

# SCIENTIFIC REPORTS



OPEN

## Integrated transcriptomic and proteomic analyses of a molecular mechanism of radular teeth biomineralization in *Cryptochiton stelleri*

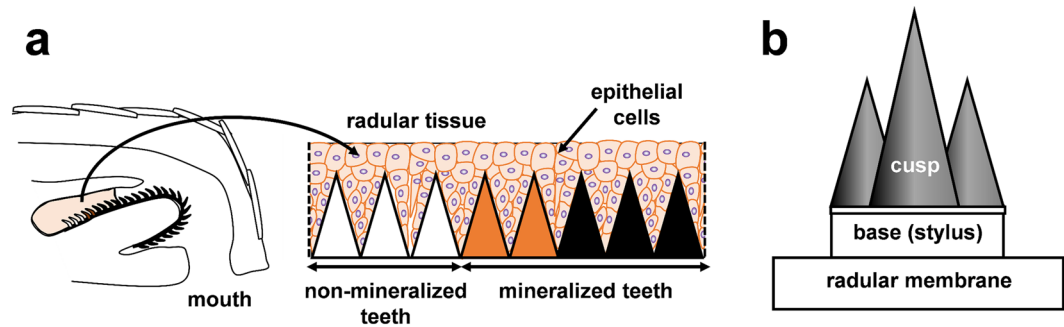
Michiko Nemoto<sup>1</sup>, Dongni Ren<sup>2</sup>, Steven Herrera<sup>2</sup>, Songqin Pan<sup>3</sup>, Takashi Tamura<sup>1</sup>, Kenji Inagaki<sup>1</sup> & David Kisailus<sup>2,4</sup>

Many species of chiton are known to deposit magnetite (Fe<sub>3</sub>O<sub>4</sub>) within the cusps of their heavily mineralized and ultrahard radular teeth. Recently, much attention has been paid to the ultrastructural design and superior mechanical properties of these radular teeth, providing a promising model for the development of novel abrasion resistant materials. Here, we constructed *de novo* assembled transcripts from the radular tissue of *C. stelleri* that were used for transcriptome and proteome analysis. Transcriptomic analysis revealed that the top 20 most highly expressed transcripts in the non-mineralized teeth region include the transcripts encoding ferritin, while those in the mineralized teeth region contain a high proportion of mitochondrial respiratory chain proteins. Proteomic analysis identified 22 proteins that were specifically expressed in the mineralized cusp. These specific proteins include a novel protein that we term radular teeth matrix protein1 (RTMP1), globins, peroxidases, antioxidant enzymes and a ferroxidase protein. This study reports the first *de novo* transcriptome assembly from *C. stelleri*, providing a broad overview of radular teeth mineralization. This new transcriptomic resource and the proteomic profiles of mineralized cusp are valuable for further investigation of the molecular mechanisms of radular teeth mineralization in chitons.

Biologically formed magnetite (Fe<sub>3</sub>O<sub>4</sub>) has been found from several organisms including bacteria, chiton, homing pigeons, honeybees and salmon<sup>1–5</sup>. In the last few decades, significant information has been accumulated regarding bacterial magnetite formation<sup>6–9</sup>. In contrast, little is known about molecular mechanisms of magnetite biomineralization in eukaryotes. The magnetite-rich teeth of chitons was first reported in 1962 and its formation mechanism has been of interest for several decades<sup>1</sup>. Chitons have several dozen rows of teeth that are attached to a ribbon-like structure called a radular membrane. Each tooth is composed of a mineralized cusp and base (also referred to as the stylus) supporting the mineralized cusp (Fig. 1b). Magnetite is deposited only in the cusp region. Recently, ultrastructural and mechanical characterization of fully mineralized chiton teeth from *Cryptochiton stelleri* showed that it has the maximum hardness and stiffness of any known biomineral<sup>10,11</sup>. Although some preliminary analysis provides critical structure-mechanical property relationships in these teeth<sup>12,13</sup>, a further understanding of the formation machinery used to construct this ultrahard biomaterial will allow us to develop a novel abrasion resistant material.

We have studied magnetite biomineralization in the world's largest species of chiton, *C. stelleri*. As other chitons, *C. stelleri* use their teeth to scrape the algae growing on intertidal rocks. When these teeth become worn, they are replaced by a new row of teeth that are continuously formed from inside a radular sac<sup>14</sup>. Therefore, it

<sup>1</sup>Graduate School of Environmental and Life Science, Okayama University, Okayama, 700-8530, Japan. <sup>2</sup>Materials Science and Engineering Program, University of California, Riverside, CA, 92521, USA. <sup>3</sup>W. M. Keck Proteomics Laboratory, Institute for Integrative Genome Biology, University of California, Riverside, CA, 92521, USA. <sup>4</sup>Department of Chemical and Environmental Engineering, University of California, Riverside, CA, 92521, USA. Correspondence and requests for materials should be addressed to M.N. (email: [mnemoto@okayama-u.ac.jp](mailto:mnemoto@okayama-u.ac.jp)) or D.K. (email: [david@engr.ucr.edu](mailto:david@engr.ucr.edu))



**Figure 1.** Schematics of radular tissues used in this study. (a) Radular tissues of the giant pacific chiton, *Cryptochiton stelleri*. (b) Individual radular teeth, which are composed of mineralized cusp, base and radular membrane.

is possible to observe all stages of tooth development within one radula. In the radular teeth of *C. stelleri*, no mineral deposits are found on the first 8–12 rows of teeth. These non-mineralized teeth are transparent and are mainly composed of  $\alpha$ -chitin and proteins. The next 2–5 rows of teeth turn a reddish-brown color, representing a partial mineralization of the teeth with an amorphous or weakly nanocrystalline iron oxide (ferrihydrite). The subsequent rows of teeth have black-colored cusps, indicative of a transformation from ferrihydrite to magnetite towards fully mineralized teeth. Detailed observations of developing teeth revealed that the mineral aggregates of ferrihydrite first deposited on alpha-chitin fibers continuously grow in size as teeth maturation proceeds. Phase transformation to magnetite occurred within a few rows of teeth after the onset of ferrihydrite formation. Although radular membrane, base and cusp are all composed of alpha-chitin, iron-based mineral was deposited only in the cusp region. From these results, we assume that there may be several proteins that dictate specific iron oxide nucleation in the cusp region as well as transformation to magnetite in chiton teeth<sup>15,16</sup>. To reveal the proteins involved in the formation of radular teeth, we previously conducted proteome analysis of mineralized teeth of *C. stelleri* using a *de novo* peptide sequencing technology<sup>15</sup>. However, because of the lack of genomic sequence information, only partial sequences of mineralized teeth proteins were obtained.

In this study, to better understand the molecular mechanisms of magnetite biomineralization, we use RNA sequencing (RNA-seq) to construct the transcriptome data set of radular tissue in *C. stelleri*. RNA-seq based transcriptome analysis reveals highly expressed transcripts in both non-mineralized and mineralized teeth regions. Furthermore, the transcriptome data was used to identify the proteins in mineralized teeth.

## Results

**RNA-seq and *de novo* transcriptome assembly.** Radular teeth are formed within the radular tissue. Here, teeth are surrounded by epithelial cells (Fig. 1a) and it is likely that the proteins required for teeth formation may be supplied from these cells. To construct expressed sequence data sets from the radular tissue, RNA was extracted from this tissue (including radular teeth and epithelial cells). After dissection, radular tissue from the non-mineralized teeth region (first 8 to 12 rows of transparent teeth) and those from the mineralized teeth region (following 2 to 5 rows of partially-mineralized reddish-brown colored teeth plus 5 to 6 rows of mineralized magnetic teeth) (Fig. 1a) were separated and RNA was extracted from each sample. RNA-seq was conducted on the extracted RNA. The reads obtained from the non-mineralized teeth region and the mineralized teeth region were then assembled together into 187,980 transcripts. The average length of transcripts is 705 bp and the maximum length is 16,738 bp (with a N50 length of 1,110 bp, Table 1).

**Highly expressed transcripts in mineralized and non-mineralized teeth regions.** To detect highly expressed transcripts in mineralized and non-mineralized teeth region, each set of reads from those samples was first aligned to the assembled 187,980 transcripts.

Among the twenty most highly expressed transcripts in the non-mineralized teeth region, five transcripts are isoforms of Cs17717|c0\_g1, which did not show similarity to any known proteins (Supplementary Table S1). There are two transcripts encoding peritrophin-like proteins: Cs79475|c0\_g1\_i1 and Cs70642|c0\_g2\_i1. Peritrophin and peritrophin-like proteins have been found in both insects and crustaceans. These proteins have chitin binding domains and have been thought to be involved in immune defense or food digestion<sup>17</sup>. Cs25220|c1\_g1\_i1 shares a similarity to a calcium carbonate biomineralization related protein, Pif, in mollusks<sup>18</sup>. Four proteins, including peritrophin-like and Pif-like proteins, described above have a type-2 chitin-binding domain<sup>19</sup>. The transcripts of a cytochrome c oxidase subunit I (Cs76629|c0\_g1\_i1) and a ferritin subunit 2 (Cs75144|c0\_g1\_i1) that are involved in iron accumulation were included in highly expressed transcripts in the non-mineralized teeth region. Furthermore, four transcripts encoding radular teeth proteins, which are mentioned below, were also included.

Distinct profiles were observed in the twenty most highly expressed transcripts in the mineralized teeth region (Supplementary Table S2). The mitochondrial respiratory chain enzymes constitute 35% of the highly expressed transcripts. Those enzymes include cytochrome c oxidase subunit I, cytochrome c oxidase subunit II, cytochrome c oxidase subunit III, ATP synthase F0 subunit 6, NADH dehydrogenase subunit 5 and NADH dehydrogenase subunit 1. Those are encoded in mitochondrial genome of *C. stelleri*<sup>20</sup>. In addition, the chitin binding beak protein

Sequencing summary	
Number of reads	
non-mineralized teeth region	43,802,270
mineralized teeth region	45,141,430
Assembly summary	
Number of transcripts	
Average length of transcripts (bp)	705
N50 length (bp)	1,110
Maximum length of transcript (bp)	16,738
Mapping summary	
Total mapped reads	
non-mineralized teeth	35,343,878
mineralized teeth	35,511,782
% of Total mapped reads	
non-mineralized teeth (%)	80.69
mineralized teeth (%)	78.67

**Table 1.** Summary of *C. stelleri* RNA-seq results.

3 (Cs12250|c0\_g1\_i1), which has been found from squid beaks was included. Five transcripts (25%) are isoforms of Cs22243|c0\_g7, whose function is unknown.

**Analysis of ferritin genes.** Ferritin is an iron transport and storage protein which assembles to form a protein cage containing up to 4500 iron atoms in the form of a hydrous ferric oxide-based mineral. Since ferritins were detected and isolated from epithelial cells of the radular tissue in *C. stelleri*<sup>21</sup>, ferritin has been thought to be involved in iron transport for tooth mineralization. In this study, the transcripts showing homology with ferritin genes were screened from the 134,993 transcripts. As a result, the deduced amino acid sequences of the four transcripts share homology with ferritin protein from the chiton *Liolophura japonica*. Among four ferritin homologs, two are identical except one amino acid substitution (Cs22563|c0\_g1\_i1 and Cs75144|c0\_g1\_i1) (Supplementary Fig. S1). The sequence of Cs17042|c0\_g1\_i1 lacked the 5' terminus of the full-length cDNA. Until now, two types of ferritins, cytoplasmic ferritins and secreted ferritins, have been isolated from mollusks<sup>22</sup>. The results of a homology search indicated that Cs90734|c0\_g1\_i1 shared a homology with secreted ferritins, while Cs22563|c0\_g1\_i1, Cs75144|c0\_g1\_i1 and Cs17042|c0\_g1\_i1 are more similar to cytoplasmic ferritins. The phylogenetic analysis was conducted using the sequences of known secreted and cytoplasmic ferritins from Mollusca and Arthropoda. As a result, Cs90734|c0\_g1\_i1 ferritin clustered with secreted ferritins and the other three ferritins clustered with cytoplasmic ferritins (Supplementary Fig. S2). Evidently, compared to other ferritins in *C. stelleri*, Cs90734|c0\_g1\_i1 has distinct structural features commonly observed in secreted-type ferritins such as: (1) a ~40 amino acid residue insertion located in the middle of the putative loop region, (2) the putative metal binding residues, which are highly conserved in cytoplasmic ferritins, are replaced with other amino acids. However, it does not contain signal sequences that exist in other secreted ferritins. Among four ferritin genes, Cs75144|c0\_g1\_i1 showed a predominant expression, suggesting its importance in radular teeth formation (Supplementary Table S3). Notably, Cs75144|c0\_g1\_i1 was more highly expressed in the non-mineralized teeth region than in the mineralized teeth region.

**Proteome analysis of radular teeth proteins.** In chitons, each radular tooth consists of a mineralized cusp, base (also referred to as stylus) supporting the mineralized cusp and a radular membrane (Fig. 1b). Previously, the mineralized cusps were physically separated from base plus radular membrane and the proteins were extracted from each part of radular teeth using boiling SDS<sup>15</sup>. Extracted proteins were digested into peptides and analyzed with nano-LC-MS. In that work, the lack of available genetic information on *C. stelleri* required a MS/MS ion search against the NCBI nr database and *de novo* peptide sequencing of MS/MS spectra in order to identify the proteins.

In this study, a full six-frame translated database was initially generated from assembled transcripts of *C. stelleri*. Then, the previously obtained MS/MS spectra of teeth proteins were analyzed by a MS/MS ion search using the translated database. Data sets of MS/MS spectra from replicate analyses ( $n = 2$ ) were used for the search. As a result, the peptide sequences from the mineralized cusp fraction were assigned to 232 transcripts (Supplementary Table S4) and those from the base plus membrane fraction were assigned to 114 transcripts (Supplementary Table S5). A high level of redundancy was present in the results as many of the peptides were shared by different proteins. This could be due to the presence of isoforms or artificial redundancy caused by sequencing errors or assembly errors<sup>23</sup>. The proteins matching the same set of peptides are listed in Supplementary Tables S4 and S5. From FPKM expression data, the proteins encoded by the transcripts which showed the greatest expression were selected for further analysis. Annotations of the proteins are based on the results of BlastP analyses.

In total, 31 proteins were identified from the base plus membrane fraction (Supplementary Table S5). From this, 22 proteins were identified from 1<sup>st</sup> analysis and 24 proteins were identified from 2<sup>nd</sup> analysis. From the cusp fraction, a total of 77 proteins were identified. 61 proteins were identified from the 1<sup>st</sup> analysis and 54 proteins were identified from the 2<sup>nd</sup> analysis. The proteins identified from the cusp fraction include four proteins:

Transcript ID	Annotation <sup>a</sup>	Best hit organisms	E-value	FPKM (mineralized)	FPKM (non mineralized)	1st analysis		2nd analysis	
						Mascot scores	Peptides	Mascot scores	Peptides
Cs77024 c0_g1_i1	Neuroglobin-like	<i>Actinia tenebrosa</i>	2.00E-25	2323.65	2020.22	845	4	428	4
Cs77196 c0_g2_i1	Myoglobin	<i>Liolophura japonica</i>	3.00E-84	874.48	5345.03	387	4	234	3
Cs75674 c0_g4_i1	protein usf-like	<i>Octopus bimaculoides</i>	1.00E-101	67.74	2144.29	331	5	54	1
Cs82664 c0_g1_i1	Arginine kinase	<i>Liolophura japonica</i>	0	104.69	2469.32	189	4	178	2
Cs18149 c2_g1_i5	Peroxidasin	<i>Crassostrea gigas</i>	0	12.02	0.18	173	3	195	3
Cs68435 c0_g1_i1	No hits found			309.95	75	126	2	88	1
Cs83384 c0_g1_i1	protein yellow-like	<i>Lingula anatina</i>	3.00E-103	18.69	2.05	109	1	90	1
Cs18149 c2_g1_i2	Peroxidasin	<i>Mizuhopecten yessoensis</i>	1.00E-19	1.98	0	102	2	111	3
Cs62719 c1_g8_i1	gastric intrinsic factor-like protein 2	<i>Hyriopsis cumingii</i>	2.00E-35	28.05	0.08	114	1	56	1
Cs58153 c0_g2_i5	apoptosis inducing factor-3	<i>Haliotis discus discus</i>	0.00E+00	62.87	16.52	95	1	98	2
Cs22616 c0_g1_i1	copper transport protein ATOX1-like	<i>Exaiptasia pallida</i>	3.00E-27	2324.41	144.53	109	1	104	1
Cs28725 c1_g1_i1	hillarin	<i>Crassostrea virginica</i>	0	15.52	62.53	83	1	61	1
Cs84529 c0_g2_i1	muscle-specific protein 20-like isoform X2	<i>Lingula anatina</i>	7.00E-91	999.04	345.27	97	1	74	1
Cs59400 c0_g1_i1	Eukaryotic initiation factor 4A-II, partial	<i>Stegodyphus mimosarum</i>	0	145.13	141.93	95	1	85	1
Cs74134 c0_g2_i2	histone H2B 8-like	<i>Protobothrops mucrosquamatus</i>	2.00E-68	6.85	51.47	91	1	127	1
Cs46312 c2_g1_i1	hephaestin-like protein	<i>Stylophora pistillata</i>	0	0.27	179.74	90	1	84	2
Cs54305 c0_g1_i2	filamin-A-like isoform X1	<i>Biomphalaria glabrata</i>	0	125.51	91.37	75	1	147	2
Cs59318 c0_g1_i1	peroxiredoxin 6	<i>Cristaria plicata</i>	4.00E-119	274.95	561.11	77	2	93	2
Cs71263 c1_g1_i1	Peroxidasin	<i>Crassostrea gigas</i>	0	9.37	0.08	55	1	70	1
Cs69741 c0_g1_i1	superoxide dismutase (Cu-Zn)	<i>Drosophila sechellia</i>	6.00E-46	46.09	22.35	76	1	190	2
Cs53984 c1_g2_i8	filamin-A-like isoform X5	<i>Crassostrea virginica</i>	2.00E-104	10.13	4.09	69	1	132	2
Cs28687 c2_g1_i1	cytosolic malate dehydrogenase	<i>Mytilus galloprovincialis</i>	3.00E-176	34.88	81.1	61	1	59	1

**Table 2.** List of mineralized cusp-specific proteins in *C. stelleri*. <sup>a</sup>Annotation is based on the results of BlastP analysis.

Myoglobin, actin, elongation factor 1 alpha and Arginine kinase, which are encoded by the top 20 most highly expressed transcripts in the non-mineralized teeth region (Cs77196|c0\_g2\_i1, Cs47470|c2\_g3\_i1, Cs24354|c0\_g1\_i1 and Cs82664|c0\_g1\_i1; Supplementary Table S1). An iron transporting protein, ferritin, whose gene was also highly expressed in non-mineralized teeth region, was not identified. The 19 proteins were identified from both cusp and base plus membrane fractions. It includes a large number of structural proteins such as actin, filamin-A, tubulin and fibrillin. Elongation factor 1 alpha (Cs24354|c0\_g1\_i1) was identified from a blank sample and is thought to be a contaminant.

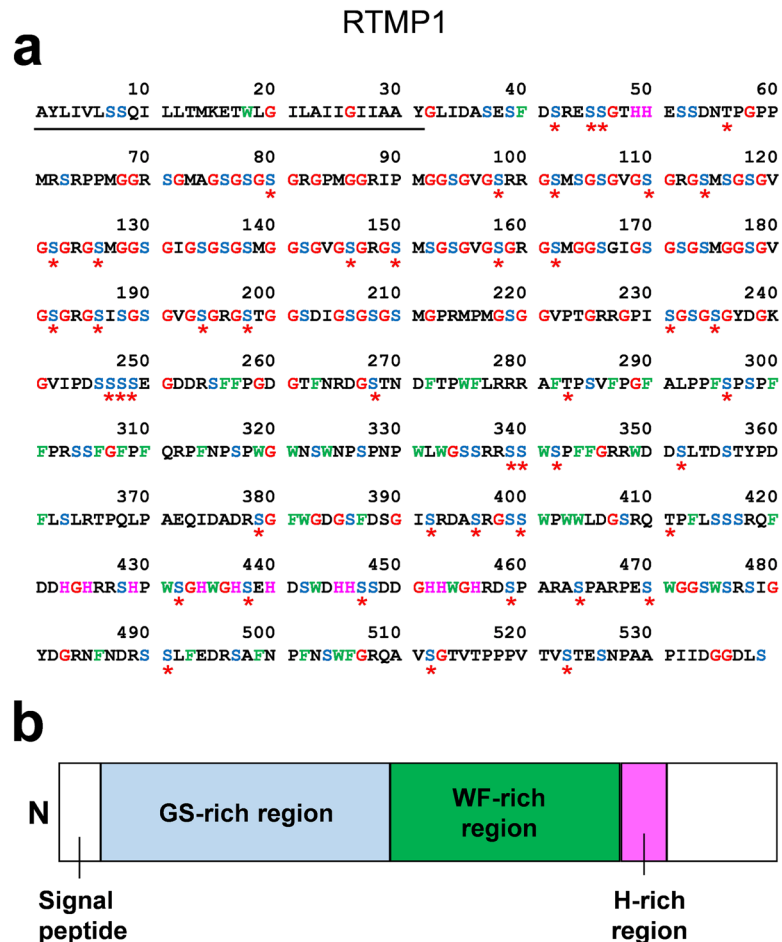
**Mineralized cusp-specific proteins.** The mineralized cusp proteins, which were identified in both 1<sup>st</sup> and 2<sup>nd</sup> analyses (and not identified from the base plus membrane fraction), were listed as mineralized cusp-specific proteins (Table 2).

Two globin-like proteins were identified as mineralized cusp-specific proteins with the highest MASCOT scores (Cs77196|c0\_g2\_i1 and Cs77024|c0\_g1\_i1). The sequences of both proteins contain heme-binding domains of globin proteins. Cs77196|c0\_g2\_i1 showed the best match with radular muscle myoglobin of *Liolophura japonica*<sup>24</sup>. This protein is abundant in the cusp region and its N-terminal sequence was determined in our previous study<sup>15</sup>. Cs77024|c0\_g1\_i1 showed a similarity with a neuroglobin-like protein. An acidic peptide sequence detected in the previous study which was not assigned to any protein at that time was the part of Cs77024|c0\_g1\_i1<sup>15</sup>. Sequence alignment with other neuroglobins revealed that the acidic region of Cs77024|c0\_g1\_i1 was unique to this protein.

Three proteins shared a similarity with peroxidase. Among them, two were isoforms: Cs18149|c2\_g1\_i5 and Cs18149|c2\_g1\_i2. All peroxidase possess N-terminal signal sequences and heme-dependent peroxidase domains (Supplementary Table S6). Furthermore, a chitin-binding domain was found at the C-terminus of Cs18149|c2\_g1\_i5. A number of peroxidase homologues are in the transcriptome data, and although there are 3'- or 5'- truncated forms, only Cs18149|c2\_g1\_i5 has a chitin-binding domain.

Two proteins correspond to antioxidant enzymes, peroxiredoxin 6 (Cs59318|c0\_g1\_i1) and superoxide dismutase (Cu-Zn) (Cs69741|c0\_g1\_i1). As a result of subcellular localization analysis, peroxiredoxin 6 was predicted as a cytoplasmic protein, while superoxide dismutase (Cu-Zn) was predicted as an extracellular protein (Supplementary Table S6).

Two proteins were homologous to the metal transport proteins, such as a copper transport protein ATOX1-like (Cs22616|c0\_g1\_i1) and a hephaestin-like protein (Cs46312|c2\_g1\_i1). Hephaestin is a protein



**Figure 2.** Amino acid sequence of a novel mineralized cusp-specific protein, radular teeth matrix protein 1 (RTMP1). **(a)** The entire amino acid sequence of RTMP1 deduced from the cDNA sequence. It consists of 539 amino acids and abundant in glycine (red) and serine (cyan). The underline indicates a putative signal sequence. The putative phosphorylation sites are indicated by asterisks. **(b)** Schematic of the domain structure of RTMP1. GS-rich repetitive region, WF-rich, and H-rich regions are represented by pale blue, green, and pink boxes.

that possesses ferroxidase activity and is believed to mediate intestinal iron absorption in mammals<sup>25,26</sup>. Both membrane-associated and soluble forms of hephaestin have been reported<sup>27,28</sup>. A domain search predicted the presence of a chitin-binding domain in hephaestin-like protein (Cs46312|c2\_g1\_i1). It has three predicted  $\beta$ -sheet structures and several aromatic amino acids that are characteristic to a type-3 chitin-binding domain<sup>29</sup>. Cs22616|c0\_g1\_i1 has a heavy metal-associated domain of ATOX1, which contain two cysteine residues for metal ion binding. ATOX1 is a copper chaperone that plays a role in incorporating Cu into the copper-dependent enzymes such as hephaestin<sup>30</sup>.

Other proteins include Arginine kinase (Cs82664|c0\_g1\_i1), hillarin (Cs28725|c1\_g1\_i1) and cytosolic malate dehydrogenase (Cs28687|c2\_g1\_i1), which are predicted to localize in the mitochondria (Supplementary Table S6). Furthermore, nuclear proteins (Eukaryotic initiation factor 4A-II, partial and histone H2B 8-like) and cytoskeletal proteins (muscle-specific protein 20-like isoform X2 and filamin-A-like isoforms) were identified. Protein usf-like (Cs75674|c0\_g4\_i1) that has a dienelactone hydrolase domain was identified with a high MASCOT score and a 33% sequence coverage.

There is only one protein (Cs68435|c0\_g1\_i1) that shared no sequence similarity with any known proteins. This protein possesses a unique repetitive sequence and an unusually high proportion of glycine and serine residues. Further analysis was conducted on this novel protein. Since the 5' terminus of the Cs68435|c0\_g1\_i1 coding sequence was truncated, its full-length sequence was determined by 5'-RACE. The overall sequence consists of a glycine- and serine-rich repetitive region, followed by a tryptophan- and phenylalanine-rich region and a histidine-rich region (Fig. 2). Structure prediction indicates that a disordered structure is present in a glycine- and serine-rich region (Supplementary Fig. S4). A sequence alignment of a WF-rich region in RTMP1 with the chitin-binding domains of four bacterial chitinases indicated that three of five putative chitin interacting residues in the chitin-binding domains were also conserved in RTMP1 (Supplementary Fig. S5)<sup>29</sup>. Three  $\beta$ -strands were predicted in the corresponding region of RTMP1. Based on results of *in silico* analyses, a number of serine residues of Cs68435|c0\_g1\_i1 are predicted to be phosphorylated. Analyses by several different prediction software

indicated that Cs68435|c0\_g1\_i1 contains an N-terminal signal sequence (Fig. 2, Supplementary Table S6). This novel protein is termed RTMP1 (radular teeth matrix protein 1).

**Gene expression profiles of mineralized cusp-specific proteins.** The expression profiles of the transcripts encoding mineralized cusp-specific proteins in different mineralization stages were analyzed based on FPKM values. Expression levels of several transcripts (e.g., Neuroglobin-like, peroxidase 6 and superoxide dismutase (Cu-Zn)) differed in abundance by less than 4-fold between the mineralized teeth and non-mineralized teeth regions (Supplementary Fig. S3). On the other hand, the transcripts whose expression increased more than 4-fold in the non-mineralized teeth region (e.g., protein usf-like, Arginine kinase, Myoglobin and hephaestin-like protein) or mineralized teeth region (e.g., RTMP1, Peroxidase and copper transport protein ATOX1-like) are thought to be differentially expressed during teeth mineralization (Supplementary Fig. S3).

## Discussion

In this study, a comprehensive transcriptome data from the radular tissue of *C. stelleri* has been developed for the first time. The constructed transcriptome data provides new insights into the genes highly expressed in radular tissue during teeth formation in *C. stelleri*. Furthermore, by using the transcriptome data, a proteome profile of mineralized cusp-specific proteins were determined. Those transcriptome and proteome data represent invaluable resources for the investigation of the magnetite biomineralization (including iron oxide nucleation, growth and phase transformation) in chitons.

Four homologues of ferritin were identified from the transcriptome data. Among the four ferritin homologs, the cytoplasmic ferritin, Cs75144|c0\_g1\_i1 is highly expressed in radular tissue. Previously, it was demonstrated that ferritin-containing granules first appear in the epithelial cells located 1–3 rows anterior to the tooth in which iron oxide is first deposited. The amount of granules increases greatly over a few rows, and then gradually decreases<sup>14</sup>. Those observations were correlated with the expression pattern of cytoplasmic ferritins (Supplementary Table S3). Since the iron transport is the earliest event in the tooth formation process, the expression of the ferritin genes might be strongly induced in the non-mineralized teeth region prior to mineralization.

Contrary to the above data, previous research using other chiton species suggested that secreted ferritins accumulated in the epithelial cells of radular tissue during teeth mineralization<sup>14,31</sup>. In order to determine which ferritin plays a role in teeth mineralization, further studies including protein analysis of epithelial cells are needed.

Although the ferritin gene, Cs75144|c0\_g1\_i1 was included in the twenty most highly expressed transcripts in the non-mineralized teeth region (Supplementary Table S1), ferritin proteins were not identified from mineralized cusps. This result was consistent with the previous observation that no ferritin granules were detected inside the teeth<sup>32</sup>. Based on these results, it is likely that iron in ferritin is released and transported inside the teeth in a soluble form, and re-deposited as ferrihydrite. Shaw *et al.* reported the existence of ferritin like granules in the cells located inside of the base (also referred to as stylus)<sup>33</sup>. However, in this study, we did not detect ferritin from the base plus membrane fraction (Supplementary Table S5). It is possible that sample preparation of the base plus membrane fraction that included repeated wash with 70% ethanol and water<sup>15</sup>, could result in the loss of ferritin proteins.

The mitochondrial respiratory chain enzymes were overrepresented among the top-20 most highly expressed transcripts in the mineralized teeth region (Supplementary Table S2). Electron microscopic analysis of partially mineralized teeth have previously shown the presence of numerous mitochondria in the cells surrounding the partially mineralized teeth<sup>14,34</sup>. It has been suggested that a large amount of energy required for iron transport or transformation of ferrihydrite to magnetite may be provided by those mitochondria<sup>14,34</sup>.

Proteomic analysis using the *de novo* constructed transcriptome data of radular tissue in *C. stelleri* led to the identification of three-fold more mineralized cusp-specific proteins (22 in total) than our previous report<sup>15</sup>. Furthermore, a novel protein that could not be detected by the previous method was identified in this study.

Cytoskeletal proteins (e.g., filamin-A-like isoforms, muscle-specific protein 20-like isoform X2) or nuclear proteins (e.g., Eukaryotic initiation factor 4A-II, partial, histone H2B 8-like), which were apparently unrelated to teeth formation, were identified as mineralized cusp-specific proteins (Table 2). It is likely that those highly expressed proteins were occluded in the teeth during mineralization and identified as mineralized cusp proteins.

Two globin-like proteins were identified as mineralized cusp-specific proteins. Significant amounts of myoglobin, which help supply oxygen to muscle<sup>35</sup>, have been known to exist in the radular muscles of chiton<sup>24,36,37</sup>. Therefore, it is likely that the highly expressed myoglobin could contaminate the mineralized cusp fraction. Conversely, as the oxygen concentration strongly affects the redox state of iron, iron release from ferritin and magnetite formation, those proteins could have essential roles for teeth formation and mineral phase transformation. Although the function of the neuroglobin-like protein, Cs77024|c0\_g1\_i1, is unknown, it has been suggested to play a role in O<sub>2</sub> control as myoglobin<sup>38</sup>. Of note, the neuroglobin-like protein, Cs77024|c0\_g1\_i1, has a unique acidic region that was not present in other neuroglobin-like proteins. The acidic domain of the magnetite morphology-modulating protein, Mms6 has been shown to be essential for its function<sup>39</sup>. Therefore, Cs77024|c0\_g1\_i1, which has a unique acidic amino acid rich region, might have a specialized function for magnetite formation.

3 out of 22 mineralized cusp-specific proteins (Cs18149|c2\_g1\_i5, Cs18149|c2\_g1\_i2 and Cs71263|c1\_g1\_i1) showed similarities to peroxidase. Peroxidase is known to have a hybrid structure consisting of a peroxidase domain and extracellular matrix motifs<sup>40</sup>. It is secreted extracellularly and mediates the cross-link formation in collagen for tissue development<sup>41</sup>. Notably, among all peroxidase-like proteins in transcriptome data, only Cs18149|c2\_g1\_i5 has a chitin binding domain in its C-terminus. It is possible that Cs18149|c2\_g1\_i5 is secreted extracellularly to bind chitin and catalyze the chemical bond formation to consolidate a scaffold for biomineralization. Interestingly, peroxidase has been identified from other calcium-based biominerals, including pearl oyster, and has been suggested to be important for biomineral formation<sup>42–44</sup>.

The biomineralization protein Silicatein, which was isolated from sponge spicules, has been known to catalyze the hydrolysis and polycondensation of silicon alkoxide to form silica<sup>45,46</sup>. We speculate that the usf-like protein (Cs75674|c0\_g4\_i1), containing a diene lactone hydrolase domain, could also have a catalytic role contributing to hydrolysis of iron oxide precursors to form magnetite<sup>47</sup>.

Two antioxidant proteins, peroxiredoxin 6 (Cs59318|c0\_g1\_i1) and superoxide dismutase (Cu-Zn) (Cs69741|c0\_g1\_i1) were predicted to be localized both inside and outside of the cells (Supplementary Table S6). Since the proposed models of iron release pathways from ferritins involved superoxide radicals<sup>48</sup>, superoxide dismutase (Cu-Zn) (Cs69741|c0\_g1\_i1) could control excess amounts of superoxide radicals. As we mentioned in our previous work, peroxiredoxin 6 (Cs59318|c0\_g1\_i1) may function to control the reduction of iron species and therefore control the initial phase transformation from ferrihydrite to magnetite by adsorption of Fe<sup>2+</sup> ions with subsequent solid-state transformation<sup>16</sup>.

A hephaestin-like protein (Cs46312|c2\_g1\_i1) is a ferroxidase protein and may be involved in ferrous iron oxidation into ferric iron. Of note, hephaestin-like protein contains an N-terminal signal sequence (Supplementary Table S6) and a chitin-binding domain. It suggests that extracellularly secreted hephaestin-like proteins may bind with chitin and catalyze ferrous iron oxidation.

A novel mineralized cusp-specific protein, RTMP1, has a distinctive glycine and serine-rich region (Fig. 2). Similar glycine- and serine-rich domains were found from other biomineral-associated proteins<sup>49–51</sup>. The secondary structure of a glycine and serine-rich region was predicted to be disordered and thus might form a flexible structure like the aforementioned biomineral proteins (Supplementary Fig. S4). It is possible that the extended conformation of RTMP1 might provide more mineral binding sites than tightly packed globular proteins<sup>52</sup>. The WF-rich region in RTMP1 resembles a type-3 chitin binding domain of bacterial chitinase, which have  $\beta$ -sheets as well as aromatic and hydrophobic-rich sequences (Supplementary Fig. S5). Thus, similar to those bacterial chitinases, it is likely that RTMP1 binds to crystalline forms of chitin<sup>29</sup>. A number of phosphorylated serines were inferred to reside in RTMP1. Of note, amino acid analysis of matrix proteins in the radula teeth of the chiton *Acanthopleura hirtosa* indicated that there were appreciable amounts of phosphorylated serines<sup>53</sup>. In other biominerals, phosphorylation is often required for mineral-protein interaction<sup>54,55</sup>. These results support the idea that the extracellularly secreted RTMP1, which is likely to be phosphorylated, may interact with chitin and promote iron oxide nucleation on chitin fibers.

Expression analysis of the transcripts encoding the mineralized cusp-specific proteins provides additional insight into the functions of those proteins for teeth mineralization. Except for the aforementioned possible contaminants, the transcripts highly expressed throughout teeth mineralization (e.g., Neuroglobin-like, peroxiredoxin 6 and superoxide dismutase (Cu-Zn)) may be important for the entire process of mineralization (Supplementary Fig. S3). Conversely, the transcripts highly expressed in the non-mineralized teeth region (e.g., protein usf-like, Arginine kinase, hephaestin-like protein) may play a role in the former processes, including initial iron transport (Supplementary Fig. S3). The transcripts highly expressed in the mineralized teeth region (e.g., RTMP1, Peroxidases, copper transport protein ATOX1-like) may be involved in latter processes, such as ferrihydrite nucleation, growth, and phase transformation to magnetite (Supplementary Fig. S3).

By analyzing gene deletion mutants, several proteins important for magnetite mineralization in magnetotactic bacteria have been reported<sup>6–9</sup>. Those proteins include redox-active proteins that may control oxidation and reduction of iron for magnetite mineralization<sup>56,57</sup>, magnetotactic bacteria-specific proteins related to maturation of magnetite crystals<sup>6,7</sup>, cation diffusion facilitators<sup>58</sup> and proteins with protease domains<sup>59</sup>. When compared to the magnetotactic bacterial proteins, redox-active proteins (peroxiredoxin-6, superoxide dismutase (Cu-Zn) and hephaestin-like protein) were also identified from *C. stelleri* as mineralized cusp-specific proteins. These proteins may play roles in oxidation and reduction of iron for transport, nucleation and magnetite mineralization.

## Conclusions

In this study, we constructed the first comprehensive transcriptome data sets from radular tissue in the gumbboot chiton, *Cryptochiton stelleri*. Transcript expression analysis revealed that the most expressed transcripts in the non-mineralized teeth region include ferritin genes, while those in the mineralized teeth region include genes of a number of mitochondrial respiratory chain enzymes. These results concur with the previous microscopic observation of developing radular teeth in chitons and suggest the importance of these genes in teeth mineralization. Furthermore, using proteomic analysis, 22 mineralized cusp specific proteins including a novel protein were identified. These proteins are good candidates for future research of iron oxide mineralization in chitons. The results obtained from this study are valuable in terms of providing the first overview of radular teeth formation.

## Methods

**RNA isolation and cDNA library construction.** Live specimens of *C. stelleri* were collected and maintained as previously described<sup>16</sup>. Collection of *Cryptochiton stelleri* is permitted by the Department of Fish and Game and the California Natural Resources Agency. Radular tissues were isolated from freshly dissected *C. stelleri*. After radular tissue isolation, non-mineralized teeth region (first 8 to 12 rows of transparent teeth) and mineralized teeth region (following 2 to 5 rows of partially-mineralized reddish-brown colored teeth plus 5 to 6 rows of mineralized magnetic teeth) were separated from whole radulae and frozen immediately in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until used. Total RNA was extracted from each radular tissue region separately using TRI reagent (Molecular Research Center), and purified using RNeasy mini kit (QIAGEN) following the manufacturer's instructions. Integrity and quantity of extracted RNA was assessed using Agilent 2100 Bioanalyzer (Agilent Technologies). cDNA Libraries were prepared using the TruSeq Stranded mRNA Sample Prep Kit (Illumina).

**Sequencing and *de novo* transcriptome assembly.** The constructed libraries were sequenced on an Illumina HiSeq 2500 platform and paired-end reads were generated. From raw reads, adapter sequences were

removed using cutadapt<sup>60</sup>. Reads were then quality trimmed using Trimmomatic<sup>61</sup>. The trimmed reads from two samples (non-mineralized teeth region and mineralized teeth region) were pooled and assembled *de novo* into a transcriptome reference using Trinity<sup>62</sup>.

**Bioinformatic analysis.** The trimmed reads obtained from two samples (non-mineralized teeth region and mineralized teeth region) were aligned to the above constructed transcriptome reference using Bowtie<sup>63</sup>. Based on the mapping results, the expression levels of transcripts were estimated by fragments per kilobase of transcript per million fragments mapped value (FPKM) using RSEM<sup>64</sup>. A homology search was conducted using the BLAST program<sup>65</sup> against NCBI nr database for functional annotation of transcripts and proteins. Homologues of ferritin genes were searched from the transcriptome data of *C. stelleri* using local BLAST. The conserved domains of protein sequences were searched with SMART<sup>66</sup>. MEGA version 6<sup>67</sup> was used for phylogenetic analysis of ferritin proteins with the maximum likelihood (ML) method. The following ferritin proteins from Mollusca and Arthropoda were used: LjFer (*Liolophura japonica*, BAA21810), MmxFer (*Meretrix meretrix*, AAZ20754), ScFer (*Sinonovacula constricta*, ACZ65230), AiFer2 (*Argopecten irradians*, AEN71561), AiFer3 (*Argopecten irradians*, AEN71564), PyFer1 (*Patinopecten yessoensis*, AGK92812), PyFer2 (*Patinopecten yessoensis*, AGK92813), PyFer4 (*Patinopecten yessoensis*, AGK92815), LsFer (*Lymnaea stagnalis*, CAA40097), CrFer-H1 (*Carcinoscorpius rotundicauda*, AAW22505), CrFer-H2 (*Carcinoscorpius rotundicauda*, AAW22506), HddFer (*Haliotis discus discus*, ABG88845), EsFer1 (*Eriocheir sinensis*, ADF87490), EsFer2 (*Eriocheir sinensis*, ADF87491), PyFerS1 (*Patinopecten yessoensis*, AHH31562), PyFerS2 (*Patinopecten yessoensis*, AHH31563), SdFer (*Suberites domuncula*, CAC84556). The analysis was run with the Jones-Taylor-Thornton (JTT) model. Bootstrap trials were replicated 1000 times. The putative phosphorylation sites of RTMP1 was predicted in silico using NetPhos 3.1 Server (<http://www.cbs.dtu.dk/services/NetPhos/>)<sup>68</sup>. The potential phosphorylation sites with scores higher than 0.9 were considered to be positive. Disordered regions in RTMP1 were predicted using the DISOPRED server<sup>69</sup>. Jpred<sup>70</sup> was used to predict the secondary structures of the WF-rich region in RTMP1. Subcellular localization of identified proteins were predicted using the following software: SignalP 4.1<sup>71</sup>, TargetP1.1<sup>72</sup>, WOLF PSORT<sup>73</sup>, MultiLoc2<sup>74</sup> and Signal-BLAST<sup>75</sup>. Amino acid sequence alignments were conducted using Clustal Omega software<sup>76</sup>.

**Proteomic analysis.** A detailed method of protein extraction from the radular teeth of *C. stelleri* and following LC-MS/MS analysis were described in our previous work<sup>15</sup>. In this study, first, a six frame translated database was created from a transcriptome data of *C. stelleri*. Then the previously obtained MS/MS spectra of radular teeth proteins were searched against a six-frame translated database with the MASCOT algorithm for protein identification. The parameters used for MASCOT were same as previously described<sup>15</sup>. The protein identification was based on two technical replicates.

**5' Rapid amplification of cDNA ends (RACE).** 5' rapid amplification of cDNA ends (RACE) was performed with 5'-Full RACE Core Set (Takara) to determine the full length sequences of Cs68435|c0\_g1\_i1 and Cs59318|c0\_g1\_i1. Total RNA was extracted from radular tissue as described above. 5'-RACE products were cloned into pMD20-T vector (Takara) and sequenced.

## Data Availability

The reads derived from Illumina Hiseq of the analyzed samples have been deposited in DDBJ Sequence Read Archive with accession number DRA007106.

## References

- Lowenstam, H. A. Magnetite in Denticle Capping in Recent Chitons. *Geological Society of America Bulletin* **73**, 435–438 (1962).
- Blakemore, R. Magnetotactic bacteria. *Science* **190**, 377–379 (1975).
- Gould, J. L., Kirschvink, J. L. & Deffeyes, K. S. Bees have magnetic remanence. *Science* **201**, 1026–1028, <https://doi.org/10.1126/science.201.4360.1026> (1978).
- Kirschvink, J. L., Walker, M. M., Chang, S.-B., Dizon, A. E. & Peterson, K. A. Chains of single-domain magnetite particles in chinook salmon, *Oncorhynchus tshawytscha*. *J Comp Physiol A* **157**, 375–381 (1985).
- Walcott, C., Gould, J. L. & Kirschvink, J. L. Pigeons have magnets. *Science* **205**, 1027–1029 (1979).
- Tanaka, M., Mazuyama, E., Arakaki, A. & Matsunaga, T. MMS6 Protein Regulates Crystal Morphology during Nano-sized Magnetite Biomineralization in Vivo. *Journal of Biological Chemistry* **286**, 6386–6392, <https://doi.org/10.1074/jbc.M110.183434> (2011).
- Arakaki, A., Yamagishi, A., Fukuyo, A., Tanaka, M. & Matsunaga, T. Co-ordinated functions of Mms proteins define the surface structure of cubo-octahedral magnetite crystals in magnetotactic bacteria. *Molecular Microbiology* **93**, 554–567, <https://doi.org/10.1111/mmi.12683> (2014).
- Murat, D., Quinlan, A., Vali, H. & Komeili, A. Comprehensive genetic dissection of the magnetosome gene island reveals the step-wise assembly of a prokaryotic organelle. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5593–5598, <https://doi.org/10.1073/pnas.0914439107> (2010).
- Lohsse, A. *et al.* Genetic Dissection of the mamAB and mms6 Operons Reveals a Gene Set Essential for Magnetosome Biogenesis in *Magnetospirillum gryphiswaldense*. *Journal of Bacteriology* **196**, 2658–2669, <https://doi.org/10.1128/jb.01716-14> (2014).
- Weaver, J. C. *et al.* Analysis of an ultra hard magnetic biomineral in chiton radular teeth. *Materialstoday* **13**, 42–52 (2010).
- Grunenfelder, L. K. *et al.* Stress and Damage Mitigation from Oriented Nanostructures within the Radular Teeth of *Cryptochiton stelleri*. *Advanced Functional Materials* **24**, 6093–6104, <https://doi.org/10.1002/adfm.201401091> (2014).
- de Obaldia, E. E., Jeong, C., Grunenfelder, L. K., Kisailus, D. & Zavattieri, P. Analysis of the mechanical response of biomimetic materials with highly oriented microstructures through 3D printing, mechanical testing and modeling. *Journal of the Mechanical Behavior of Biomedical Materials* **48**, 70–85, <https://doi.org/10.1016/j.jmbbm.2015.03.026> (2015).
- de Obaldia, E. E., Herrera, S., Grunenfelder, L. K., Kisailus, D. & Zavattieri, P. Competing mechanisms in the wear resistance behavior of biomineralized rod-like microstructures. *Journal of the Mechanics and Physics of Solids* **96**, 511–534, <https://doi.org/10.1016/j.jmps.2016.08.001> (2016).
- Nesson, M. H. & Lowenstam, H. A. In *Magnetite Biomineralization and Magnetoreception in Organisms* (eds J. L. Kirschvink, D. S. Jones, & B. J. MacFadden) Ch. 16, 333–363 (Springer US, 1985).



15. Nemoto, M. *et al.* Proteomic analysis from the mineralized radular teeth of the giant Pacific chiton, *Cryptochiton stelleri* (Mollusca). *Proteomics* **12**, 2890–2894, <https://doi.org/10.1002/pmic.201100473> (2012).
16. Wang, Q. Q. *et al.* Phase Transformations and Structural Developments in the Radular Teeth of *Cryptochiton Stelleri*. *Advanced Functional Materials* **23**, 2908–2917, <https://doi.org/10.1002/adfm.201202894> (2013).
17. Tellam, R. L., Wijffels, G. & Willadsen, P. Peritrophic matrix proteins. *Insect Biochemistry and Molecular Biology* **29**, 87–101, [https://doi.org/10.1016/s0965-1748\(98\)00123-4](https://doi.org/10.1016/s0965-1748(98)00123-4) (1999).
18. Suzuki, M. *et al.* An Acidic Matrix Protein, Pif, Is a Key Macromolecule for Nacre Formation. *Science* **325**, 1388–1390, <https://doi.org/10.1126/science.1173793> (2009).
19. Elvin, C. M. *et al.* Characterization of a major peritrophic membrane protein, peritrophin-44, from the larvae of *Lucilia cuprina* - cDNA and deduced amino acid sequences. *Journal of Biological Chemistry* **271**, 8925–8935, <https://doi.org/10.1074/jbc.271.15.8925> (1996).
20. Irisarri, I., Eernisse, D. J. & Zardoya, R. Molecular phylogeny of Acanthochitonina (Mollusca: Polyplacophora: Chitonida): three new mitochondrial genomes, rearranged gene orders and systematics. *Journal of Natural History* **48**, 2825–2853, <https://doi.org/10.1080/00222933.2014.963721> (2014).
21. Towe, K. M., Lowenstam, H. A. & Nesson, M. H. Invertebrate ferritin: Occurrence in Mollusca. *Science* **142**, 63–64 (1963).
22. Vondaril, M., Harrison, P. M. & Bottke, W. cDNA cloning and deduced amino acid sequence of two ferritins: soma ferritin and yolk ferritin, from the snail *Lymnaea stagnalis* L. *European Journal of Biochemistry* **222**, 353–366 (1994).
23. Nesvizhskii, A. I. & Aebersold, R. Interpretation of shotgun proteomic data - The protein inference problem. *Molecular & Cellular Proteomics* **4**, 1419–1440, <https://doi.org/10.1074/mcp.R500012-MCP200> (2005).
24. Suzuki, T., Furukohri, T. & Okamoto, S. Amino acid sequence of myoglobin from the chiton *Liolophura japonica* and a phylogenetic tree for molluscan globins. *Journal of Protein Chemistry* **12**, 45–50, <https://doi.org/10.1007/bf01024913> (1993).
25. Fuqua, B. K. *et al.* The Multicopper Ferroxidase Hephaestin Enhances Intestinal Iron Absorption in Mice. *Plos One* **9**, <https://doi.org/10.1371/journal.pone.0098792> (2014).
26. Vulpe, C. D. *et al.* Hephaestin, a ceruloplasmin homologue implicated in intestinal iron transport, is defective in the sla mouse. *Nature Genetics* **21**, 195–199 (1999).
27. Ranganathan, P. N., Lu, Y., Fuqua, B. K. & Collins, J. F. Immunoreactive Hephaestin and ferroxidase activity are present in the cytosolic fraction of rat enterocytes. *Biomaterials* **25**, 687–695, <https://doi.org/10.1007/s10534-012-9527-9> (2012).
28. Martins, R., Silva, B., Proenca, D. & Faustino, P. Differential HFE Gene Expression Is Regulated by Alternative Splicing in Human Tissues. *Plos One* **6**, <https://doi.org/10.1371/journal.pone.0017542> (2011).
29. Ikegami, T. *et al.* Solution structure of the chitin-binding domain of *Bacillus circulans* WL-12 chitinase A1. *Journal of Biological Chemistry* **275**, 13654–13661, <https://doi.org/10.1074/jbc.275.18.13654> (2000).
30. Pufahl, R. A. *et al.* Metal ion chaperone function of the soluble Cu(I) receptor Atx1. *Science* **278**, 853–856, <https://doi.org/10.1126/science.278.5339.853> (1997).
31. Okoshi, K. & Hamaguchi, M. In *KAIYOUSEIBUTU NO KINOU (Function of marine organism)* (ed. Yoshio Takei) 345–356 (Tokai Daigaku Shuppan, 2005).
32. Lowenstam, H. A. Minerals formed by organisms. *Science* **211**, 1126–1131 (1981).
33. Shaw, J. A. *et al.* The Chiton Stylus Canal: An Element Delivery Pathway for Tooth Cusp Biomineralization. *Journal of Morphology* **270**, 588–600, <https://doi.org/10.1002/jmor.10705> (2009).
34. Shaw, J. A. *et al.* Ultrastructure of the Epithelial Cells Associated with Tooth Biomineralization in the Chiton *Acanthopleura hirtosa*. *Microscopy and Microanalysis* **15**, 154–165, <https://doi.org/10.1017/s1431927609090230> (2009).
35. Suzuki, T. & Imai, K. Evolution of myoglobin. *Cellular and Molecular Life Sciences* **54**, 979–1004, <https://doi.org/10.1007/s00180050227> (1998).
36. Terwilliger, R. C. & Read, K. R. H. The radular muscle myoglobins of the amphineuran mollusc, *Acanthopleura granulata* Gmelin. *Comparative Biochemistry and Physiology* **29**, 551–560 (1969).
37. Terwilliger, R. C. & Read, K. R. H. The radular muscle myoglobins of the amphineuran molluscs, *Katharina tunicata* Wood, *Cryptochiton stelleri* Middendorf, and *Mopalia muscosa* Gould. *International Journal of Biochemistry* **1**, 281–291 (1970).
38. Burmester, T. & Hankeln, T. What is the function of neuroglobin? *Journal of Experimental Biology* **212**, 1423–1428, <https://doi.org/10.1242/jeb.000729> (2009).
39. Yamagishi, A., Narumiya, K., Tanaka, M., Matsunaga, T. & Arakaki, A. Core Amino Acid Residues in the Morphology-Regulating Protein, Mms6, for Intracellular Magnetite Biomineralization. *Scientific Reports* **6**, <https://doi.org/10.1038/srep35670> (2016).
40. Nelson, R. E. *et al.* Peroxidase: a novel enzyme-matrix protein of *Drosophila* development. *Embo Journal* **13**, 3438–3447 (1994).
41. Bhave, G. *et al.* Peroxidase forms sulfimine chemical bonds using hypohalous acids in tissue genesis. *Nature Chemical Biology* **8**, 784–790, <https://doi.org/10.1038/nchembio.1038> (2012).
42. Liu, C. *et al.* In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. *Scientific Reports* **5**, <https://doi.org/10.1038/srep17269> (2015).
43. Germer, J., Mann, K., Worheide, G. & Jackson, D. J. The Skeleton Forming Proteome of an Early Branching Metazoan: A Molecular Survey of the Biomineralization Components Employed by the Coralline Sponge *Vaceletia* Sp. *Plos One* **10**, <https://doi.org/10.1371/journal.pone.0140100> (2015).
44. Luo, Y. J. *et al.* The *Lingula* genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nature Communications* **6**, <https://doi.org/10.1038/ncomms9301> (2015).
45. Cha, J. N. *et al.* Silicatein filaments and subunits from a marine sponge direct the polymerization of silica and silicones *in vitro*. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 361–365, <https://doi.org/10.1073/pnas.96.2.361> (1999).
46. Shimizu, K., Cha, J., Stucky, G. D. & Morse, D. E. Silicatein alpha: Cathepsin L-like protein in sponge biosilica. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6234–6238, <https://doi.org/10.1073/pnas.95.11.6234> (1998).
47. Bharde, A. *et al.* Extracellular biosynthesis of magnetite using fungi. *Small* **2**, 135–141, <https://doi.org/10.1002/sml.2005000180> (2006).
48. Melman, G. *et al.* Iron release from ferritin by flavin nucleotides. *Biochimica Et Biophysica Acta-General Subjects* **1830**, 4669–4674, <https://doi.org/10.1016/j.bbagen.2013.05.031> (2013).
49. Sarashina, I. & Endo, K. The complete primary structure of molluscan shell protein 1 (MSP-1), an acidic glycoprotein in the shell matrix of the scallop *Patinopecten yessoensis*. *Marine Biotechnology* **3**, 362–369, <https://doi.org/10.1007/s10126-001-0013-6> (2001).
50. Shen, X., Belcher, A. M., Hansma, P. K., Stucky, G. D. & Morse, D. E. Molecular cloning and characterization of lustrin A, a matrix protein from shell and pearl nacre of *Haliotis rufescens*. *J Biol Chem* **272**, 32472–32481 (1997).
51. Kroger, N., Deutzmann, R. & Sumper, M. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. *Science* **286**, 1129–1132, <https://doi.org/10.1126/science.286.5442.1129> (1999).
52. Magdalena, W., Piotr, D. & O., A. In *Advanced Topics in Biomineralization* (ed. Jong Seto) (2012).
53. Evans, L. A., Macey, D. J. & Webb, J. Distribution and composition of matrix protein in the radula teeth of the chiton *Acanthopleura hirtosa*. *Marine Biology* **109**, 281–286, <https://doi.org/10.1007/bf01319396> (1991).
54. Kroger, N., Lorenz, S., Brunner, E. & Sumper, M. Self-assembly of highly phosphorylated silaffins and their function in biosilica morphogenesis. *Science* **298**, 584–586, <https://doi.org/10.1126/science.1076221> (2002).
55. Kwak, S. Y. *et al.* Role of 20-kDa Amelogenin (P148) Phosphorylation in Calcium Phosphate Formation *In Vitro*. *Journal of Biological Chemistry* **284**, 18972–18979, <https://doi.org/10.1074/jbc.M109.020370> (2009).

56. Raschdorf, O., Muller, F. D., Posfai, M., Plitzko, J. M. & Schuler, D. The magnetosome proteins MamX, MamZ and MamH are involved in redox control of magnetite biomineralization in *Magnetospirillum gryphiswaldense*. *Molecular Microbiology* **89**, 872–886, <https://doi.org/10.1111/mmi.12317> (2013).
57. Jones, S. R. *et al.* Genetic and biochemical investigations of the role of MamP in redox control of iron biomineralization in *Magnetospirillum magneticum*. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 3904–3909, <https://doi.org/10.1073/pnas.1417614112> (2015).
58. Uebe, R. *et al.* The cation diffusion facilitator proteins MamB and MamM of *Magnetospirillum gryphiswaldense* have distinct and complex functions, and are involved in magnetite biomineralization and magnetosome membrane assembly. *Molecular Microbiology* **82**, 818–835, <https://doi.org/10.1111/j.1365-2958.2011.07863.x> (2011).
59. Yang, W. *et al.* mamO and mamE genes are essential for magnetosome crystal biomineralization in *Magnetospirillum gryphiswaldense* MSR-1. *Research in Microbiology* **161**, 701–705, <https://doi.org/10.1016/j.resmic.2010.07.002> (2010).
60. Martin, M. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**, 10–12, <https://doi.org/10.14806/ej.17.1.200> (2011).
61. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
62. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**, 1494–1512, <https://doi.org/10.1038/nprot.2013.084> (2013).
63. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, <https://doi.org/10.1186/gb-2009-10-3-r25> (2009).
64. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *Bmc Bioinformatics* **12**, <https://doi.org/10.1186/1471-2105-12-323> (2011).
65. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402, <https://doi.org/10.1093/nar/25.17.3389> (1997).
66. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 5857–5864, <https://doi.org/10.1073/pnas.95.11.5857> (1998).
67. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725–2729, <https://doi.org/10.1093/molbev/mst197> (2013).
68. Blom, N., Gammeltoft, S. & Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology* **294**, 1351–1362, <https://doi.org/10.1006/jmbi.1999.3310> (1999).
69. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138–2139, <https://doi.org/10.1093/bioinformatics/bth195> (2004).
70. Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. & Barton, G. J. JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**, 892–893, <https://doi.org/10.1093/bioinformatics/14.10.892> (1998).
71. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785–786, <https://doi.org/10.1038/nmeth.1701> (2011).
72. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**, 1005–1016, <https://doi.org/10.1006/jmbi.2000.3903> (2000).
73. Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic Acids Research* **35**, W585–W587, <https://doi.org/10.1093/nar/gkm259> (2007).
74. Blum, T., Briesemeister, S. & Kohlbacher, O. MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *Bmc Bioinformatics* **10**, <https://doi.org/10.1186/1471-2105-10-274> (2009).
75. Frank, K. & Sippl, M. J. High-performance signal peptide prediction based on sequence alignment techniques. *Bioinformatics* **24**, 2172–2176, <https://doi.org/10.1093/bioinformatics/btn422> (2008).
76. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**, <https://doi.org/10.1038/msb.2011.75> (2011).

## Acknowledgements

This work was supported by Program to Disseminate Tenure Tracking System from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan (to M.N.), Kato Memorial Bioscience Foundation (to M.N.) and Army Research Office: W911NF-15-1-0306 (to D.K.).

## Author Contributions

M.N. and D.K. designed the research; M.N., D.R., S.H. and S.P. performed the experiments; M.N., T.T., K.I. and D.K. analyzed the data; M.N. and D.K. wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-37839-2>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019