DOI: 10.1002/anie.200702413

Biomineralization

A Code for Lysine Modifications of a Silica Biomineralizing Silaffin Protein**

Manfred Sumper, Robert Hett, Gerhard Lehmann, and Stephan Wenzl*

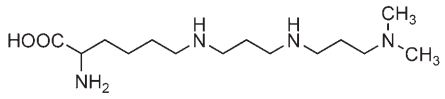
The biochemical processes guiding the formation of complex silica skeletons have mainly been studied in sponges[1–4] and diatoms.[5,6] Diatoms are a key taxon of marine phytoplankton and a major contributor to global carbon dioxide fixation.[7] Their silica-based cell walls display fine structures in the nanometer range which are precisely reproduced in each generation confirming a genetic control over this process.[8] Extreme mechanical stability is one of the remarkable properties of these constructions.[9] Silica formation and patterning in diatoms can be attributed to the influence of very specialized biopolymers. Silaffins, as well as long-chain polyamines, have been shown to be constituents of biosilica and to promote the formation of silica.[10,11] A model based on phase-separation processes involving polyamines is able to explain important aspects of silica patterning in *Coscinodiscus* diatoms.[12,13]

The species specificity in silica architectures may be reflected in species-specific structures of silaffins and polyamines. Indeed, species-specific structures of long-chain polyamines have been documented for a number of diatom species.[14] In contrast, the complete chemical structure of a silaffin protein is known only for a single diatom species. In this silaffin, all the serine residues are phosphorylated and all the lysine residues are either methylated or covalently linked with polyamines.[15] This set of modifications creates a highly zwitterionic polypeptide with the capability for self-assembly, the process is driven by ionic interactions and forms the matrix for silica formation.

Additional lysine derivatives modified by the formal addition of two aminopropyl units to the ε-amino group have been characterized in silaffins from another diatom species.[16] The extremely high degree of posttranslational modification present in silaffins prevented the collection of amino acid sequence data required for the cloning of additional silaffin genes. Therefore, the recent sequencing of the genome of the diatom *Thalassiosira pseudonana* greatly simplified the identification of silaffin genes and qualifies this species as a model organism in future research on silica biomineralization.[17,18] However, the chemical characterization of the posttranslational modifications in silaffins remains a challenging task. Therefore, we characterized the structures of all the modified lysines within the most abundant silaffin polypeptide (silaffin-3) from this species.

Previous work demonstrated the presence of unknown lysine derivatives in acidic hydrolysates from purified *T. pseudonana* silaffins.[18] Besides ε-*N*,*N*-dimethyl-lysine, two series of lysine derivatives have been found. Members within both series differ by 14 mass units indicating different degrees of methylation. The first series includes lysine derivatives with *m*/*z* 275, 289, 303, and 317. The second series displayed *m*/*z* values of 319, 333, and 347.

Electrospray ionization mass spectrometry of an acid hydrolysate from silaffin-3 revealed the dominance of the *m*/*z* 289 species and lesser amounts of the *m*/*z* 333 component (see Figure S1 in the Supporting Information). The mass of the former derivative could result from the addition of two aminopropyl units along with two methyl groups to generate lysine modification **1**. To verify this interpretation, the *m*/*z* 289 compound was converted into its fully methylated



m/*z* 289

1

derivative by reductive methylation with CD_2O. Methylation should convert this species into derivative **2** with *m*/*z* 353, which corresponds to the introduction of four CD_2H groups. Upon MS/MS analysis the pattern of fragments confirmed the assumed structure **2** (Figure 1a). Similarly, the other main lysine derivative (*m*/*z* 333) could result from the attachment of two aminopropyl units along with four methyl groups to a hydroxylysine basis. Hydroxylysine is found in diatom cell walls.[19] To verify this interpretation, the *m*/*z* 333 compound was also subjected to reductive methylation with CD_2O and subsequent acetylation. The resulting derivative displays a peak at *m*/*z* 423, which corresponds to the introduction of both an acetyl and three CD_2H groups. Upon MS/MS, the resulting fragments confirmed the existence of structure **3** (Figure 1b). This derivative contains a quaternary ammonium group at the site of the ε-amino group of the hydroxylysine moiety. Fragmentation preferentially occurs at the site of the positively charged nitrogen atom.

[*] Prof. Dr. M. Sumper, R. Hett, G. Lehmann, Dr. S. Wenzl
Lehrstuhl Biochemie I
Universität Regensburg
93040 Regensburg (Germany)
Fax: (+49) 941-943-2936
E-mail: manfred.sumper@vkl.uni-regensburg.de
Homepage: http://www.biologie.uni-regensburg.de/Biochemie/Sumper/

[**] The work was supported by Volkswagenstiftung and the Fonds der Chemischen Industrie. We thank R. Deutzmann and E. Hochmuth for taking the MS spectra and peptide sequencing.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

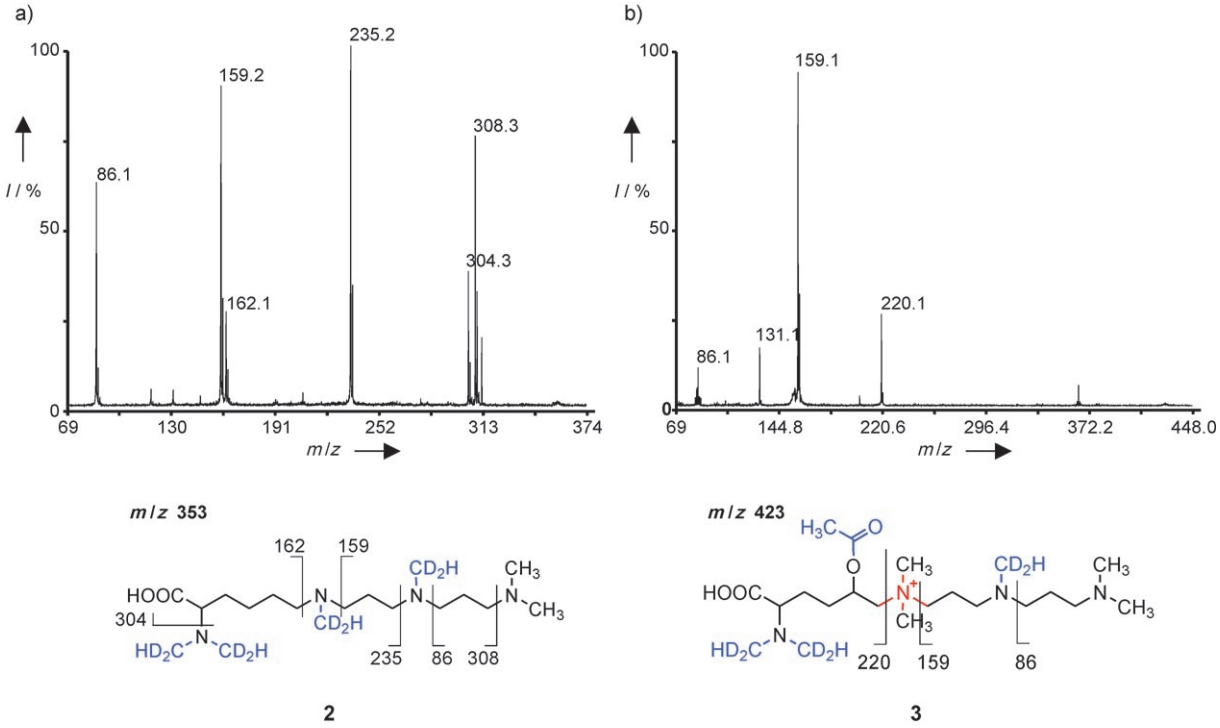


Figure 1. Structural analysis by mass spectrometry (MALDI-TOF-TOF fragmentation) of the lysine derivatives *m*/*z* 289 and 333 found in acid hydrolysates from silaffin-3 after reductive methylation with CD_2O and acetylation. a) Lysine derivative *m*/*z* 289 was converted into *m*/*z* 353 indicating the introduction of four CD_2H groups. b) Lysine derivative *m*/*z* 333 was converted into *m*/*z* 423 indicating the introduction of three CD_2H groups and one acetyl group. The interpretations of the observed mass fragments are given in the structures shown (**2** and **3**).

The polypeptide chain of silaffin-3 contains a total of 33 lysine residues. Thirty of these residues are embedded in a defined tetrapeptide sequence K-(A/S/Q)-X-K (boxed in Figure 2). This regularity suggests the existence of rules for the introduction of a given type of lysine modification. To decipher this possible code, we undertook a detailed chemical analysis of a large collection of silaffin-3 peptides using mass spectrometry as well as Edman sequencing. Peptides were obtained by cleavage of the polypeptide chain by treatment with the methionine specific reagent cyanbromide (BrCN) as well as by cleavage with the site specific endoproteinases Asp-N (and subsequent chymotrypsin cleavage in a few cases) and Glu-C. The peptides which were chosen for a detailed analysis are indicated in Figure 2. From this set of peptides, 14 out of the 15 tetrapeptide sequences K-(A/S/Q)-X-K could be characterized with respect to their modification pattern.

Representative for this type of analysis, data collection, and interpretation for two typical peptides (peptide 26, *m*/*z* 1148, and peptide 7, *m*/*z* 1439; see Figure 2 and Supporting Information, Table S1) from silaffin-3 are shown in Figure 3. Peptide 26 contains the *m*/*z* 289 lysine derivative, whereas peptide 7 contains the quaternary lysine derivative (*m*/*z* 333). All the fragments obtained from peptide 26 can be interpreted by the structure **4** shown in Figure 3a. In this peptide, the N-terminal lysine in the K-A-E-K motif is modified by two aminopropyl units (4,8-diazaoctanyl-residue), whereas the C-terminal lysine is converted into ε-*N*,*N*-dimethyllysine. This common rule also holds for the structures determined for seven more peptides containing this tetrapeptide motif. However, six out of the fifteen K-(A/S/Q)-X-K sequences contain the lysine derivative with the quaternary ammonium

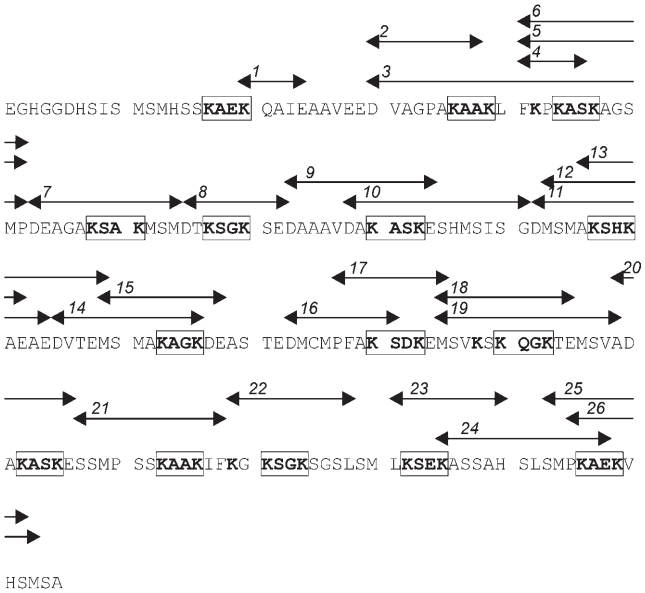


Figure 2. The amino acid sequence of silaffin-3 from *T. pseudonana*. The K-(A/S/Q)-X-K tetrapeptide motif is boxed. Arrows and numbers indicate all the peptides which were obtained and purified by reverse-phase HPLC for the characterization of the lysine modifications present in silaffin-3. The properties of these peptides are summarized in Table S1 of the Supporting Information.

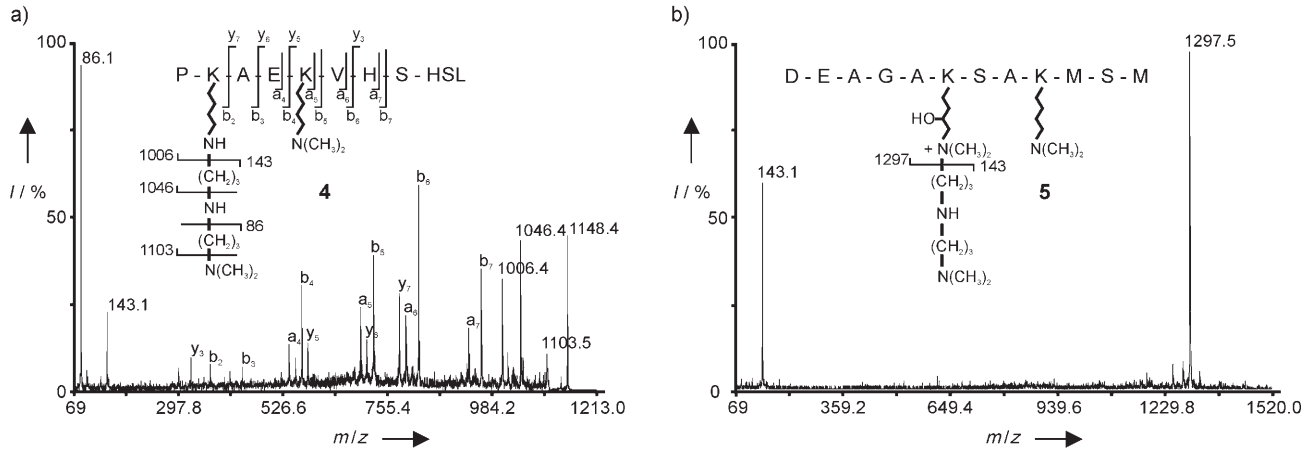


Figure 3. a) MALDI-TOF-TOF analysis of peptide 26. The fragmentation pattern and its interpretation are shown for the peptide *m/z* 1148 (**4**). b) From a set of silaffin-3 peptides obtained by cleavage with Asp-N endoproteinase, the peptide 7, *m/z* 1439 was analyzed by MALDI-TOF-TOF. The interpretation of the two fragments produced by this peptide containing a quaternary ammonium modification is shown (**5**).

group. These peptides produce only two main fragments in the MALDI-TOF-TOF analysis because the quaternary ammonium group is the preferred site of cleavage. Internal proton transfer from the adjacent secondary amino group favors the specific site of cleavage.[16] This fact is demonstrated in Figure 3b for peptide 7 (structure **5**). Therefore, all these particular peptides were also subjected to Edman degradation for their unequivocal identification. Table S1 of the Supporting Information summarizes the data of all the peptides analyzed.

All structural data obtained are schematically summarized in Figure 4. The type of lysine modification at a given position within the silaffin-3 polypeptide is indicated by a color code. From these results, rules emerge for the action of the modification machinery. 1) In the K-(A/S/Q)-X-K tetrapeptide motifs, the N-terminal lysine is modified by the diazaoctanyl moiety. This rule holds for 13 out of 14 analyzed motifs (except for position 182 in the polypeptide chain). 2) For tetrapeptide motifs which are not clustered (i.e. at least separated by six amino acid residues), the C-terminal lysine becomes dimethylated. This rule holds for 7 out of 8 corresponding motifs (except for position 116 in the polypeptide chain). 3) If a single lysine is located close to a tetrapeptide motif (i.e. separated by no more than one or two amino acids), both lysines of the adjacent tetrapeptide are modified by the diazaoctanyl moiety. This rule holds for 5 out of 5 corresponding motifs. 4) If two tetrapeptide motifs are clustered (i.e. separated by less than six amino acid residues) both the terminal lysine residues in this cluster are converted into the *m/z* 333 lysine variant bearing the quaternary ammonium group. This situation is found at positions 36/47, 58/70, and 163/174 (Figure 4).

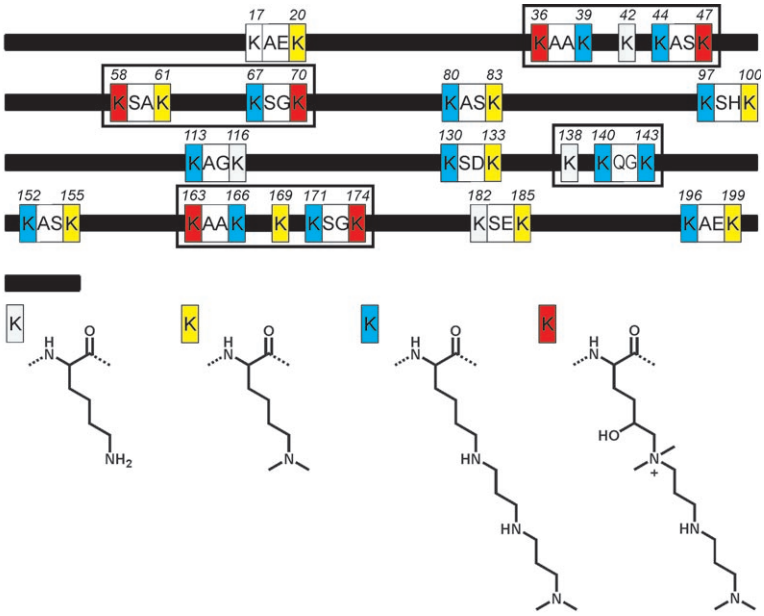


Figure 4. The nature and distribution of lysine modification within the silaffin-3 polypeptide. The black line represents the polypeptide chain and the positions of all the lysines are labeled. The type of lysine modification is indicated by the color code explained by the chemical structures shown (see text for details).

The existence of such rules implicates the action of an enzymatic machinery in diatoms which transforms the amino acid sequence information within a silaffin polypeptide into the corresponding modification pattern. Thus, the primary structure of a silaffin polypeptide dictates the site-specific introduction and spacing of positive charges. The charge density in turn would influence the physicochemical behavior of these molecules. The modification machinery introduces clusters of positive charges provided by polyamine residues. These charges are likely to influence the (self-)assembly behavior by electrostatic interactions with anionic phosphate groups also present in silaffins.[15,18] Since, at least in vitro,

silaffin assemblies are able to guide silica formation, this code may represent a link between species-specific silica morphologies and their genetic control. To some extent, this scenario is reminiscent of the pattern of histone modifications, called the histone code, which influences self-assembly processes causing transitions between chromatin states.[20–22]

Experimental Section

Culture Conditions: Thalassiosira pseudonana clone CCMP1335 was grown in an artificial seawater medium according to the recipe from the North East Pacific Culture Collection (http://www3.botany.ubc.ca/cccm/index.html) at 18 °C in a 14 h light/10 h dark cycle at 5000–10000 lux.

Purification of silaffin-3: Silaffin-3 was purified as described elsewhere[18] with the modifications described in the Supporting Information. Deglycosylation was performed by HF-treatment as described elsewhere.[23]

Analysis of Lysine Derivatives: HF-treated silaffin-3 corresponding to 20 mg diatom shells was hydrolyzed in 1 mL 6 N HCl (Pierce; sequanal grade) at 110 °C for 20 h. The dried hydrolysate was dissolved in 50 mM sodium phosphate, pH 7.0, and subjected to reductive methylation with sodium cyanoborohydride and CD_2O as described elsewhere.[24] To remove salts and low-molecular-weight reaction products the methylated derivatives were applied to a Superdex-Peptide HR 10/30 column. The eluate containing the modified lysine residues was dried, dissolved in triethylamine (400 μL) and acetic anhydride (50 μL) and incubated at 30 °C for 80 min.

Peptide Mapping: HF-treated silaffin-3 (15 μg) was dissolved in 50 mM NH_4HCO_3, supplemented with specific endoproteinase (1 μg; Roche; sequencing grade) and incubated under the conditions recommended by the distributor. Additional cleavage was performed with BrCN by incubating HF-treated silaffin-3 (15 μg) in 70 % formic acid (100 μL) containing BrCN (20 mg mL^{-1}; 24 h; 25 °C). Peptides were separated by reverse-phase HPLC on a C18 column (EC 250/2 Nucleosil 100–5 C18MPN, Machery-Nagel) by application of an gradient (buffer A: 0,1 % trifluoroacetic acid in H_2O; buffer B: 0,09 % trifluoroacetic acid in 50 % acetonitrile; gradient 0–100 % buffer B in 50 min). The material of selected peaks was re-chromatographed, and sequenced by Edman degradation using an automated gas-phase sequencer (Applied Biosystems).

Mass Spectrometry: The molecular mass and fragmentation pattern of modified lysine residues and peptides were analyzed by MALDI-TOF-TOF (Applied Biosystems 4700 Proteomics Analyzer). Amino acid hydrolysates were analyzed by Electrospray ionization MS using an Ion Trap ESQUIRE LC instrument (Bruker, Billerica, MA). Samples were infused by a nanospray source in 50 % methanol, 0.5 % acetic acid.

Received: June 4, 2007
Revised: July 27, 2007
Published online: September 26, 2007

Keywords: biomineralization · diatoms · mass spectrometry · posttranslational modifications · silaffin proteins

[1] J. C. Weaver, D. E. Morse, *Microsc. Res. Tech.* **2003**, *62*, 356.
[2] J. Aizenberg, J. C. Weaver, M. S. Thanawala, V. C. Sundar, D. E. Morse, P. Fratzl, *Science* **2005**, *309*, 275.
[3] W. E. Muller, S. I. Belikov, W. Tremel, C. C. Perry, W. W. Gieskes, A. Boreiko, H. C. Schroder, *Micron* **2006**, *37*, 107.
[4] J. C. Weaver, J. Aizenberg, G. E. Fantner, D. Kisailus, A. Woesz, P. Allen, K. Fields, M. J. Porter, F. W. Zok, P. K. Hansma, P. Fratzl, D. E. Morse, *J. Struct. Biol.* **2007**, *158*, 93.
[5] M. Hildebrand, B. E. Volcani, W. Gassmann, J. I. Schroeder, *Nature* **1997**, *385*, 688.
[6] M. Sumper, E. Brunner, *Adv. Funct. Mater.* **2006**, *16*, 17.
[7] R. C. Dugdale, F. P. Wikerson, *Nature* **1998**, *391*, 270.
[8] F. Round, R. Crawford, D. Mann, *The Diatoms*, Cambridge University Press, Cambridge, **1990**.
[9] C. E. Hamm, R. Merkel, O. Springer, P. Jurkojc, C. Maier, K. Prechtel, V. Smetacek, *Nature* **2003**, *421*, 841.
[10] N. Kröger, R. Deutzmann, M. Sumper, *Science* **1999**, *286*, 1129.
[11] N. Kröger, R. Deutzmann, C. Bergsdorf, M. Sumper, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 14133.
[12] M. Sumper, *Science* **2002**, *295*, 2430.
[13] M. Sumper, *Angew. Chem.* **2004**, *116*, 2301; *Angew. Chem. Int. Ed.* **2004**, *43*, 2251.
[14] M. Sumper, G. Lehmann, *ChemBioChem* **2006**, *7*, 1419.
[15] N. Kröger, S. Lorenz, E. Brunner, M. Sumper, *Science* **2002**, *298*, 584.
[16] S. Wenzl, R. Deutzmann, R. Hett, E. Hochmuth, M. Sumper, *Angew. Chem.* **2004**, *116*, 6059; *Angew. Chem. Int. Ed.* **2004**, *43*, 5933.
[17] E. V. Armbrust, J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kroger, W. W. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, D. S. Rokhsar, *Science* **2004**, *306*, 79.
[18] N. Poulsen, N. Kröger, *J. Biol. Chem.* **2004**, *279*, 42993.
[19] T. Nakajima, B. E. Volcani, *Biochem. Biophys. Res. Commun.* **1970**, *39*, 28.
[20] B. M. Turner, *Cell* **1993**, *75*, 5.
[21] S. Henikoff, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5308.
[22] J. Mellor, *Cell* **2006**, *126*, 22.
[23] A. J. Mort, D. T. Lamport, *Anal. Biochem.* **1977**, *82*, 289.
[24] N. Jentoft, D. G. Dearborn, *Methods Enzymol.* **1983**, *91*, 570.

