Published in final edited form as:

Connect Tissue Res. 2018 December ; 59(SUP1): 20–29. doi:10.1080/03008207.2017.1408605.

The unique biomineralization transcriptome and proteome of *Lytechinus variegatus* teeth

Keith Alvares^a, Caroline J DeHart^b, Paul M Thomas^b, Neil L Kelleher^b, and Arthur Veis^{a,c}

^aFeinberg School of Medicine, Department of Cell and Molecular Biology, Northwestern University, Chicago, IL, USA

^bProteomics Center of Excellence, Northwestern University, Evanston, IL, USA

^cFeinberg School of Medicine, Department of Cell, Molecular and Structural Biology, Northwestern University, Chicago, IL, USA

Abstract

Background—Matrix-regulated biomineralization involves the specific nucleation and growth of mineral phases within or upon preformed structured organic matrices. We hypothesized that there might be a general mechanism whereby anionic, phosphorylated mineral ion-binding proteins assist in specifically locating the mineral ions with respect to the mineralizing structural organic matrix. Here we extended these studies to invertebrate mineralization in *Lytechinus variegatus* (Lv) teeth.

Materials and Methods—The tooth proteins were extracted and the phosphoproteins occluded in the mineral were enriched by passage through a ProQ Diamond phosphoprotein enrichment column, and subjected to MS/MS analysis. A Lv RNA-seq derived transcriptome database was generated. The MS/MS data found 25 proteins previously classified as “Predicted uncharacterized proteins” and many of the spicule matrix proteins. As these 25 proteins were also identified with the transcriptome analysis, and were thus no longer “hypothetical” but real proteins in the Lv tooth. Each protein was analyzed for the presence of a signal peptide, an acidic pI 4, and the ability to be phosphorylated. Results: Four new Lv tooth specific Pro-Ala-rich proteins were found, representing a new class of proteins.

Conclusion—The tooth is different from the spicules and other urchin skeletal elements in that only the tooth contains both “high” and “very high” magnesium calcite, $[\text{Ca}(1-X)\text{Mg}(X)\text{CO}_3]$, where X is the mole fraction of Mg. We speculate that our newly discovered proline-alanine rich proteins, also containing sequences of acidic amino acids, may be involved in the formation of high magnesium and very high magnesium calcite.

Keywords

Sea Urchin teeth; biomineralization-related proteins; RNAseq; MS/MS; Mg Calcites

Introduction

The Echinoderms occupy an interesting place in evolutionary development, as they are invertebrates belonging to the superphylum Deuterostomia, which includes the phyla: Chordata, Hemichordata, Echinodermata, and Xenacoelomorpha. The chordata represent all vertebrate animals which have bony structures based on calcium phosphate (phosphatitic) mineral deposits, whereas the hemichordates and Echinoderms do not have “bony” structures, although they do have mineralized skeletal elements based on calcium carbonates (calcitic). As shown by Lowenstam (1) all living systems, bacteria, plants, and animals seem to have the capacity to generate mineralized elements of various types. Indeed biogenic silicates are present in greater mass than either calcitic or phosphatitic biominerals. Nevertheless, our laboratory began its study of biomineralization from the very narrow perspective of the mineralization mechanisms in the collagenous matrices of bone and tooth dentin. We focused, in those pre-molecular biology days, on two questions: (1) Were the collagens of bone and dentin and the soft tissues skin and tendon identical, or were there intrinsic differences in collagen structure and composition? and (2) Were there differences in the non-collagenous components of those tissues that might be related to the deposition of apatite within them? We elected to study dentin rather than bone because, once formed, dentin did not turn over, whereas in bone there was always the additional process of remodeling to consider. Further, in practical terms, we could obtain abundant supplies of fresh unerupted bovine teeth.

The early work (2–8) showed unequivocally that the dentin collagen contained covalently bound phosphate present both as directly attached to the collagen (2,6), and that a highly phosphorylated macromolecular protein, unique to dentin, could be extracted from the dentin (5,7,8). Detailed studies of the dentin collagen, via analysis of the cyanogen bromide cleaved peptides, showed that the dentin collagen was type I, identical to that of the soft tissue collagen (6,8). This shifted our emphasis from the collagen to the role of the phosphorylated proteins in mediating the collagen matrix mineralization. We showed that we could use the same methods of extraction of the dentin specific protein, called dentin phosphophoryn (DPP), to extract a bone specific phosphoprotein, bone sialophosphoprotein (BSP, distinct in protein sequence from DPP) from bovine bone (9–12). Further studies showed that the DPP would localize at the mineralization front (13–15) in rat incisor dentin. Isolated DPP was shown to bind strongly to Ca^{2+} ions by many methods (16–18). Thus it was possible to hypothesize (19–21) that there might be a general mechanism for biomineralization involving specific deposition of mineral crystals within preformed structural organic matrices guided by non-structural anionic mineral ion-binding proteins (NCP in bone and dentin) that specifically located the mineral ions with respect to the structural organic matrix. It seemed logical to explore this very general hypothesis by examining the proteins involved in the formation of the mineralized elements of the most closely related invertebrate mineralized tissues, the echinoderms (22), and we began with the heavily mineralized calcitic echinoderm teeth, which had been studied extensively in terms of structure and mineralization, but not biochemically. We had prepared a polyclonal, but highly specific, antibody to rat DPP (23) and used that antibody to identify any cross-reactive protein in the extracted proteins obtained during the demineralization of *Lytechinus*

variegatus (*Lv*) teeth, and to our surprise (24) found a vigorous, but specifically localized, immunoreactivity. Later, the principal urchin target of anti-DPP antibody was isolated, sequenced and identified as P16, a transmembrane protein binding at the surfaces of the urchin odontoblast cell syncytia and displaying a phosphorylated dentin DPP-like sequence into the extracellular space (25,26). These works demonstrated explicitly that one or more sea urchin proteins, possibly involved in the formation of carbonate minerals such as calcite, contained epitopes similar in sequence to DPP, the vertebrate protein demonstrated to be involved in the mineralization of tooth dentin.

Much of the work done on the characterization of sea urchin biomineralization proteins has been done on the spicules of the larval urchin (27). The mineralized spicules give structure to the larvae but are transient, and not present in the post-larval maturing animal. The other mineralized structures of the growing urchin are the test, spines, lantern stereom, and the teeth (28). The teeth are especially interesting to us because they grow continuously in a vectorial fashion, and thus show all phases of tooth development. The preodontoblasts originating at the aboral plumula arise from a mixed population of coelomocytes. The individual monocytes condense at the outer surface of the plumula (29,30) just under the epithelial layer, where they fuse and become multinucleated cells which form sheet-like syncytial layers (31). Calcium carbonate mineral forms at calcification sites in between the cellular layers, with the mineral deposition related to the syncytial plasma membranes (21,32,33). Another important difference between the tooth and other mineralizing elements is that only the tooth contains “very high” magnesium calcite. The *Lv* tooth Mg calcite has the composition $(\text{Ca}_{1-x}\text{Mg}_x\text{CO}_3)$, where x is the mole fraction of Mg content (34). The first stage “high Mg calcite” elements (plates, prisms) have $x \sim 0.13$ while the second stage “very high Mg calcite”, which forms later at the tooth midshaft and there begins to cement the first stage material together into a mechanically functional tooth (35), has $x \sim 0.33$ (36,37). Calcite appears very early during post-larval tooth development (36), and adjacent plates grow in discrete compartments, separated by cellular syncytia, but all have the same crystallographic orientation. Clearly the mineral composition changes with location and matrix maturity within the heavily mineralized portion of the tooth. This prompted us to question whether there were different sets of proteins regulating mineral composition in the tooth, compared to the spicules and spines. The present paper describes the identification of a novel set of proline–alanine rich proteins containing interspersed domains of acidic amino acid-rich sequences. Which of these domains might be involved in regulation of the formation of high magnesium and very high magnesium calcites remains to be determined.

Methods

Isolation of the *L. variegatus* tooth proteins

Lv teeth were obtained and the proteins were isolated as described earlier (26). Briefly, teeth were collected from 100 mature urchins and immediately submerged in liquid nitrogen. Nine gram of teeth samples were thawed and washed in 100 mL of 150 mM NaCl containing a freshly prepared protease inhibitor mixture (PIM) consisting of 5 mM PMSF, 5 mM NEM, 5 mM EDTA, 50 mM Caproic acid, and 2 mM Benzamidine, with slow stirring for 2 hr at 4°C. The washed teeth were suspended in 100 mL of 6 M guanidine-HCl plus

PIM. The teeth were crushed to a fine powder with pestle and mortar, and extracted with gentle stirring for 20 hr at 4°C. The solvent accessible extracted material (Gdn-HCl extract) was separated from the insoluble powder by centrifugation. The residual mineralized powder pellet was taken up in 150 mL of 0.6N HCl plus PIM and the contents were stirred for 4 hr. Treatment with 0.6 N HCl completely demineralized the powder leaving a slightly turbid suspension. The mixture was centrifuged, and the supernatant (HCl-extract) was collected, neutralized and the phosphoproteins present were enriched by passage over a Pro-Q Diamond phosphoprotein enrichment column. The mixture of phospho-proteins finally eluted from the Pro-Q Diamond column served as the starting points for subsequent analysis, first by gel electrophoresis to see the distribution of molecular mass and calcium ion binding properties of the individual components, and then finally analyzed by MS/MS.

Gel electrophoresis and specific staining procedures for phosphoproteins and calcium ion binding

Each gel was divided into two parts. One part was stained for phosphoproteins using ProQ Diamond Phosphoprotein gel stain (Thermo Fisher Scientific) according to the manufacturer's instructions. The other half was transferred to nitrocellulose and assayed for calcium binding done by the Ca^{2+} overlay method [39]. After detection of the phosphoproteins, the gel was washed with deionized water followed by 20% isopropanol, and then stained with Stains- All to detect acidic proteins.

Sample preparation for mass spectrometry

Proteins were precipitated in cold acetone to remove salts and resuspended in the suggested loading buffer for the type of experiment being performed. For in-gel digestions, proteins were resuspended in Laemmli buffer, separated using traditional SDS-PAGE, and visualized by silver nitrate stain prior to gel band excision and subsequent in-gel digestion (38). For in-solution digestion, proteins were resuspended in 8M urea, reduced in 100 mM DTT, alkylated in 300 mM iodoa-cetamide, and diluted with 100 mM NH_4HCO_3 , pH 7.5, prior to digestion. Because protein amount is difficult to quantify for acidic proteins, a fixed volume of eluted phosphoproteins was used. Proteins were digested with endopeptidases trypsin, chymotrypsin, and Glu-C in separate experiments, each following the manufacturer's protocol. The resulting peptides were concentrated and desalted by C18-Zip-Tip prior to mass spectrometry.

Mass spectrometry

Peptides were resuspended in solvent A (95% water, 5% acetonitrile, and 0.2% formic acid) and 5 μL were injected onto a trap column (150 μm ID $\times$ 3 cm) coupled with a nanobore analytical column (75 μm ID $\times$ 15 cm). The trap and analytical columns were packed with C18 reversed phase (Magic C18, Bruker) media (3 μm , 100 Å pore size). Samples were separated by UPLC using a linear gradient of solvent A and solvent B (5% water, 95% acetonitrile, 0.2% formic acid) delivered over 60–90 minutes by a Dionex UltiMate 3000 nano-LC (Thermo). MS data were obtained on either a 7T LTQ-FT or a 12T Velos-FT (Thermo) mass spectrometer fitted with a custom nanospray ionization source. Intact mass spectrometry data were obtained at an FT resolving power of 60,000 (at m/z 400). The top 5 or 10 m/z species were isolated within the ion trap and fragmented using collisionally

induced dissociation. Data were analyzed with Mascot 2.4 (Matrix Science) and Peaks 7.0 (Bioinformatics Solutions, Inc.) using the custom *Lv* RNASeq database prepared as described below. False discovery rates were adjusted to 1% at the peptide level using Scaffold 4.3 (Proteome Software). The amino acid sequences of proteins from which peptides were confidently identified were searched using BLAST (39) against the nonredundant protein database “nr” to find homologous proteins in other organisms.

Rnaseq analysis: library construction and sequencing

Library construction and sequencing were performed at the Genomics Core facility at the University of Chicago. RNA quality and quantity were first determined with the Agilent Bioanalyzer 2100, accepting RNA integrity numbers (RIN) of > 7 and quantities of 100 nanograms or more per sample. Samples were enriched for mRNA using an oligo-dT column. A directional 75 × 75 paired-end mRNA library was prepared by using Illumina TruSeq mRNA Sample Preparation Kits per manufacturer’s instructions. Briefly, polyadenylated mRNAs were captured from total RNA using oligo-dT selection. Next, the sample was converted to cDNA by reverse transcription, and then ligated to Illumina sequencing adapters containing unique barcode sequences. The sample was amplified by PCR and the resulting cDNA library was quantified using qPCR. Finally, the cDNA library was sequenced on the Illumina MiSeq *de novo* Transcriptome Assembly. The quality of DNA reads, in fastq format, was evaluated using FastQC. The dataset consisted of 16.3 million reads. Adapters were removed and reads of poor quality were filtered out. *de novo* transcript assembly was performed using Trinity (version trinityrnaseq_r20140413p1) with default parameters. There were 46,519 individual transcripts in the assembly with an average length of 572 nt and a total of 26,641,918 nt. (Trinity File is: Transcript_Assembly_Trinity.fasta). For mass spectrometry searches, the program “getorf” (40) was used in all six frames of translation to analyze the 46,519 assembled transcripts. Filtering out potential reads fewer than 10 amino acids, ~1.72 million potential amino acid sequences were predicted. To this large dataset, entries within the common repository of adventitious proteins (<http://www.thegpm.org/crap/>) were appended to filter out proteases, trypsin and other common laboratory protein contaminants.

Results and discussion

Co-localization of acidic phosphoproteins and calcium binding: identification of mineral occluded protein

Figure 1 shows the results of a typical gel electrophoresis experiment. The majority of the Stains-All staining proteins also stain with the phosphoprotein stain. Of special interest is the broad but intense band around ~ 45 kDa which shows a high intensity of phosphostaining [Figure1 B2, C] and also a high level of calcium binding. In separate experiments the phosphoprotein enriched fraction run on a PAGE gel was cut into sections along its length, digested *in situ* with a combination of enzymes (trypsin, chymotrypsin, and Glu-C) and analyzed by MS/MS.

Mass spectrometry and transcriptome analysis

Many of the *Lv* peptides recovered after in-gel digestion did not match identically with any proteins in the *Strongylocentrotus purpuratus* (*Sp*) database. This led us to conclude that perhaps the amino acid sequences of *Sp* and *Lv* differ significantly. To overcome this shortcoming in the use of the *Sp* database, the *Lv* tooth RNA was isolated and sent to the University of Chicago for RNA-seq analysis. We now have the entire transcriptome analysis of the *Lv* tooth: RNA-seq identified 3744 genes expressed in the tooth. When the initial MS/MS data of the mineral occluded phosphoprotein enriched fraction was analyzed against the newly created RNA-seq derived data base, the number of peptide spectral matches rose from 1815 (*Sp* database) to 8208 (*Lv* RNASeq-derived database). Table 1 shows that many of the “biomineralization-associated proteins” identified in the *Sp* spicules are also present in the tooth of *Lv* although, as expected, the percentage of homology varied greatly between the two species, except in the case of extracellular matrix protein 3. In addition to the known spicule biomineralization proteins we identified 25 proteins that were designated as “Predicted uncharacterized proteins”, corresponding to the genome analysis of *Sp*. However, since these proteins have now been identified by both MS/MS analysis and with the transcriptome analysis, as a result they are no longer “hypothetical” proteins but rather proteins that actually exist within the *Lv* tooth. Appendix Table 1A shows the list of these previously uncharacterized hypothetical proteins identified by MS/MS in the mineral occluded *Lv* tooth extract. For completeness Appendix Table 1B lists the other predicted proteins found in the *Lv* HCl extract which could be identified with cellular and matrix proteins in the *Sp* data base.

Identification of novel proline–alanine rich proteins

Since our hypothesis is that biomineralization proteins are acidic phosphoproteins present in the extracellular matrix, the entries listed in Appendix Table 1A were analyzed *in silico*. Using the criteria of the presence of a signal peptide, an acidic isoelectric point between 3.5 and 4.5, and the ability to be phosphorylated, four such *Lv* tooth specific proteins, marked in the table with an asterisk, were found. The complete amino acid sequences of these four proteins, with Identifying Numbers 1540, 8526, 8382, and 8444, in order of increasing isoelectric point (pI), are shown in Figure 2. There are three outstanding features of these proteins. First, the overall amino acid compositions, provided in Table 2, show Ala, Pro, and Gly to be the major constituents, and all have extensive predicted phosphorylations on Ser, Tyr, and Thr residues (quantified in Table 2), which along with the Asp and Glu and lesser content of Arg and Lys make these molecules quite acidic. Predicted phosphorylation sites were determined by the program <http://www.cbs.dtu.dk/services/NetPhos/>. Second, the Ala-Pro are gathered in extended Ala-Pro sequences, while the acidic residues are likewise in separate domains, most clearly evident in the 1540 and 8526 sequences, present in alternating patches. In each case the Ala-Pro domains seem to terminate with GG or GGG sequences, yet the Ala-Pro proteins have some specific compositional sequence preferences. For example, 1540 has prominent APAPPA and GGGANGDTSGET domains; 8526 has PAAPHAPA, APHPPA, GGVHGDTS, and GETGYTPDLGY domains; 8380 has a PPARAPA domain; and 8444 has PRAPGRPRAP, SSDQQSDAQ, and GSSGSDESSGSGEDY domains. Finally, all four proteins have extended acidic N-terminal

and C-terminal regions containing extended D-rich sequences suggesting that the terminal regions may have important structural functions.

Discussion

What is next?

The living keeled urchin tooth is an excellent model with which to examine the mechanism of the role of matrix proteins in the matrix mediated initiation of mineralization, since we know that the tooth organic matrix structure is formed before the onset of mineral crystallization to produce the formation of the first stage low Mg calcite plates [34]. The mechanical toughening is achieved via the interplate fusion of the plates, with very high Mg calcite as the tooth elongates in the adoral direction [3]. It should be noted that the tooth is different from the embryonic calcitic spicules in that only the mature tooth contains the high and very high magnesium calcite, $\text{Ca}_{(1-x)}\text{Mg}_x\text{CO}_3$. The composition of the first stage high Mg mineral elements (plates, prisms) has $x = 0.13$ and the second stage mineral (very high Mg), which first forms at the tooth midshaft and there begins to cement the first stage calcite plates together into a functional tooth, has $x = 0.33$.

The present finding and complete sequencing of a new and unexpected set of proteins among the urchin tooth mineral-related matrix, fitting the presumed requirements, opens the way for localizing the proteins, and for their structural analysis, and perhaps gaining fresh insight into the generality of the mechanisms of biomineralization.

In addition to the alteration in Pro-Ala rich sequences and acidic sequences, it is evident that the four proteins have other sequence differences in both Pro-Ala and the acidic domains. As pointed out above, 1540 has prominent APAPPA and GGGANGDTSGET repeats, while 8526 has PAAPHAPA and APHPA, GGVHGDTS, and GETGYTPDLGY; 8380 has PPARAPA and 8444 has PRAPGRPRAP and SSDQQSDAQ, GSSGSDESSGSGEDY repeats, emphasizing that both types of extended domains may have important roles in determining structure. The Ala-Pro domains seem to terminate with GG or GGG.

Proline-rich regions are also found in amelogenins that are responsible for the mineralization of enamel. This strengthens our hypothesis that there are common motifs shared by all biomineralization proteins, whether they are from vertebrates or invertebrates. We believe that we have identified a new set of proteins, not studied so far, that may be involved in urchin tooth mineralization.

Extended structures and domains

In order to go further into the structures of these Pro-Ala rich proteins it will be necessary to undertake diffraction or other detailed structural analysis studies. So it is premature to make any serious statements but preliminary computational studies suggest that the PAPAP, PAAPAP, and PAPPAP can fold into polyproline II like structures, while the Asp-rich sequences can form into alpha-helix structures. Certainly, one cannot assume that the high Pro, Gly content drives these proteins to intrinsically disordered IDP status.

One could speculate that these novel urchin proline-alanine rich proteins containing stretches of acidic amino acids may be involved in the inter-syncytial formation of the very high Mg calcites, important in mechanical strengthening of the echinoderm teeth.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Special thanks go to Dr. Jason Wan of NIDCR for his encouragement and support during the past few years. Proteomics work was performed in the Northwestern Proteomics core facility. Dr. Matthew Schipma from the Northwestern University NUSeq facility is thanked for his assistance with RNASeq analysis. The technical assistance of Elizabeth Lux is greatly appreciated.

Funding

All of the new work reported here was supported by grant R01-DE 01374 from the National Institute of Dental and Craniofacial Research to AV.

References

1. Lowenstam HA. Minerals formed by organisms. *Science* (80-). 1981; 211:1126–1131.
2. Veis A, Schlueter R. Presence of phosphate-mediated cross-linkages in hard tissue collagens. *Nature*. 1963; 197:1204.
3. Veis A, Schlueter RJ. The macromolecular organization of dentine matrix Collagen. I. Characterization of dentine collagen. *Biochemistry*. 1964; 3:1650–1656. [PubMed: 14235325]
4. Schlueter RJ, Veis A. The macromolecular organization of dentine matrix Collagen. II. Periodate degradation and carbohydrate cross-linking. *Biochemistry*. 1964; 3:1657–1665. [PubMed: 14235326]
5. Veis A, Perry A. The phosphoprotein of the dentin matrix. *Biochemistry*. 1967; 6:2409–2416. [PubMed: 6049465]
6. Veis A, , Sabsay B. Bone and dentin, insights into mineralization strategies. In: Westbroek P, , deJong EW, editors *Biom mineralization and biological metal accumulation* Dordrecht, Holland: D. Reidel Pub. Co; 1983 273284
7. Mann K, Poustka AJ, Mann M. In-depth, high-accuracy proteomics of sea urchin tooth organic matrix. *Proteome Sci*. 2008 Dec 9.6:33.doi: 10.1186/1477-5956-6-33 [PubMed: 19068105]
8. Volpin D, Veis A. The isolation and characterization of the CNBr peptides of the $\alpha 1$ and $\alpha 2$ chains of bovine collagen. *Biochemistry*. 1971; 10:1751–1755. [PubMed: 5563760]
9. Carmichael DJ, Veis A, Wang ET. Dentin matrix collagen: evidence for a covalently linked phosphoprotein attachment. *Calcif Tiss Res*. 1971; 7:331–344.
10. Volpin D, Veis A. Differences between CNBr peptides of soluble and insoluble bovine collagens. *Biochem Biophys Res Commun*. 1971; 44:804–812. [PubMed: 5125227]
11. Veis A, Spector AR, Zamosciany H. The isolation of an EDTA-soluble phosphoprotein from mineralizing bovine dentin. *Biochim Biophys Acta*. 1972; 257:404–413. [PubMed: 4623341]
12. Shuttleworth A, Veis A. The isolation of anionic phosphoproteins from bovine cortical bone via the periodate solubilization of bone collagen. *Biochim Biophys Acta*. 1972; 257:414–420. [PubMed: 4336720]
13. Dimuzio MT, Veis A. The biosynthesis of phosphophoryns and dentin collagen in the continuously erupting rat incisor. *J Biol Chem*. 1978; 253:6845–6852. [PubMed: 690128]
14. Dimuzio MT, Veis A. Phosphophoryns - major non-collagenous proteins of the rat incisor dentin. *Calcif Tiss Res*. 1978; 25:169–178.

15. Maier GD, , Lechner JH, , Veis A. Evidence for the conjugation of collagen to the phosphophoryn of rat incisor dentin in relation to the mineralization front. In: Veis A, editor *Chemistry and biology of mineralized connective tissues* North-Holland: Elsevier; 1981 477-487
16. Veis A, , Maier GD, , Lechner JH. The kinetics of formation of collagen-phosphophoryn conjugates in relation to the advancement of the mineralization front in rat incisor dentin. In: Akeson WH, Bornstein P, , Glimcher MJ, editors *Heritable disorders of connective tissue metabolism* St. Louis, MO: C.V. Mosby; 1982 165-172
17. Maier GD, Lechner JH, Veis A. The dynamics of formation of a collagen-phosphophoryn conjugate in relation to the passage of the mineralization front in rat incisor dentin. *J Biol Chem.* 1983; 258:1450–1455. [PubMed: 6822521]
18. Lee SL, Veis A, Glonek T. Dentin phosphoprotein: an extracellular calcium binding protein. *Biochemistry.* 1977; 16:2971–2979. [PubMed: 406911]
19. Lee SL, Veis A. Studies on the structure and chemistry of dentin collagen-phosphophoryn covalent complexes. *Calc Tiss Res.* 1980; 31:123–134.
20. Lee SL, Veis A. Cooperativity in calcium ion binding to repetitive, carboxylate-serylphosphate polypeptides and the relationship of this property to dentin mineralization. *Int J Peptide Protein Res.* 1980; 16:231–240. [PubMed: 6780483]
21. Veis A, , Sabsay B. Bone and dentin, insights into mineralization strategies. In: Westbroek P, , deJong EW, editors *Biom mineralization and biological metal accumulation* Dordrecht, Holland: D. Reidel Pub. Co; 1983 *Lytechinus variegatus* *J Exptl Zool.* 1986; 240:35–46. 273–284. [PubMed: 3095485]
22. Adoutte A, Balavoine G, Lartillot N, Lespinet O, Prud'homme B, De Rosa R. The new animal phylogeny: reliability and implications. *Proc Natl Acad Sci U S A.* 2000; 97:4453–4456. [PubMed: 10781043]
23. Rahima M, Tsay T-G, Andujar M, Veis A. Localization of phosphophoryn in rat incisor dentin using immunocytochemical techniques. *J Histo Cytochem.* 1988; 36:153–157.
24. Veis DJ, Albinger TM, Clohisy J, Rahima M, Sabsay B, Veis A. Matrix proteins of the teeth of the sea urchin *Lytechinus variegatus*. *J Exp Zool.* 1986; 240(1):35–46. [PubMed: 3095485]
25. Veis A, , Dahl T, , Barss J. The progress of mineral deposition Matrix proteins of the teeth of the sea urchin *Lytechinus variegatus*. within the developing tooth of the sea urchin *Lytechinus variegatus* and its relation to specialized matrix proteins. In: Heinzeller T, , Niebelsick JH, editors *Echinoderms: munchen, IEC2003* London: Taylor and Francis Group; 2004 365-370
26. Alvares K, Dixit SN, Lux E, Veis A. Echinoderm phosphorylated matrix proteins UTMP16 and UTMP19 have different functions in sea urchin tooth mineralization. *J Biol Chem.* 2009; 284:26149–26160. [PubMed: 19596854]
27. Wilt FH. Matrix and mineral in the sea urchin larval skeleton. *J Struct Biol.* 1999; 126:216–226. [PubMed: 10475684]
28. Wang RZ, Addadi L, Weiner S. Design strategies of sea urchin teeth: structure, composition and micromechanical relations to function. *Philos Trans R Soc Lond B Biol Sci.* 1997; 352(1352):469–480. [PubMed: 9163824]
29. Kniprath E. Ultrastructure and growth of the sea urchin tooth. *Calcif Tissue Res.* 1974; 14:211–228. [PubMed: 4843789]
30. Märkel K, Roser U, Mackendstedt U, Klosterman M. Ultrastructural investigation of matrix-mediated bio-mineralization in echinoids. *Zoomorphology.* 1986; 106:232–243.
31. Alvares K, Dixit SN, Lux E, Barss J, Veis AJ. The proteome of the developing tooth of the sea urchin, *Lytechinus variegatus*: mortalin is a constituent of the developing cell syncytium. *Exptl Zoolog B Mol Dev Evol.* 2007; 308:357–370.
32. Ameye L, Hermann R, Killian C, Wilt F, Dubois P. Ultrastructural localization of proteins involved in sea urchin biomineralization. *J Histochem Cytochem.* 1999; 47:1189–1200. [PubMed: 10449540]
33. Veis A, Barss J, Dahl T, Rahima M, Stock S. Mineral-related proteins of sea urchin teeth: *lytechinus variegatus*. *Microsc Res Tech.* 2002; 59:342–351. [PubMed: 12430165]

34. Stock SR, Barss J, Dahl T, Veis A, Almer JD. X-ray absorption microtomography (microCT) and small beam diffraction mapping of sea urchin teeth. *J Struct Biol.* 2002; 139(1):1–12. [PubMed: 12372315]
35. Stock SR, Ignatiev K, Lee PL, Almer JD. Calcite orientations and composition ranges within teeth across echinoidea. *Connect Tissue Res.* 2014; 55(Suppl 1):48–52. [PubMed: 25158180]
36. Stock SR, Veis A, Xiao X, Almer JD, Dorvee JR. Sea urchin tooth mineralization: calcite present early in the aboral plumula. *J Struct Biol.* 2012; 180(2):280–289. [PubMed: 22940703]
37. Ma YR, Weiner S, Addadi L. Mineral deposition and crystal growth in the continuously forming teeth of sea urchins. *Adv Funct Mater.* 2007; 17:2693–2700.
38. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc.* 2006; 1(6):2856–2860. [PubMed: 17406544]
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215:403–410. [PubMed: 2231712]
40. Rice P, Longden I, Bleasby A. EMBOSS: The European molecular biology open software suite. *Trends Genet.* 2000 Jun; 16(6):276–277. [PubMed: 10827456]

Figure 1.

Gel electrophoresis analysis of phosphoprotein enriched HCl extract.

Gel A -Stained with Stains All:

Lane 1. Molecular Mass Standard Markers.

Lane 2. The phosphoprotein enriched HCl extract. Gel B—Stained with ProQ Diamond Phosphoprotein gel stain:

Lane 1. Phosphorylated Protein Standard Markers.

Lane 2. The phosphoprotein enriched HCl extract.

Gel C—Nitrocellulose overlay of the phosphoprotein enriched HCl extract gel, demonstrating ^{45}Ca -binding.

*The red asterisk marks the broad heavily stained band at about Mr 47 kDa in Lanes A2, B2, and C, demonstrating the presence of a heterogeneous set of similar mass acidic, phosphorylated, calcium binding proteins in the mineral phase of the mature Lv urchin tooth.

Table 1

Biom mineralization proteins present in spicules also identified in the mineral occluded fraction of the Lv tooth *

The sequences of Lv P16 and P19 determined in the present work are compared with previously isolated and sequenced Lv P16 and P19, described in detail in [24, Alvares et al. J. Biol. Chem.284:26149–26160 (2009)]. The 100% identities of these older, completely independent analyses determined by different methods, confirm the validity of our present RNASeq and MS/MS approach.

Best blast hit (BBH)	% identity With best blast hit
Biom mineralization protein P16 (<i>Lv</i>)	100
Tooth matrix protein P19 (<i>Lv</i>)	99
27kDa primary mesenchymal-specific spicule protein (<i>Sp</i>)	85
SM29 Spicule matrix protein (<i>Sp</i>)	84
30kDa spicule matrix protein (<i>Sp</i>)	74
SM32 spicule matrix protein (<i>Sp</i>)	78
34kDa spicule matrix protein (<i>Sp</i>)	89
SM37 spicule matrix protein (<i>Sp</i>)	75
Extracellular matrix protein 3 (<i>Sp</i>)	100
Mesenchymal-specific cell surface glycoprotein (<i>Sp</i>)	74
MSP130 primary mesenchymal-specific protein (<i>Sp</i>)	68

Table 2

Amino acid compositions of the 4 Proline-Alanine acidic proteins predicted mineralization-related selected from Appendix Table 1A. The most abundant residues, Pro, Ala and Gly are colored red. The intrinsic acidic residues are in green. The post-translationally predicted phosphorylated groups (as determined by <http://www.cbs.dtu.dk/services/NetPhos/>) are in green in parentheses. The intrinsic positive groups are highlighted in blue, to show their very sparse numbers. However, it should be noted that the bulk of the Arg residues reside among the Pro-Ala (neutral) regions.

Residues per secreted molecule	LOC10089 1540 pl 3.5	LOC10088 8526 pl 3.7	LOC10088 8382 pl 4.3	LOC10088 8444 pl 4.5
A Ala	109	76	99	80
P Pro	89	67	82	103
G Gly	83	65	92	90
D Asp	39	57	25	28
E Glu	17	37	13	6
S Ser (SerP)	32 (21)	20 (12)	49 (25)	34 (12)
T Thr (ThrP)	12 (2)	11 (2)	15 (3)	9 (5)
Y Tyr (TyrP)	27 (12)	21 (7)	15 (5)	19 (6)
K Lys	7	12	1	2
R Arg	12	3	27	35
L Leu	5	4	3	2
H His	5	16	3	10
N Asn	11	18	14	9
M Met	0	0	0	2
Q Gln	7	8	7	14
F Phe	4	6	7	5
I Ile	5	4	4	3
V Val	22	25	21	16
W Trp	0	0	0	1
Total AA	498	458	483	472

Note: Colored version of the table can be seen online.

Appendix Table 1A

Predicted uncharacterized mineral occluded proteins in the sea urchin (Lv) tooth identified by MS/MS. Highlighted (Bold, asterisk) proteins are those that are phosphorylated and contain a signal peptide, acidic domains and have acidic pI and were selected for detailed sequencing analysis on the basis that they matched our criteria for direct involvement in the mineralization process. Note that all have masses in the 47 to 48 kDa range. As seen in Figure1C they all are clumped together in the broad most intense Ca binding region.

Translated Transcript size	Best Blast hit (BBH): Proteins detected in the Lv MS/MS analysis. In comparison to the uncharacterized proteins of previously proposed as mineralization related components of the Sp teeth.	% Identity with best blast hit
9 kDa	Predicted uncharacterized protein LOC593369 (<i>Sp</i>)	62
30 kDa	Predicted uncharacterized protein LOC579383 (<i>Sp</i>)	91
20 kDa	Predicted uncharacterized protein LOC100893973 (<i>Sp</i>)	73
22 kDa	Predicted uncharacterized protein LOC764925(<i>Sp</i>)	66
9 kDa	Predicted uncharacterized protein LOC100888382 (<i>Sp</i>)	66
24 kDa	Predicted uncharacterized protein LOC575627 (<i>Sp</i>)	40
40 kDa	Predicted uncharacterized protein LOC100891313 (<i>Sp</i>)	54
36 kDa	Predicted uncharacterized protein LOC100888659 (<i>Sp</i>)	70
47 kDa	Predicted uncharacterized protein LOC100891540 (<i>Sp</i>)*	83
28 kDa	Predicted uncharacterized protein ZK1073.1-like (<i>Sp</i>)	71
40 kDa	Predicted uncharacterized protein LOC762462 (<i>Sp</i>)	90
15 kDa	Predicted uncharacterized protein LOC587978 (<i>Sp</i>)	73
24 kDa	Predicted uncharacterized protein LOC577267 (<i>Sp</i>)	79
95 kDa	Predicted uncharacterized protein LOC582269 (<i>Sp</i>)	78
86 kDa	Predicted uncharacterized protein LOC585121(<i>Sp</i>)	63
47 kDa	Predicted uncharacterized protein LOC100888382 (<i>Sp</i>)*	75
50 kDa	Predicted uncharacterized protein LOC579383 (<i>Sp</i>)	83
48 kDa	Predicted uncharacterized protein LOC100888526 (<i>Sp</i>)*	89
49 kDa	Predicted uncharacterized protein LOC577685 (<i>Sp</i>)	76
47 kDa	Predicted uncharacterized protein LOC100888444 (<i>Sp</i>)*	61

Appendix Table IB

Other proteins detected in the mineral occluded portion of *Lytechinus variagatus* (Lv) tooth. Predicted proteins identified by MS/MS

Proteins identified in the HCl extract	% BBH Identity of Lv proteins to Sp
PREDICTED: filamin-C isoform 3 [Sp]	98
PREDICTED: prostaglandin reductase 1-like, partial [Sp]	83
PREDICTED: deleted in malignant brain tumors 1 protein-like [Sp]	67
PREDICTED: myosin light chain 1/3, skeletal muscle isoform-like [Sp]	85
PREDICTED: L-rhamnose-binding lectin CSL3-like [Sp]	78
PREDICTED: myosin regulatory light polypeptide 9-like [Sp]	95
PREDICTED: LOW QUALITY PROTEIN: 78 kDa glucose-regulated protein-like [Sp]	89
PREDICTED: gelsolin-like protein 1-like [Sp]	85
PREDICTED: 60S ribosomal protein L22-like [Sp]	99
PREDICTED: protein app1-like [Diaphorina citri]	52
PREDICTED: cofilin-like [Sp]	30
PREDICTED: matrix metalloproteinase-24-like [Sp]	83
histone H1-delta [Sp]	86
PREDICTED: calmodulin-like [Sp]	99
PREDICTED: grancalcin-like isoform 2 [Sp]	82
PREDICTED: epithelial chloride channel protein-like [Sp]	80
PREDICTED: myosin heavy chain, striated muscle-like isoform 2 [Sp]	93
myosin heavy chain, partial [Lv]	100
PREDICTED: ADP,ATP carrier protein-like [Sp]	90
RecName: Full = Profilin	83
PREDICTED: annexin A7-like [Sp]	92
PREDICTED: eukaryotic translation initiation factor 5A-1-like [Sp]	77
alpha-1 collagen precursor [Sp]	95
fibrillar collagen alpha 120 and 140 chains, partial [Hemicentrotus pulcherrimus]	78
PREDICTED: actin-related protein 2/3 complex subunit 5-like protein-like [Sp]	75
PREDICTED: matrix metalloproteinase-24-like [Sp]	83
PREDICTED: high mobility group protein 1 homolog [Sp]	79
translation elongation factor 1B beta subunit [Sp]	85
PREDICTED: annexin A4-like [Sp]	87
PREDICTED: tubulin beta chain-like [Sp]	99
PREDICTED: tubulin beta-4 chain-like isoform 2 [Sp]	96
PREDICTED: muscle-specific protein 20-like isoform 5 [Sp]	89
PREDICTED: tropomyosin-like isoform 2 [Sp]	89
actin related protein 1 [Sp]	98
RecName: Full = Actin, cytoskeletal; AltName: Full = M; Flags: Precursor	99
mitochondrial ATP synthase alpha subunit precursor [Sp]	96

Proteins identified in the HCl extract	% BBH Identity of Lv proteins to Sp
PREDICTED: arginine kinase-like [<i>Sp</i>]	88
PREDICTED: LOW QUALITY PROTEIN: alpha-enolase-like [<i>Sp</i>]	88
PREDICTED: 14-3-3 protein zeta-like [<i>Sp</i>]	73
hypothetical protein CAEBREN_18196 [<i>Caenorhabditis brenneri</i>]	48
PREDICTED: alpha-actinin, sarcomeric-like isoform 1 [<i>Sp</i>]	88
PREDICTED: matrix metalloproteinase-24-like [<i>Sp</i>]	83
PREDICTED: FK506-binding protein 15 [<i>Sp</i>]	86
calreticulin precursor [<i>Sp</i>]	84
PREDICTED: lithostathine-like [<i>Sp</i>]	87
PREDICTED: brevican core protein-like [<i>Sp</i>]	85
ER calcistorin precursor [<i>Sp</i>]	79
PREDICTED: matrix metalloproteinase-9-like [<i>Sp</i>]	79
PREDICTED: carbonic anhydrase 14 isoform 1 [<i>Sp</i>]	84
sodium/potassium ATPase alpha subunit [<i>Strongylocentrotus purpuratus</i>]	94
heat shock protein gp96 precursor [<i>Sp</i>]	89
PREDICTED: protein BMH2 isoform 1 [<i>Sp</i>]	80
PREDICTED: activated RNA polymerase II transcriptional coactivator p15-like [<i>Sp</i>]	77