Mar Biotechnol (2011) 13:810–819
DOI 10.1007/s10126-010-9343-6

ORIGINAL ARTICLE

Occurrence of a Silicatein Gene in Glass Sponges (Hexactinellida: Porifera)

Galina N. Veremeichik · Yuri N. Shkryl · Victor P. Bulgakov · Sergey V. Shedko · Valery B. Kozhemyako · Svetlana N. Kovalchuk · Vladimir B. Krasokhin · Yuri N. Zhuravlev · Yuri N. Kulchin

Received: 29 January 2010 / Accepted: 3 October 2010 / Published online: 23 December 2010

Abstract Silicatein genes are involved in spicule formation in demosponges (Demospongiae: Porifera). However, numerous attempts to isolate silicatein genes from glass sponges (Hexactinellida: Porifera) resulted in a limited success. In the present investigation, we performed analysis of potential silicatein/cathepsin transcripts in three different species of glass sponges (*Pheronema raphanus*, *Aulosaccus schulzei*, and *Bathydorus levis*). In total, 472 clones of such transcripts have been analyzed. Most of them represent cathepsin transcripts and only three clones have been found to represent transcripts, which can be related to silicateins. Silicatein transcripts were identified in *A. schulzei* (Hexactinellida; Lyssacinosida; Rosselidae), and the corresponding gene was called *AuSil-Hexa*. Expression of *AuSil-Hexa* in *A. schulzei* was confirmed by real-time PCR. Comparative sequence analysis indicates high sequence identity of the *A. schulzei* silicatein with demosponge silicateins described previously. A phylogenetic analysis indicates that the AuSil-Hexa protein belongs to silicateins. However, the AuSil-Hexa protein contains a catalytic cysteine instead of the conventional serine.

Keywords Silicatein · Cathepsin · Sponge · Gene expression · Biosilification · Hexactinellida

Electronic supplementary material The online version of this article (doi:10.1007/s10126-010-9343-6) contains supplementary material, which is available to authorized users.

G. N. Veremeichik · Y. N. Shkryl · V. P. Bulgakov (✉) · S. V. Shedko · Y. N. Zhuravlev
Institute of Biology and Soil Science,
Far East Branch of Russian Academy of Sciences,
Vladivostok, Russia
e-mail: bulgakov@ibss.dvo.ru

V. B. Kozhemyako · S. N. Kovalchuk · V. B. Krasokhin
Pacific Institute of Bioorganic Chemistry,
Far East Branch of Russian Academy of Sciences,
Vladivostok, Russia

Y. N. Kulchin
Institute for Automation and Control Processes,
Far East Branch of Russian Academy of Science,
Vladivostok, Russia

Introduction

Demosponges (Demospongiae: Porifera) and hexactinellids (Hexactinellida: Porifera) possess network-based skeletons as well as spicules (macro- and microscleres) comprised mainly of amorphous hydrated silica (Uriz et al. 2003). Silicatein genes are involved in spicule formation in Demospongiae (e.g., see for review Schröder et al. 2008). Of late years, the study of silicateins has taken much attention because of their potential use in bionanotechnology and materials science (Brutchey and Morse 2008). An analysis of sequence similarity revealed that silicatein α is a member of the cathepsin L subfamily of papain-like cysteine proteases (Shimizu et al. 1998). Silicateins seem to represent a group of enzymes possessing bi-functional activity; in addition to the silica-condensing activity, they possess a proteolytic (cathepsin-like) activity (Schröder et al. 2008). Recently, Fairhead et al. (2008) investigated mutations that could confer a silica-condensing activity on cathepsin L. These authors showed that replacing the catalytic cysteine with a serine and replacing adjacent amino acids (SSW→ASY; the catalytic serine is underlined) in the engineered cathepsin created a sufficient pocket to allow recognition of a $Si(OH)_4$ molecule. The engineered cathepsin possesses high silica-condensing activity (Fairhead et al. 2008). Although the role of silicateins in silica polymerization is well-documented in vitro (Schröder et al. 2008), doubt has been expressed as to

whether silicateins are responsible for the formation of silica-based structures in all sponges (Ehrlich and Worch 2007; Heinemann et al. 2007; Ehrlich et al. 2010). It has been suggested that silicateins resemble cathepsins, which are known to be collagenolytic and capable of attacking the triple helix of fibrillar collagens. The occurrence of silicatein genes in non-spicule-forming demosponges favors this hypothesis (Kozhemyako et al. 2010).

Glass sponges (Hexactinellida: Porifera) are the most ancient extant metazoans, and they represent probably the oldest lineage of animals alive on earth today (Krautter 2002). Hexactinellids are especially important for biomimetic studies because only these sponges form a meter long and high flexible anchoring spicules. Therefore, one of the principal question in biochemistry of Porifera is whether or not glass sponges contain conventional silicateins. Whereas silicateins have been isolated from many demosponges (Schröder et al. 2008), their isolation from glass sponges has offered a difficult problem.

At the same time, several cathepsin genes were isolated from the *Aphrocallistes vastus* and *C. meyeri* glass sponges (Müller et al. 2008b). This could suggest that glass sponges might not contain conventional silicatein genes, and thus the process of spiculogenesis in glass sponges might not be based only on the presence of silicateins. Indeed, Ehrlich and coworkers reported recently about finding of collagen (Ehrlich et al. 2006; Ehrlich and Worch 2007; Ehrlich et al. 2010) as well as chitin (Ehrlich et al. 2007) within skeletal formations of different Hexactinellida species. Both biopolymers showed well silicification activity in experiments, which has been carried out in vitro.

Müller et al. (2008a) reported the identification of a silicatein-related protease in the giant spicules of the glass sponge *Monorhaphis chuni*. This 27 kDa protein reacted with polyclonal antibodies raised against *Suberites domuncula* silicatein and showed a proteolytic activity. In a subsequent publication, the authors isolated a 24 kDa protein from spicules of another hexactinellid *Crateromorpha meyeri*, which reacted with anti-silicatein antibodies in western blots (Müller et al. 2008b). The corresponding cDNA was isolated, and a partial amino acid sequence of the protein named SILCA_CRAME was deduced (Müller et al. 2008b).

Our analysis of silicatein/cathepsin cDNAs in western Pacific demosponges (*Latrunculia oparinae*) has shown the presence of conventional silicateins genes *LoSilA1*, *LoSilA1a*, *LoSilA2*, and *LoSilA3* (silicatein-α group), *LoSilB* (silicatein-β group), as well as one cathepsin gene, *LoCath* (Kozhemyako et al. 2009). *LoCath* was shown to represent an intermediate form between silicatein and cathepsin genes in demosponges (Kozhemyako et al. 2009). The aim of the present study was to investigate in detail the structure of silicatein and cathepsin gene transcripts in the following glass sponges species: *Pheronema raphanus* (Hexactinellida: Amphidiscosida: Pheronematidae), *Aulosaccus schulzei*, and *Bathydorus levis* (Hexactinellida; Lyssacinosida; Rosselidae).

Materials and Methods

Sponges

Marine sponges were collected during expeditions on the research vessel "Akademik Oparin". Samples of the marine sponge *P. raphanus* (Hexactinellida: Amphidiscosida: Pheronematidae) were collected in South China Sea (Ladd reef; 08°41.2 N, 111°42.3 E, depth 300 m, May 2007). The body of the sponge is hemispherical (15 cm in diameter) with long bundle spicules from sponge basal to substrate. The composition and size of macroscleres and microscleres corresponded to the first description by Schulze, 1894. The voucher sample (PIBOC O34-156) is on deposit in the Marine Sponges Collection at the Pacific Institute of Bioorganic Chemistry, Vladivostok, Russia. Specimens of the marine sponge *A. schulzei* (Hexactinellida: Lyssacinosa: Rossellidae; PIBOC O36-113) were collected by dredging near Urup Island (Kuril Islands, Okhotsk Sea, 45°40.2 N, 150°05.4 E, depth 611 m) in August 2008. The sponge has a vase or barrel form with thick walls (2.5–5 cm) and up to 20 cm in height. The surface is tubercular. The sponge has an unusual thick (200 μm) spindle form diactines up to 3 cm length. Hypodermal pentactines with the proximal ray up to 3 mm length. Microscleres are discohexasters and oxehexasters with high reduction of main spines rays. Specimens of the marine sponge *B. levis* (Hexactinellida: Lyssacinosa: Rossellidae) (PIBOC O36-213) were collected by dredging in the same place (Urup Isl., Kurile Isl., Okhotsk Sea, 46°16,3 N, 150°15,6 E, depth 450 m) in August 2008. The sponge body is saccular with thin walls (1–3 mm). The external and internal surfaces are smooth with rare hairiness of separate long spicules (1 cm). The diactines with sharp ends of main skeleton are located to parallel to surface (up to 15 mm length). Hypodermal pentactines with proximal ray are up to 2 mm length. Microscleres are presented in only one type, oxehexasters with main rays of 6 μm. These rays are to ramify into two to three roughness and lightly curved branches.

RNA Isolation and cDNA Synthesis

Total RNA was isolated from 0.3 g of the apical part of the sponges immediately after collection (onboard of research vessel) by the guanidine thiocyanate/phenol/chloroform method (Chomczynski and Sacchi 1987). cDNAs were synthesized from total RNA and subjected to 25 cycles of

amplification (10 s at 95°C, 20 s at 63°C, and 90 s at 72°C) using a SMART cDNA Amplification Kit according to the manufacturer's instructions (Clontech, USA).

Gene Cloning and Sequencing

Polymerase chain reaction (PCR) with degenerate primers was used to isolate DNA fragments of silicatein and cathepsins from sponge cDNA. cDNAs (produced using the SMART kit, above) were diluted 25-fold with water, and 1 μl of this solution was used as the PCR template. Degenerate PCR primers targeting conserved regions within silicatein/cathepsin proteins were developed based on alignment of the following sequences from the GenBank database at NCBI: *Petrosia ficiformis* (GenBank accession no. AAO23671), *Tethya aurantia* (α isoform; AAC23951), *T. aurantia* (β isoform; AAF21819), *Hymeniacidon perlevis* (α isoform; ABC94586), *Lubomirskia baicalensis* (α1 isoform; CAI43319), *L. baicalensis* (α2 isoform; CAI91571), *L. baicalensis* (α3 isoform; CAI91572), *L. baicalensis* (α4 isoform; CAI91573), *S. domuncula* (α isoform; CAI46305), *S. domuncula* (β isoform; CAH04635) and *Ephydatia fluviatilis* (M1 isoform; BAE54434), as well as cathepsins of *H. perlevis* (ABB91778), *L. baicalensis* (CAI46307), and *S. domuncula* (CAH04632). Table 1 lists degenerate PCR primers used in this study. PCR conditions were as follows: an initial denaturation at 96°C for 3 min, 40 amplification cycles consisting of denaturation at 96°C for 20 s, annealing using a temperature gradient from 51°C to 55°C (0.1°C increase every cycle) for 30 s, extension at 72°C for 30 s and a final extension at 72°C for 10 min.

The resulting DNA fragments were separated by electrophoresis and isolated from low melt agarose gel (Sigma) using standard phenol extraction. Purified fragments were cloned into the pTZ57R/T cloning vector (Fermentas, Lithuania) and sequenced using an ABI 3130 Genetic Analyzer (Applied Biosystems, USA) and a Big Dye Terminator Cycle Sequencing Kit (Applied Biosystems).

To obtain full-length sequences, we performed rapid amplification of cDNA ends (RACE) using step-out PCR technology according to Matz et al. (1999). Reverse gene-specific primers were designed to perform 5′ RACE and forward gene-specific primers were used to perform 3′ RACE of the *AuSil-Hexa*, *PhrCat1*, and *PhrCat2* (Table 1). The fragments obtained by RACE were cloned into the pTZ57R/T and sequenced.

Quantitative Real-time RT-PCR Analysis

To perform quantitative real-time RT-PCR (qPCR) analysis of silicatein and cathepsin genes expression, two cDNA samples were independently synthesized from total RNA isolated from *A. schulzei* as described above. RNA was treated with 1 U of RNase-free DNase I (Sileks M, Russia) according to the manufacturer's protocol. Following digestion, DNase was removed from the solution with sorbent BlueSorb (Sileks M) and RNA was precipitated with 2 volumes of 96% ethanol. The resulting RNA pellet was washed with ice-cold 70% ethanol and resuspended in 20 μl of DEPC-treated water. RNA concentration and 28S/18S ratio were determined using an RNA StdSens LabChip® kit and Experion™ Automated Electrophoresis Station (Bio-Rad Laboratories, Inc., USA) with Experion™ Software System Operation and Data Analysis Tools (version 3.0) following the manufacturer's protocol and recommendations. The samples with 28S/18S ribosomal RNA between 1.5 and 2.0 and an RNA Quality Indicator (RQI) above 9.0 were used for real-time PCR analysis.

The first strand of cDNA was synthesized from 2.5 μg of RNA using 0.5 ng of oligo-d$(T)_{15}$ primer. Solutions containing the RNA and primer were preheated (5 min at 72°C) and cooled on ice. Reverse transcription was performed in a 50 μl volume, containing 1X M-MLV buffer, 0.24 mM dNTP mix and 200 U M-MLV reverse transcriptase (Sileks M). The reaction was carried out for 1 h at 36°C followed by 10 min at 72°C. The cDNAs were diluted 1:5 with nuclease-free water. As an additional negative control, we performed RT-RCR reactions with RNA sample without adding the M-MLV enzyme (RNA-RT control).

The qPCR analysis was performed using the Bio-Rad CFX96 Real-Time System (Bio-Rad Laboratories, Inc., USA) with 2.5× SYBR green PCR master mix containing ROX as a passive reference dye (Syntol, Russia). Reactions were done in a 25 μl volume containing 300 nM of each primer, 1 μl of the diluted cDNA sample and 2.5 mM $MgCl_2$. All PCR reactions were performed under the following conditions: 5 min at 95°C, followed by 35 cycles of 10 s at 95°C and 30 s at 60°C in a 96-well reaction plate. Two biological replicates were used for analysis and three technical replicates were analyzed for each biological replicate. No-template controls and RNA-RT controls were included in the analysis to verify the absence of contamination. The absence of non-specific products or primer-dimer artefacts in the samples was confirmed by melting curve analysis at the end of each run and by product visualization using electrophoresis on a 1% agarose gel stained with ethidium bromide.

The gene-specific primer pairs used in the qPCR are listed in Table 1. Data were analyzed using CFX Manager Software (Version 1.5) (Bio-Rad Laboratories).

Molecular Phylogenetic Analysis

Sequences of sponge silicateins and cathepsins were retrieved from GenBank using BLASTP (Altschul et al.

Table 1 Sequences of degenerate and gene-specific primers used in this study

Target genes	Forward primer (5′-3′)	Reverse primer (5′-3′)	Expected size (bp)
Degenerate PCR primers targeting conserved regions within silicatein/cathepsin proteins			
Silicatein/ Cathepsin	GGNTTYACNYTNGCNATGAA (GFTLAM[a]) —"F1"	CCANSWRTTYTTNACNARCCAR (WLVKNSW[a])—"R"	708
Silicatein/ Cathepsin	GAYTGGMGNACNAARGGNGC (DWRTKG[a]) —"F2"		549
Silicatein	TGYGGNGCNAGYTAYGCNTT (CGASYA[a]) —"F3"		495
Degenerate PCR primers targeting distinctive regions within *AuSil-Hexa*			
AuSil-Hexa	GGNWSNGTNAARAAYCARTGG (GSVKNQW[a])	TARTTYTTNCCYTGRTANGC (AYQGKNY[a])	504
Gene-specific primers for 3′ and 5′ RACE PCR			
AuSil-Hexa	ATCAAGAGTGGGGACGAG	TCCCCCTGTAGTTATCAC	
	CAATCGCTGTGGCTGTGG	ATGAAAGAGTTGTAGTGAT	
	CACCTCCACTAACCTCGC	CGCCTTCCAATGCTCCCAC	
PhrCat1	CGCTTTTAGATACTGGGAA	TCCCAGTATCTAAAAGCGT	
	CAAGATGCTGTAAGAACTAT	AGTCCTCCTTTGCAGCCATT	
	CGGTGTTCTTGTTGTTGGAT	TATCTTCTTGAGCAGTCG	
PhrCat2	GCCAGTCATAATAGCTTCC	CCATCACAGCCTTCATTTC	
	GTATCTATGACCCCAAGCG	GCACTTCGCATCTCTTCCTA	
	GTGTTAGTAGTGGGGTATG	GTATTTTCTTGTAACCAG	
Gene-specific primers for real-time PCR			
AuSil-Hexa	GCTACTGCTCTGGCTAAGGG	ATGAAAGAGTTGTAGTGAT	116
AuCat1	GGATGGAAGTTGTCACGCTG	CGGTATGGCGAGAGGTGGG	156
AuCat2	CGTGTGTGGGTCCTGCTAC	CCATCCGCCAGAGCAGTCCC	148

[a] Conserved amino acid motives used to design the degenerate primers

1997) in a search carried out using the amino acid sequence of AuSil-Hexa from *A. schulzei* against the non-redundant protein database. GenBank sequences and sequences described in this work (in total 61) were aligned using the MAFFT (Katoh et al. 2002) program (v.6.708) with default settings. The multiple sequence alignment was manually edited to remove redundant sequences. Only mature protein sequences were used. Phylogenetic trees were generated using the Bayesian (BA), the maximum likelihood (ML), and the maximum parsimony (MP) methods of tree construction. Bayesian Markov chain Monte Carlo analysis was conducted with the parallel version of MrBayes v.3.2 (Ronquist and Huelsenbeck 2003). Prior to the analysis, the ProtTest program (Abascal et al. 2005) was used to determine the most appropriate amino acid substitutions that best fit the data. ProtTest v.2.2 tests for 112 amino acid substitution models. The best substitution matrix suggested by ProtTest was WAG+I+G, and this matrix was applied during the Bayesian analysis. A Bayesian search was run for 3,000,000 generations, with 6 chains, sampling every 500 generations. The distribution of log likelihood scores was examined to determine stationarity and burn-in time for the search. Stationarity in log likelihood scores was observed after approximately 60,000–80,000 generations. Trees from the first 500,000 generations (1,000 trees) were discarded to ensure that all burn-in trees were excluded. This analysis left 5,000 trees, which were then used to construct the 50% majority-rule consensus. Branch support for each clade was based on posterior probability values, indicated by the frequency of occurrence of each clade among the trees retained after the initial burn-in topologies were discarded.

For maximum likelihood analysis, the Phyml program (v 3.0) was used (Guindon and Gascuel 2003) with the best substitution model (WAG+I+G) suggested by ProtTest. We used a heuristic search strategy of 40 random addition series, each followed by SPR branch swapping. Two hundred bootstrap cycles were carried out to obtain statistical support for the phylogeny. Parsimony analysis was conducted in PAUP*4.0b10 (Swofford 2002) using a heuristic search strategy of 100 random addition series, each followed by TBR branch swapping. Analyses were conducted with a stepmatrix derived from the Blosum62 amino acid transition matrix (downloaded from the PAUP website). Bootstrap branch support was estimated in PAUP using a stepmatrix and an abbreviated search strategy of 10

random additions+TBR per each of the 400 bootstrap replicates. In all of these analyses, cathepsin from *S. domuncula* (GenBank accession number CAH04633) was used as the out-group.

Results and Discussion

Analysis of Silicatein and Cathepsin Transcripts

To analyze silicatein and cathepsin transcripts in *A. schulzei*, we designed degenerate primers corresponding to conserved parts of known silicatein and cathepsin genes. Three primer sets (F1-R, F2-R, and F3-R) were used in PCR reactions with amplified cDNA from *A. schulzei* (Table 1). Fragments of DNA corresponding to the predicted lengths were amplified using the degenerate primers and cloned separately into pTZ57R/T. A total of 201 cDNA clones were sequenced. A silicatein gene and two cathepsin genes were identified using the BLAST algorithm. These genes were designated as *AuSil-Hexa, AuCath1* (GenBank accession number ***GQ387055***), and *AuCath2* (GenBank accession number ***GQ387056***).

The results of analysis of the number and percentage of clone distribution among the sequenced genes are presented in Table 2. All sets of degenerate primers revealed the *A. schulzei* cathepsin genes. *AuSil-Hexa* clones were found only in the F3-R-derived sequences. It was not possible to reveal the S/C catalytic amino acid in the AuSil-Hexa sequence by using the S primer because it was designed against the conserved GASYAF domain of silicatein proteins. To recover the full-length cDNA of the silicatein gene, RACE PCR was performed. The complete sequence of the *AuSil-Hexa* transcript was determined and submitted to the GenBank under accession number ***GQ387054***.

Alignment of the deduced AuSil-Hexa amino acid sequence with silicatein and cathepsin proteins is presented in Fig. 1. The sequences corresponding to mature proteins are included there. It was found that mature AuSil-Hexa shares the highest amino acid sequence identity with *Latrunculia oparinae* silicatein-A3 and *S. domuncula* silicatein α-isoform (CAI46305) (both 65% identity). The silicatein of *A. schulzei* also displayed a high homology with the *L. baicalensis* silicatein-α (CAI43319), *E. fluviatilis* silicatein-M2 (BAG74343), *L. oparinae* silicatein-A1 (ACG63793), and *T. aurantia* silicatein-α (AAC23951) (64%, 63%, 62%, and 61% identity, respectively). Principally, however, is that AuSil-Hexa has cysteine instead of a serine as the predicted catalytic amino acid residue (Fig. 1). AuCath1 and AuCath2 are most similar to *A. vastus* cathepsin L (CAI91577), because each of them shows 63% identity at the amino acid level.

Analysis of Transcripts in *B. levis*

B. levis cDNA was analyzed with degenerate primer sets F1-R, F2-R, and F3-R as described for *A. schulzei*. The number and percentage of clones is presented in Table 2. *BoCath1* (GenBank accession number ***GQ387057***) and *BoCath2* (GenBank accession number ***GQ387058***) were most similar to *A. vastus* cathepsin L (CAI91577), showing 64% and 48% identity at the amino acid level, respectively. We failed to detect any silicatein-like sequences in the analyzed clones.

Analysis of Transcripts in *P. raphanus*

Amplified cDNA of *P. raphanus* was analyzed with the F1-R and F3-R degenerate primer sets using the same strategy as above. The number and percentage of the sequenced clones are presented in Table 2. After RACE PCR, full-length cDNA sequences of the *P. raphanus* cathepsin genes were submitted to GenBank under accession numbers ***GQ380497*** (*PhrCat1*) and ***GQ380498*** (*PhrCat2*). It was found that PhrCath1 and PhrCath2 share the highest amino acid sequence identity with *A. vastus* cathepsin L (CAI91577), showing 61% and 58% identity at the amino acid level, respectively. In this case, silicatein-like transcripts were also not found.

Table 2 Number and percentage of clones corresponding to the silicatein/cathepsin genes from *Aulosaccus schulzei, Bathydorus levis* and *Pheronema raphanus*

A. schulzei Clones number (%)	*B. levis* Clones number (%)	*P. raphanus* Clones number (%)
Cathepsin-related transcripts		
AuCath1 132 (65.7)	*BoCath1* 90 (52)	*PhrCath1* 59 (60)
AuCath2 3 (1.5)	*BoCath2* 6 (3.5)	*PhrCath2* 39 (40)
Cathepsin-like transcripts[a] 63 (31.3)	Cathepsin-like transcripts[a] 77 (44.5)	Not found
Silicatein-related transcripts *AuSil-Hexa* 3 (1.5)	Not found	Not found
Total number of clones 201 (100)	Total number of clones 173 (100)	Total number of clones 98 (100)

[a] Transcripts of cathepsin-like genes containing insertions or deletions

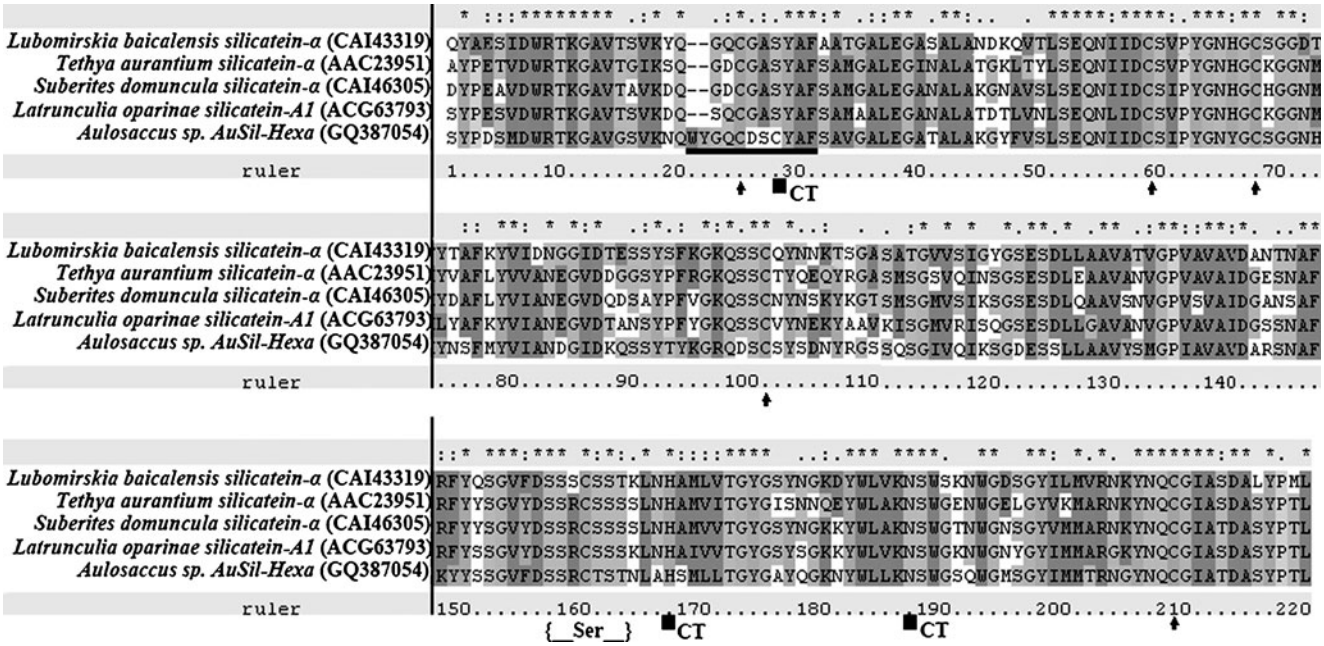

Fig. 1 The deduced amino acid sequence of *Aulosaccus schulzei AuSil-Hexa* cDNA. The sequence was aligned with homologous sequences including *L. baicalensis* silicatein α isoform (CAI43319), *Tethya aurantia* silicatein α isoform (AAC23951), *Suberites domuncula* silicatein α isoform (CAI46305), and *L. oparinae* silicatein α1 isoform (ACG63793). The mature protein sequences are presented. *Asterisks, dots,* and *double dots* indicate identical residues and amino acid substitutions with low and high similarity, respectively. The characteristic sites within the silicatein sequences are indicated: amino acids forming the catalytic triad (CT; Cys/Ser, His, and Asn); *arrows* indicate the cysteine residues potentially involved in the formation of disulfide bonds; *Ser* serine cluster

Structure of AuSil-Hexa

Analysis of the AuSil-Hexa protein from *A. schulzei* showed that the protein definitively contains a cysteine in the catalytic triad Cys-His-Asn (Fig. 1), which is characteristic for cathepsins (Berti and Storer 1995). Fairhead et al. (2008) showed an important role of the catalytic cysteine and surrounding residues for silica-condensing activity of a mutated cathepsin. The role of residues flanking the other two catalytic amino acids (His and Asn) has been shown to be less prominent. Therefore, we concentrated our attention on the analysis of the residues adjacent to the catalytic cysteine.

Sequences surrounding the catalytic cysteine (**C**) in AuSil-Hexa (CDS**C**YAF) contain a YAF motif (Fig. 1). By comparing available sequences of silicateins, we noted that the YAF motif is conventional for all known silicateins. Most silicateins also contain a CGA motif before the catalytic serine (**S**), such as CGA**S**YAF in *T. aurantia* (Shimizu et al. 1998). Our sequence comparisons indicate that, in the CGA triad, the first residue is an invariantly conserved residue for all demosponge silicateins and cathepsins described so far. The G residue is also characteristic for silicateins and cathepsins, excluding LoCath from *L. oparinae*, where lysine occupies this position (Kozhemyako et al. 2010). The A residue is not conventional for all silicateins; this residue is replaced by S in such proteins as SilB of *Acanthodendrilla* sp. (Kozhemyako et al. 2010), SilA3 of *L. baicalensis* (Wiens et al. 2006) and SilB1 and SilB2 of *S. domuncula* (Schröder et al. 2004; Schröder et al. 2005). Therefore, AuSil-Hexa differs from the demosponge silicateins and cathepsins by the presence of a D residue near the catalytic residue. We want to note here, that the serine cluster CSSS located on the surface of the silicatein molecule, which likely acts as a template for silica deposition (Schröder et al. 2008) is absent in AuSil-Hexa.

The AuSil-Hexa protein shows 52% identity of amino acid residues with the hexactinellid *C. meyeri* SILCA_CRAME protein (accession no. CAP49202, Müller et al. 2008b). The similarity with *C. meyeri* cathepsins is lower: 38%, 34%, 40%, 23%, and 45% identity of amino acid residues with cathepsin-like proteins 1 (CAP17584), 2 (CAP17585), 3 (CAP17586), 4 (CAP17587), and 5 (CAP17588), respectively.

Analysis of Glass Sponge cDNAs with Degenerate Primers Targeting AuSil-Hexa

In order to identify *AuSil-Hexa*-homologous transcripts or "conventional" silicatein gene(s) in the glass sponges, degenerate primers were designed that target distinctive regions within *AuSil-Hexa* (Table 1). PCR reactions with the degenerate primers and amplified cDNA from *A.*

schulzei, *B. levis*, and *P. raphanus* were performed. PCR products of the predicted size (504 bp) were obtained with cDNA samples from *A. schulzei* sponge only. The corresponding amplicon was isolated from an agarose gel and cloned into the pTZ57R/T vector. Approximately 30 clones were randomly sequenced. The analysis showed that most of the sequenced clones belong to *AuSil-Hexa* transcripts, and remaining transcripts were shown to relate to non-specific amplifications. Conventional silicatein genes and any *AuSil-Hexa*-related genes were not found by this analysis in the studied sponge species.

Expression of Silicatein and Cathepsin Genes in Mature Sponge *A. schulzei*

Analysis of silicatein/cathepsin clone frequency (Table 2) in *A. schulzei* was complemented with real-time PCR (Fig. 2). Expression of *AuSil-Hexa*, *AuCath1*, and *AuCath2* genes was studied with gene-specific primers indicated in Table 1. This analysis confirmed the data presented in Table 2, i.e., the *AuCath2* transcripts was shown to represent the most abundant transcripts; *AuCath1* and *AuSil-Hexa* were expressed ten and 50 times weaker, respectively (Fig. 2a). In spite of expression of the *AuCath1* and *AuSil-Hexa* genes was low, corresponding real-time PCR transcripts could be visualized in agarose gels (Fig. 2b).

Phylogenetic Analysis

We found that over 60 unique sequences of sponge silicateins and cathepsins exist in the sequence databases. GenBank sequences and sequences described in this work were aligned and further manually edited to remove redundant sequences (see the Electronic Supplementary Material (Fig. 1)). We analyzed these sequences by different methods (the Bayesian analysis, the maximum likelihood analysis, and the maximum parsimony analysis) to validate constructed phylogenetic trees. A combined result of these analyses is presented in Fig. 3. In the supplementary figures, we presented original trees resulted from the Bayesian analysis (Electronic Supplementary Material (Fig. 2), the maximum likelihood analysis (Electronic Supplementary Material (Fig. 3), and the maximum parsimony analysis (Electronic Supplementary Material (Fig. 4)). To root the trees, cathepsin X/O from *S. domuncula* (GenBank accession number CAH04633) was used as the out-group. In general, all of the used methods yielded very similar results, thus confirming reliability of the analysis. A well-resolved SILICATEIN clade with high posterior (1.0) and bootstrap supports (82 for the ML-tree and 80 for the MP-tree) was revealed in all trees (Fig. 3 and Electronic Supplementary Materials (Figs. 2, 3, and 4)). The silicatein-like cathepsin LoCath of *L. oparinae*

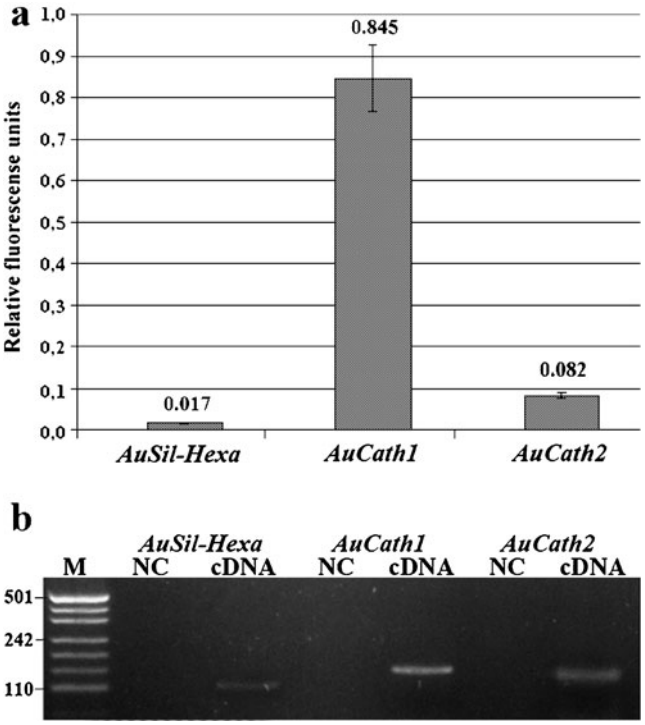

Fig. 2 Expression of *AuSil-Hexa* in *Aulosaccus schulzei* mature sponges in comparison with cathepsin genes. **a** Quantitative expression of *AuSil-Hexa* (real-time PCR data), presented as mean of relative fluorescent units±standard error from two independent cDNA samples with three replicates each. **b** Visualization of real-time PCR products in an agarose gel

described in our previous work (Kozhemyako et al. 2010), occupies an outside position in the clade that confirms its intermediate status between sponge silicateins and cathepsins. The SILICATEIN clade is subdivided into three subclades (I, II, and III) as indicated in Fig. 3. The first two clades consist of silicateins from marine sponges. Silicateins from fresh-water sponges form the third subclade. Analysis indicates that the silicateins of marine sponges (II subclade) and silicateins of fresh-water sponges (III subclade) form two clearly distinguished clusters on the tree with high posterior (1.0) and bootstrap supports (78 for both the ML- and MP-trees) (Fig. 3; see also the Electronic Supplementary Materials (Figs. 2, 3, and 4). This result is an agreement with the phylogenetic analysis performed by Mohri et al. (2008). The authors also demonstrated that fresh-water sponge silicateins and marine sponge silicateins formed two separate clusters. Silicatein α and β isoforms do not form individual clusters; they are randomly distributed between subclades of the SILICATEIN clade.

AuSil-Hexa was consistently placed by all methods of analysis in the silicatein group of proteins (SILICATEIN clade, II subclade). This indicates that AuSil-Hexa belongs to silicateins and not cathepsins. It forms a cluster with the silicatein LoSilA3 from *L. oparinae*. It should be noted that LoSilA3 differs from conventional silicateins by the AY

insertion located four residues before the catalytic serine (Kozhemyako et al. 2010). The functionality of such an insertion is unknown, but one can suggest accelerated evolution of *LoSilA3* and acquisition of a new function by the protein. The deduced SILCA_CRAME protein (accession no. CAP49202) described by Müller et al. (2008b) forms a cluster with silicatein β of *L. oparinae* with a good bootstrap probability (LM90/PM70, Fig. 3).

A cluster of cathepsins present at the bottom of the tree (Fig. 3) is poorly resolved. It consists of most divergent cathepsins and presented in the tree by the cathepsins from *C. meyeri* and *A. vastus*, as well as the cathepsins of *A. schulzei*, *B. levis*, and *P. raphanus* described in this work. An additional analysis, in which numerous non-sponge cathepsin sequences were included, shows that this clade contains sequences from divergent living organisms, including mammals (data not shown).

Because information about silicatein genes in glass sponges is very limited, it is difficult to explain the origin of the *AuSil-Hexa* gene. The suggestion that demosponge silicateins have been evolved from cathepsins (Schröder et al. 2008; Krasko et al. 2000; Müller et al. 2007) is supported by our phylogenetic analysis. One can hypothesize that *AuSil-Hexa* has been originated from an ancestral

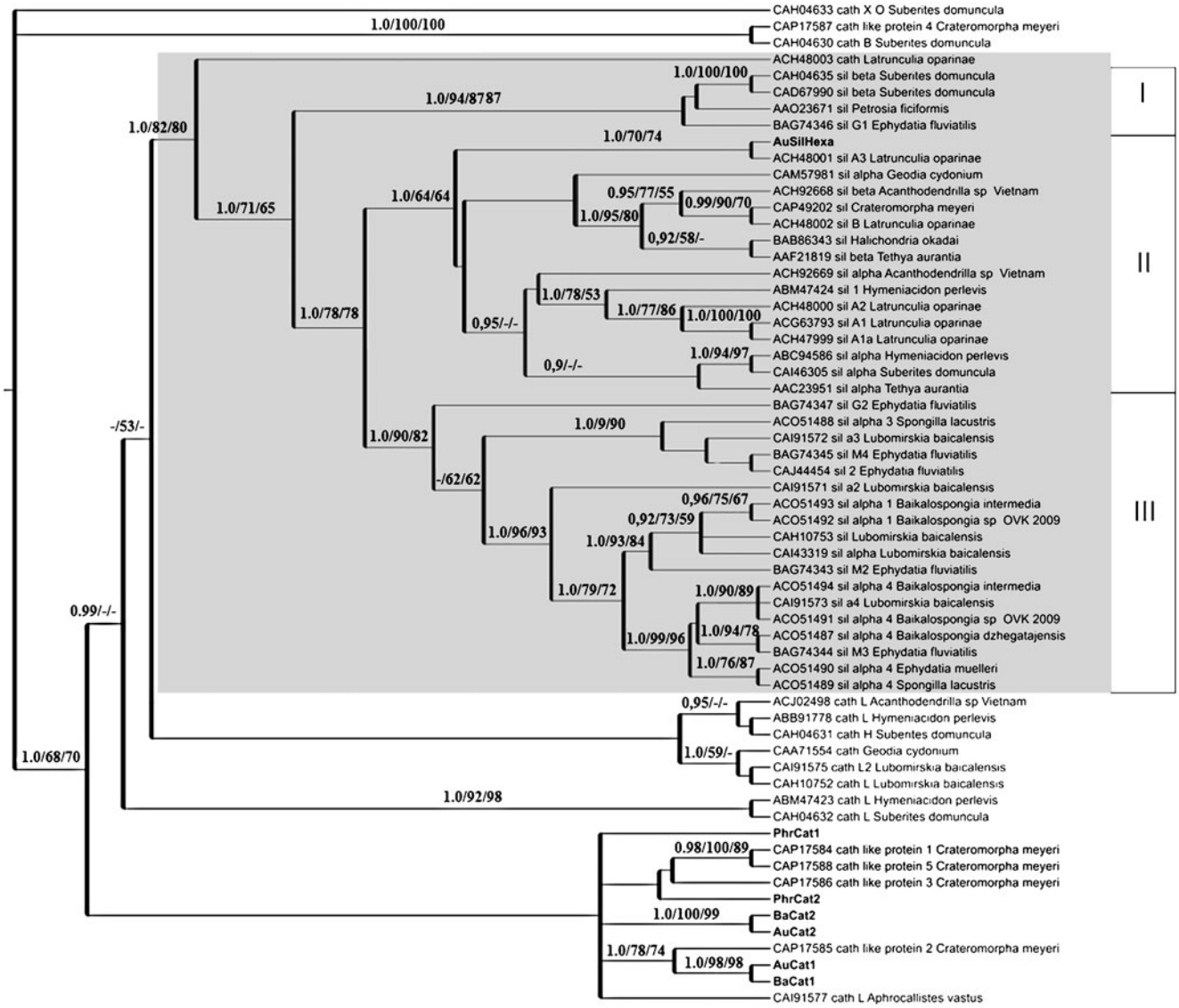

Fig. 3 Bayesian cladogram (50% majority-rule consensus tree) of sponge silicateins and cathepsins. The substitution model WAG+I+G was used. Posterior probabilities (PP; only PP>0.90 shown) and bootstrap proportions (%) for the maximum likelihood tree and maximum parsimony tree are given. *Dashes* indicate supports <50%. The GenBank accession numbers of the proteins are given in the cladogram. The sequences obtained in this work are in *bold*. The *grey box* represents the SILICATEIN clade, which includes also the silicatein-like cathepsin of *Latrunculia oparinae*. The SILICATEIN clade is subdivided into three subclades (I, II, and III)

silicatein gene or silicatein-like gene. In this case, serine in the catalytic triad Ser-His-Asn was replaced to cysteine.

Conclusions

By investigating silicatein-related transcripts in three species of marine glass sponges, we did not find transcripts corresponding to conventional silicatein genes. The absence of silicatein transcripts in the glass sponges *B. levis* and *P. raphanus* indicates that if these sponges do contain silicatein genes, their expression must be negligible. The sponge *A. schulzei* contains silicatein gene transcripts with a low frequency (1.5% of total silicatein/cathepsin transcripts). There is only one other known example of a silicatein cDNA from glass sponges that has been recently identified in *C. meyeri* (Müller et al. 2008b). Given the low abundance of silicatein transcripts in hexactinellids, one can suggest a different organization of the biosilicification mechanisms in demosponges and hexactinellids. It is also possible that active spicule-producing cells are not abundant in Hexactinellid sponges, taking into account they very slow growth (Dayton 1979; Gatti 2002). At present, only one complete sequence of a silicatein cDNA of glass sponges is known, namely *AuSil-Hexa*. The functionality of the corresponding protein is not clear. Further investigation of the catalytic activity of recombinant AuSil-Hexa would shed light on its biological function. It would also be of interest to screen and isolate silicatein genes from other species of glass sponges and to determine, which amino acids with catalytic activity they encode.

Acknowledgments This work was supported by a grant from the Grant Program "Principles of Basic Investigations of Nanotechnologies and Nanomaterials" of the Russian Academy of Sciences.

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Berti P, Storer A (1995) Alignment/phylogeny of the papain superfamily of cysteine proteases. J Mol Biol 246:273–283

Brutchey RL, Morse DE (2008) Silicatein and the translation of its molecular mechanism of biosilicification into low temperature nanomaterial synthesis. Chem Rev 108:4915–4934

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159

Dayton PK (1979) Observations of growth, dispersal and population dynamics of some sponges in McMurdo Sound, Antarctica. In: Lévi C and Boury-Esnault N (eds.) *Colloques internationaux du C.N.R.S. 291. Biologie des spongiaires*. E′ ditions du Centre National de la Recherche Scientifique, Paris. pp. 271–282

Ehrlich H, Worch H (2007) Collagen: a huge matrix in glass sponge flexible spicules of the meter-long *Hyalonema sieboldi*. In: Bäuerlein E (ed) Handbook of biomineralization. Wiley, Weinheim, pp 23–41

Ehrlich H, Ereskovsky A, Drozdov A, Krylova D, Hanke T, Meissner H, Heinemann S, Worch H (2006) A modern approach to demineralisation of spicules in the glass sponges (Hexactinellida: Porifera) for the purpose of extraction and examination of the protein matrix. Russ J Mar Biol 32:186–193

Ehrlich H, Krautter M, Hanke T, Simon P, Knieb C, Heinemann S, Worch H (2007) First evidence of the presence of chitin in skeletons of marine sponges. Part II. Glass sponges (Hexactinellida: Porifera). J Exp Zool (Mol Dev Evol) 308B:473–483

Ehrlich H, Deutzmann R, Brunner E, Cappellini E, Koon H, Solazzo C, Yang Y, Ashford D, Thomas-Oates J, Lubeck M, Baessmann C, Langrock T, Hoffmann R, Wörheide G, Reitner J, Simon P, Tsurkan M, Ereskovsky AV, Kurek D, Bazhenov VV, Hunoldt S, Mertig M, Vyalikh DV, Molodtsov SL, Kummer K, Worch H, Smetacek V, Collins MJ (2010) Mineralization of the metre-long biosilica structures of glass sponges is templated on hydroxylated collagen. Nat Chem 12:1084–1088

Fairhead M, Johnson K, Kowatz T, McMahon S, Carter L, Oke M, Liu H, Naismith J, van der Walle C (2008) Crystal structure and silica condensing activities of silicatein alpha-cathepsin L chimeras. Chem Commun (Camb) 15:1765–1767

Gatti S (2002) High Antarctic carbon and silicon cycling—how much do sponges contribute? VI International Sponge Conference. In: Book of Abstracts, Bollettino dei Musei Instituti Biologici, vol. 66–67 University of Genoa. p. 76

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Heinemann S, Ehrlich H, Knieb C, Hanke T (2007) Biomimetically inspired hybrid materials based on silicified collagen. Int J Mat Res 98:603–608

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066

Kozhemyako VB, Veremeichik GN, Shkryl YN, Kovalchuk SN, Krasokhin VB, Rasskazov VA, Zhuravlev YN, Bulgakov VP, Kulchin YN (2010) Silicatein genes in spicule-forming and non-spicule-forming Pacific demosponges. Mar Biotechnol 12:403–409

Krasko A, Lorenz B, Batel R, Schröder HC, Müller IM, Müller WEG (2000) Expression of silicatein and collagen genes in the marine sponge *Suberites domuncula* is controlled by silicate and myotrophin. Eur J Biochem 267:4878–4887

Krautter M (2002) Fossil Hexactinellida: an overview. In: Hooper JNA, van Soest RWM (eds) Systema Porifera: a guide to the classification of sponges. Plenum, New York, pp 1211–1223

Matz M, Shagin D, Bogdanova O, Lukyanov S, Diatchenko L, Chenchik A (1999) Amplification of cDNA ends based on template-switching effect and step-out PCR. Nucleic Acids Res 2:1558–1560

Mohri K, Nakatsukasa M, Masuda Y, Agata K, Funayama N (2008) Toward understanding the morphogenesis of siliceous spicules in freshwater sponge: differential mRNA expression of spicule-type-specific silicatein genes in *Ephydatia fluviatilis*. Dev Dyn 237:3024–3039

Müller WEG, Boreiko A, Wang X, Belikov SI, Weins M, Grebenjuk VA, Schloßmacher U, Schröder HC (2007) Silicateins, the major biosilica forming enzymes present in demosponges: protein analysis and phylogenetic relationship. Gene 395:62–71

Müller WEG, Boreiko A, Schloßmacher U, Wang X, Eckert C, Kropf K, Li J, Schröder HC (2008a) Identification of a silicatein(-related) protease in the giant spicules of the deep-sea hexactinellid *Monorhaphis chuni*. J Exp Biol 211:300–309

Müller WEG, Wang X, Kropf K, Boreiko A, Schloßmacher U, Brandt D, Schröder HC, Wiens M (2008b) Silicatein expression in the

hexactinellid *Crateromorpha meyeri*: the lead marker gene restricted to siliceous sponges. Cell Tissue Res 333:339–351

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Schröder H, Perović-Ottstadt S, Wiens M, Batel R, Müller IM, Müller WEG (2004) Differentiation capacity of epithelial cells in the sponge *Suberites domuncula*. Cell Tissue Res 316:271–280

Schröder H, Perović-Ottstadt S, Grebenjuk VA, Engel S, Müller IM, Müller WEG (2005) Biosilica formation in spicules of the sponge *Suberites domuncula*: synchronous expression of a gene cluster. Genomics 85:666–678

Schröder H, Wang X, Tremel W, Ushijima H, Müller W (2008) Biofabrication of biosilica-glass by living organisms. Nat Prod Rep 25:455–474

Shimizu K, Cha J, Stucky G, Morse D (1998) Silicatein α: cathepsin L-like protein in sponge biosilica. Proc Natl Acad Sci USA 95:6234–6238

Swofford DL (2002) PAUP* Phylogenetic analysis using parsimony. Version 4.0b10. Sinauer Associates, Massachusetts

Uriz M-J, Turon X, Becerro MA, Agell G (2003) Siliceous spicules and skeleton frameworks in sponges: origin, diversity, ultrastructural patterns, and biological functions. Microsc Res Tech 62:279–299

Wiens M, Belikov SI, Kaluzhnaya OV, Krasko A, Schröder HC, Perović-Ottstadt S, Müller WEG (2006) Molecular control of serial module formation along the apical-basal axis in the sponge *Lubomirskia baicalensis*: silicateins, mannose-binding lectin and mago nashi. Dev Genes Evol 216:229–242