# Biochemistry and Cell Biology of Silica Formation in Sponges

WERNER E.G. MÜLLER,\* ANATOLI KRASKO, GAËL LE PENNEC, AND HEINZ C. SCHRÖDER Institut für Physiologische Chemie, Abteilung Angewandte Molekularbiologie, Universität, D-55099 Mainz, Germany

*KEY WORDS* sponges; spicules; silica; silicatein

ABSTRACT The main inorganic material forming the skeletal elements in Demospongiae as well as in Hexactinellida, the spicules, is amorphous silica. The spicules occur in the cytoplasm and the extracellular space and also in the nucleus (as silicate crystals) of some sponge cells; the function in the latter compartment is unknown. Recent evidence shows that the formation of spicules is mediated by the enzyme silicatein. The cDNA as well as the gene encoding this enzyme was cloned from Suberites domuncula. The recombinant silicatein catalyzes the synthesis of amorphous silicate using tetraethoxysilane as substrate. The enzyme is dependent on ferric iron. Silicatein also has proteolytic (cathepsin-like) activity. Incubation of primmorphs, a special form of 3D-cell aggregates, with silicon resulted in a strong increase of their size from 1–7 mm to  $\sim$ 10 mm. The morphogenetic activity of silicate is underscored by the finding that this ion increases gene expression of silicatein and collagen. Based on these findings, it is concluded that both iron and silicate stimulate the activity of silicatein. Furthermore, it is proposed that the growing spicules are surrounded by the scavenger receptor which might be considered as a docking molecule for the collagen matrix into which the spicules are embedded. Microsc. Res. Tech. 62:368-377, 2003. © 2003 Wiley-Liss, Inc.

## **INTRODUCTION**

Sponges (Porifera) represent the oldest extant metazoan phylum (Müller, 1995, 2001). The earliest fossil records indicate that soft-bodied, presumably multicellular animals, including sponges (from the Ediacara fauna), appeared  $\sim 600$  Myr ago (MYA) (Mehl et al., 1998; Brasier, 2000). However, the paleontological evidence for possible metazoan fossils is still under debate, like the recent claims about metazoan embryos from the Doushanto Formation (Xiao et al., 1999), which had been attributed to Volvox-like chlorophytes (Xue et al., 1998). However, molecular evidence appears to be reliable; based on analyses of informative protein sequences it had been proposed that the phylum Porifera evolved more than 600 MYA from the hypothetical Urmetazoa (reviewed in Müller, 1998). Furthermore, molecular biology techniques also provide evidence that, among the three classes within the phylum Porifera, the Hexactinellida evolved first, followed by the Demospongiae, and finally the Calcarea (reviewed in Müller et al., 1998). In this context it is notable that the Hexactinellida and the Demospongiae contain siliceous spicules as skeletal elements, while the spicules of the Calcarea are of calcareous material. Isotopic data revealed that siliceous biomineralization preceded calcareous biomineralization by more than 10 Myr (Brasier et al., 1997), thus further supporting this assumption.

It is surprising that during the Cambrian/Precambrian transition an explosive evolution of the biota occurred, characterized by the appearance of organisms with phosphate and also siliceous skeletons. During the Cambrian/Precambrian period, a stepwise increase in oxygen occurred (Brasier, 2000); this process was followed by a depletion of dissolved iron in the marine environment. In the present day oceans, the seawater contains 1 nM of dissolved iron and <3 to 150  $\mu$ M of dissolved silicon (silicate) (Kennish, 1994). The surface water is even poorer in silicon, with concentrations often below 3  $\mu$ M (Nelson et al., 1995). In contrast, the concentration of silicon in human blood is  $\approx$ 140  $\mu$ M and that of iron  $\approx$ 15  $\mu$ M (Geigy Scientific Tables, 1984).

Recent studies reported that the expression of the gene for silicatein, which has been assumed to be the key enzyme for the synthesis of spicules in demosponges *Tethya aurantium* (Shimizu et al., 1998 as *T. aurantia*) and *Suberites domuncula* (Krasko et al., 2000), is controlled by extracellular silicate (Krasko et al., 2000). At  $\approx 1 \,\mu$ M of soluble silicate in seawater, the expression of silicatein is low; at 60  $\mu$ M of silicate expression increases dramatically (Krasko et al., 2000). Moreover, extracellular silicate also acts on a second gene which is involved in spicule formation, the collagen gene (Krasko et al., 2000). The expression of this latter is further upregulated when sponge cells are incubated with an extracellular signaling factor, the

Contract grant sponsor: the Deutsche Forschungsgemeinschaft; Contract grant number: Mü 348/12-5; Contract grant sponsors: the Bundesministerium für Bildung und Forschung (project: Center of Excellence BIOTECmarin), the European Commission (project: SPONGE), and the International Human Frontier Science Program; Contract grant number: RG-333/96-M.

<sup>\*</sup>Correspondence to: Prof. Dr. W.E.G. Müller, Institut für Physiologische Chemie, Abteilung Angewandte Molekularbiologie, Universität, Duesbergweg 6, 55099 Mainz; Germany. E-mail: wmueller@mail.uni-mainz.de and http://www.biotecmarin.de

Received 11 July 2002; accepted in revised form 18 March 2003 DOI 10.1002/jemt.10402

Published online in Wiley InterScience (www.interscience.wiley.com).

cardiac myotrophin-related molecule, isolated from *S. domuncula* (Schröder et al., 2000).

We established the growth conditions for cells from sponges in vitro (Custodio et al., 1998; Müller et al., 1999). The 3D-aggregates formed were termed primmorphs. After dissociation, the cells start to proliferate (proliferation phase) when they are brought into contact with adjacent cells (Custodio et al., 1998; Müller et al., 1999). During this process, the spicules which contribute to control of morphogenesis are formed (Schröder et al., 2000; Krasko et al., 2000).

#### SILICA FORMATION: SILICATEIN Silicatein Sequence From S. domuncula

**Protein.** The S. domuncula cDNA (termed: SUBDO-SILICA) encoding the sponge silicatein was isolated and characterized (Krasko et al., 2000). The 1,169 bp sequence has an open reading frame of 993 nt; the predicted translation product of 330 amino acids (aa), SILICA\_SUBDO, (Fig. 1A) has a calculated  $M_r$  of 36,306 Da.

As outlined before (Shimizu et al., 1998; Krasko et al., 2000), silicatein is a new member of the cathepsin subfamily. The *S. domuncula* silicatein shares 70% identical and 79% similar (identical plus physicochemical-related aa) aa with the *T. aurantium/aurantia* silicatein polypeptide (Shimizu et al., 1998); while an alignment with the corresponding *Geodia cydonium* cathepsin protein (Krasko et al., 1997) shows a lower identity/similarity of 43%/59%. The three amino acids Cys, His, and Asn, which form the catalytic triad of cysteine proteases, are present in the sponge cathepsin at the characteristic sites:  $Cys_{125}$ , His<sub>164</sub>, and Asn<sub>184</sub> (Krasko et al., 2000). As reported before for the *T. aurantium* silicatein (Shimizu et al., 1998), the cluster comprising the characteristic hydroxy amino acids (serine) is also present in the corresponding *S. domuncula* molecule (Fig. 1A).

A processing step is required for the formation of active cathepsins, which is mediated either by autolysis or by a second protease. Based on published comparisons with cathepsins, the characteristic cleavage site is also present in the sponge silicatein sequences, suggesting a processing of a proenzyme with an  $M_r$  of 36,306 to the mature form with an  $M_r$  of 23,125 (Fig. 1A).

**Phylogenetic Analysis.** The highest similarity of the sponge silicatein is shared with the sponge cathepsin of the cathepsin-L type. This conclusion is drawn from the alignment (Fig. 1A) and the resulting unrooted tree (Fig. 1B). The *S. domuncula* silicatein branches off from the same root which also comprises the cathepsin-L from human (Fig. 1B). The other known human cathepsin molecules are more distantly related to the silicatein/cathepsin-L molecules; the sponges also contain other cathepsins, e.g., cathepsin B, as indicated in Figure 1.

**Silicatein Gene.** The complete gene from which the cDNA for silicatein derives was cloned from the cDNA library of *S. domuncula*. The total length of the gene is 2,280 nt (Fig. 2). It comprises six introns which are—as is usual for sponge sequences—short (Gamulin et al., 1997); the intron borders follow the "GT–AG" rule. Intron number 2 has 166 nt and intron number 5 (the longest) comprises 375 nt (Fig. 2). The cathepsin-spe-

cific motifs 1 to -3 are not interrupted by introns. Motif 1 is present between intron-2 and -3, motif 2 between and intron-5 and -6 and motif 3 is also in this gene segment.

**Recombinant Silicatein.** As recently reported (Krasko et al., 2002), the sequence including the putative catalytic segment, from  $aa_{17}$  to  $aa_{330}$  ( $nt_{49}$  to  $nt_{993}$ ), was used for expression in *Escherichia coli*. The cDNA was inserted into the bacterial oligohistidine expression vector pQE-30 (Qiagen, Chatsworth, CA) as summarized in Figure 3A. E. coli strain XL1-blue was transformed with this plasmid and expression of fusion protein was induced for 6 hours at 37°C with 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG). The fusion protein was purified. The purity of the material was checked by electrophoresis on 10% polyacrylamide gels containing 0.1% sodium dodecyl sulfate. The purified silicatein, termed r-silicatein, was digested with 5 U of recombinant enterokinase. The recombinant silicatein was cleaved from the His-tag by digestion with recombinant enterokinase; the purified long form of the r-silicatein preparations had an M<sub>r</sub> of 35 kDa (calculated size: 34,663) and the short form of M<sub>r</sub> 26 kDa (22,149) (Fig. 3B, lane a). These two recombinant r-silicatein preparations were used for analysis of enzyme activity.

Silicatein Enzyme Assay. In a total volume of 500  $\mu$ l, 100  $\mu$ g of r-silicatein was added to the 25 mM Tris-HCl buffer (pH 7.2; 10 mM of DL-dithiothreitol, 100 mM NaCl). As substrate, 4.5 mM tetraethoxysilane was used. Where indicated, 5 mM MgCl<sub>2</sub> and/or 5 mM (final concentration) of Fe(+++) citrate was added. The reaction was performed at 22°C for 90 minutes and terminated by centrifugation (20,000g; 4°C, 15 min). The sediment obtained was washed three times in ethanol and an aliquot of 250  $\mu$ l was determined for released silicic acid using the "Silicon Test" colorimetric assay kit (Merck, Darmstadt, Germany; 1.14794). Based on the absorbance values at 810 nm, the absolute amounts of silicic acid were calculated after construction of a calibration curve using a silicon standard (Merck; 1.09947).

**Enzymatic Activity: Silicate Formation.** Both forms of the recombinant silicatein (r-silicatein) were found to increase the amount of polymerized silica significantly from 2.2 (assays without enzyme) to 5.9 ng/mg for the long form of the enzyme (2.4-6.1 ng/mg for the short form) compared to the assays without r-silicatein. If ferric iron [Fe(+++)] was added to the reaction mixture, a 3-fold increase of the reaction product was observed in both the assay with the long form of the enzyme and that with the short form. In contrast, Mg(++) had no effect on the reaction (Fig. 4).

These experiments and those already published (Krasko et al., 2000, 2002; Müller et al., 2002), together with those reported for the *T. aurantium* model (Shimizu et al., 1998), provide strong evidence that silicatein is the major enzyme involved in silicate production required for the formation of spicules.

**Enzymatic Activity: Proteolytic Activity.** As outlined above (Fig. 1), the deduced protein sequence of the sponge silicatein shows high sequence similarity to lysosomal enzymes, the cathepsins. We therefore explored whether the recombinant silicatein also has proteolytic activity, in addition to the ability to form amor-

370



Fig. 1. Silicatein polypeptide. A: The deduced as sequence from the S. domuncula cDNA encoding silicatein (termed SILICA\_SUBDO) was aligned with the cathepsin sequences from sponges; the S. domuncula cathepsin B (CATB\_SUBDO) and the G. cydonium cathepsin L (CATL\_GEOCY, accession number Y10527). Residues conserved (similar or related with respect to their physicochemical properties) in all sequences are shown in white on black and those in at least two sequences in black on gray. The characteristic three active sites aa are indicated ( $\blacksquare$ ); the cleavage site for the processing to the active enzymes is indicated ( $\blacktriangleright \rightarrow$ ). Furthermore, the three motifs, motif-1 to -3, are highlighted. The serien-rich section in the silicatein sequence is

delimited above the alignment. B: These deduced sponge protein sequences were aligned with the different human cathepsin sequences: cathepsin B (CATB\_HUMAN, P07858), cathepsin C (CATC\_HUMAN; P53634), cathepsin H (CATH\_HUMAN; P09668), cathepsin K (CATK\_HUMAN; P43235), cathepsin L (CATL\_HUMAN; P07711), cathepsin O (CATO\_HUMAN; P43234), cathepsin S (CATS\_HUMAN; P25774), and cathepsin W (CATW\_HUMAN; P56202). The phylogenetic relationships of the sponge sequences within the human sequences is given in an unrooted tree. Scale bar indicates an evolutionary distance of 0.1 aa substitutions per position in the sequence.

-

CTAACTAAAT	TCTATACTGT	TTGTTATCAT	CCTTTATCAT	ATACATGCAC	TCACACAGGA	60
MLV	тvv	VLGL	LGF	ASA	AQPK	20
CAATGCTTGT	CACAGTGGTA	GTACTGGGTC	TACTGGGGTT	TGCTTCTGCA	GCCCAGCCCA	120
7 3 7	VEE	W O L W	кст	HSK	MYES	40
ልርጥጥጥርልልጥጥ	TGTAGAAGAA	TCCCACCTCT	GGAAGTCCAC	TCACTCTAAG	ATGTACGAGT	180
	F L F	PHI. T	W T. S	NKK	YTEO	60
						240
CACAGIIAAI	GGAACICGAA	AGACAICIGA		M N O		240
H N V	N S H		T L A			200
AACACAATGT	CAACTCACAC	ATTTTCGGTT	TTACTCTGGC	AATGAACCAG	TTTGGAGATC	300
⊧ intron-	1					
TGGTAAGATT	GGCATACTAT	TGTAGCATAA	TAATTATACA	AACTCCATGC	ATATATAAGT	360
ATAGGCTATA	TATAGGTCTA	ATTTAAGAGA	TTTGGTTGTG	ACAACATTTC	AAAAACTTGT	420
CATGAAGGCT	CAATGTGTAT	GAACCACCAA	TTCCAATCTT	ACGTATCCTT	TACTGTATAA	480
		intron-1 🚽	SELE	YAD	YLG	90
TACCAGCATT	CACCATATTT	ATTGTTTTAG	AGTGAATTGG	AGTATGCTGA	CTATCTTGGC	540
QYRI	EDK	K S G	NYSK	TFQ	RDP	110
CAGTATCGCA	TTGAGGATAA	AAAATCTGGC	AACTACTCAA	AGACTTTTCA	GCGTGATCCT	600
	DEA	VDW	R T K C	Δ V T	A V K	130
		TOTACACTICC	ACAACCAAAC	CACCTCTCAC	CCCTCTCAAC	660
	ACCCIGAAGC	IGIAGACIGG	AGAACCAAAG	GAGCIGICAC	GGCIGICANG	122
				30003333000		132
GACCAGGTAT	TCTTCATAAA	AATGATTATT	CTTTCTATGC	ATGGAAACCT	GAATTGTTAA	720
TATCACGAGC	CCAGTAAGCA	TATCTAAAGG	CAGCTTTCAT	TTGTACAACG	ATTTATAGGG	780
				intron-	2 ╣G D C	135
TGAATCTAGT	CATATTTCAT	ATTTGAAACT	CATTAATGAC	GTATCAATCC	AGGGAGACTG	840
GAS	YAFS	A M G	ALE	GANA	LAK	155
TGGTGCTAGC	TATGCTTTTA	GTGCTATGGG	TGCTTTGGAG	GGTGCTAATG	CTTTAGCCAA	900
GNA	VSLS	EQN	IID	C S I⊫in	tron-3	171
GGGAAATGCA	GTATCTCTCA	GTGAACAGAA	CATCATTGAT	TGCTCGAGTT	AGTTCTATAA	960
GACGGTGACC	ACATGCATAC	GATAAGTTAT	TTCCATTTGT	ATAACTTTCG	AGGTCATATG 1	020
TATTTATATA	CATACATTC	GTTCATCCAT	ΔΤΔΔΤΤΔΤΤC	ͲͲͲϹͲͲͲͲͲϪ	тсаттсасса 1	020
	V C N U	C C U				100
Incol P	IGNH	всп				
000300000		ACCOMMONOA				140
CGCAGTTCCT	TACGGTAACC	ACGGTTGTCA	TGGAGGCAAT	ATGTATGATG	CTTTTTTGTA 1	140
CGCAGTTCCT VIA	TACGGTAACC N E G V	ACGGTTGTCA D Q D	TGGAGGCAAT S A Y	ATGTATGATG P F V G	FLI CTTTTTTGTA 1 K  = in-4	189 140 207
CGCAGTTCCT VIA TGTCATCGCT	TACGGTAACC N E G V AACGAGGGGGG	ACGGTTGTCA D Q D TCGATCAGGA	TGGAGGCAAT S A Y CAGTGCATAT	ATGTATGATG P F V G CCATTTGTTG	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1	189 140 207 1200
CGCAGTTCCT VIA TGTCATCGCT TAATAATTAT	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC	ACGGTTGTCA DQD TCGATCAGGA AACAGTAATT	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA	FLI CTTTTTTGTA 1 K    in-4 GAAAGGTGAA 1 ACCAACTGTA 1	189 207 1200 1260
CGCAGTTCCT VIA TGTCATCGCT TAATAATTAT ATGTCTGCAA	TACGGTAACC N E G V AACGAGGGGGG TCTTCATCAC CTCTTAGAAT	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGT	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT	F L I CTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1	189 140 207 1200 1260 1320
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA	TACGGTAACC N E G V AACGAGGGGGG TCTTCATCAC CTCTTAGAAT	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGT i	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 ╢Q	FLI CTTTTTGTA 1 K Fin-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C	189 140 207 1200 1260 1320 211
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGT i ACTTCCCTGC	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 ╢Q ACACACAGCA	F L I CTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1	189 207 200 200 200 200 200 211 2380
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTGT i ACTTCCCTGC M S G M	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 ╢Q ACACACAGCA V S I	F L I CTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G	189 207 207 200 260 320 211 380 231
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACCTATTCT TACTTTAGCC G T S AGGTACATCA	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTGT ACTTCCCTGC M S G M ATGTCGGGTA	A TGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\parallel Q$ ACACACAGCA V S I TGGTGTCAAT	F L I CTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1	189 140 207 1200 1260 1320 211 1380 231 1440
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L O A	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTG ACTTCCCTGC M S G M ATGTCGGGTA N V G P	A TGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\exists Q$ ACACACAGCA V S I TGGTGTCAAT V S V	F L I CTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D	189 140 207 1200 1260 1320 211 1380 231 1440 251
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGGT ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC	A T T D A ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\exists$ Q ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT	F L I CTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1	189 140 207 200 1260 1320 211 1380 231 1440 251
CGCAGTTCCT VIA TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT NYNS AACTATAACA SESD AGTGAGTCTG CANS	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA L introp=5	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGGT i ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC	ATGTATGATG PFVG CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 flQ ACACACAGCA VSI TGGTGTCAAT VSV CTGTATCTGT	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1	189 140 207 200 220 231 380 231 440 251 500 258
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA F intron-5 GGTAGCACAG	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGT i ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\exists Q$ ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1	139 207 200 220 231 380 231 440 251 500 258
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA \$\\$ intron-5 GGTAAGACAG	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGT i ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA	A TGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\exists Q$ ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1	139 207 200 2200 231 380 231 440 251 500 258 560
CGCAGTTCCT VIA TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT NYNS AACTATAACA SESD AGTGAGTCTG GANS GGTGCTAACA CGTATGGTTG	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA F intron-5 GGTAAGACAG ATGTTTGCCT	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGT i ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT	A TGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $  Q$ ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1	139 140 207 200 220 231 380 231 440 251 500 258 560 620
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTC	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTAC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA F intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGT i ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 #Q ACACACAGCA V S I TGGTGTCCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCCGGGCAAT 1	139 140 207 200 220 231 380 231 440 251 500 258 560 620
CGCAGTTCCT VIA TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT NYNS AACTATAACA SESD AGTGAGTCTG GANS GGTGCTAACA CGTATGGTTG CCGACTTTC TTGATAGTTA	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGGTGTTTCA F intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGT i ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 #Q ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCCGGGCAAT 1 AAAGCCTATA 1	139 140 207 2200 220 231 380 231 440 251 500 258 560 620 680 680
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTC TTGATAGTTA CTATTGTTTC	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGGTGTTTCA F intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA CTTTTTCGAA	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTGT ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG	A TGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 ∜Q ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTG	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1 AACATGACTG 1	139 140 207 200 220 231 380 231 440 251 500 258 560 620 680 740 800
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTCC TTGATAGTTA CTATTGTTTC	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TAAATGCATG	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGGTGTTTCA # intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA GTGTATCCGT	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTGT ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG AGGTTCATAC	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 ¶Q ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTG GTTAATTGTG	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 ATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AACATGACTG 1 TTTTGATTGA 1	139 140 207 200 220 231 380 231 440 251 500 258 560 620 680 1680 1800 189 189 189 140 190 190 190 190 190 190 190 19
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTTC TTGATAGTTA CTATTGTTTC TCTATTGCTT	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TAAATGCATG TGTGTACATA	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGGTACATCA I intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA CTTTTTCGAA GTGTATCCGT	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTGT ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 ⋕F Y	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 ¶Q ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTG GTTAATTGTG Y S G	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1	139 140 207 200 260 231 380 231 440 251 500 258 560 620 680 1680 1860 266
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTTC TTGATAGTTA CTATTGTTTC TCTATTGCTT	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TAAATGCATG TGTGTACATA	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGGTACATCA # intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA GTGTATCCGT int ATACACATAC	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTG ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 ╣F Y ATACAGTTT	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\parallel Q$ ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTG GTTAATTGTG Y S G ACTACAGTGG	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AACATGACTG 1 TTTTGATTGA 1 V Y D TGTCTATGAC 1	139 140 207 1200 1320 231 1380 231 1440 251 1500 258 1560 1620 1680 1740 1860 266 1920
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTTC TTGATAGTTA CTATTGTTTC TCTATTGCTT S S R C	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TAAATGCATG TGTGTACATA	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA # intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA CTTTTTCGAA GTGTATCCGT int ATACACATAC S L N	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTG ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 # F Y ATACAGTTTT H A M V	A TGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\parallel Q$ ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTG GTTAATTGTG Y S G ACTACAGTGG V T G	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AACATGACTG 1 TTTTGATTGA 1 V Y D TGTCTATGAC 1 Y G S	139 140 207 1200 1320 231 1380 231 1440 251 1500 258 1560 1
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTTC TTGATAGTTA CTATTGTTTC TCTATTGCTT S S R C TCATCACGAT	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TAAATGCATG TGTGTACATA TGTTCTGTTC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA # intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA GTGTATCCGT int ATACACATAC S L N TAGTCTTAAC	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTG ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 #F Y ATACAGTTTT H A M V CACGCAATGG	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\exists$ Q ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTG GTAATTGTG Y S G ACTACAGTGG V T G TAGTCACTGG	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1	139 140 207 200 220 2320 231 231 231 231 231 250 258 560 620 258 560 620 258 560 266 266 266 266 266 266 266 2
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTTC TTGATAGTTA CTATTGTTTC TCTATTGCTT S S R C TCATCACGAT Y N G K	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TAAATGCATG TGTTCTGTTC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA # intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA CTTTTTCGAA GTGTATCCGT int ATACACATAC S L N TAGTCTTAAC L A K	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTG ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 ╣F Y ATACAGTTTT H A M V CACGCAATGG N S ╠ intr	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\exists Q$ ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTG GTTAATTGTG Y S G ACTACAGTGG V T G TAGTCACTGG on-6	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 AATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1 AAAGCCTATGA 1	139 140 207 200 220 230 231 230 231 231 233 2440 2500 258 1560 1620 1660 266 1920 286 1920 286 1920 298
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTTC TTGATAGTTA CTATTGTTTC TCTATTGCTT S S R C TCATCACGAT Y N G K	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TAAATGCATG TGTTCTGTTC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA # intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA GTGTATCCGT int ATACACATAC S L N TAGTCTTAAC L A K GCTGGCCAAG	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTG ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 ╣F Y ATACAGTTTT H A M V CACGCAATGG N S ⊫ intr AATAGGTTCG	A TGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\exists$ Q ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTG GTAACTGTG Y S G ACTACAGTGG V T G TAGTCACTGG on-6 CTCAGCTGTC	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1 AAAAGCCTATA 1 AAAAGCCTATA 1 AACATATGAC 1 Y G S ATACGGATCA 1	139 140 207 200 220 230 231 230 231 230 250 250 250 256 1560 1660 266 266 286 286 286 286 286 286
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTC TTGATAGTTA CTATTGTTC TGGAGGTCCT S S R C TCATCACGAT Y N G K TACAATGGGA	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TGGCACCAAG TAAATGCATG TGTTCTGTTC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA # intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA GTGTATCCGT int ATACACATAC S L N TAGTCTTAAC L A K GCTGGCCAAG TAATTCACAC	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTG ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GACAGTGGAT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 ╣F Y ATACAGTTTT H A M V CACGCAATGG N S ╠ intr AATAGGTTCG ACATATACAG	ATGTATGATG PFVG CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\exists$ Q ACACACAGCA VSI TGGTGTCAAT VSV CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTGTG GTAACTGTG YSG ACTACAGTGG VTG TAGTCACTGG on-6 CTCAGCTGTC CATATACAGC	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1 AACATGACTG 1 Y Y D TGTCTATGAC 1 Y G S ATACGGATCA 1 ACATATTGTT 2 ATCACTCACC 2	139 140 207 200 220 230 231 230 231 230 231 230 231 250 256 256 256 256 256 256 256 266 26
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTC TTGGAGGTCCT S S R C TCATCACGAT Y N G K TACAATGGGA CTGGGTTGC CGCATGTACA	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TGATCTGTTC S S S GCTCTAGTAG K Y W AAAAATACTG CCCTCTCAC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA # intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA GTGTATCCGT int ATACACATAC S L N TAGTCTTAAC L A K GCTGGCCAAG TAATTCACAC	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGTG ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GACATGTACTT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 ╣F Y ATACAGTTTT H A M V CACGCAATGG N S ╠ intr AATAGGTCG ACATATACAG ATTGAGACTT	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\exists$ Q ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTGTG Y S G ACTACAGTGG V T G TAGTCACTGG On-6 CTCAGCTGTC CATATACAGC	F L I CTTTTTTGTA 1 K $\parallel$ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1 AAAGCCTATGA 1 TTTGATTGA 1 V Y D TGTCTATGACT 1 Y G S ATACGGATCA 1 ACATATTGTT 2 ATCACTCACC 2 ATACCGAC	139   140   207   1200   1200   1200   1200   1200   1200   1200   1200   1300   1300   1300   1300   1300   1300   1300   1300   1300   1300   1300   1300   1500   1500   1500   1500   1500   1500   1500   1500   1500   1500   1500   1500   1600   1000   1000   1000   1000
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTTC TTGATAGTTA CTATTGTTTC TCTATTGCTT S S R C TCATCACGAT Y N G K TACAATGGGA CTGGGTTGC CGCATGTACA	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TAAATGCATG TGTGTCTGTTC S S S GCTCTAGTAG K Y W AAAAATACTG CCCTCTCAC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA # intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA CTTTTTCGAA GTGTATCCGT int ATACACATAC S L N TAGTCTTAAC L A K GCTGGCCAAG TAATTCACAC	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGT ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 ╣F Y ATACAGTTTT H A M V CACGCAATGG N S ╠ intr AATAGGTTCG ACATATACAG ATTGAGACTT N S C Y	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\exists$ Q ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTGTG Y S G ACTACAGTGG V T G TAGTCACTGG on-6 CTCAGCTGTC CATATACAGC TTACCTTCTT V M M	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATGA 1 V Y D TGTCTATGACT 1 V Y D TGTCTATGAC 1 Y G S ATACGGATCA 1 ACATATTGTT 2 ATCACTCACC 2 ATATCTGACT 2	139   140   207   200   2200   230   231   380   231   440   251   560   620   680   258   660   258   660   2980   2980   2980   2160   314
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTTC TTGATAGTTA CTATTGTTTC TCTATTGCTT S S R C TCATCACGAT Y N G K TACAATGGGA CTGGGTTGC CGCATGTACA intron-6	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TAAATGCATG TGTTCTGTTC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGGTGTTTCA F intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA CTTTTTCGAA GTGTATCCGT int ATACACATAC S L N TAGTCTTAAC L A K GCTGGCCAAG TAATTCACAC CAATATGCCG N W G	TGGAGGCAAT S A Y CAGTGCATAT ATCGAACAGT GTTGTTGTGT ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 ⋕ F Y ATACAGTTTT H A M V CACGCAATGG N S ⊫ intr AATAGGTCG ACTTATACAG N S ∯ intr AATAGGTTCG ACATATACAG	A TGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTG ntron-4 $\exists$ Q ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTGTG Y S G ACTACAGTGG V T G TAGTCACTGG On-6 CTCAGCTGTC CATATACAGC TTACCTTCTT V M M	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 TTCGGGCAAT 1 AAAGCCTATGA 1 TTTGATTGA 1 V Y D TGTCTATGAC 1 Y G S ATACGGATCA 1 ACATATTGTT 2 ATCACTCACC 2 A R N CCCCCCADC 2	139   140   207   1200   1200   1200   1380   231   140   251   1500   1600   2980   20100   2160   2160   2160   2160   2160   2160   2160
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTTC TTGATAGTTA CTATTGTTTC TCTATTGCTT TGGAGGTCCT S S R C TCATCACGAT Y N G K TACAATGGGA CTTGGGTTGC CGCATGTACA intron-6 CATACACACA	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TAAATGCATG TGTTCTGTTC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGGTGTTTCA F intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA CTTTTTCGAA GTGTATCCGT int ATACACATAC S L N TAGTCTTAAC L A K GCTGGCCAAG TAATTCACAC CAATATGCCG N W G TAACTGGGGGT	TGGAGGCAAT S A Y CAGTGCAATAT ATCGAACAGT GTTGTTGTGTG ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 ∜F Y ATACAGTTTT H A M V CACGCAATGG N S ∲ intr AATAGGTCG ACATATACAG ATTGAGACTT N S G Y AACAGTGGCT	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\exists$ Q ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTG GTAATTGTG Y S G ACTACAGTGG V T G TAGTCACTGG On-6 CTCAGCTGTC CATATACAGC TTACCTTCTT V M M ATGTGATGAT	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTTGGGGCAAT 1 AAAGCCTATA 1 AAAGCCTATGA 1 V Y D TGTCTATGACT 2 ATACGGATCA 1 ACATATTGTT 2 ATCACTCACC 2 ATATCTGACT 2 A R N GGCCCGCAAC 2	139   140   207   200   2200   231   380   231   440   251   560   258   560   258   560   286   2980   2980   2160   314   220
CGCAGTTCCT V I A TGTCATCGCT TAATAATTAT ATGTCTGCAA TTGTGTTTGT N Y N S AACTATAACA S E S D AGTGAGTCTG G A N S GGTGCTAACA CGTATGGTTG CCGACTTTC TTGATAGTTA CTATTGTTTC TCTATTGCTT TGGAGGTCCT S S R C TCATCACGAT Y N G K TACAATGGGA CTTGGGTTGC CGCATGTACA intron-6 CATACACACA K Y N Q	TACGGTAACC N E G V AACGAGGGGG TCTTCATCAC CTCTTAGAAT AGACTATGTT K Y K GTAAATACAA L Q A ACTTACAAGC A F R GTGCCTTCAG CCGTTACTTA CACGCATGTC TTGCACCAAG TAAATGCATG TGTTCTGTTC	ACGGTTGTCA D Q D TCGATCAGGA AACAGTAATT TACCTATTCT TACTTTAGCC G T S AGGTACATCA A V S AGCTGTTTCA F intron-5 GGTAAGACAG ATGTTTGCCT AGTTACAAAA TCGCTATTCA CTTTTTCGAA GTGTATCCGT int ATACACATAC S L N TAGTCTTAAC L A K GCTGGCCAAG TAATTCACAC CAATATGCCG N W G TAACTGGGGT A T D	TGGAGGCAAT S A Y CAGTGCAATAT ATCGAACAGT GTTGTTGTGTG ACTTCCCTGC M S G M ATGTCGGGTA N V G P AACGTTGGCC ACTCACTCTA TAATTACTTT GAAGACGTTT GGCATGAAAA TGCATGTATG AGGTTCATAC ron-5 ╣F Y ATACAGTTTT H A M V CACGCAATGG N S ╠ intr AATAGGTTCG ACATATACAG ATTGAGACTT N S G Y AACAGTGGCT A S Y P	ATGTATGATG P F V G CCATTTGTTG GTCCTAATAA TTGTTTGTTT ntron-4 $\parallel Q$ ACACACAGCA V S I TGGTGTCAAT V S V CTGTATCTGT ATTTAGTAAT CCACTCTACT TCACATTGTG GAAGCACAAA CCATGAATTG GTAATTGTG Y S G ACTACAGTGG V T G TAGTCACTGG On-6 CTCAGCTGTC CATATACAGC TTACCTTCTT V M M ATGTGATGAT T L - CCACCCTTT	F L I CTTTTTTGTA 1 K ⊫ in-4 GAAAGGTGAA 1 ACCAACTGTA 1 GTTTGTTTGT 1 S S C ATCCAGCTGC 1 K S G CAAAAGTGGT 1 A I D TGCTATTGAT 1 TAGCTGTTAG 1 ATTAATGACG 1 TTCGGGCAAT 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATA 1 AAAGCCTATGAC 1 V Y D TGTCTATGAC 1 Y G S ATACGGATCA 1 ACATATTGT 2 ATCACTCACC 2 ATATCTGACT 2 A R N GGCCCGCAAC 2	139   140   207   200   2200   231   380   231   440   251   560   258   560   286   2920   286   2980   211   202   211   251   560   258   560   2920   286   2980   2160   314   2200   3200

Fig. 2. Gene organization of silicatein. The deduced translation of the nt sequence into the polypeptide is given above the nucleotides. The borders of the introns are indicated.



Fig. 3. Preparation of the recombinant silicatein. A: The S. domuncula silicatein was inserted into the expression vector pQE-30 via the BamHI (5'-end) and the HindIII (3'-end). Between oligohistidine (6xHis) and silicatein a nucleotide stretch corresponding the enterokinase recognition site [EK-rec. seq.] with the cleavage terminus [Lys  $\downarrow$ ] was inserted. The start ATG as well as the stop TAA codon for transcription are indicated. The position at which the enterokinase cleaved the recombinant fusion protein is indicated. Two constructs of silicatein cDNA have been inserted: For the preparation of the long form of silicatein (comprising  $aa_{17}$  to  $aa_{330}$ ), the stretch spanning nt<sub>49</sub>-nt<sub>993</sub> was selected, while for the short form of silicatein (aa<sub>113</sub> to

aa<sub>330</sub>) the nt<sub>337</sub>-nt<sub>993</sub> part of the sequence was used (Fig. 1A). **B**: Expression of the His-silicatein fusion protein. The expression of the short form of silicatein is shown. Aliquots of lysates from bacteria, which remained noninduced (-IPTG; lane d) or had been induced with IPTG (+IPTG; lane c) as well as the purified fusion protein (lane a) were analyzed by NaDodSO<sub>4</sub>/PAGE (15% gels) as described in text. In lane b the fraction which had not been adsorbed to the Ni-NTAagarose resin was analyzed. The fusion protein was subjected to digestion with recombinant enterokinase. The arrow points to the band corresponding to the recombinant silicatein protein. Marker proteins (M) were run in parallel.

phous silicate. To examine whether r-silicate in hydrolyses protein, the in situ activity after separation of the protein by NaDodSO<sub>4</sub>/PAGE was monitored. Gelatin was used as substrate for the detection of enzymatic activity.

The short form of r-silicatein was separated by NaDodSO<sub>4</sub>/PAGE using 10% polyacrylamide gels and 0.1% NaDodSO<sub>4</sub> according to Laemmli (1970). Bacterial extracts, obtained from *E. coli* which had been induced with IPTG (Fig. 3B, lane c) or had not been induced (Fig. 3B, lane d), were size-separated. Two types of gels were used: 1) a gel which did not contain gelatin (Fig. 5A) or 2) a gel which contained 0.1% gelatin (Fig. 5B). The gel without gelatin was immediately stained with Coomassie brilliant blue (Fig. 5A) to demonstrate the appearance of the recombinant sili-

catein in the extracts from IPTG-treated *E. coli* (Fig. 5A, lane a). In parallel, a zymogram was prepared under the same gel separation conditions using a gel which was supplemented with 0.1% gelatin (Fig. 5B); after incubation and subsequent staining of the gel the band corresponding to the r-silicatein protein was unstained (Fig. 5B, lane a), indicating that the gelatin was digested at that position. This result indicates that r-silicatein also has proteolytic activity.

Inhibition studies were performed to support this conclusion. As outlined before (Krasko et al., 1997), following the protocols described (Beynon and Bond, 1989), the zymograms were incubated in Tris buffer, supplemented with the cysteine protease inhibitors *p*-chloromercuriphenylsulfonic acid (1 mM), iodoacetic acid (1 mM), *N*-ethylmaleimide (1 mM), E-64 (10  $\mu$ M;

CELL BIOLOGY OF SPONGE SILIFICATION



Fig. 4. Dependence of the activity of silicatein on Fe(+++). The assay was performed with the two r-silicatein preparations (long as well as short form) using 4.5 mM tetraethoxysilane (TEOS) as substrate and, where indicated, MgSO<sub>4</sub> or Fe(+++)citrate. After incubation the amount of polymerized silica was quantified after treatment with NaOH; the released silicic acid was determined (n = 5; ±SD). The values are given in g of silicon (polymerized), and if enzyme was added, correlated with 1 mg of protein; means of five separate experiments are presented.

Fig. 5C, lane c) or cystatin (1 mg/ml; lane b). All inhibitors reduced the proteolytic activity of silicatein in the zymogram assay. From these data we concluded that silicatein has the characteristic activity of cysteine peptidase(s).

## BIOLOGICAL ROLE OF SILICON Localization in Sponges

Silica is stored in sponges in the form of spicules. Two intracellular compartments which contain silica have been identified: the extranuclear and the intranuclear one.

**Spicules.** The spicules are formed by scleroblasts (reviewed in Simpson, 1984). The formation starts around an axial filament (Fig. 6A). Spicules can grow at up to 5  $\mu$ m/h (Weissenfels, 1989). The first microscopical study was performed by Lieberkühn in 1856 (Fig. 6B).

**Nucleus.** The fact that siliceous spicules also occur in the nuclei of cells of the freshwater sponge *Spongilla lacustris* (Fig. 6D) was surprising (Imsiecke et al., 1995). Digestion studies showed that the crystals, observed by Lieberkühn (1856) (Fig. 6C), consist of silica; they are not digested by proteinase K but by hydrofluoric acid (Imsiecke et al., 1995). The physiological role of the intranuclear silicon rods is unknown. Since they show a high metabolic rate they are assumed to be involved in the scavenging of other ions or in the maintenance of osmotic pressure within the cell (Imsiecke et al., 1995).

#### **Effect on Size of Primmorphs**

**Primmorphs.** Primmorph cultures have been established from the demosponges S. domuncula (Custodio et al., 1998; Müller et al., 1999), Dysidea avara (Müller et al., 2000), and occasionally G. cydonium. In brief, tissue samples were dissociated to single cells in  $Ca^{2+}$ -



Demonstration of the proteolytic activity of silicatein. Fig. 5. NaDodSO<sub>4</sub>/PAGE (10% gel) was performed in the presence of NaDodSO<sub>4</sub>. The gels were prepared in the absence (**A**) or together with 0.1% gelatin (Sigma, St. Louis, MO; G-2500) (**B**). Bacterial extracts, after expression of the short form of silicatein (see Fig. 3), were applied to the gels at a concentration of 20  $\mu g$  per lane. The samples were obtained either from bacteria which had been induced with IPTG for 4 hours (+IPTG; lane a) or were cultivated in the absence of this inducer for 4 or 6 hours (-IPTG; lanes b, c). A: The gels which did not contain gelatin were immediately stained with Coomassie brilliant blue; in parallel, protein markers were separated. B: Alternatively, the bacterial samples were applied in the same order onto the gelatin-containing gels and separated. After the run the gels were washed in Tween-20 solution and twice in a 100 mM Tris/HCl buffer (pH 8.0; 0.01 M  $CaCl_2$ ). Finally, the gels were incubated for 18 hours in this Tris buffer at 20°C and then stained with Coomassie brilliant blue. The arrow indicates the location where the gelatin had been digested and remained unstained by Coomassie blue. C: Inhibition studies. The extracts from bacteria which had been induced with IPTG were applied at equal concentrations to the three slots and electrophoresed; lanes a-c. After size separation the zymograms were cut into slices and incubated separately in the above-mentioned 100 mM Tris/HCl buffer in the absence of any further component (lane a), or in this buffer supplemented with 1 mg/ml of cystatin (lane b) or  $10~\mu M$  of E-64 (lane c). Then the zymograms were stained in Coomassie brilliant blue.

and  $Mg^{2+}$ -free artificial seawater containing EDTA. Subsequently, the cells were transferred to natural seawater and kept at 16°C. When aggregates formed, they were incubated in seawater containing 0.2% of RPMI1640-medium (Le Pennec et al., 2003), where they form primmorphs.

The primmorphs are defined by the presence of proliferating cells and a characteristic histology (Custodio et al., 1998; Müller et al., 1999). During the phase of primmorph formation the aggregates contract to round-shaped bodies, 1–5 mm in diameter, leaving be-







Fig. 7. Formation of primmorphs from dissociated cells of *S. domuncula*. A: When the cells are transferred to seawater, supplemented with 0.2% of RPM11640, primmorphs of a size between 1–7 mm are formed (incubation period 5 days). The large-size primmorphs (one specimen is shown) are very rare. B: Incubation for 5 days of the primmorphs in this medium, supplemented with 60  $\mu$ M silicate and 30  $\mu$ M Fe(+++) (as ferric citrate); size of the primmorphs ranges

from 6–12 mm. C: Incubation of the primmorphs in seawater/ RPMI1640/silicate/30  $\mu$ M Fe(+++) in the current chamber at a water velocity of 20 cm/sec. The primmorphs under these conditions often show an oscule-like opening (arrowhead). Scale bar = 5 mm. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

hind detritus and dead cells. In the initial phase the primmorphs remain round-shaped but after 3-4 weeks many of them attach to the bottom of the culture dish. Cross sections through the primmorphs revealed that the inner cells are surrounded by an almost complete single-cell layer of epithelial-like cells. The cells forming the epithelium of the primmorphs are pinacocytes, as judged from their flattened, fusiform extensions and their prominent nucleus. In the absence of any additional factor, the diameter of the primmorphs ranged from 1-7 mm (Fig. 7A); the usual diameter of the round-shaped primmorphs is  $\sim 4 \text{ mm}$ .

The growth conditions were optimized by supplementing the natural seawater, including 0.2% of RPMI1640-medium, with silicate and Fe(+++). Furthermore the primmorphs were kept in a current chamber (Le Pennec et al., 2003). If the culture medium was enriched with 60  $\mu$ M silicate and 30  $\mu$ M Fe(+++), the primmorphs grew; 6–12 mm-long 3D-

aggregates were usually obtained (Fig. 7B). When the incubation was performed under higher water current, the primmorphs frequently formed oscule-like openings (Fig. 7C).

**Morphogenetic Activity.** Two inorganic factors have morphogenetic activity on primmorphs: silicate and ferric iron.

Effect of Silicate. In a preliminary series of experiments the effect of silicate on spicule formation in vitro was determined semiquantitatively (Krasko et al., 2000, 2002). When the primmorphs were incubated in the absence of exogenous silicate for 5 days no spicules were seen. In contrast, when the primmorphs were incubated together with  $50-250 \mu$ M of silicate, bundles of spicules were found. The content of polymerized silicon also increased in primmorphs in the presence of silicate. In the absence of additional silicate in the medium the primmorphs contained little polymerized silicon (0.1 mg/0.4 g aliquots<sup>-1</sup>); however, when they



Fig. 8. Effect of silicate and ferric iron on the steady-state expression level of selected genes in primmorphs. Single cells were transferred after dissociation into Ca<sup>2+</sup> and Mg<sup>2+</sup>-free artificial seawater containing EDTA into seawater, enriched with 0.2% of RPMI1640. During primmorph formation the cultures remained either without any additional component for 0 (control; "Con") to 5 days or they were incubated together with silicate 60  $\mu$ M silicate (**A**,**B**). In a further series, the seawater/RPMI1640/silicate medium was additionally supplemented with 30  $\mu$ M Fe(+++) (**C**-**E**). Primmorphs were collected, RNA was extracted, and the level of transcripts was estimated by Northern blot analysis, as described previously (Krasko et al., 2000,

were cultured in seawater medium enriched with  $60 \mu$ M silicate the content increased to 4.4 mg.

Several genes are expressed before or during spicule formation (Krasko et al., 2000, 2002). The effect of silicate on the steady-state expression level of silicatein and collagen in primmorphs has been described (Krasko et al., 2000). The genes for silicatein and collagen were cloned and used for Northern blot studies.

The concentration of silicate in the natural seawater used for the experiments was  $\approx 1 \ \mu$ M. In this medium the expression of silicate in primmorphs is hardly detectable (Fig. 8A). However, when the concentration of silicate was raised to 60  $\mu$ M, the expression of silicate strongly increased after 1 day, to reach a maximum after 5 days (Fig. 8A). Furthermore, the expression of the gene encoding collagen (Schröder et al., 2000) also responded to silicate (Fig. 8B). In natural seawater the expression of collagen was relatively low and amounted to  $\approx 15\%$  of that seen in the presence of 60  $\mu$ M exogenous silicate.

2002). The following cDNAs were used as probes: (A) silicatein, (B) collagen, (C) ferritin, (D) septin, or (E) scavenger. After blotting, the intensities of the hybridized bands were semiquantitatively determined using the GS-525 Molecular Imager (Bio-Rad, Hercules, CA). The relative degree of expression was correlated with that seen for the maximal expression after 5 days in the presence of silicate, which was set to 1.0 (A,B). In the studies to estimate the effect of Fe(+++) (C-E), the values are correlated with the expression at day 5 in primmorphs cultured in the presence of silicate and Fe(+++); modified according to Krasko et al. (2002).

**Effect of Ferric Iron.** As outlined above, addition of Fe(+++) results in the formation of larger primmorphs (Fig. 7).

In addition, ferric iron affects the expression of the following genes: *ferritin*, the iron storage protein (reviewed in Chasteen and Harrison, 1999), a putative septin polypeptide, and a scavenger receptor. *Septin* molecules are involved in the completion of cytokinesis (reviewed in Xie et al., 1999). Since the spicules in some Demospongiae are formed extracellularly (Uriz et al., 2000) it has been assumed that cell surface receptors are involved in the growth guidance/control of the siliceous spicules (unpublished). However, the expression of the sponge integrin receptor did not depend on extracellular iron, while a considerable upregulation of a scavenger receptor was observed after addition of iron to primmorph cultures. This *scavenger receptor* was also cloned (Krasko et al., 2002).

The effect of iron on the expression level of these genes, *ferritin*, *septin*, and *scavenger receptor*, in prim-



Fig. 9. Proposed biochemical steps involved in the spicule formation in Demospongiae. It can be suggested that Fe(++) is taken up by sponge cells and undergoes oxidation to Fe(+++) via the ferroxidase of ferritin. Ferritin/Fe(+++) is translocated to the axial filament, where it is coated by silicatein. This enzyme, which is mediating the formation of amorphous silica, is very likely activated by iron. This skeletal element, amorphous silica, is surrounded by the scavenger receptor and finally embedded into the collagen matrix.

morphs was also determined by Northern blot experiments. Primmorphs were transferred to seawater and supplemented with 60  $\mu$ M silicate and 30  $\mu$ M Fe(+++). Without Fe(+++) the expression of the 0.9 kb *ferritin* increased slightly after 5 days (Fig. 8C). In contrast, if RNA was isolated from primmorphs which had been cultivated in the presence of Fe(+++), *ferritin* expression increased after 3 days, and continued to rise for 5 days. In parallel, samples were hybridized with the *septin* probe. Again, the expression of the *septin* gene strongly increased in cultures treated with Fe(+++), but not in its absence (Fig. 8D). In addition, increased *scavenger receptor* expression was seen after exposure to iron for 3 and 5 days, but not in the iron-free controls (Fig. 8E).

These data indicate that, besides silicate, iron is also essential for proliferation and morphogenesis of sponge primmorph cells. This view was also substantiated by in situ hybridization studies (Le Pennec et al., 2003). Evidence showed that in the initial phase of reorganization iron enhances proliferation followed by a morphogenetic effect via its accelerative and supportive influence on spicule synthesis.

#### CONCLUSION

It is well established that silicon and iron are major elements in sponge spicules (Vinogradov, 1953). Following the studies performed with *T. aurantium* (Cha et al., 1999) and *S. domuncula* (Krasko et al., 2000), it appears to be well established that silicatein is the major enzyme involved in spicule formation.

Applying the primmorph system from *S. domuncula*, the data (Krasko et al., 2002) suggested that besides silicon, iron ions play an important role in the initiation of spicule formation. Iron upregulates ferritin in sponge cells, which may bind the intracellular Fe(++). Subsequently, Fe(++) is oxidized to Fe(+++) under consumption of oxygen (Ha et al., 1999); this process is mediated by the ferroxidase site of ferritin. During this process apoferritin is converted to ferritin. Finally, iron may be deposited in ferritin as  $5Fe_2O_3 \cdot 9H_2O$  and then as the stable mineral  $\alpha$ -Fe<sub>2</sub>O<sub>3</sub> (see Chasteen and Harrison, 1999). The deposited mineralized iron, the ferric oxide nanoparticles, remain surrounded by the ferritin protein shell. In the iron-free state the ferritin molecules (apoferritin) comprise hollow spheres with an outer diameter of 12 nm and an inner diameter of 8 nm (Lawson et al., 1991). Therefore, ferritin can be considered as a shell that causes both the oxidation of iron ions and controls the growth of the iron mineral.

Recent studies applying the technique of in situ hybridization with S. domuncula revealed that ferritin is abundant in cells surrounding the canals within the sponge body (Le Pennec et al., 2003). The function of ferritin in those cells remains to be established. Based on analytical (Vinogradov, 1953) and functional data (Krasko et al., 2002) the deposition of iron in ferritin spheres may also be involved in the initiation of spicule formation (Fig. 9). Since the diameter of the axial filament, which also comprises the silicatein, is  $\approx 1 \ \mu m$ (Uriz et al., 2000) there is ample space for the ferritin shell. It may be assumed that these spheres may be covered by silicatein, which results in the initiation of spiculogenesis. Silicatein is finally activated by iron [as Fe(+++)]. This assumption is supported by the data, which indicate that Fe(+++) stimulates the activity of silicatein. In addition, iron within the ferritin shell is mobilized after reduction (Chasteen, 1998). Another gene, coding for a scavenger receptor, is apparently involved in the spicule formation. The expression studies suggest that after or during the synthesis of amorphous silica, the growing spicules are surrounded by the scavenger receptor, which might be considered a docking molecule for the collagen matrix into which the spicules are embedded (Fig. 9).

### ACKNOWLEDGMENT

We thank Claudia I. Müller for help in the analysis of the cathepsin.

## REFERENCES

Beynon RJ, Bond JS. 1989. Proteolytic enzymes—a practical approach. Oxford: IRL Press.

Brasier MD. 2000. The Cambrian explosion and the slow burning fuse. Science Progress 83:77–92.

- Brasier M, Green O, Shields G. 1997. Ediacarian sponge spicule clusters from southwestern Mongolia and the origin of the Cambrian fauna. Geology 25:303-306.
- Cha JN, Shimizu K, Zhou Y, Christianssen SC, Chmelka BF, Stucky GD, Morse DE. 1999. Silicatein filaments and subunits from a marine sponge direct the polymerization of silica and silicones in vitro. Proc Natl Acad Sci USA 96:361-365.
- Chasteen ND. 1998. Ferritin: uptake, storage, and release of iron. In: Sigel A, Sigel H, editors. Metal ions in biological systems. New York: Marcel Dekker. p 479–514. Chasteen ND, Harrison PM. 1999. Mineralization in ferritin: an effi-
- cient means of iron storage. J Struct Biol 126:182-194.
- Custodio MR, Prokic I, Steffen R, Koziol C, Borojevic R, Brümmer F, Nickel M, Müller WEG. 1998. Primmorphs generated from dissociated cells of the sponge Suberites domuncula: a model system for studies of cell proliferation and cell death. Mech Ageing Dev 105: 45 - 59
- Gamulin V, Skorokhod A, Kavsan V, Müller IM, Müller WEG. 1997. Experimental indication against blockwise evolution of metazoan protein molecules. Example: receptor tyrosine kinase gene from the sponge Geodia cydonium. J Mol Evol 44:242-252.
- Geigy Scientific Tables. 1984. Physical chemistry-composition of blood, vol. 3. Basel: Ciba-Geigy. Ha Y, Shi D, Small GW, Theil EC, Allewell NM. 1999. Crystal struc-
- ture of bullfrog M ferritin at 2.8 Å resolution: analysis of subunit interactions and the binuclear metal center. J Biol Inorg Chem 4:243 - 256
- Imsiecke G, Müller WEG. 1995. Unusual presence and intranuclear storage of silica crystals in the freshwater sponges Ephydatia muelleri and Spongilla lacustris (Porifera: Spongillidae). Cell Mol Biochem 41:827-832.
- Kennish MJ. 1994. Practical handbook of marine science. Boca Raton, FL: CRC Press.
- Krasko A, Gamulin V, Seack J, Steffen R, Schröder HC, Müller WEG. 1997. Cathepsin, a major protease of the marine sponge Geodia cydonium: purification of the enzyme and molecular cloning of cDNA. Mol Mar Biol Biotechnol 6:296–307.
- Krasko A, 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.
- Krasko A, Schröder HC, Batel R, Grebenjuk VA, Steffen R, Müller IM, Müller WEG. 2002. Iron induces proliferation and morphogenesis in primmorphs from the marine sponge Suberites domuncula. DNA Cell Biol 21:67-80.
- Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685.
- Lawson DM, Artymiuk PJ, Yewdall SJ, Smith JMA, Livingstone JC, Treffry A, Luzzago A, Levi S, Arosio P, Cesareni G, Thomas CD, Shaw WV, Harrison PM. 1991. Solving the structure of human H-ferritin by genetically engineering intermolecular crystal contacts. Nature 349:541-544.
- Le Pennec G, Perovic S, Ammar SMA, Grebenjuk VA, Steffen R, Brümmer F, Müller WEG. 2003. Cultivation of primmorphs from the marine sponge Suberites domuncula: morphogenetic potential of silicon and iron. A review. J Biotechnol 100:93-108.

- Lieberkühn N. 1856. Zur Entwicklungsgeschichte der Spongillen. (Nachtrag). Müller's Archive. p $399-\bar{4}1\bar{4}.$
- Mehl D, Müller I, Müller WEG. 1998. Molecular biological and palaeontological evidence that Eumetazoa, including Porifera (sponges), are of monophyletic origin. In: Watanabe Y, Fusetani N, editors. Sponge science-multidisciplinary perspectives. Tokyo: Springer. p 133 - 156
- Müller WEG. 1995. Molecular phylogeny of Metazoa (animals): monophyletic origin. Naturwiss 82:321-329.
- Müller WEG. 1998. Molecular phylogeny of Eumetazoa: genes in sponges [Porifera] give evidence for monophyly of animals. Prog Mol Subcell Biol 19:89-132.
- Müller WEG. 2001. How was metazoan threshold crossed: the hypothetical Urmetazoa. Comp Biochem Physiol 129(A):433-460.
- Müller WEG, Kruse M, Koziol C, Leys SP. 1998. Evolution of early Metazoa: phylogenetic status of the Hexactinellida within the phylum of Porifera [sponges]. Prog Mol Subcell Biol 21:141-156.
- Müller WEG, Wiens M, Batel R, Steffen R, Borojevic R, Custodio MR. 1999. Establishment of a primary cell culture from a sponge: primmorphs from Suberites domuncula. Mar Ecol Prog Ser 178:205–219.
- Müller WEG, Böhm M, Batel R, De Rosa S, Tommonaro G, Müller IM, Schröder HC. 2000. Application of cell culture for the production of bioactive compounds from sponges: synthesis of avarol by primmorphs from Dysidea avara. J Nat Prod 63:1077-1081.
- Müller WEG, Lorenz B, Krasko A, Schröder HC. 2002. Silicateinvermittelte Synthese von amorphen Silikaten und Siloxanen und ihre Verwendung. Offenlegungsschrift (Deutsches Patentamt); C 12 P3/00 [14.2.2002].
- Nelson DM, Tréguer P, Brezinski MA, Leynaert A, Quéguiner B. 1995. Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. Glob Biochem Cycles 9:359-372
- Schröder HC, Krasko A, Batel R, Skorokhod A, Pahler S, Kruse M, Müller IM, Müller WEG. 2000. Stimulation of protein (collagen) synthesis in sponge cells by a cardiac myotrophin-related molecule from Suberites domuncula. FASEB J 14:2022-2031.
- Shimizu K, Cha J, Stucky GD, Morse DE. 1998. Silicatein alpha: cathepsin L-like protein in sponge biosilica. Proc Natl Acad Sci USA 95:6234-6238.
- Simpson TL. 1984. The cell biology of sponges. New York: Springer.
- Uriz MJ, Turon X, Becerro MA. 2000. Silica deposition in Demospongiae: spiculogenesis in Crambe crambe. Cell Tissue Res 301:299-309.
- Vinogradov AP. 1953. The elementary chemical composition of marine organisms. New Haven: Sears Foundation.
- Weissenfels N. 1989. Biologie und Mikroskopische Anatomie der Süßwasserschwämme (Spongillidae). Stuttgart: Gustav Fischer.
- Xiao S, Yun Z, Knoll AH. 1999. Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. Nature 391:553-558.
- Xie H, Surka M, Howard J, Trimble WS. 1999. Characterization of the mammalian septin H5: distinct patterns of cytoskeletal and membrane association from other species. Cell Motil Cytoskel 43:52-62.
- Xue YS, Zho C, Tang T. 1998. "Animal embryos," a misinterpretation of Neoproterozoic microfossils. Acta Micropalaeontol Sin 16:1-4.