High-throughput identification of heavy metal binding proteins from the byssus of chinese green mussel (*Perna viridis*) by combination of transcriptome and proteome sequencing

Xinhui Zhang\(^1\)\(^2\)*, Huiwei Huang\(^1\)*, Yanbin He\(^3\)*, Zhiqiang Ruan\(^2\)*, Xinxin You\(^2\), Wanshun Li\(^3\), Bo Wen\(^3\), Zizheng Lu\(^4\), Bing Liu\(^1\), Xu Deng\(^1\)*, Qiong Shi\(^1\)\(^2\)*

\(^1\) Shenzhen Key Laboratory of Marine Bioresource and Eco-Environmental Science, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China, \(^2\) Shenzhen Key Lab of Marine Genomics, Guangdong Provincial Key Lab of Molecular Breeding in Marine Economic Animals, BGI Academy of Marine Sciences, BGI Marine, BGI, Shenzhen, China, \(^3\) BGI-Shenzhen, BGI, Shenzhen, China, \(^4\) Shenzhen Horus Marine Technology Co. Ltd., Shenzhen, China, \(^5\) Laboratory of Aquatic Bioinformatics, BGI-Zhenjiang Institute of Hydrobiology, BGI Marine, BGI, Zhenjiang, China

* These authors contributed equally to this work.
* shiqiong@genomics.cn (QS); dengxu@suzu.edu.cn (XD)

Abstract

The Byssus, which is derived from the foot gland of mussels, has been proved to bind heavy metals effectively, but few studies have focused on the molecular mechanisms behind the accumulation of heavy metals by the byssus. In this study, we integrated high-throughput transcriptome and proteome sequencing to construct a comprehensive protein database for the byssus of Chinese green mussel (*Perna viridis*), aiming at providing novel insights into the molecular mechanisms by which the byssus binds to heavy metals. Illumina transcriptome sequencing generated a total of 55,670,668 reads. After filtration, we obtained 53,047,718 clean reads and subjected them to de novo assembly using Trinity software. Finally, we annotated 73,264 unigenes and predicted a total of 34,298 protein coding sequences. Moreover, byssal samples were analyzed by proteome sequencing, with the translated protein database from the foot transcriptome as the reference for further prediction of byssal proteins. We eventually determined 187 protein sequences in the byssus, of which 181 proteins are reported for the first time. Interestingly, we observed that many of these byssal proteins are rich in histidine or cysteine residues, which may contribute to the byssal accumulation of heavy metals. Finally, we picked one representative protein, Pvfp-5-1, for recombinant protein synthesis and experimental verification of its efficient binding to cadmium (Cd\(^{2+}\)) ions.

Introduction

Next-generation sequencing (NGS) technologies have been employed at a large scale for molecular studies of non-model organisms [1]. They have promoted the development of...
transcriptome sequencing, which usually presents a complete set of transcripts in a tissue or cell for revealing molecular bases of functional responses at specific developmental stages or to environmental changes [2, 3]. Many molecular changes of an organism upon environmental stress can be interpreted in a comprehensive way through high-throughput transcriptomes [4]. Proteome sequencing by liquid chromatography tandem mass spectrometry (LC-MS/MS) is another effective technique for the high-throughput identification of proteins, and it has proved to be an effective tool to characterize protein structures in model or non-model species [5–7]. In contrast to conventional methods, proteome sequencing allows for the identification of a large number of proteins in one sample.

Many metal ions are essential in organisms for various physiological roles, but they become toxic at high concentrations. Anthropogenic activities and products (such as waste, sewage, and industrial wastewater) release heavy metals into aquatic environments and generate a serious threat to ecosystems [8]. Heavy metal ions are very difficult to remove from aquatic environments by using physical, chemical, or biological methods. However, some organisms have attracted increasing attention due to the effective accumulation of heavy metals in their bodies; they can be used directly or indirectly for decontamination of heavy metals from aquatic environments. For example, certain algae and bacteria can be used for the clean-up of environments contaminated with heavy metals [9, 10]. Mussels have also been extensively applied to environmental monitoring programs [11]. Many Mytilidae mussels have been employed as biomonitors throughout the Indo-Pacific region for assessing chemical and heavy metal pollutants [12, 13]. They are useful due to their widespread distribution and sedentary life style, and they grow enough tissue for studying the accumulation of heavy metals.

Mussels can generate high-performance natural adhesives, which have been applied for surgery, cell culture, immunohistochemistry, sealants, coatings, and anchoring purposes [14, 15]. The mussel byssus has a strong adhesive capacity, which keeps the mussel stably stuck to rocks or growing substrates in strongly flowing waters. The molecular mechanisms of adhesion in mussels have been well studied before [16–18]. We previously reported that the majority of heavy metals accumulate in the byssus, and even after separation from the mussels, the byssus still contains heavy metals [19, 20]. In this study, we tried to reveal the composition of the byssus of the Chinese green mussel (Perna viridis), aiming at providing novel insights into the molecular mechanisms of byssal binding to heavy metals. Therefore, we combined transcriptome and proteome sequencing to explore the diversity of byssal proteins in this mussel species. Through this integrative approach, we identified many novel protein sequences that have not been previously reported in any public protein database, and we provide basic data for in-depth studies on novel byssal proteins. Our ultimate goal is to combine our knowledge about the molecular structures and the mechanical features of the byssus and to design byssal-protein-based biomaterials for the removal of heavy metal pollutants from aquatic environments.

Materials and methods

Sample collection and total RNA extraction

Fresh specimens of P. viridis (30 individuals, shell length 6–8 cm) were collected from a local market in Yantian District, Shenzhen, Guangdong Province, China. The foot areas of 5 mussels (near the foot gland; Fig 1A) were collected and snap frozen in liquid nitrogen before storage at −80°C. Total RNA of each sample was extracted using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer’s instructions. After treatment with RNase-Free DNase I (Thermo Fisher Scientific, Waltham, MA, USA) to eliminate genomic DNAs, the extracted mRNAs were reverse transcribed to construct a cDNA library for further transcriptome sequencing.
Transcriptome sequencing and data analysis

The cDNA library was sequenced using a HiSeq2000 sequencing platform (Illumina, San Diego, CA, USA) with the 90-bp paired-end (PE) sequencing module. We subsequently filtered raw reads to remove adapter sequences and reads with more than 5% of non-sequenced (N) bases or with a quality value below 20. We then employed Trinity software [21] to assemble clean reads to obtain contigs and unigenes. Functions of these unigenes were further predicted on the basis of sequence similarity searches with several public databases, including the NCBI non-redundant protein database (Nr), NCBI non-redundant nucleotide database (Nt), Kyoto Encyclopedia of Genes and Genomes (KEGG), Swiss-Port, and Clusters of orthologous groups of proteins (COG).
We also employed Blast2GO [22] to predict unigenes and obtain gene ontology (GO) annotation for each unigene. Subsequently, we performed GO functional classification of these unigenes using WEGO [23]. KEGG annotation was also applied to obtain pathway annotation for these unigenes. We searched unigene sequences against the public databases using BLASTX (E-value ≤ 1.0e^{-5}), with a priority order of Nr, Swiss-Port, KEGG, and COG. The alignment results were subsequently used to determine coding sequences of the unigenes and translate them into amino acid sequences. If unigenes had no hit in any known protein database, their coding sequences were predicted using ESTScan [24], and also translated into the corresponding protein sequences.

**Protein fractionation and mass-spectrometry (MS) analysis**

Twenty of the collected mussels were cultured in a glass tank at 26–28˚C, where they generated threads and plaques overnight. Threads (0.5 g; pooled from 10 mussels) and plaques (0.3 g; pooled from 10 mussels) were harvested (Fig 1A) for further grinding in liquid nitrogen. After the addition of acetic acid (1 ml, 5%) and treatment by ultrasound for 3 min, the protein lysates were centrifuged at 19,160 × g for 15 min at 4˚C to remove debris. After the addition of 100 μl of L3 Buffer (7 M urea, 2 M thiourea, 50 mM Tris-HCl, pH 8.0) to each lysate, the supernatants were used as plaque (1.02 μg/μl) and thread (5.91 μg/μl) protein extracts, respectively.

The obtained protein solutions were subjected to SDS-PAGE (Fig 1C) followed by in-gel digestion with trypsin [25] in 10 μl of 50 mM NH₄HCO₃ for 12 h at 37˚C. Subsequently the pooled mixtures of peptides were fractionated into 10 portions using SCX chromatography (GE, Boston, MA, USA). The fractionated peptides were further separated by LC-20AD (Shimadzu, Kyoto, Japan) high-pH reverse-phase chromatography and analyzed by LTQ-Orbitrap Velos (Thermo Fisher Scientific) [26].

The acquired MS data were converted to MGF files by Proteome Discoverer 1.4 (Thermo Fisher Scientific), and then the exported MGF files were searched using Mascot (v2.3.02; MatrixScience, London, UK) against the byssal-transcriptome-annotated database. Mascot parameters were set as follows. Trypsin was selected as the specific enzyme with a maximum of 1 missed cleavage permitted per peptide; fixed modifications of carbamidomethyl (C); variable modifications consisting of oxidation (M), deamidation (N, Q) and Gln->pyro-Glu (N-term Q); peptide charge, 2+, 3+, and 4+; 20 ppm of peptide mass tolerance; 0.05 Da of fragment mass tolerance. The automatic Mascot decoy database search was performed, and the Mascot results were processed by IQuant [27]. MascorPercolator was utilized to re-score the peptide spectrum matches (PSMs) [28, 29]. The identified peptide sequences were subsequently assembled into a set of confident proteins using the Occam’s razor approach implemented in IQuant. Finally, the false discovery rate (FDR) was set at 1%, at both the PSM and the protein levels [30].

**Reverse-transcription PCR (RT-PCR)**

Total RNA was extracted as described above. Reverse transcription of cDNA was subsequently performed with 2 μg of DNase-treated total RNA using the M-MuLV First Strand cDNA Synthesis Kit (Sangon, Shanghai, China). We randomly selected 6 byssal protein coding genes and designed primer pairs using Primer Premier 5.0 (S1 Table) for PCR validation. The primary RT-PCR reactions were carried out in a volume of 50 μl, containing 0.5 μl of rTaq DNA Polymerase (Toyobo, Osaka, Japan), 0.5 μl of cDNA (1,000 ng), 1×PCR reaction buffer, 0.2 μM of forward and reverse primers, and 200 μM of each dNTP. DNA amplification on an ABI 9700 thermal cycler (Thermo Fisher Scientific) was performed with the following cycling conditions: initial denaturation at 94˚C for 5 min; then 35 cycles of 94˚C for 30 sec, 55˚C for 30 sec
and 72˚C for 1 min; final extension at 72˚C for 10 min. All PCR amplicons were analyzed by 1.5% agarose gel electrophoresis for further sequencing validation.

**Pvfp-5-1: Cloning, protein expression and purification**

The protein sequence of Pvfp-5-1, a byssal protein, was obtained from the LC-MS/MS analysis. Molecular cloning and standard recombinant DNA techniques were applied to clone the Pvfp-5-1 gene into *E. coli*. Codon adaptation of the amino acid sequences of Pvfp-5-1 was carried out by online codon optimization software of the Codon Adaptation Tool (JACT) [31]. Forward and reverse primers containing *BamHI* and *XhoI* restriction sites (5’-GGATCCTACGATTACCGTGA-3’ and 5’-CTCGAGGTAGTATTTACCCAG-3’) were designed, respectively, using the modified Pvfp-5-1 nucleotide sequence (S2 Table).

The Pvfp-5-1 plasmid was mixed with competent *E. coli* cells that were subsequently cultured on LB supplemented with 100 μg/ml of ampicillin overnight at 37˚C. Sequencing was performed to identify Pvfp-5-1-positive colonies. After the colony confirmation, we used a Prime Prep Plasmid DNA Isolation Kit (GeNet Bio, Cheonan, South Korea) to extract the Pvfp-5-1 and pET-32a vectors and digested them with *BamHI* and *XhoI* at 37˚C for 4 h. The Pvfp-5-1 construct was separated on a 1% agarose gel, purified with a Prime Prep Gel Purification Kit (GeNet Bio), and then ligated into the multiple cloning site (MCS) of the T7lact promoter expression plasmid pET-32a with T4 DNA ligase (Thermo Fisher Scientific). To confirm the successful cloning of the full length of Pvfp-5-1 into the pET-32a vector, we extracted and sequenced these recombinant plasmids. Only the validated pET-32a-Pvfp-5-1 plasmid was transformed into *E. coli* BL21 (DE3) to obtain purified cells for expression of the Pvfp-5-1 gene. The cells were cultured in 50 ml of liquid LB, incubated in a shaker at 37˚C for 12–16 h, and then inoculated in 200 ml of liquid LB at a ratio of 1: 100. After incubation at 37˚C until an OD of 0.5~0.7 was reached, IPTG was added to the cell culture at a final concentration of 1 mM, and continuous shaking was performed for 4 more hours. Subsequent centrifugation at 1,532 × g for 15 minutes (4˚C) was carried out, and the cells were collected and stored at −20˚C until further use.

Moreover, we collected 200 μl of the upper bacterial supernatant for SDS-PAGE analysis. We added 25 μl of distilled water and 25 μl of 2× protein loading buffer to each sample before boiling at 100˚C for 10 minutes. After a short centrifugation, the protein products were separated by standard SDS-PAGE [32].

**Enrichment experiment of Cd^{2+} by the recombinant Pvfp-5-1 protein**

Cadmium solutions (50 and 100 μg/l) were prepared by dissolving cadmium chloride (CdCl₂) in double distilled H₂O (ddH₂O). A CdCl₂ concentration of 50 μg/ml (experimental groups 5A, 5B, and 5C) or 100 μg/ml (groups 10A, 10B, and 10C) was used. In each experiment group, 100 μl, 300 μl, or 500 μl of recombinant Pvfp-5-1 solution was added to 3 ml of CdCl₂ solution. In the corresponding control groups, the same volume of pET-32a was added to the CdCl₂ solution (Table 1). Cd^{2+} quantification was realized using inductively coupled plasma
mass spectrometry (ICP-MS) with a NexION 300X (PerkinElmer, Boston, MA, USA) for the calculations, following the manufacturer’s instructions. Each experiment was repeated three times. We used the Student’s t test for statistical analysis, where \( P < 0.05 \) was considered statistically significant.

**Results**

**Data summary for the high-throughput transcriptome sequencing and de novo assembly**

We sequenced a foot transcriptome of *P. viridis* (Fig 1A) and generated a total of 55,670,668 raw reads. After filtration, we subjected the 53,047,718 clean reads to subsequent de novo assembly using Trinity software. Finally, we obtained 73,571 unigenes. Lengths of the assembled unigenes ranged from 200 bp to 14,157 bp, with an average of 599 bp and an N50 of 794 bp (S3 Table).

**Functional annotation of the predicted unigenes**

BLASTX alignment (E-value \( \leq 1.0\text{e}^{-5} \)) was performed for these unigenes to search public protein databases. The results (S4 Table) indicate that within the total 73,571 unigenes, 29,973 were annotated against the Nr, 18,615 against the KEGG, 9,466 against the GO, 22,988 against the Swiss-Prot, and 6,721 against the Nt.

Based on the COG annotation, 8,834 unigenes were predicted and classified into 25 functional categories (S1 Fig). “General function prediction only” was the most popular group (19.72%), followed by “Replication, recombination and repair” (9.10%) and “Translation, ribosomal structure and biogenesis” (7.45%). For the GO annotation, 9,466 unigenes were assigned GO terms and categorized into 51 subcategories (S2 Fig) belonging to 3 main categories.

“Binding and catalytic activity” was the largest group in the category of molecular function. In the category of biological processes, “cellular process” was obviously the most dominant; however, in the cellular component, “cell part” was the largest representative. According to the KEGG annotation results, 18,615 unigenes were annotated and assigned to 241 KEGG pathways. The most common classifications include “metabolic pathway” (2,295 unigenes), “focal adhesion” (955 unigenes), “pathway in cancer” (852 unigenes), and “regulation of actin cytoskeleton” (838 unigenes). For the KEGG annotation, we observed that 955 unigenes were annotated in the focal adhesion pathway, which is related to the adhesive function of the byssus. Jointly, the annotations of GO terms and KEGG pathways provide a useful resource for further identification of specific cellular structures, pathways, processes, and protein functions in the Chinese green mussel.

In summary, we employed BLAST searches against the important public databases (Nr, Swissi-Prot, KEGG, GO, COG, and Nt) to show that a total of 31,710 assembled unigenes were annotated to known biological functions (see more details in S4 Table).

**Byssal proteins revealed by the LC-MS/MS analysis**

Proteomic analysis of the *P. viridis* byssus has previously been reported, but few byssal proteins were identified [33, 34]. In order to uncover the complexity of the byssus, we determined the byssal proteins on a more sensitive Prominence Nano-HPLC system coupled with Q-Exactive. After separation of the total byssal proteins using SDS-PAGE, we obtained 14 (named as S1–S14) and 17 (named as P1–P17) protein bands from the byssal thread and plaque, respectively (Fig 1C).
The total 31 protein bands were cut out individually and digested by trypsin for subsequent LC-MS/MS determination. The generated data were analyzed by Mascot software (v2.3.02) with the byssus-transcriptome-based protein database (i.e., translated from the transcriptome-based transcripts) as the reference for protein prediction. A total of 1,031 unique peptides were identified, and 187 protein sequences were predicted (S5 Table), in which 130 proteins matched with multiple peptides and 57 proteins matched with only one peptide. Interestingly, the numbers of peptides and proteins from the byssal thread are higher than those from the byssal plaque (S5, S6 and S7 Tables).

Detailed information about the identified foot proteins was listed in S6 and S7 Tables, including identified peptide sequences, unique peptide numbers, and protein coverage. The spectra of all unique peptides labeled with PDV software (https://github.com/wenbostar/PDV) are provided in S3 Fig; the precursor m/z, mass error, and expect value for each spectrum are presented in S8 Table.

We subsequently used the CD-HIT program [35] to remove redundant sequences, and we finally identified 187 protein sequences (S9 Table). Among these predicted proteins, 181 proteins showed only partial sequence similarity to known proteins, implying that most of these byssal proteins are novel. Many byssal proteins were only partially resolved in our present work, possibly due to their low abundance.

Among the identified 187 byssal protein sequences, 113 sequences were assigned to 79 KEGG pathways (S10 Table), in which “Focal adhesion” was the most common group (15.9%). To validate the accuracy of these predicted byssal protein sequences, we randomly picked 6 sequences for validation by RT-PCR (Fig 2) with subsequent Sanger sequencing.

Content and distribution of histidine and cysteine residues in byssal proteins

Histidine (His, H) and cysteine (Cys, C) residues play important roles in heavy metal binding peptides and/or proteins [36–38]. In particular, the metal binding properties make cysteine an important component of many proteins and a key catalytic component of enzymes [39]. As is well known, cysteine-rich metallothioneins (MTs) are important metal binding proteins, in which the Cys-Cys, Cys-X-X-Cys, and Cys-X-Cys motifs (X denotes any amino acid) are remarkable [36, 40, 41].

In our present work, through protein structural analysis, we observed that several byssal proteins are rich in histidine residues or cysteine residues or contain a cysteine-rich domain. A cysteine content of >10% and 5%–10% was found in 32 and 37 byssal proteins, respectively; the histidine content was mainly in the range of 1% to 5%, and one protein contained more than 10% (see more details in Fig 3). In the byssal proteins of our interest (i.e., Pvfp-2, -3, -5-1, -5-2, and -6), cysteine residues or Cys-X-Cys motifs are abundant (Table 2).

Foot proteins of P. viridis

Using known foot protein sequences from other mussels (such as Mefp1–Mefp6 from Mytilus edulis; downloaded from the NCBI database) as the queries to perform BLAST homology searches against our newly established transcriptome database and byssal protein database, we identified 7 foot protein sequences (named as Pvfp-1, -2, -3, -4, -5-1, -5-2, and -6 respectively; Tables 2 and 3) in P. viridis. Interestingly, Unigene22875_2A (Table 3) is similar to Mcfp-4 (from Mytilus californianus); hence, we renamed it Pvfp-4 (although the sequence is only partially available; Fig 4). Despite that only 2 foot protein sequences have been confirmed (Pvfp-4 and -6) in the public protein databases, we should pay attention to the low sequence homology between our predicted Pvfps and previously reported foot proteins from other mussels. The
significant species differences may be due to various environmental conditions, such as water temperature, salinity, water flow, and microbial influences [33, 43].

Other byssus proteins: Precollagen and tyrosinase in *P. viridis*

The byssus contains 3 peculiar collagen proteins, named preCol-NG, preCol-D, and preCol-P [44]. It was reported that preCol-D localizes to the stiff distal portion, preCol-P is present in the proximal portion, while preCol-NG is evenly distributed [45]. By homology searches against our proteome database, we identified 3 preCols (Table 3), among which preCol-P is novel. Homology was predominantly found in the conserved central domain with several pentapeptide repeat sequences, XGXPG, where X denotes a glycine or hydrophobic residue (red underlined in Fig 4); the glycine residues of the mature proteins are highly conserved between *P. viridis* and *Mytilus* species [44, 46]. Interestingly, these identified collagen proteins exhibited subtle but substantial species-specific modifications, compared with those from other mussels.
Fig 3. Content and distribution of histidine (H) and cysteine (C) residues in the byssal protein sequences of *P. viridis*. The x-axis represents the content of histidine (red) and cysteine (blue) in each protein. The y-axis represents the number of proteins.

https://doi.org/10.1371/journal.pone.0216605.g003

Table 2. Identified byssal proteins from the Chinese green mussel.

| Foot proteins | Search conditions | Identified foot protein sequences and their characteristics |
|---------------|-------------------|----------------------------------------------------------|
| Pvfp-1 (partial) | Homology sequences | MARNMNILTFLAVUMASASAVHPPSWTAKAPKWAKVPPFAMAWKHAPWATWAKMP |
| Pvfp-2 (partial) | Homology sequences | (CXXXPCXGXXGXXXXXXYXXCKXXGYTGXXCX) n High content of Y and C |
| Pvfp-3 | Homology sequences | MKCTLSISVFPVFAISMYVYVAQSLTCFTTDGFTIDGQNCGQSFCRDRNSPSYG6ECCNGNQCCPC6CTYG Cys content > 15% |
| Pvfp-4 (partial) | LC-MS/MS | HETFIA RMW = 34kDa; High content of Y and H |
| Pvfp-5-1 | Homology sequences | MLKFLAVVLCAFPYMYDDPDKFPRFPCVNGTCCRKSSYTCRXYGYGNCQ1NCSFPCQKNTGCCKLGGDKFKCYC KGYKGYKGYQGPDCTNDNGTCATAYLYLPGKYYKCMGYK |
| Pvfp-5-2 (byssal thread) | Homology sequences | MLKLVYVYVQYQYQYDNYLCPNCRYGTGCRSGLIFGYKFCITNGKCHQFNACTPNPCNGGCTCALKYQFPPYQC SCPGYGYTKCEFVRHYDRCGCCNGLCNDYSGYKRCVCPGYYGKRC1DPYY MW = 13.6kDa; C content 13%; Y content 14.3; 7% peptide coverage |
| Pvfp-6 (byssal plaque) | Homology sequences | MISAVCIYFFLVQIQAGYIPYEPKPGPCVTRQGIPCVVCP ENFECDFNSNCPGAMKCCDFGCGGKNKCPFPVPSLPQCYNYQ YYP1GASVDSNVGTCYCNDDTGMCTLKACGYYK MW = 11kDa; C content > 14%; 20.6% peptide coverage |

Note: Sequences of the identified byssal proteins were searched by BLAST homology from the foot transcriptome database. Underlined sequences are signal peptides. Bold area are mass spectrum-matched peptide sequences. Prediction of signal peptides was performed with SignalP 4.0 software [42]. MW, molecular weight; X, any amino acid other than cysteine.

https://doi.org/10.1371/journal.pone.0216605.t002
Tyrosinase, a copper-containing enzyme [47], can convert tyrosine into adhesive DOPA residues [48]. It has been recognized as a key component of byssal adhesion proteins [49]. By BLASTX homology searches against our transcriptome and proteome databases, we identified 5 tyrosinases (Table 3) from the transcriptome and proteome data. Homologous sequences of these tyrosinases are largely localized in the conserved active sites (comprising 7 histidine residues), which contain 2 copper binding sites, Cu(A) and Cu(B) [33, 50, 51]. Interestingly, tyrosinases have been reported to bind copper directly, and the Cu(A) and Cu(B) sites are both required to bind copper for catalytic activity [51].

**Accumulation of Cd\(^{2+}\) by the recombinant Pvf-p-5-1 protein**

Our previous studies demonstrated that the byssus can bind heavy metals effectively [20]. In order to examine the heavy metal enrichment ability of byssal proteins, we employed

![Fig 4. Comparison of partial preCol-P sequences between *P. viridis* and *Mytilus* species. Red underlined sequences are XGXPG repeats.](https://doi.org/10.1371/journal.pone.0216605.g004)
recombinant Pvfp-5-1 (159 mg/l) to study its binding to Cd^{2+}. Our results (Fig 5) show that the Cd^{2+} concentrations decreased significantly ($P < 0.05$) after addition of the purified recombinant Pvfp-5-1 protein to the initial solution. With increasing Pvfp-5-1 concentrations, the final Cd^{2+} concentration decreased. In summary, these data obviously proved the enrichment ability of our recombinant Pvfp-5-1 for heavy metals.

**Discussion**

The mussel byssus is composed of many byssal proteins, which present differences in function and biological activity. Several byssal proteins have been identified before, including foot proteins, precollagens, tyrosinases, and proximal thread matrix proteins [37, 46, 52, 53]. It was reported that different byssal proteins, with differential biological functions, make the byssus a valuable resource. For example, natural foot proteins from various Mytilius species have been used as a resource for underwater coatings and adhesives [33, 43, 54]. Interestingly, foot proteins (Fp-1–Fp-6) that presumably act as adhesives can also bind heavy metals [53, 55]. Hence, in the future, we may be able to design novel byssal-protein-based biomaterials to remove heavy metal pollution from aquatic environments. This is our main drive to examine the diversity of the byssal proteins in *P. viridis*, i.e., to deal with heavy metal pollution and radioactive waste from local factories.

Proteome sequencing is an efficient and widely used technique for identification of functional proteins. In this research, we combined proteome sequencing with transcriptome sequencing to construct a comprehensive library of *P. viridis* byssal proteins. Thousands of peptide fragments and 187 proteins were identified by LC-MS/MS. Six proteins had been reported before, and 181 are novel.

![Fig 5. Accumulation of Cd^{2+} by the recombinant Pvfp-5-1 protein.](https://doi.org/10.1371/journal.pone.0216605.g005)
Metal ions are essential for organisms, but excessive metal ions produce toxic effects. In the face of heavy metal stress, organisms protect themselves by various defense systems, such as synthesis of metal binding proteins or peptides. Histidine and cysteine residues play important roles in heavy metal binding proteins or peptides [38, 56]. In this study, we analyzed the content of cysteine and histidine in byssal proteins, and we observed that several novel byssal proteins are rich in histidine residues or cysteine residue or contain a cysteine-rich domain. For example, Antistasin-like protein (ALP, Unigene24116_2A; Fig 6A) is a novel protein in the byssus of *P. viridis*, containing internal repeats of a 30-aa sequence with a highly conserved pattern of 6 cysteine (Cys) and 2 glycine (Gly) residues; however, no similar sequences have been identified in other mussels. Over 20% of amino acids in the mature sequence of ALP are cysteine residues, with Cys-X-Cys and Cys-X-X-Cys motifs similar to MTs, indicating that this new protein may be able to bind metals.

Two more novel protein sequences (Unigene23933_2A and Unigene24349_2A; Table 3), with molecular weights of 35 kDa (30% peptide coverage) and 13 kDa (17% peptide coverage), respectively, have remarkably high contents of cysteine residues and homology with serine protease inhibitor like (SPI-like) protein and Oikosin-like protein, respectively. The mature peptide sequence of SPI-like protein contains 6 kazal domains of duplication (6 highly conserved cysteine residues, Fig 6B). The sequence of Oikosin-like protein (Unigene62001_2A) is rich in aspartic acid (11.9%) and histidine (12.4%) residues. It comprises 3 active Cu-Zn superoxide dismutase (SOD) domains of obvious sequence duplication (Fig 6C).

---

**Fig 6.** Sequence comparisons of several important byssal proteins. (a) antistasin-like protein (ALP). (b) SPI-like protein, which contains 6 repeated regions. (c) Oikosin-like protein; (d) Pernin precursor protein, which contains 3 repeated regions (Cu-Zn SODs in the red boxes). Note that the underlined regions are signal sequences. The cysteine (Cys, C) and Histidine (His, H) residues are highlighted in red and blue, respectively. Yellow areas are the identified peptides by LC-MS/MS.  

https://doi.org/10.1371/journal.pone.0216605.g006
Aspartic acid and histidine are known to participate in the binding of many metal cations [57]. The pernin precursor (Unigene62001_2A) has a high histidine content and contains 3 Cu-Zn SOD domains (Fig 6D), which might explain its remarkable metal binding capacity. Interestingly, our previous studies have confirmed that, under Cd stress conditions, expression of these byssal protein coding genes (including ALP, Pvfp-1, Pvfp-5-1, Pvfp-5-2, and Pvfp-6) are upregulated [20].

Mussel foot proteins have been applied in underwater experiments and for medicinal purposes. However, the process to extract byssal proteins from the mussel byssus is labor-intensive and inefficient, and approximately 10,000 mussels are required for isolation 1 mg of adhesive proteins [58]. E. coli can effectively be used for the expression of adhesive proteins, and the microscale assay showed purified recombinant Mgfp-5 has significant adhesive activity [59]. However, not all the foot proteins can be expressed by E. coli. For example, the recombinant Fp-1 protein has to be decoded in a yeast expression system [60, 61]. The failure in E. coli system may be due to the highly biased amino acid composition, the long amino acid sequence, or the different codon usage preference between the mussel and E. coli [62]. In this study, hence, we cloned and expressed recombinant Pvfp-5-1 with sequence modifications, and we confirmed that the newly recombinant Pvfp-5-1 has the capacity to bind Cd$^{2+}$ ions. Our results suggest that the recombinant Pvfp-5-1 could be developed into a commercial product for the removal of heavy metals and/or radioactive waste from aquatic environments.

Conclusions

In this study, we performed a combination of transcriptome and proteome sequencing to investigate protein components in the foot and byssus (threads and plaques) of the Chinese green mussel. By BLAST homology searches of known sequences from other mussel species against our generated transcriptome and proteome databases, we could rapidly predict and identify a collection of protein sequences in a high-throughput way. Since the mussel byssus has been proved to accumulate heavy metals effectively, we chose several byssal proteins that are rich in cysteine and/or tyrosine residues for structural analysis. Metal binding experiments were further performed to prove the Cd$^{2+}$ binding ability of recombinant Pvfp-5-1. In summary, we have established a valuable resource for the identification of more important proteins, engineering of more recombinant proteins, and development and processing of biomaterials for the removal of heavy metals and/or radioactive waste from aquatic environments.

Supporting information

S1 Fig. COG classification of all unigenes in the P. viridis transcriptome. (PDF)

S2 Fig. GO annotation of all unigenes in the P. viridis transcriptome. (PDF)

S3 Fig. The labeled spectra with MS identification information of all identified unique peptides. (PDF)

S1 Table. Nucleotide sequences of primer pairs for the RT-PCRs. (DOCX)

S2 Table. Nucleotide sequence of the modified Pvfp-5-1. (DOCX)
S3 Table. Summary of the assembled foot transcriptome of \textit{P. viridis}.

(SDOCX)

S4 Table. Statistics of functionally annotated unigenes in the foot of \textit{P. viridis}.

(SDOCX)

S5 Table. Summary of the proteome data from the byssal samples of \textit{P. viridis}.

(SDOCX)

S6 Table. Byssal thread proteins identified from \textit{P. viridis}.

(XLSB)

S7 Table. Byssal plaque proteins identified from \textit{P. viridis}.

(XLSB)

S8 Table. The precursor mass, mass error, and E-value of partial unique peptides from identified proteins.

(SDOCX)

S9 Table. Byssal protein sequences identified from \textit{P. viridis}.

(DOCX)

S10 Table. The KEGG pathway annotation of byssal proteins.

(XLSX)

Acknowledgments

We thank Chengye Yang and Jintu Wang, employees of BGI-Shenzhen, China, for their assistance in sample preparation and data analysis.

Author Contributions

Conceptualization: Xu Deng, Qiong Shi.

Data curation: Xinhui Zhang, Huiwei Huang, Yanbin He, Zhiqiang Ruan, Wanshun Li, Bo Wen, Zizheng Lu, Bing Liu.

Formal analysis: Huiwei Huang, Yanbin He, Zhiqiang Ruan, Xinxin You, Wanshun Li, Bo Wen, Zizheng Lu, Bing Liu.

Investigation: Xinhui Zhang.

Methodology: Xu Deng.

Supervision: Qiong Shi.

Writing – original draft: Xinhui Zhang, Qiong Shi.

Writing – review & editing: Xinxin You, Xu Deng, Qiong Shi.

References

1. Perez-Enciso M, Ferretti L. Massive parallel sequencing in animal genetics: wherefroms and wheretos. Anim Genet. 2010; 41(6):561–569. https://doi.org/10.1111/j.1365-2052.2010.02057.x PMID: 20477787

2. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10(1):57–63. https://doi.org/10.1038/nrg2484 PMID: 19015660

3. Suarez-Ulloa V, Fernandez-Tajes J, Manfrin C, Gerdol M, Venier P, Eirin-Lopez JM. Bivalve omics: state of the art and potential applications for the biomonitoring of harmful marine compounds. Mar Drugs. 2013; 11(11):4370–4389. https://doi.org/10.3390/md11114370 PMID: 24189277
4. Leung PT, Ip JC, Mak SS, Qiu JW, Lam PK, Wong CK, et al. De novo transcriptome analysis of Perna viridis highlights tissue-specific patterns for environmental studies. BMC Genomics. 2014; 15:804. https://doi.org/10.1186/1471-2164-15-804 PMID: 25239240

5. Casanova A, Carrascal M, Abian J, Lopez-Tejero MD, Llobera M. Discovery of lipoprotein lipase pl isoforms and contributions to their characterization. J proteomics. 2009; 72(6):1031–1039. https://doi.org/10.1016/j.jprot.2009.06.002 PMID: 19527804

6. Vergani L, Grattarola M, Grasselli E, Dondero F, Viarengo A. Molecular characterization and function analysis of MT-10 and MT-20 metallothionein isosforms from Mytilus galloprovincialis. Arch Biochem Biophys. 2007; 465(1):247–253. https://doi.org/10.1016/j.abb.2007.05.023 PMID: 17601485

7. Maltez HF, Villanueva Tagle M, Fernandez de la Campa Mdel R, Sanz-Medel A. Metal-metallothioneins like proteins investigation by heteroatom-tagged proteomics in two different snails as possible sentinel organisms of metal contamination in freshwater ecosystems. Anal Chim ACTA. 2009; 650(2):234–240. https://doi.org/10.1016/j.aca.2009.07.052 PMID: 19720198

8. Mosleh YY, Paris-Palacios S, Biaggiati-Risbourg S. Metallothioneins induction and antioxidative response in aquatic worms Tubifex tubifex (Oligochaeta, Tubificidae) exposed to copper. Chemosphere. 2006; 64(1):121–128. https://doi.org/10.1016/j.chemosphere.2005.04.045 PMID: 16330073

9. Gin KY, Tang YZ, Aziz MA. Derivation and application of a new model for heavy metal biosorption by algae. Water Res. 2002; 36(5):1313–1323. PMID: 11902786.

10. Kostal J, Yang R, Wu CH, Mulchandani A, Chen W. Enhanced arsenic accumulation in engineered bacterial cells expressing ArsR. Appl Environ Microb. 2004; 70(8):4582–4587. https://doi.org/10.1128/AEM.70.8.4582–4587.2004

11. Livingstone DR, Chipman JK, Lowe DM, Minieri JN, Waite JH. Development of biomarkers to detect the effects of organic pollution on aquatic invertebrates: recent molecular, genotoxic, cellular and immunological studies on the common mussel (Mytilus edulis L.) and other mytilids. Int J Environ Pollut. 2000; 13(1–6):56–91. https://doi.org/10.1504/IJEP.2000.002311

12. Nicholson S, Lam PK. Pollution monitoring in Southeast Asia using biomarkers in the mytilid mussel Perna viridis (Mytilidae: Bivalvia). Environ Int. 2005; 31(1):121–32. https://doi.org/10.1016/j.envint.2004.05.007 PMID: 15607786

13. Pinto R, Acosta V, Segnini MI, Brito L, Martinez G. Temporal variations of heavy metals levels in Perna viridis, on the Chacopata-Bocaripo lagoon axis, Sucre State, Venezuela. Mar Pollut Bull. 2015; 91 (2):418–423. https://doi.org/10.1016/j.marpolbul.2014.09.059 PMID: 25444616

14. Ninan L, Monahan J, Stroshine RL, Wilker JJ, Shi R. Adhesive strength of marine mussel extracts on porcine skin. Biomaterials. 2003; 24(22):4091–4099. PMID: 12834605

15. Lee BP, Messersmith PB, Israilachvili JN, Waite JH. Mussel-Inspired Adhesives and Coatings. Annu Rev Mater Res. 2011; 41:99–132. https://doi.org/10.1146/annurev-matsci-062910-100429 PMID: 22058660

16. Holten-Andersen N, Waite JH. Mussel-designed protective coatings for compliant substrates. J Dent Res. 2008; 87(7):701–709. https://doi.org/10.1177/154082210800700808 PMID: 18650539

17. Holten-Andersen N, Fantner GE, Hohlbauch S, Waite JH, Zok FW. Protective coatings on extensible biofibres. Nat Mater. 2007; 6(9):669–672. https://doi.org/10.1038/nmat1956 PMID: 17618290

18. Hennebert E, Wattiez R, Waite JH, Flammang P. Characterization of the protein fraction of the temporary adhesive secreted by the tube feet of the sea star Asterias rubens. Biofouling. 2012; 28(2):409–4099. PMID: 12834605

19. Lee BP, Messersmith PB, Israilachvili JN, Waite JH. Mussel-Inspired Adhesives and Coatings. Annu Rev Mater Res. 2011; 41:99–132. https://doi.org/10.1146/annurev-matsci-062910-100429 PMID: 22058660

20. Yap C, Ismail A, Tan S, Omar H. Accumulation, depuration and distribution of cadmium and zinc in the green-lipped mussel Perna viridis (Linnaeus) under laboratory conditions. Hydrobiologia. 2003; 498 (1):151–160. https://doi.org/10.1023/A:1026221930811

21. Zhang X, Ruan Z, You X, Wang J, Chen J, Peng C, et al. De novo assembly and comparative transcriptome analysis of the foot from Chinese green mussel (Perna viridis) in response to cadmium stimulation. PloS one. 2017; 12(5):e0176777. https://doi.org/10.1371/journal.pone.0176777 PMID: 28520756

22. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat protoc. 2013; 8(8):1494–1512. https://doi.org/10.1038/nprot.2013.084 PMID: 23845962

23. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, et al. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 2006; 34(Web Server issue):W293–W297. https://doi.org/10.1093/nar/gkl031 PMID: 16845012
24. Iseli C, Jongeneel CV, Bucher P. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proceedings for International Conference on Intelligent Systems for Molecular Biology; ISMB International Conference on Intelligent Systems for Molecular Biology. 1999:138–148. https://doi.org/10.1016/j.cell.2009.01.041 PMID: 19345192

25. Xu P, Duong DM, Seyfried NT, Cheng D, Xie Y, Robert J, et al. Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. Cell. 2009; 137(1):133–145. https://doi.org/10.1016/j.cell.2009.01.041 PMID: 19950988

26. Song C, Ye M, Han G, Jiang X, Wang F, Yu Z, et al. Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. Anal Chem. 2009; 82(1):53–56. https://doi.org/10.1021/ac9023044 PMID: 19950988

27. Wen B, Zhou R, Feng Q, Wang Q, Wang J, Liu S. IQuant: an automated pipeline for quantitative proteomics based upon isobaric tags. Proteomics. 2014; 14(20):2280–2285. https://doi.org/10.1002/pmic.201300361 PMID: 25069810

28. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence database using mass spectrometry data.pdf. Electrophoresis. 1999; 20(18):3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-EPLPS3551>3.0.CO;2-2 PMID: 10612281

29. Feng J, Naiman DQ, Cooper B. Probability-based pattern recognition and statistical framework for randomization: modeling tandem mass spectrum/peptide sequence false match frequencies. Bioinformatics. 2007; 23(17):2210–2217. https://doi.org/10.1093/bioinformatics/btm267 PMID: 17510167

30. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods. 2007; 4(3):207–214. https://doi.org/10.1038/nmeth1019 PMID: 17327847

31. Grote A, Hiller K, Scheer M, Munch R, Nortemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res. 2005; 33(Web Server issue):W526–W531. https://doi.org/10.1093/nar/gki376 PMID: 15980527

32. Kinoshita-Kikuta E, Kinoshita E, Koike T. Neutral Phosphate-Affinity SDS-PAGE System for Profiling of Protein Phosphorylation. In: Posch A, editor. Proteomic Profiling: Methods and Protocols. New York, NY: Springer New York; 2015. p. 323–354.

33. Guerette PA, Hoon S, Seow Y, Raida M, Masic A, Wong FT, et al. Accelerating the design of biomimetic materials by integrating RNA-seq with proteomics and materials science. Nat Biotechnol. 2013; 31(10):908–915. https://doi.org/10.1038/nbt.2671 PMID: 24031986

34. Qin C, Pan Q, Qi Q, Fan M, Sun J, Li N, et al. In-depth proteomic analysis of the byssus from marine mussel Mytilus coruscus. Journal of Proteomics. 2016; 144(Supplement C):87–98. https://doi.org/10.1016/j.jprot.2016.06.014

35. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012; 28(23):3150–3152. https://doi.org/10.1093/bioinformatics/bts565 PMID: 23060610

36. Mejaře M, Bűlow L. Metal-binding proteins and peptides in bioremediation and phytoremediation of heavy metals. Trends Biotechnol. 2001; 19(2):67–73. PMID: 11164556.

37. Quig D. Cysteine metabolism and metal toxicity. Altern Med Rev. 1998; 3:262–270. PMID: 9727078.

38. Hara M, Fujinaga M, Kuboi T. Metal binding by citrus dehydrin with histidine-rich domains. J Exp Bot. 2005; 56(420):2695–2703. https://doi.org/10.1093/jxb/eri262 PMID: 16131509

39. Giles NM, Watts AB, Giles GI, Fry FH, Littlechild JA, Jacob C. Metal and redox modulation of cysteine protein function. Chem Biol. 2003; 10(8):677–693. PMID: 12954327.

40. Cobbett C, Goldsborough P. Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. Annu Rev Plant Biol. 2002; 53:159–182. https://doi.org/10.1146/annurev.arplant.53.100301.135154 PMID: 12221971

41. Hamer DH. Metallothionein. Annu Rev Biochem. 1986; 55(1):913–951. https://doi.org/10.1146/annurev.bi.55.070186.004405 PMID: 3527054

42. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011; 8(10):785–786. https://doi.org/10.1038/nmeth.1701 PMID: 21959131

43. Lu Q, Danner E, Waite JH, Israelachvili JN, Zeng H, Hwang DS. Adhesion of mussel foot proteins to different substrate surfaces. J R Soc Interface. 2013; 10(79):20120759. https://doi.org/10.1098/rsif.2012.0759 PMID: 23173195

44. Waite JH, Qin X-X, Coyne KJ. The peculiar collagens of mussel byssus. Matrix Biol. 1998; 17(2):93–106. PMID: 9694590.
45. Qin X-X, Coyne KJ, Waite JH. Tough tendons mussel byssus has collagen with silk-like domains. J Biol Chem. 1997; 272(51):32623–32627. PMID: 9405478.

46. Coyne KJ. Extensible collagen in mussel byssus: A natural block copolymer. Science. 1997; 277(5333):1830–1832. https://doi.org/10.1126/science.277.5333.1830 PMID: 9292575

47. Aguilera F, McDougall C, Degnan BM. Evolution of the tyrosinase gene family in bivalve molluscs: independent expansion of the mantle gene repertoire. Acta Biomater. 2014; 10(9):644–652. https://doi.org/10.1016/j.actbio.2014.03.031 PMID: 24704693

48. Sanchez-Ferrer A, Rodriguez-Lopez JN, Garcia-Canovas F, Garcia-Carmona F. Tyrosinase: a comprehensive review of its mechanism. Bioch bioph Acta. 1995; 1247(1):1–11. PMID: 7873577.

49. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011; 29(7):644–652. https://doi.org/10.1038/nbt.1883 PMID: 21572440

50. Goldfeder M, Kanteev M, Isaschar-Ovdat S, Adir N, Fishman A. Determination of tyrosinase substrate-binding modes reveals mechanistic differences between type-3 copper proteins. Nat Commun. 2014; 5:4505. https://doi.org/10.1038/ncomms5505 PMID: 25074014

51. Spritz RA, Ho L, Furumura M, Hearing VJ Jr. Mutational analysis of copper binding by human tyrosinase. J Invest Dermatol. 1997; 109(2):207–212. PMID: 9242509.

52. Suhre MH, Gertz M, Steegborn C, Scheibel T. Structural and functional features of a collagen-binding matrix protein from the mussel byssus. Nat Commun. 2014; 5:3392. https://doi.org/10.1038/ncomms4392 PMID: 24569701

53. Waite JH. Adhesion a la moule. Integr Comp Biol. 2002; 42(6):1172–1180. https://doi.org/10.1093/icb/42.6.1172 PMID: 21680402

54. Lin Q, Gourdon D, Sun C, Holten-Andersen N, Anderson TH, Waite JH, et al. Adhesion mechanisms of the mussel foot proteins mfp-1 and mfp-3. P Natl Acad Sci USA. 2007; 104(10):3782–3786. https://doi.org/10.1073/pnas.0607852104 PMID: 17360430

55. Hedlund J, Andersson M, Fant C, Bitton R, Bianco-Peled H, Elwing H, et al. Change of colloidal and surface properties of Mytilus edulis foot protein 1 in the presence of an oxidation (NaIO4) or a complex-binding (Cu2+) agent. Biomacromolecules. 2009; 10(4):845–849. https://doi.org/10.1021/bm801325j PMID: 19209903

56. Hempe JM, Cousins RJ. Cysteine-rich intestinal protein binds zinc during transmucosal zinc transport. P Natl Acad Sci USA. 1991; 88(21):9671–9674. https://doi.org/10.1073/pnas.88.21.9671 PMID: 1946385.

57. Scotti PD, Dearing SC, Greenwood DR, Newcomb RD. Pernin: a novel, self-aggregating haemolymph protein from the New Zealand green-lipped mussel, Perna canaliculus (Bivalvia: Mytilidae). Comp Biochem Physiol B Biochem Mol Biol. 2001; 128(4):767–779. https://doi.org/10.1016/s1096-4959(01)00301-3 PMID: 11290459

58. Morgan D. Two firms race to derive profits from mussels glue: despite gaps in their knowledge of how the mollusk produces the adhesive, scientists hope to recreate it. Scientist. 1990; 4:1.

59. Hwang DS, Yoo HJ, Jun JH, Moon WK, Cha HJ. Expression of functional recombinant mussel adhesive protein Mgf1-5 in Escherichia coli. Appl Environ Microb. 2004; 70(6):3352–3359. https://doi.org/10.1128/AEM.70.6.3352–3359.2004

60. Filipula DR, Lee SM, Link RP, Strausberg SL, Strausberg RL. Structural and functional repetition in a marine mussel adhesive protein. Biotechnol Progr. 1990; 6(3):171–177. https://doi.org/10.1021/bp00003a001 PMID: 1367451

61. Salerno AJ, Goldberg I. Cloning, expression, and characterization of a synthetic analog to the bioadhesive precursor protein of the sea mussel Mytilus edulis. Appl Microbiol Biot. 1993; 39(2):221–226. PMID: 7763730.

62. Kitamura M, Kawakami K, Nakamura N, Tsumoto K, Uchiyama H, Ueda Y, et al. Expression of a model peptide of a marine mussel adhesive protein in Escherichia coli and characterization of its structural and functional properties. J Polym Sci A Pol Chem. 1999; 37(6):729–736.