our study revealed a much greater proportion of 28% uncharacterized proteins. While partially attributed to sample size, it is most likely due to quality of genomic data available, since stony corals are non-model organisms and their genomic libraries are far from being complete, resulting in incomplete databases on which to map the proteome and many uncharacterized genes.

Conclusion:

In this study we have considered the differential effects of coral skeletal protein extract preparation as well as the method by which these proteins, or parts thereof, are transported from intracellular to extracellular locations. When preparing coral skeletal proteomes, we propose that a multi-method approach to cleaning, demineralization, and protein extraction should be used. Our results showed that each protein preparation protocol yielded exclusive sets of proteins with little overlap between ACT and CF fractions. While CF protocols yielded many more proteins than did ACT methods, use of a single protocol to clean and concentrate coral skeletal proteins results in a significant amount of data loss, and it is therefore of crucial importance to consider alternative and complementary methods to obtain a fully comprehensive

skeletal proteome. We showed that while the role of TM domain proteins cannot be overlooked, many of the proteins detected in the *S. pistillata* skeletal proteomes as well as in that of other species point toward other secretory or vesicular pathways. Our categorization method, supported by data from other recent studies, also suggests that corals use an alternative secretory pathway, such as vesicles, and much work is required in order to determine the calcium deposition pathway and the proteins involved. Our study provides a large set of new uncharacterized coral skeletal proteins as well as others of purported function but that have not been observed before in the coral SOM. These data expand the current knowledge of the SOM in corals and will help, in future studies, to resolve corals' calcium deposition mechanism and the various roles of the proteins involved.

Abbreviations

acetone precipitation of proteins (ACT)

acid soluble matrix (ASM)

acid insoluble matrix (AIM)

centrifugation ultrafiltration of proteins (CF)

false discovery rate (FDR)

gene ontology (GO)

glycosylphosphatidylinositol (GPI)

peptide spectra matches (PSM)

phosphate buffered saline (PBS)

skeletal organic matrix (SOM)

skeletal organic matrix proteins (SOMP)

transmembrane (TM)

Declarations

Ethics approval and consent to participate

Coral nubbins were collected under a special permit from the Israeli Natural Parks Authority.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated during the current study are available in the ProteomeXchange repository under file number PXD017891 (https://www.ebi.ac.uk/pride/archive). The reviewer login username and password have been provided to the reviewers. Data will be made publicly available without a required login in ProteomeXchange upon publication of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Israel Science Foundation (Grant 312/15), and the European Research Commission (ERC; Grant # 755876). JLD was supported by the Zuckerman STEM Leadership Program.

Author contributions

YP, JD, and TM designed the study. YP and RA prepared the samples. YP, RA, AM, ML, DM, and JD analyzed the data. All authors wrote the manuscript and approve this submission.

Acknowledgements

We thank staff at the Bioinformatics Core Unit, University of Haifa, and The Crown Genomics Institute of the Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science.

References

1. Cairns SD: **Deep-water corals: An overview with special reference to diversity and distribution of deep-water scleractinian corals.** *Bull Mar Sci* 2007, **81:**311-322.
2. Reaka-Kudla ML: **The global biodiversity of coral reefs: a comparison with rain forests.** *Biodiversity II: Understanding and Protecting our Biological Resources* 1997, **2:**551.
3. Knowlton N, Brainard RE, Fisher R, Moews M, Plaisance L, Caley MJ: **Coral reef biodiversity.** *Life in the world's oceans: Diversity distribution and abundance* 2010**:**65-74.
4. Erwin DH, Laflamme M, Tweedt SM, Sperling EA, Pisani D, Peterson KJ: **The Cambrian conundrum: Early divergence and later ecological success in the early history of animals.** *Science* 2011, **334:**1091-1097.
5. Tambutté S, Tambutté E, Zoccola D, Allemand D: **Organic matrix and biomineralization of scleractinian corals.** *Handbook of Biomineralization: Biological Aspects and Structure Formation* 2007**:**243-259.
6. Cuif J-P, Dauphin Y, Berthet P, Jegoudez J: **Associated water and organic compounds in coral skeletons: Quantitative thermogravimetry coupled to infrared absorption spectrometry.** *Geochemistry Geophysics Geosystems* 2004, **5:**Q11011.
7. Wainwright SA: **Skeletal organization in the coral, *Pocillopora damicornis*.** *Quarterly Journal of Microscopical Science* 1963, **s3-104:**169-183.
8. Von Euw S, Zhang Q, Manichev V, Murali N, Gross J, Feldman LC, Gustafsson T, Flach C, Mendelsohn R, Falkowski PG: **Biological control of aragonite formation in stony corals.** *Science* 2017, **356:**933-938.
9. Mann S: *Biomineralization: Principles and Concepts in Bioinorganic Materials Chemistry*. New York: Oxford University Press; 2001.
10. Weiner S, Dove PM: **An overview of biomineralization processes and the problem of the vital effect.** *Reviews in Mineralogy and Geochemistry* 2003, **54:**1-29.
11. Schüler D, Frankel RB: **Bacterial magnetosomes: microbiology, Biomineralization and biotechnological applications.** *Appl Microbiol Biotechnol* 1999, **52:**464-473.
12. Westbroek P, Young J, Linschooten K: **Coccolith production (biomineralization) in the marine alga *Emiliania huxleyi*.** *The Journal of Protozoology* 1989, **36:**368-373.
13. Addadi L, Joester D, Nudelman F, Weiner S: **Mollusk shell formation: a source of new concepts for understanding biomineralization process.** *Chemical European Journal* 2006, **12:**980-987.
14. Levine P, Glimcher M: **The isolation and amino acid composition of the organic matrix and neutral soluble proteins of developing rodent enamel.** *Archives of Oral Biology* 1965, **10:**753-IN751.
15. Beniash E: **Biominerals—hierarchical nanocomposites: the example of bone.** *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* 2011, **3:**47-69.
16. Lowenstam HA, Weiner S: *On Biomineralization.* New York, New York: Oxford University Press; 1989.
17. Karakostis K, Zanella-Cléon I, Immel F, Guichard N, Dru P, Lepage T, Plasseraud L, Matranga V, Marin F: **A minimal molecular toolkit for**

mineral deposition? Biochemistry and proteomics of the test matrix of adult specimens of the sea urchin *Paracentrotus lividus*. *Journal of Proteomics* 2016, **136:**133-144.

18. Shashikant T, Khor JM, Ettensohn CA: **From genome to anatomy: The architecture and evolution of the skeletogenic gene regulatory network of sea urchins and other echinoderms.** *Genesis* 2018, **56:**e23253.

19. Nebelsick JH, Dynowski JF, Grossmann JN, Tötzke C: **Echinoderms: hierarchically organized light weight skeletons.** In *Evolution of Lightweight Structures*. Springer; 2015: 141-155

20. Killian CE, Wilt FH: **Molecular aspects of biomineralization of the echinoderm endoskeleton.** *Chemical Reviews* 2008, **108:**4463-4474.

21. Weiner S, Hood L: **Soluble protein of the organic matrix of mollusk shells: a potential template for shell formation.** *Science* 1975, **190:**987-989.

22. Mann K, Edsinger-Gonzales E, Mann M: **In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*.** *Proteome Science* 2012, **10:**1-18.

23. Politi Y, Mahamid J, Goldberg H, Weiner S, Addadi L: **Asprich mollusk shell protein: in vitro experiments aimed at elucidating function in $CaCO_3$ crystallization.** *CrystEngComm* 2007, **9:**1171-1177.

24. Bouyoucef M, Rakic R, Gómez-Leduc T, Latire T, Marin F, Leclercq S, Carreiras F, Serpentini A, Lebel J-M, Galéra P: **Regulation of extracellular matrix synthesis by shell extracts from the marine bivalve *Pecten maximus* in human articular chondrocytes—application for cartilage engineering.** *Mar Biotechnol* 2018**:**1-15.

25. Weiner S, Wagner HD: **The material bone: Structure-mechanical function relations.** *Annual Review of Materials Science* 1998, **28:**271-298.

26. Reid DG, Duer MJ, Murray RC, Wise ER: **The organic−mineral interface in teeth is like that in bone and dominated by polysaccharides: Universal mediators of normal calcium phosphate niomineralization in vertebrates?** *Chemistry of Materials* 2008, **20:**3549-3550.

27. Sharma V, Srinivasan A, Roychoudhury A, Rani K, Tyagi M, Dev K, Nikolajeff F, Kumar S: **Characterization of protein extracts from different types of human teeth and insight in biomineralization.** *Scientific Reports* 2019, **9:**9314.

28. Kamel BS, Voolstra CR, Medina M: **BioMine-DB: A database for metazoan biomineralization proteins.** *PeerJ Preprints* 2016, **4:**e1983v1982.

29. Fukuda I, Ooki S, Fujita T, Murayama E, Nagasawa H, Isa Y, Watanabe T: **Molecular cloning of a cDNA encoding a soluble protein in the coral exoskeleton.** *Biochem Biophys Res Commun* 2003, **304:**11-17.

30. Puverel S, Tambutté E, Pereira-Mouries L, Zoccola D, Allemand D, Tambutté S: **Soluble organic matrix of two Scleractinian corals: Partial and comparative analysis.** *Comparative Biochemistry and Physiology Part B* 2005, **141:**480-487.

31. Drake JL, Mass T, Haramaty L, Zelzion E, Bhattacharya D, Falkowski PG: **Proteomic analysis of skeletal organic matrix from the stony coral *Stylophora pistillata*.** *Proceedings of the National Academy of Sciences* 2013, **110:**3788-3793.

32. Ramos-Silva P, Kaandorp J, Huisman L, Marie B, Zanella-Cleon I, Guichard N, Miller DJ, Marin F: **The skeletal proteome of the coral *Acropora***

***millepora*: The evolution of calcification by cooption and domain shuffling.** *Mol Biol Evol* 2013.

33. Takeuchi T, Yamada L, Shinzato C, Sawada H, Satoh N: **Stepwise evolution of coral biomineralization revealed with genome-wide proteomics and transcriptomics.** *PLOS ONE* 2016, **11:**e0156424.

34. Mass T, Drake Jeana L, Haramaty L, Kim JD, Zelzion E, Bhattacharya D, Falkowski Paul G: **Cloning and characterization of four novel coral acid-rich proteins that precipitate carbonates in vitro.** *Curr Biol* 2013, **23:**1126-1131.

35. Puverel S, Tambutté E, Zoccola D, Dumart-Coulon I, Bouchot A, Lotto S, Allemand D, Tambutté S: **Antibodies against the organic matrix in scleractinians: A new tool to study coral biomineralization.** *Coral Reefs* 2005, **24:**149-156.

36. Clode PL, Marshall AT: **Calcium associated with a fibrillar organic matrix in the scleractinian coral *Galaxea fascicularis*.** *Protoplasma* 2003, **220:**153-161.

37. Watanabe T: **Molecular analyses of protein components of the organic matrix in the exoskeleton of two scleractinian coral species.** *Comp Biochem Phys B* 2003, **136:**767 - 774.

38. Scheltema RA, Hauschild J-P, Lange O, Hornburg D, Denisov E, Damoc E, Kuehn A, Makarov A, Mann M: **The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer.** *Molecular & Cellular Proteomics* 2014, **13:**3698-3708.

39. Drake JL, Mass T, Falkowski PG: **The evolution and future of carbonate precipitation in marine invertebrates: Witnessing extinction or documenting resilience in the Anthropocene?** *Elementa Science of the Anthropocene* 2014, **2:**000026.

40. Drake JL, Mass T, Haramaty L, Zelzion E, Bhattacharya D, Falkowski PG: **Reply to Ramos-Silva et al.: Regarding coral skeletal proteome.** *Proceedings of the National Academy of Sciences* 2013, **110:**E2147-E2148.

41. Ramos-Silva P, Marin F, Kaandorp J, Marie B: **LETTER: Biomineralization toolkit: the importance of sample cleaning prior to the characterization of biomineral proteomes.** Proceedings of the National Academy of Sciences; 2013.

42. Goffredo S, Vergni P, Reggi M, Caroselli E, Sparla F, Levy O, Dubinsky Z, Falini G: **The skeletal organic matrix from Mediterranean coral *Balanophyllia europaea* influences calcium carbonate precipitation.** *PLoS ONE* 2011, **6**.

43. Dauphin Y: **Comparative studies of skeletal soluble matrices from some scleractinian corals and molluscs.** *International Journal of Biological Macromolecules* 2001, **28:**293-304.

44. Weiner S: **Aspartic acid-rich proteins: major components of the soluble organic matrix of mollusk shells.** *Calcif Tissue Int* 1979, **29:**163-167.

45. Pereira-Mouriès L, Almeida MJ, Ribeiro C, Peduzzi J, Barthélemy M, Milet C, Lopez E: **Soluble silk-like organic matrix in the nacreous layer of the bivalve *Pinctada maxima*: A new insight in the biomineralization field.** *Eur J Biochem* 2002, **269:**4994-5003.

46. Stoll HM, Ruiz Encinar J, Ignacio Garcia Alonso J, Rosenthal Y, Probert I, Klaas C: **A first look at paleotemperature prospects from Mg in coccolith**

carbonate: Cleaning techniques and culture measurements. *Geochemistry, Geophysics, Geosystems* 2001, **2**.

47. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M: **The PRIDE database and related tools and resources in 2019: improving support for quantification data.** *Nucleic Acids Res* 2019, **47:**D442-D450.

48. Voolstra CR, Li Y, Liew YJ, Baumgarten S, Zoccola D, Flot J-F, Tambutté S, Allemand D, Aranda M: **Comparative analysis of the genomes of *Stylophora pistillata* and *Acropora digitifera* provides evidence for extensive differences between species of corals.** *Scientific Reports* 2017, **7:**17583.

49. Bertucci A, Tambutté S, Supuran C, Allemand D, Zoccola D: **A new coral carbonic anhydrase in *Stylophora pistillata*.** *Mar Biotechnol* 2011, **13:**992-1002.

50. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene Ontology: Tool for the unification of biology.** *Nat Genet* 2000, **25:**25-29.

51. Consortium GO: **Expansion of the gene ontology knowledgebase and resources.** *Nucleic Acids Res* 2017, **45:**D331-D338.

52. Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, Lee TJ, Leigh ND, Kuo T-H, Davis FG: **A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors.** *Cell Reports* 2017, **18:**762-776.

53. Krogh A, Larsson B, Heijne Gv, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes.** *J Mol Biol* 2001, **305:**567-580.

54. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H: **SignalP 5.0 improves signal peptide predictions using deep neural networks.** *Nat Biotechnol* 2019, **37:**420.

55. Nielsen H, Engelbrecht J, Brunak S, von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Protein Engineering* 1997, **10:**1-6.

56. Poisson G, Chauve C, Chen X, Bergeron A: **FragAnchor: a large-scale predictor of glycosylphosphatidylinositol anchors in eukaryote protein sequences by qualitative scoring.** *Genomics, Proteomics & Bioinformatics* 2007, **5:**121-130.

57. Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M, Fujie M, Fujiwara M, Koyanagi R, Ikuta T, et al: **Using the *Acropora digitifera* genome to understand coral responses to environmental change.** *Nature* 2011, **476:**320-323.

58. Emms DM, Kelly S: **OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy.** *Genome Biology* 2015, **16:**157.

59. Emms DM, Kelly S: **OrthoFinder: Phylogenetic orthology inference for comparative genomics.** *Genome Biology* 2019, **20:**1-14.

60. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite.** *Nucleic Acids Res* 2008, **36:**3420-3435.

61. Mass T, Drake JL, Heddleston JM, Falkowski PG: **Nanoscale visualization of biomineral formation in coral proto-polyps.** *Curr Biol* 2017, **27:**3191-3196. e3193.
62. Thongboonkerd V, Mcleish KR, Arthur JM, Klein JB: **Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation.** *Kidney International* 2002, **62:**1461-1469.
63. Bern MW, Kil YJ: **Two-dimensional target decoy strategy for shotgun proteomics.** *Journal of Proteome Research* 2011, **10:**5296-5301.
64. Fondrie WE, Noble WS: **A machine learning strategy that leverages large datasets to boost statistical power in small-scale experiments.** *BioRxiv* 2019**:**849331.
65. Sykes GA, Collins MJ, Walton DI: **The significance of a geochemically isolated intracrystalline organic fraction within biominerals.** *Org Geochem* 1995, **23:**1059-1065.
66. Mann K, Wilt FH, Poustka AJ: **Proteomic analysis of sea urchin (*Strongylocentrotus purpuratus*) spicule matrix.** *Proteome Science* 2010, **8:**33.
67. Von Heijne G, Abrahmsèn L: **Species-specific variation in signal peptide design Implications for protein secretion in foreign hosts.** *FEBS Lett* 1989, **244:**439-446.
68. McGrath JP, Varshavsky A: **The yeast STE6 gene encodes a homologue of the mammalian multidrug resistance P-glycoprotein.** *Nature* 1989, **340:**400-404.
69. Nickel W, Rabouille C: **Mechanisms of regulated unconventional protein secretion.** *Nature Reviews Molecular Cell Biology* 2009, **10:**148-155.
70. Rabouille C: **Pathways of unconventional protein secretion.** *Trends Cell Biol* 2017, **27:**230-240.
71. Rabouille C, Malhotra V, Nickel W: **Diversity in unconventional protein secretion.** *J Cell Sci* 2012, **125:**5251-5255.
72. Clode PL, Marshall AT: **Kalisomes in corals: a novel KCl concentrating organelle?** *Tissue and Cell* 2002, **34:**199-209.
73. Marshall AT, Clode PL, Russell R, Prince K, Stern R: **Electron and ion microprobe analysis of calcium distribution and transport in coral tissues.** *J Exp Biol* 2007, **210:**2453-2463.
74. Mass T, Giuffre AJ, Sun C-Y, Stifler CA, Frazier MJ, Neder M, Tamura N, Stan CV, Marcus MA, Gilbert PUPA: **Amorphous calcium carbonate particles form coral skeletons.** *Proceedings of the National Academy of Sciences* 2017, **114:**E7670-E7678.
75. Neder M, Laissue PP, Akiva A, Akkaynak D, Albéric M, Spaeker O, Politi Y, Pinkas I, Mass T: **Mineral formation in the primary polyps of pocilloporoid corals.** *Acta Biomater* 2019.
76. Mass T, Drake JL, Peters EC, Jiang W, Falkowski PG: **Immunolocalization of skeletal matrix proteins in tissue and mineral of the coral *Stylophora pistillata*.** *Proceedings of the National Academy of Sciences* 2014, **111:**12728-12733.
77. Takeuchi T, Plasseraud L, Ziegler-Devin I, Brosse N, Shinzato C, Satoh N, Marin F: **Biochemical characterization of the skeletal matrix of the massive coral, *Porites australiensis*–The saccharide moieties and their localization.** *Journal of Structural Biology* 2018, **203:**219-229.

78. Puverel S, Houlbreque F, Tambutté E, Zoccola D, Payan P, Caminiti N, Tambutté S, Allemand D: **Evidence of low molecular weight components in the organic matrix of the reef building coral, *Stylophora pistillata*.** *Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology* 2007, **147:**850-856.
79. Farre B, Cuif J-P, Dauphin Y: **Occurrence and diversity of lipids in modern coral skeletons.** *Zoology* 2010, **113:**250-257.
80. Naggi A, Torri G, Iacomini M, Colombo Castelli G, Reggi M, Fermani S, Dubinsky Z, Goffredo S, Falini G: **Structure and function of stony coral intraskeletal polysaccharides.** *ACS Omega* 2018, **3:**2895-2901.
81. Carpentier SC, Panis B, Vertommen A, Swennen R, Sergeant K, Renaut J, Laukens K, Witters E, Samyn B, Devreese B: **Proteome analysis of non-model plants: a challenging but powerful approach.** *Mass Spectrometry Reviews* 2008, **27:**354-377.
82. Dowd WW: **Challenges for biological interpretation of environmental proteomics data in non-model organisms.** *Integr Comp Biol* 2012, **52:**705-720.
83. Forné I, Abian J, Cerdà J: **Fish proteome analysis: model organisms and non-sequenced species.** *Proteomics* 2010, **10:**858-872.
84. Butler WT, Brunn JC, Qin C: **Dentin extracellular matrix (ECM) proteins: comparison to bone ECM and contribution to dynamics of dentinogenesis.** *Connect Tissue Res* 2003, **44:**171-178.
85. Qin C, Baba O, Butler WT: **Post-translational modifications of SIBLING proteins and their roles in osteogenesis and dentinogenesis.** *Critical Reviews in Oral Biology and Medicine* 2004, **15:**126-136.
86. Sarashina I, Endo K: **Skeletal matrix proteins of invertebrate animals: comparative analysis of their amino acid sequences.** *Paleontological Research* 2006, **10:**311-336.

Table and Figure Legends

Tables

Table 1: Summary of methods for protein concentration and cleaning after extraction in acetic acid. CF – Centrifugation filtration methods; ACT – acetone precipitation methods; ASM – Acid soluble matrix; AIM – Acid insoluble matrix.

Table 2: 60 coral skeletal proteins detected by LC-MS/MS across all treatments and solubility fractions. Proteins are listed in order of accession number. Geno ontology categorization is represented as [a]ECM/transmembrane and protein modification, [b]membrane processing, and [c]vesicle/secretion and metal binding.

Table 3. Orthologous coral skeletal proteins in the present work and previously published. Gene ontology categorization is represented as [a]ECM/transmembrane and

protein modification, [b]membrane processing, and [c]vesicle/secretion and metal binding.

Figures

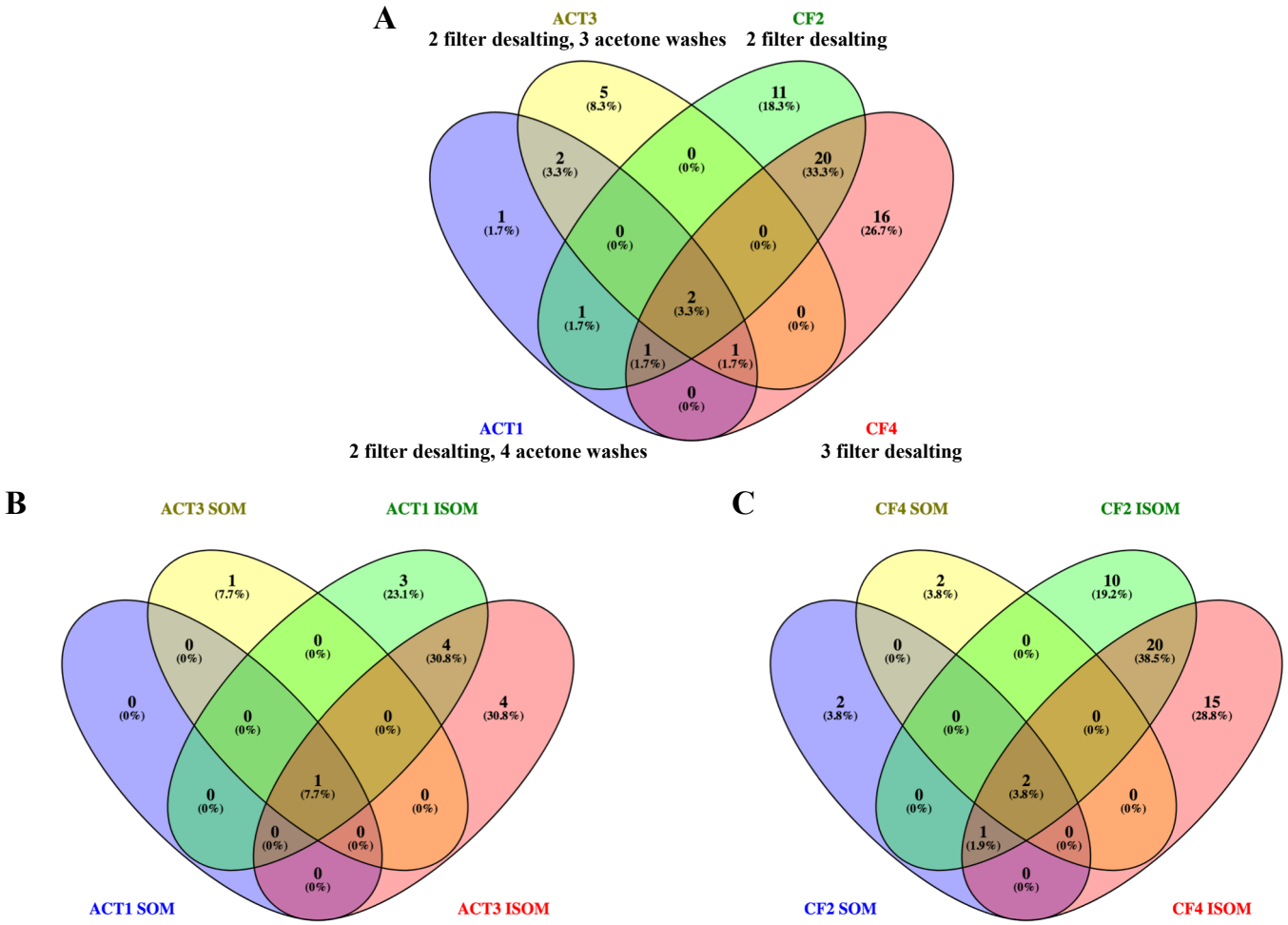
Figure 1: Distribution of proteins numbers according of the extraction methods. Distribution of all proteins (SOM and ISOM combined) by extraction methods (A). SOM and ISOM distribution by acetone precipitation methods (ACT). SOM and ISOM distribution by centrifugation and ultrafiltration method (CF). Note that CF methods yielded more proteins in total. ISOM Insoluble Skeletal Organic Matrix; SOM Skeletal Organic Matrix.

Figure 2. Suggested cellular and extracellular locations of *S. pistillata* skeletal protein based on GO terms. Large yellow and red circles represent calicoblastic cells with red cell membrane facing the skeleton. Small green circles represent vesicles. And purple squiggles represent transmembrane proteins that may or may not span the width of the calicoblastic space plus ECM proteins.

	Filter desalting (CF)				Acetone Precipitation (ACT)			
Sample Name	CF2		CF4		ACT1		ACT3	
Solubility Fraction	ASM	AIM	ASM	AIM	ASM	AIM	ASM	AIM
Filter desalting x 2	+	+	+	+	+	+	+	+
Filter desalting x 3	-	-	+	+	-	-	-	-
Acetone precipitation	-	-	-	-	+	-	+	-
1x 80% acetone wash	-	-	-	-	+	+	+	+
2x 80% acetone wash	-	-	-	-	+	+	+	+
3x 80% acetone wash	-	-	-	-	+	+	-	+
4x 80% acetone wash	-	-	-	-	-	+	-	-

Stylophora pistillata genome hit					Solubility Fraction	
gi Number	Accession Number	Annotation	Best score	Go Categorization	ASM	AIM
1263115389	PFX12726.1	Retrovirus-related Pol polyprotein from transposon 17.6 [Stylophora pistillata]	340.80	-		x
1263115517	PFX12813.1	hypothetical protein AWC38_SpisGene23165 [Stylophora pistillata]	337.40	-		x
1263116757	PFX13778.1	Sacsin [Stylophora pistillata]	267.80	-		x
1263117261	PFX14205.1	Proto-oncogene tyrosine-protein kinase receptor Ret, partial [Stylophora pistillata]	336.70	a,c		x
1263119001	PFX15740.1	Protein FAM208A [Stylophora pistillata]	373.20	-	x	x
1263119725	PFX16398.1	hypothetical protein AWC38_SpisGene19330 [Stylophora pistillata]	273.10	-		x
1263122270	PFX18785.1	Mucin-4 [Stylophora pistillata]	449.50	a,c		x
1263130352	PFX26597.1	Complement C3 [Stylophora pistillata]	283.10	a,b,c		x
1263130510	PFX26751.1	Transmembrane protease serine 9 [Stylophora pistillata]	244.20	a		x
1263131615	PFX27832.1	Poly [ADP-ribose] polymerase 11 [Stylophora pistillata]	320.10	c		x
1263134664	PFX30831.1	hypothetical protein AWC38_SpisGene4366 [Stylophora pistillata]	322.20	-		x
1263134737	PFX30903.1	hypothetical protein AWC38_SpisGene4292 [Stylophora pistillata]	548.90	-		x
1263135656	PFX31810.1	Nidogen-2 [Stylophora pistillata]	286.20	a,c		x
1270028962	XP_022783323.1	phosphopantothenoylcysteine decarboxylase subunit VHS3-like [Stylophora pistillata]	406.60	c		x
1270029544	XP_022786582.1	synapsin-2-like isoform X2 [Stylophora pistillata]	320.50	c	x	x
1270031975	XP_022799089.1	CUB domain-containing protein-like isoform X2 [Stylophora pistillata]	699.70	-		x
1270032917	XP_022804012.1	EGF and laminin G domain-containing protein-like [Stylophora pistillata]	384.10	a		x
1270036141	XP_022779720.1	vitellogenin-like [Stylophora pistillata]	416.00	b,c		x
1270037242	XP_022780303.1	LOW QUALITY PROTEIN: uncharacterized protein LOC111321626 [Stylophora pistillata]	353.60	a,b,c	x	
1270037961	XP_022780690.1	skeletal aspartic acid-rich protein 2-like [Stylophora pistillata]	723.60	a		x
1270037969	XP_022780694.1	CUB and peptidase domain-containing protein 2-like [Stylophora pistillata]	285.80	a	x	x
1270041141	XP_022782398.1	skeletal aspartic acid-rich protein 1-like [Stylophora pistillata]	709.80	-	x	x
1270042352	XP_022783044.1	uncharacterized protein LOC111323869 [Stylophora pistillata]	283.60	-		x
1270043038	XP_022783415.1	coadhesin-like isoform X3 [Stylophora pistillata]	411.20	a		x
1270044042	XP_022783952.1	collagenase 3-like [Stylophora pistillata]	262.90	a,c		x
1270045301	XP_022784623.1	cation channel sperm-associated protein subunit beta-like [Stylophora pistillata]	328.40	a,c		x
1270049598	XP_022786918.1	major yolk protein-like isoform X2 [Stylophora pistillata]	476.50	a,c		x
1270052030	XP_022788227.1	hephaestin-like protein [Stylophora pistillata]	539.80	a,c		x
1270052971	XP_022788730.1	chymotrypsin-like elastase family member 1 [Stylophora pistillata]	264.80	a,c		x
1270054595	XP_022789591.1	endothelin-converting enzyme 1-like isoform X2 [Stylophora pistillata]	94.10	a,b,c		x
1270055217	XP_022789932.1	MAGUK p55 subfamily member 7-like [Stylophora pistillata]	313.70	a		x
1270056169	XP_022790441.1	PHD finger protein 21A-like [Stylophora pistillata]	277.90	c		x
1270059479	XP_022792212.1	ras-like protein 3 [Stylophora pistillata]	289.90	a,c		x
1270063049	XP_022794122.1	galaxin-like isoform X2 [Stylophora pistillata]	309.40	-		x
1270063475	XP_022794351.1	mammalian ependymin-related protein 1-like [Stylophora pistillata]	324.80	a,c		x
1270064196	XP_022794736.1	MAM and LDL-receptor class A domain-containing protein 2-like [Stylophora pistillata]	472.20	-		x
1270068394	XP_022796981.1	uncharacterized skeletal organic matrix protein 8-like [Stylophora pistillata]	664.40	-		x
1270068396	XP_022796982.1	uncharacterized protein LOC111335364 [Stylophora pistillata]	609.30	-		x
1270071953	XP_022798902.1	low-density lipoprotein receptor-related protein 8-like [Stylophora pistillata]	257.10	c		x
1270073139	XP_022799541.1	uncharacterized protein LOC111337489 [Stylophora pistillata]	461.10	b		x
1270080553	XP_022803524.1	digestive cysteine proteinase 1-like [Stylophora pistillata]	262.20	a,c		x
1270081088	XP_022803808.1	deleted in malignant brain tumors 1 protein-like [Stylophora pistillata]	344.70	c		x
1270081207	XP_022803872.1	spore wall protein 2-like isoform X3 [Stylophora pistillata]	338.60	-		x
1270081241	XP_022803894.1	uncharacterized protein LOC111341206 [Stylophora pistillata]	390.70	b		x
1270082891	XP_022804785.1	thioredoxin reductase 1, cytoplasmic-like [Stylophora pistillata]	337.00	c		x
1270084203	XP_022805470.1	uncharacterized protein LOC111342641 [Stylophora pistillata]	256.10	-		x
1270085816	XP_022806326.1	ZP domain-containing protein-like [Stylophora pistillata]	393.90	a		x
1270086467	XP_022806664.1	protein lingerer-like [Stylophora pistillata]	250.10	b,c		x
1270086954	XP_022806928.1	SLIT-ROBO Rho GTPase-activating protein 1-like [Stylophora pistillata]	309.10	-		x
1270087345	XP_022807143.1	condensin-2 complex subunit D3-like [Stylophora pistillata]	278.10	c		x
1270087556	XP_022807256.1	uncharacterized protein LOC111344300 [Stylophora pistillata]	280.40	a,c		x
1270088573	XP_022807807.1	uncharacterized protein LOC111344812 [Stylophora pistillata]	365.30	a,b,c		x
1270089244	XP_022808163.1	uncharacterized protein LOC111345150 [Stylophora pistillata]	310.20	-		x
1270090022	XP_022808576.1	uncharacterized protein LOC111345553 isoform X2 [Stylophora pistillata]	320.10	a,c		x
1270091315	XP_022809269.1	microtubule-associated tumor suppressor 1 homolog isoform X1 [Stylophora pistillata]	259.20	a,c		x
1270091317	XP_022809270.1	microtubule-associated tumor suppressor 1 homolog isoform X2 [Stylophora pistillata]	256.40	-		x
1270093788	XP_022810585.1	von Willebrand factor D and EGF domain-containing protein-like, partial [Stylophora pistillata]	298.80	-		x
1270095516	XP_022778254.1	uncharacterized protein LOC111319781 [Stylophora pistillata]	250.70	c		x
1270095572	XP_022778283.1	uncharacterized protein LOC111319816, partial [Stylophora pistillata]	508.10	-	x	x
190710633	ACE95141.1	carbonic anhydrase [Stylophora pistillata]	495.20	a,c	x	x

NCBI Accession Number	Annotation	gi (NCBI)	Gene model id	RNA id	*Acropora digitifera* skeletal proteins (Takeuchi et al. 2016)	*Acropora millepora* skeletal proteins (Ramos-Silva et al. 2013)	*Stylophora pistillata* skeletal proteins (Drake et al. 2013)
ACE95141.1[a,c]	carbonic anhydrase	190710633	NA	NA			SeriatoporaSp_g7349; mapped to ACE95141.1 (EU532164)
PFX18785.1[a,c]	Mucin-4	1263122270	NA	NA	aug_v2a.09809.t1	JR987773	AGG36340.1
PFX30903.1	hypothetical protein AWC38_SpisGene4292	1263134737	NA	NA	aug_v2a.06327.t1	JR972076.1	AGG36350.1
XP_022779720.1[b,c]	vitellogenin-like	1270036141	gene3329	rna4552	aug_v2a.08856.t1		AGG36347.1, AGG36349.1
XP_022780690.1[a]	skeletal aspartic acid-rich protein 2-like	1270037961	gene4012	rna5498	aug_v2a.11068.t1, JT001945.1, aug_v2a.01441.t1, aug_v2a.01440.t1, JT004498.1, JR991407.1	JR991407.1, JT004498.1	AGG36358.1, AGG36357.1
XP_022780694.1[a]	CUB and peptidase domain-containing protein 2-like	1270037969	gene3981	rna5452		JR970990.1	
XP_022782398.1	skeletal aspartic acid-rich protein 1-like	1270041141	gene5242	rna7162		JT001945.1	AGG36357.1
XP_022783415.1[a]	coadhesin-like isoform X3	1270043038	gene6099	rna8310	aug_v2a.05945.t1	JT016638.1	AGG36341.1, AGG36335.1, AGE45658.1, AGG36343.1
XP_022786918.1[a,c]	major yolk protein-like isoform X2	1270049598	gene8661	rna11721			AGG36338.1
XP_022788227.1[a,c]	hephaestin-like protein	1270052030	gene9579	rna12971	aug_v2a.24015.t1	JT019463.1	
XP_022794122.1	galaxin-like isoform X2	1270063049	gene13931	rna18823	aug_v2a.18631.t1	JR976690.1	
XP_022794736.1	MAM and LDL-receptor class A domain-containing protein 2-like	1270064196	gene14425	rna19450	aug_v2a.09968.t1, aug_v2a.09969.t1	JT011118.1, JR994474.1	AGG36355.1, AGG36356.1, AGG36345.1
XP_022796981.1	uncharacterized skeletal organic matrix protein 8-like	1270068394	gene16153	rna21699		JT014391.1	AGC24391.1
XP_022796982.1	uncharacterized protein LOC111335364	1270068396	gene16156	rna21701			AGC24391.1
XP_022804012.1[a]	EGF and laminin G domain-containing protein-like	1270032917	gene2104	rna2853	aug_v2a.15580.t1, aug_v2a.06123.t1, JR980881.1, aug_v2a.24512.t1, aug_v2a.06122.t1	JR980881.1	AGG36344.1, AGG36352.1
XP_022806326.1[a]	ZP domain-containing protein-like	1270085816	gene23405	rna30870	aug_v2a.07627.t1	JN631095.1	AGG36346.1


A
ACT3
2 filter desalting, 3 acetone washes
CF2
2 filter desalting
5 (8.3%)
11 (18.3%)
2 (3.3%)
0 (0%)
20 (33.3%)
1 (1.7%)
0 (0%)
0 (0%)
16 (26.7%)
2 (3.3%)
1 (1.7%)
0 (0%)
1 (1.7%)
1 (1.7%)
0 (0%)
ACT1
2 filter desalting, 4 acetone washes
CF4
3 filter desalting
B
ACT3 SOM
ACT1 ISOM
1 (7.7%)
3 (23.1%)
0 (0%)
0 (0%)
4 (30.8%)
0 (0%)
0 (0%)
0 (0%)
4 (30.8%)
1 (7.7%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
ACT1 SOM
ACT3 ISOM
C
CF4 SOM
CF2 ISOM
2 (3.8%)
10 (19.2%)
0 (0%)
0 (0%)
20 (38.5%)
2 (3.8%)
0 (0%)
0 (0%)
15 (28.8%)
2 (3.8%)
0 (0%)
0 (0%)
1 (1.9%)
0 (0%)
0 (0%)
CF2 SOM
CF4 ISOM

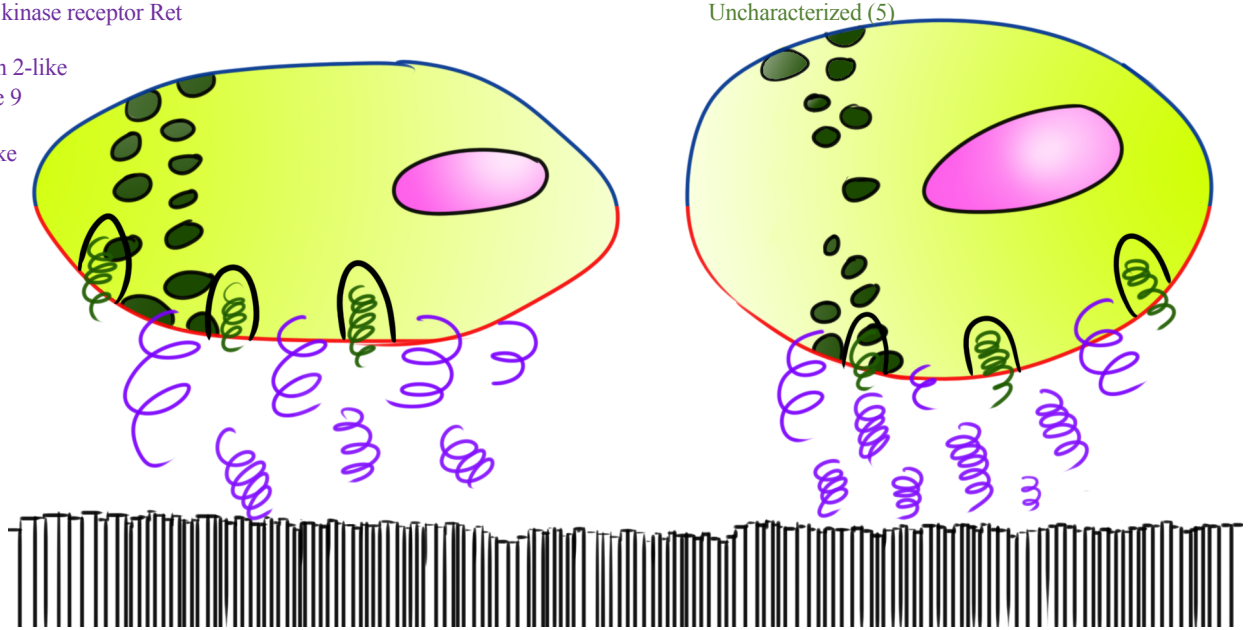
ECM/Transmembrane & Protein Modification

Carbonic anhydrase 2-like
Cation channel sperm-associated protein subunit beta-like
Chymotrypsin-like elastase family member 1
Coadhesin-like
Collagenase 3-like
Complement C3
CUB and peptidase domain-containing protein 2-like
Digestive cysteine proteinase 1-like
Endothelin-converting enzyme 1-like
EGF and laminin G domain-containing protein-like
Hephaestin-like protein
MAGUK p55 subfamily member 7-like
Major yolk protein-like
Mammalian ependymin-related protein
Mucin-4
Microtubule-associated tumor suppressor 1 homolog
Nidogen-2
Proto-oncogene tyrosine-protein kinase receptor Ret
Ras-like protein 3
Skeletal aspartic acid-rich protein 2-like
TTransmembrane protease serine 9
Uncharacterized (3)
ZP domain-containing protein-like

Membrane Processing

Complement C3
Endothelin-converting enzyme 1-like
Protein lingerer-like
Uncharacterized (6)
Vitellogenin-like

Vesicle/Secretion & Metal Binding

Carbonic anhydrase 2-like
Cation channel sperm-associated protein subunit beta-like
Chymotrypsin-like elastase family member 1
Condensin-2 complex subunit D3-like
Deleted in malignant brain tumors 1 protein-like
Digestive cysteine proteinase 1-like
Endothelin-converting enzyme 1-like
Low-density lipoprotein receptor-related protein 8-like
Mammalian ependymin-related protein 1-like
Microtuble-associated tumor suppressor 1 homolog
PHD finger protein 21A-like
Poly (ADP-ribose) polymerase 11
Phosphopantothenoylcystein decarboxylase subunit VHS3-like
Protein lingerer-like
Proto-oncogene tyrosine-protein kinase receptor Ret Ras-like protein 3
Synapsin-2-like
Thioredoxin reductase 1, cytoplasmic-like
Uncharacterized (5)

Collagenase-like
Complement C3
Hephaestin-like protein
Major yolk protein-like
Mucin-4
Nidogen-2
Vitellogenin-like