Identifying N-terminal peptides by a combination of the edman procedures with a bromine isotope tag: Application to the silicateins

G. A. Fedorova, E. S. Korneva, A. S. Belikova, and M. A. Grachev*

Limnological Institute, Siberian Branch, Russian Academy of Sciences, 3 Ulan-Batorskaya Str, Irkutsk 664033, Russia

Received 12 July 2016; Accepted 9 September 2016

DOI: 10.1002/pro.3037

Published online 12 September 2016 proteinscience.org

Abstract: Silicateins are proteins found within spicules of siliceous sponges. They are analogs of proteinases cathepsins; they catalyze the transformation of silicic acid esters into biogenic silica ($\text{SiO}_2 \cdot n\text{H}_2\text{O}$), and are believed to take part in the processes of silicification in marine and freshwater sponges. Earlier studies by Kalyuzhnaya *et al.* revealed that the Baikal Sponge *Lubomirskia baicalensis* Pallas, 1773 (*L. baicalensis*) contains a gene 1988 bp long, which hosts four sequences that encode four mRNAs giving rise to silicateins $\alpha 1$, $\alpha 2$, $\alpha 3$ and $\alpha 4$ (SIL $\alpha 1$, SIL $\alpha 2$, SIL $\alpha 3$, SIL $\alpha 4$) whose predicted amino acid sequences are similar to those of the predicted sequences of marine sponge silicateins. However, the sequences of mature silicateins of *L. baicalensis* remained unknown, since their N-terminal peptides were not identified. We found the sequences of these N-terminal peptides using a combination of the Edman procedure, which involved reaction with phenylisothiocyanate, treatment with trifluoroacetic acid and trypsinolysis followed by treatment with 4-bromine-phenylisothiocyanate performed directly within polyacrylamide gel bands, and subsequent mass spectrometry. The N-terminal peptides are YAESIDWR (SIL $\alpha 1$), YVDSIDWR (SIL $\alpha 2$ and $\alpha 4$), and YADSLDWR (SIL $\alpha 3$). All mature silicateins of *L. baicalensis* had a length 217 amino acid residues.

Keywords: mass spectrometry; Edman; 4-bromophenyl isothiocyanate (4-Br-PITC); sponge; silicateins

Abbreviations: CHCA, α -cyano-4-hydroxycinnamic acid; MALDI, Matrix-assisted laser desorption/ionisation; PAG, polyacrylamide gel; PAGE, polyacrylamide gel electrophoresis; TFA, trifluoroacetic acid; TOF, time-of-flight

Grant sponsor: Federal Agency for Scientific Organizations (FASO Russia); Grant number: 0345–2014–0006 (Program of the Presidium of RAS «Genomics and Proteomics»); Mass spectrometry was performed at Joint Instrumentation Centre «Unltramicroanalysis» of Limnological Institute, SB RAS.

*Correspondence to: MA Grachev 3, Ulan-Batorskaya, Irkutsk, 664033, P.O. box 278, Russia. E-mail: grachev@lin.irk.ru

Introduction

Silicateins are proteins found in sponge spicules. They are present in the axial canals of sponge spicules. It is postulated that silicateins play an important role in the silica deposition in sponges and in morphogenesis of spicules. *In vitro* sponge silicateins act as enzymes that catalyze the transformation of tetraethoxysilane into biogenic silica.¹ Silicateins are being tested for many bioinspired silicification procedures of potential practical importance.^{2–5}

The purpose of the present study was to elucidate the N-terminal amino acid sequences of four mature silicateins from a sponge from lake Baikal *L. baicalensis*, which were unknown up to this date. To this end we developed a new method based on chemical modification and mass spectrometry using a reagent having a Br-atom and the fact that this element consists of ^{79}Br and ^{81}Br in almost equal amounts.

Recently proposed were a few methods, which combined a chemical modification and mass-spectrometry to find N-terminal sequences of proteins. Some of these methods involve physical separation of internal peptides after chemical modification followed by mass-spectrometry of N-terminal ones.⁶⁻⁹ Others do not involve the physical separation stage and depend on recognizing isotopic signatures of modified N-terminal peptides.^{10,11}

The presence of two equally represented bromine isotopes in the modifying reagent served as a basis of a technique in which model proteins were subjected to guanidination with *o*-methylisourea followed by modification of N-termini which was performed with ethyl 5-bromopicolinimidate; the product obtained was treated with chymotrypsin, and the modified peptides were identified by the characteristic mass spectra.¹⁰ As far as we know this method was only applied to model proteins – myoglobin and haemoglobin. A disadvantage of this method is the fact that guanylation can sometimes be non-quantitative.

Another example of application of the bromine isotopic tag is modification of N-terminal amino groups of model peptides with 5-bromonicotinic acid N-hydroxysuccinimide ester (BrNANHS).¹¹ The modified peptides were subjected to mass spectrometry, and N-terminal peptides were identified using the characteristic forms of the mass spectra and MS/MS with electrospray ionization. A disadvantage of this method is the fact that, to some extent, side chains of lysine residues are modified. The reaction was performed at pH 6.5. Under these conditions, the reagent mainly modified the N-terminal amino group. However, the authors mention that the conditions have to be adjusted to ensure selective modification of the α -N-terminal amino group. The method was successfully applied to the determination of the amino acid sequences of peptides from porcine kidney aminopeptidase I produced by digestion with lysyl endopeptidase and with *Staphylococcus aureus* V8 protease.

The procedure, which we propose (cf. below), does not have the above-mentioned disadvantages.

Results and Discussion

Model studies with lysozyme

At the first step of the present study we attempted to elaborate a novel approach based on a combination of

Figure 1. Mass spectrum of a trypsin hydrolysate of a lysozyme modified at the truncated N-terminal peptide with Br-PITC according to the procedure proposed.

chemical modification and mass spectrometry using a model small protein lysozyme to confirm that our procedure gives a possibility to sequence its N-terminal peptide. Lysozyme was reduced with mercaptoethanol and alkylated with iodoacetamide by the usual procedure prior to subsequent analysis. Lysozyme was subjected to polyacrylamide gel electrophoresis (PAGE). All reactions were performed directly in a band of lysozyme cut out of this polyacrylamide gel (PAG). The protein was subjected to treatment with PITC followed by reaction with trifluoroacetic acid (TFA). This gave a possibility to remove the N-terminal amino acid and to block the side chains of lysine. The modified protein was treated with 4-Br-PITC followed by trypsinolysis and gave peptides with blocked side chains of lysine, and left intact the arginine residues. It is noteworthy that the blockade of lysine side chains was quantitative, as revealed by MS (Fig. 1). The N-terminal peptide was identified in the mass spectrum on the basis of a characteristic signature which reflected the presence of two bromine isotopes. Subsequent a MS/MS procedure (LIFT) gave an N-terminal sequence (VFGR) which is known to be present in lysozyme. It will be mentioned that our mass spectrum gave a possibility to identify the other peptides of modified lysozyme with m/z $[\text{MH}]^+$ 517.18 (TPGSR), m/z $[\text{MH}]^+$ 874.29 (HGLDNYR), m/z $[\text{MH}]^+$ 993.53 (WWCNDGR), m/z $[\text{MH}]^+$ 1184.35 (CELAAAMK⁺R), m/z $[\text{MH}]^+$ 1468.72 (CK⁺GTDVQAWIR), m/z $[\text{MH}]^+$ 1754.08 (NTDGSTDYGLQINSR), m/z $[\text{MH}]^+$ 2871.88 (GYSLGNWVCAAK⁺FESNFNTQATNR). The longest predicted trypsin peptide with m/z $[\text{MH}]^+$ 4563.0945 (NLCNIPCSALLSSDITASVNC⁺AK⁺K⁺IVSDGNGMNAWVAWR) was not found, presumably because of its

Figure 2. Part of the mass spectrum of a trypsin hydrolysate of SIL α 1, SIL α 2, SIL α 3 and SIL α 4 from *L. baicalensis*. Silicateins were modified with Br-PITC according to the method proposed.

small solubility. (K⁺ is lysine residue modified with PITC).

Hence, it was confirmed that the novel chemical modification – mass spectrometry procedure gives a possibility to find and to sequence N-terminal peptides of proteins.

Studies of silicateins

Kalyuzhnaya *et al.* studied the structure of the genome of *L. baicalensis* with primers prompted by nucleotide sequences of silicatein genes in a marine sponge *Suberites domuncula* and in DNA of a sponge from Lake Baikal *L. baicalensis*.¹² They performed amplification of the gene, its reverse transcription and analysis of the corresponding mRNA. They came to the conclusion that the genome of *L. baicalensis* contains a sequence which has a length of 1988 bp, which encodes predicted amino acid sequences of 4 silicateins, SIL α 1, SIL α 2, SIL α 3, and SIL α 4, as suggested by the distribution of terminating triplets. They subjected the mixture of proteins isolated from sponge spicules to trypsin hydrolysis, and found a few peptides corresponding to SIL α 1, SIL α 2, SIL α 3, and SIL α 4, as revealed by MS with electrospray ionization. However, N-terminal amino acid sequences of mature silicateins remained unknown.

To solve this problem we studied the structures of four mature silicateins which were isolated from spicules of *L. baicalensis*, as described under the experimental section. The mixture of products obtained from the spicules was separated by means of PAGE. The doublet of proteins having an approximate molecular mass of 25 kDa was treated with PITC followed by reaction with TFA. To introduce the bromine tag, the truncated proteins were treated with Br-PITC. The mixture obtained was subjected to treatment with trypsin. The tryptic peptides

obtained were studied by means of Matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) and MALDI-TOF/TOF. Br-containing peptides were identified by characteristic bromine signatures (Fig. 2).¹⁰ The MALDI-TOF mass spectra shown in Figure 2 suggest the presence of three modified N-terminal peptides with m/z [MH]⁺ 1075.306, 1089.318 and 1103.325. The amino acid sequences of Br-containing peptides were found with LIFT procedure (Fig. 3). The results are shown in Table I. Grey color corresponds to identical amino acid residues.

The predicted amino acid sequences and the data shown in Table I suggest: (1) N-terminal amino acid of mature SIL α 1 to SIL α 4 is tyrosine (Y); (2) peptide #2 belongs to SIL α 1, peptide #4 to SIL α 2, peptide #6 to SIL α 3 and peptide #8 to SIL α 4.

C-terminal sequences of SIL α 1 to SIL α 4 are sequences ... YPML*, ... YSML*, ... YPML*, ... YPTL* respectively, where the terminator is shown as asterisk.¹² Interestingly, the lengths of the four mature silicateins of *L. baicalensis* were the same and equal to 217 amino acid residues, however their masses are different and are equal to 22983.34 Da for SIL α 1, 23046.37 Da for SIL α 2, 22733.32 Da for SIL α 3 and 23024.37 Da for SIL α 4.

The data obtained in the present study give a possibility to prepare mature silicateins for structure functional studies by means of molecular genetic procedures or by *in vitro* translation. As for the method including tagging with Br-PITC, it is simple, reliable, cost effective and applicable to other proteins.

Experimental Section

Chemicals and materials

PITC, Br-PITC and lysozyme, as well as other chemicals and solvents were purchased from Sigma-Aldrich

Figure 3. MALDI-TOF/TOF mass spectra of truncated *N*-terminal peptides of the Br-PITC-modified SIL α 1-SIL α 4 tryptic digest. ADSLDWR is truncated *N*-terminal peptide of SIL α 3. AESIDWR is truncated *N*-terminal peptide of SIL α 1. VDSLDWR is truncated *N*-terminal peptide of SIL α 2 and SIL α 4. The modified Br-PITC truncated *N*-terminal peptides is cleaved using the LIFT procedure.

(St. Louis, USA). Trypsin was purchased from Promega (USA).

Isolation of proteins

Proteins contained in sponge spicules were isolated according to the method of Kalyuzhnaya *et al.*¹² Spicules were washed with a mixture of H₂SO₄ and HNO₃ (1:4), then washed with water and dried in the

open air. Spicules were dissolved in 2M HF - 8M NH₄F (pH 5), and the residue dialysed against deionised water. The insoluble matter was removed by centrifugation. The supernatant was subjected to PAGE.

PAGE

Gel electrophoresis was performed in the denaturing PAGE according to Kalyuzhnaya *et al.*¹² Proteins

Table I. The Results of Sequencing

1	Fragment of predicted sequence of SIL α 1 ...[10]	... YAESIDWR...
2	Truncated modified N-terminal by means of method proposed	... AESIDWR...
3	Fragment of predicted sequence SIL α 2 ...[10]	... YVDSIDWR...
4	Truncated modified N-terminal by means of method proposed	... VDSIDWR...
5	Fragment of predicted sequence SIL α 3 ...[10]	... YADSLDWR...
6	Truncated modified N-terminal by means of method proposed	... ADSLDWR...
7	Fragment of predicted sequence SIL α 4 ...[10]	... YVDSIDWR...
8	Truncated modified N-terminal by means of method proposed	... VDSIDWR...

were suspended in an Orange loading dye solution (Fermentas, Lithuania), incubated for 5 min at 95°C, and applied onto a 2-phase denaturing gel (concentrating gel – 10%, separating gel – 5%), and electrophoresis was performed at 125 V for 2 hours. The gel was stained with Coomassie Brilliant Blue G-250 (Serva, Germany).

Chemical modification and proteolysis of proteins

After PAGE, the Coomassie-stained bands corresponding to a molecular mass of 25 kDa were cut out and placed into an Eppendorf tube. They were washed with 50% acetonitrile in 25 mM ammonium bicarbonate. The gel was dehydrated with 100% acetonitrile. The gel was shortly dried in the open air and supplemented with 2% PITC in 60% aqueous pyridine (v/v) until complete swelling of the gel (the gel became translucent) and kept for 30 min at 50°C. The gel was washed with acetonitrile, dried in the open air, supplemented with dry TFA and kept for 15 min at 50°C. The gel was washed with water, followed by acetonitrile until complete dehydration (the gel shrank and became white). The dried gel was supplemented with a 2% solution of 4-Br-PITC in 60% aqueous pyridine (v/v) until complete swelling and kept for 30 min at 50°C. The gel was washed with acetonitrile. The dried gel was supplemented with a 0.1% solution of trypsin in 25 mM ammonium bicarbonate and kept for 16 hours at 37°C. Peptides were extracted with 5 µl of 50% aqueous acetonitrile – 1% TFA for 30 min at 25°C. The peptides were analyzed by means of MALDI-TOF and MALDI-TOF/TOF.

MALDI-TOF and MALDI-TOF/TOF

MALDI-TOF mass spectra were obtained using an α-cyano-4-hydroxycinnamic acid (CHCA) matrix solution (10 mg/mL). An Ultraflextreme MALDI TOF/TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) was used for MS and MS/MS measurements. Mass calibration was performed immediately prior to measurements using a commercial Peptide II Mass Standards Kit (Bruker Daltonics, Bremen, Germany). A typical mass spectrum was obtained by averaging 150 shots of the laser.

References

1. Shimizu K, Cha J, Stucky GD, Morse D (1998) Silicatein alpha: cathepsin L-like protein in sponge biosilica. *Proc Natl Acad Sci USA* 95:6234–6238.
2. Wiens M, Elkhooly TA, Schroder HC, Mohamed THA, Muller WEG (2014) Characterization and osteogenic activity of a silicatein/biosilica-coated chitosan-graft-polycaprolactone. *Acta Biomater* 10:4456–4464.
3. Muller WEG, Link T, Schroder HC, Korzhev M, Neufurth M, Brandt D, Wang XH (2014) Dissection of the structure-forming activity from the structure-guiding activity of silicatein: a biomimetic molecular approach to print optical fibers. *J Mater Chem B* 2: 5368–5377.
4. Adanyi N, Bori Z, Szendro I, Erdelyi K, Wang XH, Schroder HC, Muller WEG (2013) Bacterial sensors based on biosilica immobilization for label-free OWLS detection. *New Biotechnol* 30:493–499.
5. Muller WEG, Schroder HC, Burghard Z, Pisignano D, Wang XH (2013) Silicateins-A novel paradigm in bioinorganic chemistry: enzymatic synthesis of inorganic polymeric silica. *Chem-Eur J* 19:5790–5804.
6. Timmer JC, Enoksson M, Wildfang E, Zhu W, Igarashi Y, Denault JB, Ma Y, Dummitt B, Chang YH, Mast AE, Eroshkin A, Smith JW, Tao WA, Salvesen GS (2007) Profiling constitutive proteolytic events in vivo. *Biochem J* 407:41–48.
7. Xu G, Shin SB, Jaffrey SR (2009) Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. *Proc Natl Acad Sci USA* 106:19310–19315.
8. Brittain SM, Ficarro SB, Brock A, Peters E (2005) Enrichment and analysis of peptide subsets using fluorescent affinity tags and mass spectrometry. *Nat Biotechnol* 23:463–468.
9. Yan H, Hao FR, Cao QC, Li JB, Li NN, Tian F, Bai HH, Ren XJ, Li XY, Zhang YJ, Qian XH (2014) A novel method for identification and relative quantification of N-terminal peptides using metal-element-chelated tags coupled with mass spectrometry. *Sci China Chem* 57: 708–717.
10. Kim JS, Song JS, Kim Y, Park SB, Kim HJ (2012) De novo analysis of protein N-terminal sequence utilizing MALDI signal enhancing derivatization with Br signature. *Anal Bioanal Chem* 402:1911–1919.
11. Miyagi M, Nakao M, Nakazawa T, Kato I, Tsunasawa S (1998) A Novel derivatization method with 5-bromonicotinic acid N-hydroxysuccinimide for determination of the amino acid sequences of peptides. *Rapid Commun Mass Spectrom* 12:603–608.
12. Kaluzhnaya OV, Belikova AS, Podolskaya EP, Krasko AG, Müller WEG, Belikov SI (2007) Identification of silicateins in freshwater sponge *Lubomirskia baicalensis*. *Mol Biol* 41:554–561.