**Chemical synthesis of La1 isolated from the venom of the scorpion *Liocheles australasiae* and determination of its disulfide bonding pattern**

**Short title: Synthesis of La1 and determination of its disulfide bonding pattern**

Junya Nagao, Masahiro Miyashita*, Yoshiaki Nakagawa, and Hisashi Miyagawa

Division of Applied Life Sciences, Graduate School of Agriculture, Kyoto University, Kyoto, 606-8502, Japan

*Corresponding author: Tel/Fax: +81-75-753-6123; e-mail: miyamasa@kais.kyoto-u.ac.jp

**Abstract**

La1 is a 73-residue cysteine-rich peptide isolated from the scorpion *Liocheles australasiae* venom. Although La1 is the most abundant peptide in the venom, its biological function remains unknown. Here, we describe a method for efficient chemical synthesis of La1 using the native chemical ligation (NCL) strategy, in which three peptide components of less than 40 residues were sequentially ligated. The peptide-thioester necessary for NCL was synthesized using an aromatic N-acylurea approach with Fmoc-SPPS. After completion of sequential NCL, disulfide bond formation was carried out using a dialysis method, in which the linear peptide dissolved in an acidic solution was dialyzed against a slightly alkaline buffer to obtain correctly folded La1. Next, we determined the disulfide bonding pattern of La1. Enzymatic and chemical digests of La1 without reduction of disulfide bonds were analyzed by LC/MS, which revealed two of four disulfide bond linkages. Remaining two linkages were assigned based on MS/MS analysis of a peptide fragment containing two disulfide bonds. Consequently, the disulfide bonding pattern of La1 was found to be similar to that of a von Willebrand factor type C (VWC) domain. To our knowledge, this is the first report of the experimental determination of the disulfide bonding pattern of peptides having a single VWC domain as well as their chemical synthesis. La1 synthesized in this study will be useful for investigation of its biological role in the venom.

**Keywords:** scorpion venom, peptide synthesis, native chemical ligation, disulfide bond, MS/MS

**Abbreviations:** COMU, (1-cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate; DIC, *N,N'*-diisopropylcarbodiimide; DIEA, *N,N*-diisopropylethylamine; ECM, extracellular matrix; Fmoc-Dbz, 3-*N*-[(9*H*-fluoren-9-yl)methoxycarbonyl]-amino-4-aminobenzoic acid; HBTU, *O*-(benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; HOBt, 1-hydroxybenzotriazole; Nbz, N-acybenzimidazolinone; NCL, native chemical ligation; SVC, single von Willebrand factor type C domain; Thz, 1,3-thiazolidine-4-carboxylic acid; TIS, triisopropylsilane; VWC, von Willebrand factor type C domain.

**Introduction**

 Scorpion venom contains a wide variety of biologically active peptides ranging in size from 0.4 to 10 kDa [1]. Peptides included in the scorpion venom can be classified into two main types based on the presence of disulfide bridges. The majority of scorpion peptides containing disulfide bridges are neurotoxic, which act on various ion channels in excitable membranes [2,3]. Peptides without disulfide bridges have also been identified from the venom of various scorpion species, which mainly shows antimicrobial activity [4,5]. These peptides play important roles in the scorpion venom to capture prey or to defend against predators or pathogens, but the peptides identified and characterized so far are only a fraction of all venom components. Thus, a large number of peptides whose biological activity is unknown are included in the scorpion venom [1].

 La1, which was isolated from the *Liocheles australasiae* venom, is one example of the "neglected" components [6]. La1 is the most abundant component in the *L australasiae* venom, but its biological role has not been determined yet. Interestingly, peptides similar to La1 have been identified from the venom of various scorpion species, suggesting that La1 and its related peptides (La1-like peptides) may play an important role in the scorpion venom [7-11]. In terms of structure, it has been presumed that La1-like peptides possess a single von Willebrand factor type C (VWC) domain based on the position of eight Cys residues [12], although their disulfide bonding pattern has yet to be determined experimentally.

 To investigate the possible biological function of the neglected components including La1, milligram scale of the sample is often necessary for conducting various experiments. Since the amount of La1 available from the venom is not sufficient for this purpose, a method for its large-scale preparation must be established. La1 consists of 73 amino acid residues (Figure 1), and it could be prepared either by chemical synthesis or by recombinant expression. Chemical synthesis is generally superior to recombinant expression in terms of yields, thanks to the recently established native chemical ligation (NCL) technique [13]. NCL enables the chemoselective condensation of two unprotected peptides with readily synthesizable length. The chemical method also allows the incorporation of non-natural amino acids in the sequence. Here, we report the efficient chemical synthesis of La1 using NCL. Moreover, the disulfide bonding pattern of La1 was investigated. The pattern was determined using natural La1, and compared it with that of synthetic La1 to confirm that the folding reaction proceeded correctly. To our knowledge, this is the first report of the experimental determination of their disulfide bonding pattern of peptides having single VWC

domain (SVC) including La1 as well as their chemical synthesis.

## Materials and methods

### Materials

Natural La1 was purified from the venom of the scorpion *Liocheles australasiae* reared in our laboratory [6]. Fmoc- or Boc-protected amino acids, Fmoc-NH-SAL resin, Fmoc-NH-SAL-PEG resin, triisopropylsilane (TIS), *N,N'*-diisopropylcarbodiimide (DIC), *O*-(benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 1-hydroxybenzotriazole (HOBt), and (1-cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate (COMU) were purchased from Watanabe Chemical Industries (Hiroshima, Japan). 3-*N*-[(9*H*-fluoren-9-yl)methoxycarbonyl]-amino-4-aminobenzoic acid (Fmoc-Dbz) was synthesized according to the literature [14]. Trypsin (mass spectrometry grade) was purchased from Wako Pure Chemical Industries (Osaka, Japan). All other chemicals were commercially available.

### HPLC

Preparative RP-HPLC was carried out using a preparative column, Inertsil ODS-3, $250 \times 20$ mm ID (GL Sciences) at a flow rate of 7 ml/min, or InertSustain C18, $250 \times 14$ mm ID (GL Sciences) at a flow rate of 5 ml/min, using a 60 min linear gradient from 10 to 40% of solvent B ($CH_3CN$ containing 0.1% TFA) in solvent A ($H_2O$ containing 0.1% TFA) at 40°C. Elution was monitored by absorption at 215 nm. Analytical RP-HPLC was carried out using a micro column, Vydac Everest C18, $250 \times 1.0$ mm ID (Grace Vydac, Hesperia, CA, USA) at a flow rate of 50 µl/min, using a 50 min linear gradient from 5 to 50% of solvent B in solvent A at 40°C. Elution was monitored by absorption at 215 nm.

### Mass spectrometry

LC/MS analysis to monitor reactions was performed on an LCMS-2020 mass spectrometer (Shimadzu, Kyoto, Japan) equipped with an electrospray ion source in the positive mode. HPLC separation was carried out using a semi-micro column, InertSustatin C18, 150 x 2.1 mm ID (GL Sciences, Tokyo, Japan). The column was eluted at a flow rate of 0.2 ml/min with 0.1% formic acid in water (solvent C) and 0.1% formic acid in acetonitrile (solvent D) with a 20 min linear gradient of 5 to 50% solvent D in solvent C. Elution was also

monitored by absorption at 215 nm. LC/MS analysis for high-resolution measurements was performed on an LCMS-IT-TOF mass spectrometer (Shimadzu) equipped with an electrospray ion source in the positive mode. HPLC separation was carried out on a micro column, TSKgel ODS-100V, 150 x 1.0 mm ID (Tosoh, Tokyo, Japan). The column was eluted at a flow rate of 50 μl/min with a 25 min linear gradient of 5 to 50% solvent D in solvent C.

MALDI-TOF-MS/MS measurements were carried out on an Autoflex III mass spectrometer (Bruker Daltonics, Billerica, MA). The sample was dissolved in a matrix solution containing 10 mg/ml of α-cyano-4-hydroxycinnamic acid, 70% acetonitrile and 0.1% TFA, and 0.5 μl of the solution was spotted on the MALDI sample target and allowed to dry at room temperature. External calibration of the mass scale was carried out using peptides with known molecular masses. MS/MS measurements were carried out in the LIFT mode.


**Synthesis**

**General procedure for solid-phase peptide synthesis**

Peptides were manually synthesized by using the Fmoc-based solid-phase peptide synthesis (SPPS) technique. Fmoc deprotection was performed with 20% piperidine in *N,N*-dimethylformamide (DMF) three times, each for 3, 3, and 20 min at room temperature. Each Fmoc-protected amino acid (3 eq) were coupled in the presence of COMU (3 eq) and *N,N*-diisopropylethylamine (DIEA, 3 eq) in DMF under microwave heating (Initiator+, Biotage, Uppsala, Sweden) at 75°C for 5 min. For introduction of Cys and 1,3-thiazolidine-4-carboxylic acid (Thz), DIC (3 eq) and HOBt (3 eq) were used and reacted for 20 min with microwave heating at 50°C. The reaction was monitored by Kaiser test. After completion of coupling reaction, the resin was sequentially washed with DMF, diethyl ether and methanol, and dried *in vacuo.* Cleavage of peptides from the resin and removal of side-chain protecting groups were carried out using the solution containing TFA/triisopropylsilane/1,3-dimethoxybenzene (92.5: 2.5: 5, v/v/v), which was reacted for 2 h at room temperature. After the resin was removed by filtration, cold diethyl ether was added to the filtrate to precipitate the peptides, followed by washing with cold diethyl ether twice and drying *in vacuo*.


**Synthesis of La1(43-73)**

Fmoc-NH-SAL-PEG resin (0.24 mmol/g substitution) was used for the synthesis. After peptide chain assembly and removal of the N-terminal Fmoc group were completed, cleavage

of peptides from the resin and removal of side-chain protecting groups were performed as described above. The desired peptide was purified by HPLC and lyophilized. La1(43-73) was obtained in 27% yield (92 mg) based on the amount of the resin used. ESI-MS: observed $m/z$ 751.4 for [M+4H]$^{4+}$ (calcd. $m/z$ 751.4).

**[Thz29]La1(29-42)-thioester**

Fmoc-NH-SAL resin (0.56 mmol/g substitution) was used for the synthesis. Introduction of the C-terminal Dbz residue and the following Arg residue was performed using longer reaction time (20 min). Thz was incorporated at the position of Cys29. After peptide chain assembly was completed, the solution of $p$-nitrophenylchloroformate (2 mg/ml) in dichloromethane was added to the resin. After reaction for 1 h at room temperature, the resin was washed with dichloromethane and DMF. The resin was reacted with DIEA (0.5 M) in DMF for 30 min at room temperature, and washed with DMF and methanol. After cleavage of peptides from the resin and removal of side-chain protecting groups were performed, the peptide was treated with 2% methyl 3-mercaptopropionate in 200 mM Na$_2$HPO$_4$ (adjusted to pH 7.1 with HCl) containing 6 M guanidine•HCl for 1 h at room temperature to obtain peptide thioester. The desired peptide-thioester was purified by HPLC and lyophilized. [Thz29]La1(29-42)-thioester was obtained in 31% yield (157 mg) based on the amount of the resin used. ESI-MS: observed $m/z$ 458.1 for [M+4H]$^{4+}$ (calcd. $m/z$ 458.1).

**La1(1-28)-thioester**

Fmoc-NH-SAL-PEG resin (0.24 mmol/g substitution) was used for the synthesis. Introduction of the C-terminal Dbz residue and the following Lys residue was performed using longer reaction time (20 min). The Boc-protected amino acid was used for the N-terminal Phe residue. Further procedure was identical to that of synthesis of [Thz29]La1(29-42)-thioester. La1(1-28)-thioester was obtained in a 14% yield (92 mg) based on the amount of the resin used. ESI-MS: observed $m/z$ 805.4 for [M+4H]$^{4+}$ (calcd. $m/z$ 805.5)

**Native chemical ligation between [Thz29]La1(29-42)-thioester and La1(43-73)**

Ligations were performed in a buffer containing guanidine•HCl (6 M), tris(2-carboxyethyl)phosphine hydrochloride (TCEP•HCl, 20 mM), 4-mercaptophenylacetic acid (MPAA, 100 mM), Na$_2$HPO$_4$ (200 mM), adjusted to pH 7 with NaOH. [Thz29]La1(29-42)-thioester and La1(43-73) was dissolved in the ligation buffer at final

concentrations of approximately 1 and 0.5 mM, respectively, considering the presence of the impurities and the possibility of hydrolysis of the thioester moiety of the former peptide. The reaction mixture was stirred at room temperature for 2 h, and the progress of the ligation was monitored by LC/MS. To this mixture, methoxyamine hydrochloride (400 mM) was added to convert the N-terminal Thz to Cys. After reaction for 5 h, ligation product was purified by HPLC and lyophilized. A total of 77 mg (58-76%) of La1(29-73) was obtained. ESI-MS: observed $m/z$ 783.9 for $[M+6H]^{6+}$ (calcd. $m/z$ 784.0).

**Ligation between La1(1-28)-thioester and La1(29-73)**

La1(1-28)-thioester and La1(29-73) were dissolved in the ligation buffer at final concentrations of approximately 0.9 and 0.6 mM, respectively. Further ligation procedure was identical to that between [Thz29]La1(29-42)-thioester and La1(43-73) except that treatment with methoxyamine hydrochloride was omitted. Ligation product was purified by HPLC and lyophilized. A total of 20 mg (14-25%) of linear La1 was obtained. ESI-MS: observed $m/z$ 867.1 for $[M+9H]^{9+}$ (calcd. $m/z$ 867.1).

**Folding**

Disulfide bridge formation was carried out by the oxidative dialysis method [15]. Linear La1 (0.13 mg/ml) dissolved in 10% (v/v) acetic acid was loaded into a dialysis cassette (Slide-A-Lyzer G2, 3500 MWCO, 30 ml, Thermo Scientific, Rockford, IL) and dialyzed twice against 2 L of 100 mM phosphate buffer (pH7.7) containing NaCl (100 mM), guanidine•HCl (20 mM), and reduced (1 mM) and oxidized glutathione (0.1 mM) for 72 h. The folded peptide was purified by HPLC and lyophilized. Pure La1 was obtained in 41% (5.2 mg) yield. HR-ESI-MS: observed $m/z$ 865.63 for $[M+9H]^{9+}$ (calcd. $m/z$ 865.65)

**Determination of disulfide bonding pattern**

Natural or synthetic La1 (50 µg) was dissolved in 0.1 M $NH_4HCO_3$ buffer (adjusted to pH 6.6 with HCl), and digested with trypsin (0.5 µg/µl) at 37°C for 20 h. Digested peptides were analyzed by LC/MS, and the peptide containing three disulfide bonds was purified by HPLC and lyophilized. The peptide was dissolved in 0.1 M HCl, and reacted with an excess amount of CNBr for 20 h at room temperature. CNBr was removed by a C18 spin column (MonoSpin C18, GL Sciences), and the resultant digests were subjected to MS and MS/MS analysis using MALDI-TOF/TOF MS.

**Results and discussion**

**Synthesis of La1 using NCL**

Since the limit of the linear stepwise elongation in SPPS is around 50 residue-long, synthesis of a 73-residue peptide, La1, would be difficult using a general SPPS approach. To overcome this limitation, several techniques for synthesis of large peptides consisting of over 50 residues have been developed. Chemical ligation is one of the efficient methods to obtain large peptides, in which two or more peptide segments with readily synthesizable length are ligated [16]. Among them, an NCL approach is one of the most promising methods for the synthesis of large peptides. NCL allows the chemoselective coupling of an unprotected peptide having a C-terminal thioester moiety with another unprotected peptide bearing an N-terminal Cys residue, in which intermolecular thioester exchange occurs first between the thioester moiety and the thiol group of the Cys residue, followed by spontaneous S-N acyl transfer to form an amide bond [13]. Since a Cys residue is essential for ligation, NCL is particularly suitable for synthesis of Cys-rich peptides such as scorpion peptides. In this study, an NCL approach was, therefore, applied to the synthesis of La1.

The strategy for the synthesis of La1 is shown in Figure 2. La1 was synthesized using two sequential NCL from three segments (Table 1). The peptide having a thioester moiety at the C-terminus was synthesized through Dawson's method [14]. In this method, 3,4-diaminobenzoic acid (Dbz), which is stable under alkaline conditions, is used during peptide chain elongation using Fmoc-SPPS, and then converted into the N-acybenzimidazolinone (Nbz) moiety before cleavage of peptides from resin. The Nbz moiety is finally substituted with reactive thiols to afford a peptide thioester. Since eight Cys residues are present in the sequence of La1, there are many possible ligation sites for NCL. We initially chose the site between Arg42 and Cys43 for first ligation and that between Asp23 and Cys24 for second ligation. However, preparation of the segment La1(1-23)-Nbz necessary for second ligation was not successful, where only the peptide having a free carboxylate group at the C-terminus was obtained. This is likely because of the presence of the Asp residue adjacent to the Nbz moiety. The side chain carboxylate anion of the Asp residue could attack its carbonyl group in the main chain to remove the Nbz moiety. Therefore, we chose the site between Lys28 and Cys29 for second ligation.

**Synthesis of each segment**

The C-terminal segment, La1(43-73), was synthesized using the general Fmoc-SPPS method. After HPLC purification, the desired peptide was obtained in 27% yield (Figure 3a). The central segment [Thz29]La1(29-42)-thioester, which contains a thioester moiety at the C-terminus and a Thz residue at the N-terminus, was synthesized by Fmoc-SPPS starting from introduction of Dbz at the C-terminus. After peptide chain elongation, Dbz was converted to Nbz on the resin. The peptide-Nbz was cleaved from the resin, and treated with methyl 3-mercaptopropionate at neutral pH to obtain a peptide-thioester. After HPLC purification, the desired peptide was obtained in 31% yield (Figure 3b). In this synthesis, several impurities caused by the incorporation of amino acid residues at another amino group of the Dbz moiety were observed even after HPLC purification. However, since these peptides cannot participate in the NCL reaction due to the lack of the C-terminal thioester moiety and the N-terminal Cys residue, this was used without further purification. The synthesis of the N-terminal segment La1(1-28)-thioester was performed using the method similar to that of [Thz29]La1(29-42)-thioester. In this case, the peptide with a molecular mass of 120 Da smaller than the desired one was observed before HPLC purification. This is likely to be a peptide containing a cyclic structure formed in the C-terminal Lys residue, which could be caused by the nucleophilic attack of the side chain amino group of the Lys residue on the carbonyl carbon in the C-terminal thioester moiety. This side product was removed by HPLC. Finally the desired peptide was obtained in 14% yield (Figure 3c).

**Ligation reaction**

First ligation was carried out between [Thz29]La1(29-42)-thioester and La1(43-73). As shown in Figure 4a, reaction was completed within 2 h, which was monitored by LC/MS. To this solution, methoxyamine hydrochloride was added, and Thz was successfully converted into Cys after 5 h. The ligated peptide La1(29-73) was partially purified by HPLC in 76% yield (Figure 4b and 4c), and subjected to second ligation with La1(1-28)-thioester. The result of second ligation was shown in Figure 5a. After reaction for 2 h, the ligated product (linear La1) was observed along with the disappearance of La1(1-28)-thioester. A large amount of La1(29-73) remained even after completion of the reaction. This is because some part of La1(1-28)-thioester was converted to its side product, in which the C-terminal Lys could be intermolecularly cyclized as described above. After HPLC purification, linear La1 was successfully obtained (Figure 5b and 5c), although the yield of the second ligation was rather

low (25%).

**Folding**

Disulfide bridge formation is often carried out under slightly alkaline conditions. However, linear La1 was sparingly soluble in aqueous buffer at pH 8.5, but dissolved readily in acidic solution. This prompted us to perform disulfide bridge formation of La1 under acidic conditions. We first attempted the oxidation reaction mediated by dimethyl sulfoxide (DMSO), which is known to proceed under acidic conditions [17]. La1 was dissolved in 5% aqueous acetic acid in the presence of DMSO (20% by volume), and reacted for 48 h. LC/MS analysis of the reaction mixture showed that multiple peaks having a molecular mass of fully oxidized La1 were observed (data not shown). This indicated that peptides having incorrect disulfide linkages were mainly generated under this condition. We speculated that the acidic environment could have an adverse effect on correct folding of La1. To allow the peptide that is only soluble in an acidic environment to form disulfide bridges under alkaline conditions, a dialysis method have been developed [15]. In this method, the peptide was dissolved in acidic solution and dialyzed against alkaline buffer to promote correct folding. We applied this method to the folding of La1. Linear La1 was dissolved in 10% aqueous acetic acid, and dialyzed twice against 100 mM phosphate buffer (pH 7.7) containing oxidized and reduced glutathione for 72 h. From the result of HPLC analysis, a large main peak was observed, which has a molecular mass of fully oxidized La1 (Figure 6a). In addition to this peak, peptides with a 611 Da larger mass were observed as a broad peak. This is probably due to incorporation of glutathione molecules via a disulfide bond at two different Cys residues. The main peak was purified by HPLC in 41% yield. The identity of this peptide was confirmed by comparison with natural La1. Both synthetic and natural La1 showed the same retention time in HPLC analysis (Figure 6b). The monoisotopic molecular mass of synthetic La1 (7781.63 Da) determined by high-resolution LC/MS analysis (Figure 6c) is in good agreement with the calculated one (7781.76 Da).

**Determination of disulfide bonding pattern**

Disulfide bonding patterns of peptides are often determined by MS analysis, in which enzymatically and/or chemically digested fragments of the peptides without reduction of disulfide bonds are analyzed. To assign each disulfide linkage, the peptide chain must be cleaved at every sequence region between Cys residues, which generates peptide fragments

10

consisting of a pair of peptides cross-linked by a single disulfide bond. Based on the molecular masses of these fragments detected by MS analysis, disulfide linkages can be assigned. In the case of La1, all regions between Cys residues can be cleaved by combination of trypsin and CNBr, except for the region between Cys66 and Cys67. Since these two Cys residues are adjacently located, cleavage between these Cys residues is difficult by using commercially available enzymes or chemical methods. Therefore, we employed a MS/MS technique for assignment of the disulfide linkages involving these two Cys residues.

We first analyzed the tryptic digests of natural La1 using LC/MS (Figure 7). A fragment that corresponds to two peptide chains cross-linked by a disulfide bond between Cys5 and Cys29 was detected. In addition, a fragment consisting of four peptide chains cross-linked each other by three disulfide bonds was observed. Since this fragment contains a Met residue in the sequence between Cys67 and Cys72, it was further digested using CNBr after HPLC purification. MALDI-TOF MS analysis of the CNBr digests detected two peptide fragments; one (*m/z* 841.4) consisting of two peptide chains (SGCK and VTCK) linked by a disulfide bond between Cys52 and Cys72, and the other (*m/z* 2872.3) consisting of three peptide chains (TVPGGAGAAFPSCCPh, CATVNLK and DCALYK) linked each other by two disulfide bridges (data not shown). The latter was further subjected to MS/MS analysis using MALDI-TOF/TOF MS to assign disulfide bond linkages. As shown in Figure 8, two major product ions (*m/z* 2044.0 and 2672.0) were observed in the product ion spectrum, which correspond to the fragments cleaved at the N-terminal side of Pro residues. By careful inspection of this spectrum, fragment ions generated by cleavage of the main chain between Cys66 and Cys67 were observed, which allowed the distinction between two possible linkages of disulfide bonds. A fragment ion with *m/z* of 1047.5 corresponds to the peptide chains containing a disulfide bond between Cys43 and Cys67. The presence of another disulfide bond between Cys24 and Cys66 was also confirmed by detection of a fragment ion with *m/z* of 997.3, although its intensity was relatively low. This assignment was further supported by the absence of fragment ions containing a disulfide bond between Cys43 and Cys66 (*m/z* 1033.5) or between Cys24 and Cys67 (*m/z* 1011.4). Consequently, all four disulfide bond linkages included in La1 were assigned (Figure 9a). When these experiments were carried out using synthetic La1, the same result was obtained (data not shown), indicating that the synthetic La1 has the same disulfide bonding pattern as natural one. As expected, La1 has the disulfide bonding pattern similar to that of the VWC domain in ECM proteins, although one disulfide bond between third Cys and fifth Cys is missing in La1 as

shown in Figure 9b [18]. This suggests that La1 has similar conformation to the VWC domain in ECM proteins. Currently, X-ray structural analysis of La1 is in progress to confirm this hypothesis [19].

The VWC domain is an extremely common motif found in various extracellular matrix (ECM) proteins including von Willebrand factor, CCN (cysteine-rich protein 61, connective tissue growth factor proteins, nephroblastoma overexpressed gene), procollagen, thrombospondin, glycosylated mucins, and neuralins [20]. Although the biological roles of the VWC domain are not fully understood, this domain has been reported to be important for oligomerization or binding to other proteins [21,22]. In arthropods, a peptide family that possesses a single VWC domain (SVC) has been identified [12]. Although the biological function of the arthropod SVC peptides remains to be investigated, some of them are induced by certain environmental stresses such as pathogen infection [23]. In this regard, it is possible that La1 could play a role in defense against pathogens.

## Conclusions

La1 is a 73-residue peptide containing four disulfide bonds. In this study, we established a method for chemical synthesis of La1 using an NCL strategy, in which three peptide segments of less than 40 residues were sequentially ligated. Each peptide-thioester necessary for NCL was synthesized using an aromatic N-acylurea approach with Fmoc SPPS. Linear La1 obtained by sequential NCL was then oxidized to form the four disulfide bonds using a dialysis method, which allowed the correct folding of La1. Furthermore, the disulfide bonding pattern of La1 was determined by MS and MS/MS analysis. MS analysis of enzymatic and chemical digests of La1 without reduction revealed two of four disulfide bonds linkages. Remaining two disulfide bond linkages were assigned by MS/MS analysis of the peptide fragment containing two disulfide bonds. The disulfide bonding pattern of La1 was found to be similar to that of a VWC domain in ECM proteins. To our knowledge, this is the first report of the determination of the disulfide bonding pattern of peptides having a SVC domain. La1 synthesized by the method established in this study will be useful for future investigation of its biological roles in the venom.

## Acknowledgements

**References**

1        de la Vega R.C.R., Vidal N., Possani L.D. Scorpion Peptides. In: *Handbook of Biologically Active Peptides* (Kastin A, ed), Second Edition edn. London: Academic Press. 2013; 423-429

2        Quintero-Hernandez V., Jimenez-Vargas J.M., Gurrola G.B., Valdivia H.H., Possani L.D. Scorpion venom components that affect ion-channels function. *Toxicon* 2013; **76**: 328-342.

3        Bergeron Z.L., Bingham J.P. Scorpion toxins specific for potassium ($K^+$) channels: a historical overview of peptide bioengineering. *Toxins (Basel)* 2012; **4**: 1082-119.

4        Harrison P.L., Abdel-Rahman M.A., Miller K., Strong P.N. Antimicrobial peptides from scorpion venoms. *Toxicon* 2014; **88**: 115-137.

5        Almaaytah A., Albalas Q. Scorpion venom peptides with no disulfide bridges: A review. *Peptides* 2014; **51**: 35-45.

6        Miyashita M., Otsuki J., Hanai Y., Nakagawa Y., Miyagawa H. Characterization of peptide components in the venom of the scorpion *Liocheles australasiae* (Hemiscorpiidae). *Toxicon* 2007; **50**: 428-437.

7        Luna-Ramirez K., Quintero-Hernandez V., Vargas-Jaimes L., Batista C.V.F., Winkel K.D., Possani L.D. Characterization of the venom from the Australian scorpion *Urodacus yaschenkoi*: Molecular mass analysis of components, cDNA sequences and peptides with antimicrobial activity. *Toxicon* 2013; **63**: 44-54.

8        He Y.W., Zhao R.M., Di Z.Y., Li Z.J., Xu X.B., Hong W., Wu Y.L., Zhao H.B., Li W.X., Cao Z.J. Molecular diversity of Chaerilidae venom peptides reveals the dynamic evolution of scorpion venom components from Buthidae to non-Buthidae. *J. Proteomics* 2013; **89**: 1-14.

9        Abdel-Rahman M.A., Quintero-Hernandez V., Possani L.D. Venom proteomic and venomous glands transcriptomic analysis of the Egyptian scorpion *Scorpio maurus palmatus* (Arachnida: Scorpionidae). *Toxicon* 2013; **74**: 193-207.

10       Ramirez K.L., Hernandez V.Q., Romero E.M., Winkel K., Batista C.F., Possani L.D. Characterisation of the venom of an australian scorpion, *Urodacus yaschenkoi*: Proteome and transcriptome analysis. *Toxicon* 2012; **60**: 184-185.

11       Ma Y., He Y., Zhao R., Wu Y., Li W., Cao Z. Extreme diversity of scorpion venom peptides and proteins revealed by transcriptomic analysis: implication for proteome

evolution of scorpion venom arsenal. *J. Proteomics* 2012; **75**: 1563-1576.

12    Sheldon T.J., Miguel-Aliaga I., Gould A.P., Taylor W.R., Conklin D. A novel family of single VWC-domain proteins in invertebrates. *FEBS Lett.* 2007; **581**: 5268-5274.

13    Dawson P.E., Kent S.B.H. Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* 2000; **69**: 923-960.

14    Blanco-Canosa J.B., Dawson P.E. An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew. Chem. Int. Ed. Engl.* 2008; **47**: 6851-6855.

15    Sabatier J.M., Darbon H., Fourquet P., Rochat H., Vanrietschoten J. Reduction and reoxidation of the neurotoxin II from the scorpion *Androctonus australis* Hector. *Int. J. Pept. Protein Res.* 1987; **30**: 125-134.

16    Hojo H. Recent progress in the chemical synthesis of proteins. *Curr. Opin. Struct. Biol.* 2014; **26**: 16-23.

17    Tam J.P., Wu C.R., Liu W., Zhang J.W. Disulfide bond formation in peptides by dimethyl sulfoxide - Scope and applications. *J. Am. Chem. Soc.* 1991; **113**: 6657-6662.

18    O'Leary J.M., Hamilton J.M., Deane C.M., Valeyev N.V., Sandell L.J., Downing A.K. Solution structure and dynamics of a prototypical chordin-like cysteine-rich repeat (von Willebrand factor type c module) from collagen IIA. *J. Biol. Chem.* 2004; **279**: 53857-53866.

19    Kamachi S., Nagao J., Miyashita M., Nakagawa Y., Miyagawa H., Tada T. Crystallization and preliminary X-ray diffraction studies of La1 from *Liocheles australasiae*. *Acta Crystallogr. F* 2014; **70**: 915-917.

20    Holbourn K.P., Perbal B., Ravi Acharya K. Proteins on the catwalk: modelling the structural domains of the CCN family of proteins. *J. Cell Commun. Signal.* 2009; **3**: 25-41.

21    Ganderton T., Wong J.W., Schroeder C., Hogg P.J. Lateral self-association of VWF involves the Cys2431-Cys2453 disulfide/dithiol in the C2 domain. *Blood* 2011; **118**: 5312-5318.

22    Abreu J.G., Ketpura N.I., Reversade B., De Robertis E.M. Connective-tissue growth factor (CTGF) modulates cell signalling by BMP and TGF-β. *Nat. Cell Biol.* 2002; **4**: 599-604.

23    Chen Y.H., Jia X.T., Zhao L., Li C.Z., Zhang S., Chen Y.G., Weng S.P., He J.G. Identification and functional characterization of Dicer2 and five single VWC domain

proteins of *Litopenaeus vannamei*. *Dev. Comp. Immunol.* 2011; **35**: 661-671.

Table 1　Structure of peptide segments used for native chemical ligation

| Peptide | Sequence | Molecular mass (Da) |
|---|---|---|
| La1(1-28)-thioester | FGESCIAGRFIVPLGQQVTDQRDCALYK-CO-S(CH$_2$)$_2$COOCH$_3$ | 3217.8 |
| [Thz29]La1(29-42)-thioester | [Thz]VNYNKKFALETKR-CO-S(CH$_2$)$_2$COOCH$_3$ | 1828.2 |
| La1(43-73) | CATVNLKSGCKTVPGGAGAAFPSCCPMVTCK-NH$_2$ | 3001.7 |

**Figure legends**

Fig. 1   Amino acid sequence of La1. Asterisk indicates an amidated C-terminus. Arrows indicate the ligation sites.
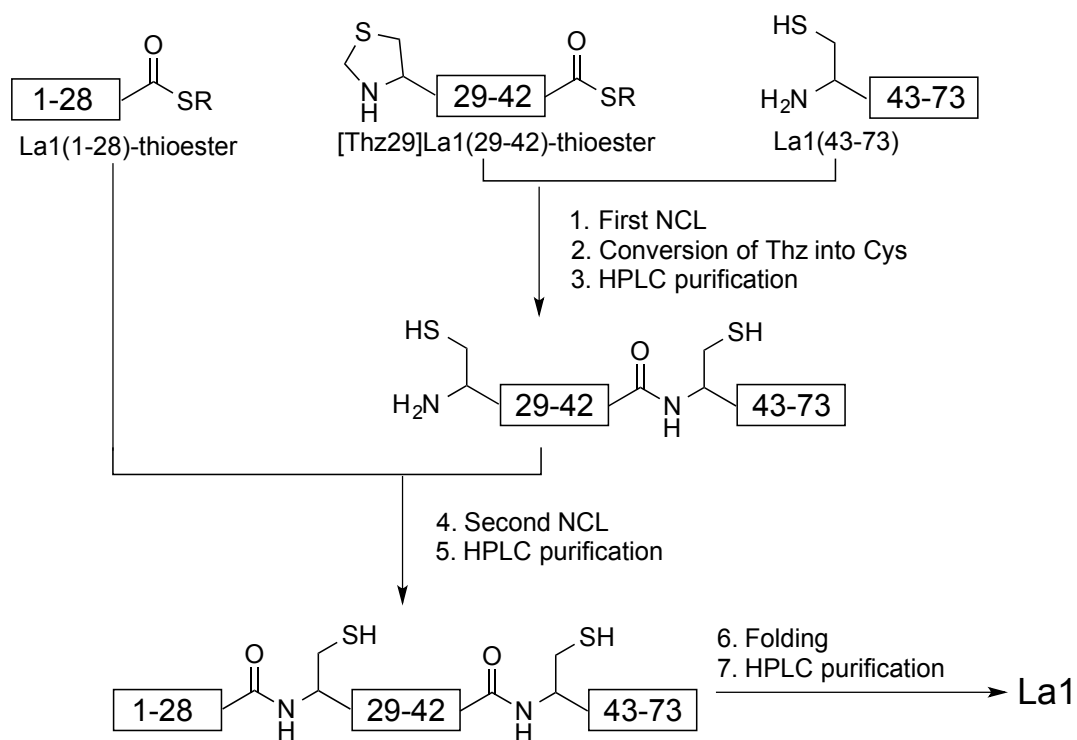
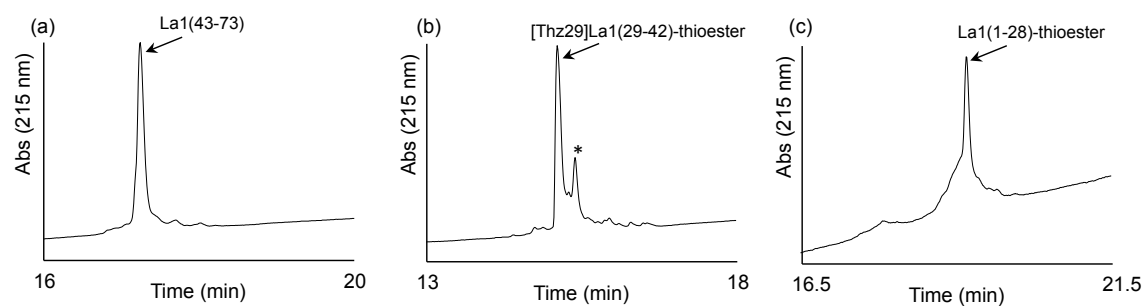Fig. 2   Synthesis of La1 through sequential NCL.

Fig 3   HPLC analysis of peptide segments after partial purification. La1(43-73) (a), [Thz29]La1(29-41)-thioester (b), and La1(1-28)-thioester (c). An asterisk indicates impurities.
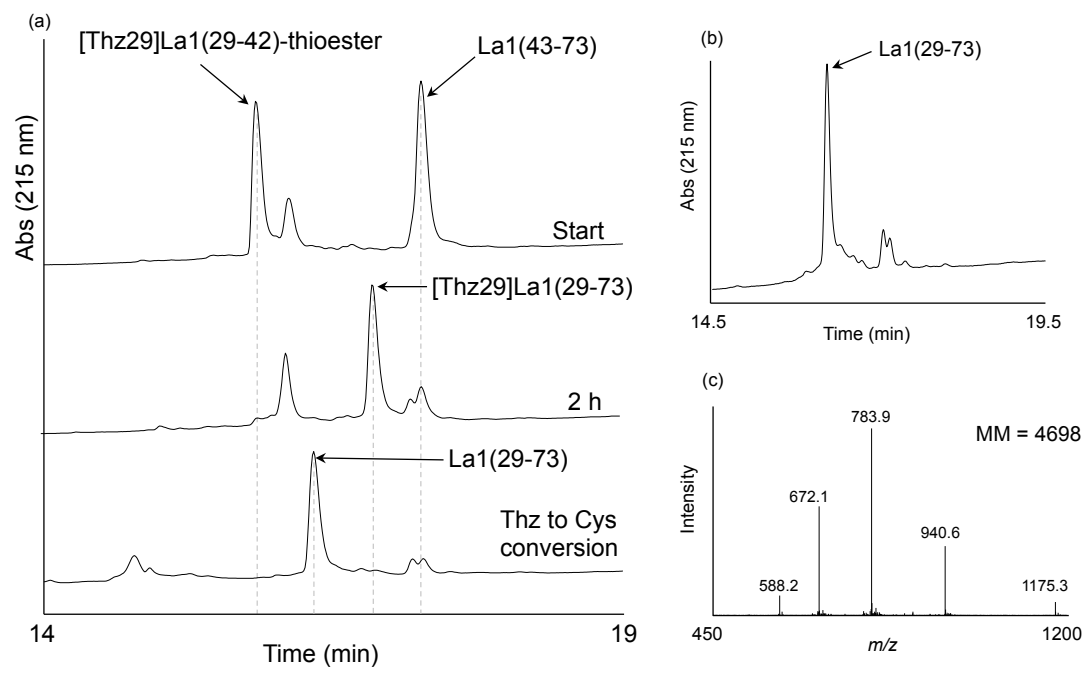
Fig 4   HPLC analysis of 1st ligation (a). UV chromatogram (b) and mass spectrum (c) of partially purified La1(29-73).

Fig 5   HPLC analysis of 2nd ligation (a). UV chromatogram (b) and mass spectrum (c) of purified linear La1

Fig 6   HPLC analysis of folding reaction (a) and comparison of retention time between synthetic and natural La1 (b). Asterisks indicate La1 containing an oxidized Met residue. MS analysis of synthetic La1 (c). The deconvoluted spectrum is shown in the inset.

Fig 7   LC/MS analysis of tryptic digests of La1 without reduction of disulfide bonds. TIC chromatogram is shown.

Fig 8   MS/MS analysis of the peptide fragment containing two disulfide bonds. The letter "h" in the peptide sequence denotes homoserine lactone.

Fig 9   Disulfide bonding pattern of La1 (a) and its comparison with that of chordin-like cysteine-rich repeat from collagen IIA (b). The disulfide bond shown in red is missing in La1.

```
            10            20            30            40            50            60            70
             |             |             |             |             |             |             |
FGES**C**IAGRFIVPLGQQVTDQRD**C**ALYK**C**VNYNKKFALETKR**C**ATVNLKSG**C**KTVPGGAGAAFPS**CC**PMVT**C**K*
                              ↑                        ↑
```

Fig. 1  Amino acid sequence of La1. Asterisk indicates an amidated C-terminus.
Arrows indicate ligation sites.

Fig. 2  Synthesis of La1 through sequential NCL

Fig 3 HPLC analysis of peptide segments after partial purification. (a) La1(43-73), (b) [Thz29]La1(29-41)-thioester, and (c) La1(1-28)-thioester. An asterisk indicates impurities.

Fig. 4  HPLC analysis of 1st ligation (a). UV chromatogram (b) and mass spectrum (c) of partially purified La1(29-73).
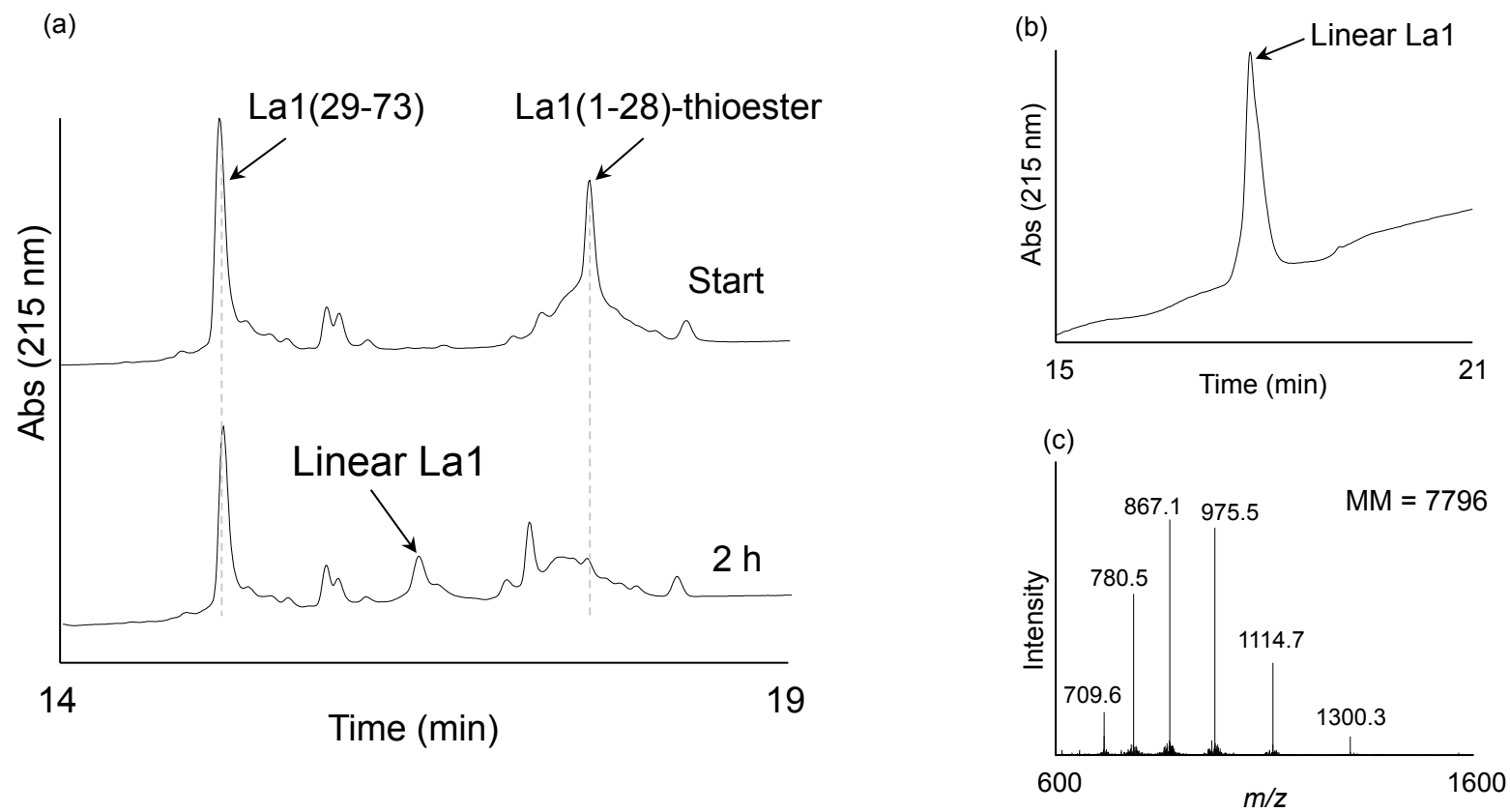
Fig. 5   HPLC analysis of 2nd ligation (a). UV chromatogram (b) and mass spectrum (c) of purified linear La1
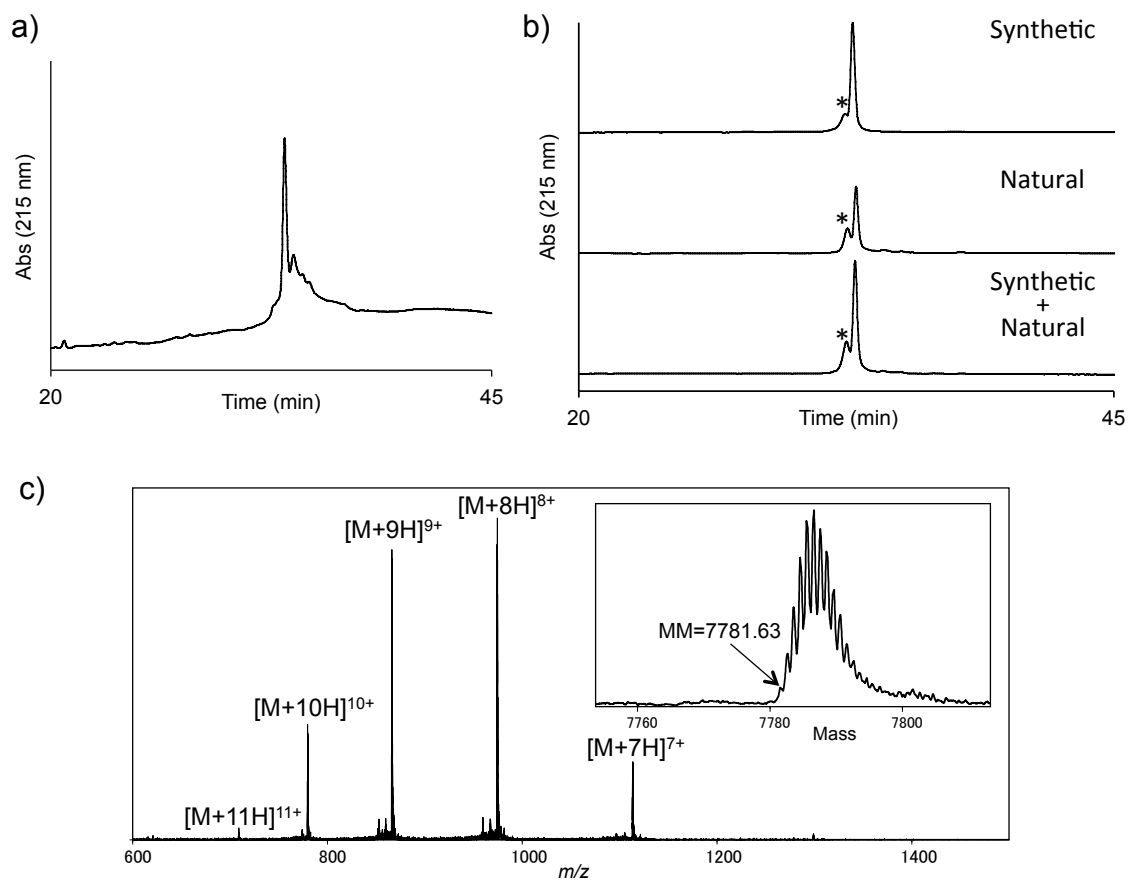
Fig 6  HPLC analysis of folding reaction (a) and comparison of retention time between synthetic and natural La1 (b). Asterisks indicate La1 containing an oxidized Met residue. (c) MS analysis of synthetic La1. The deconvoluted spectrum is shown in the inset.
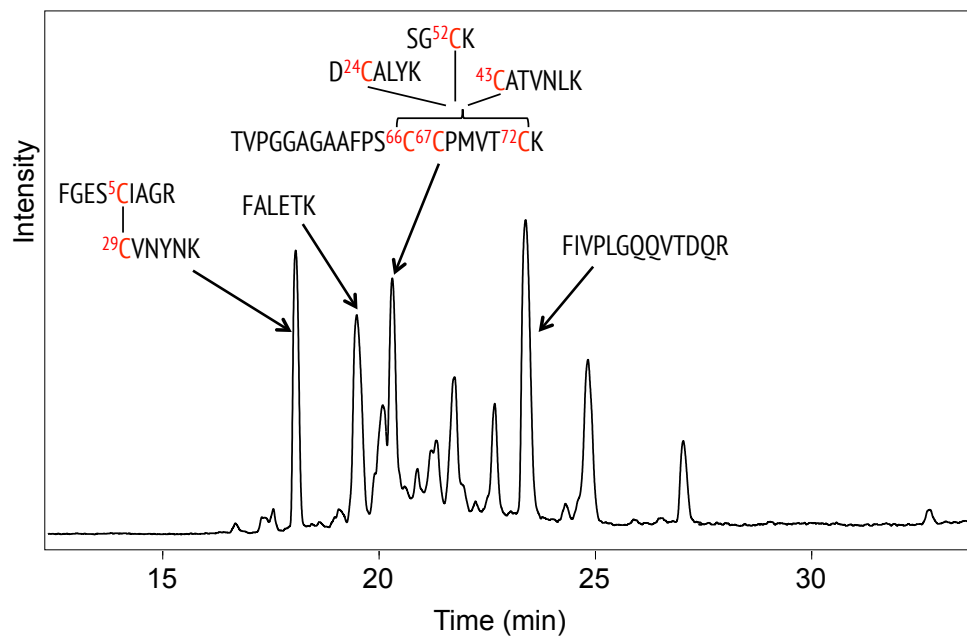
Fig 7  LC/MS analysis of tryptic digests of La1 without reduction of disulfide bonds. TIC chromatogram was shown.
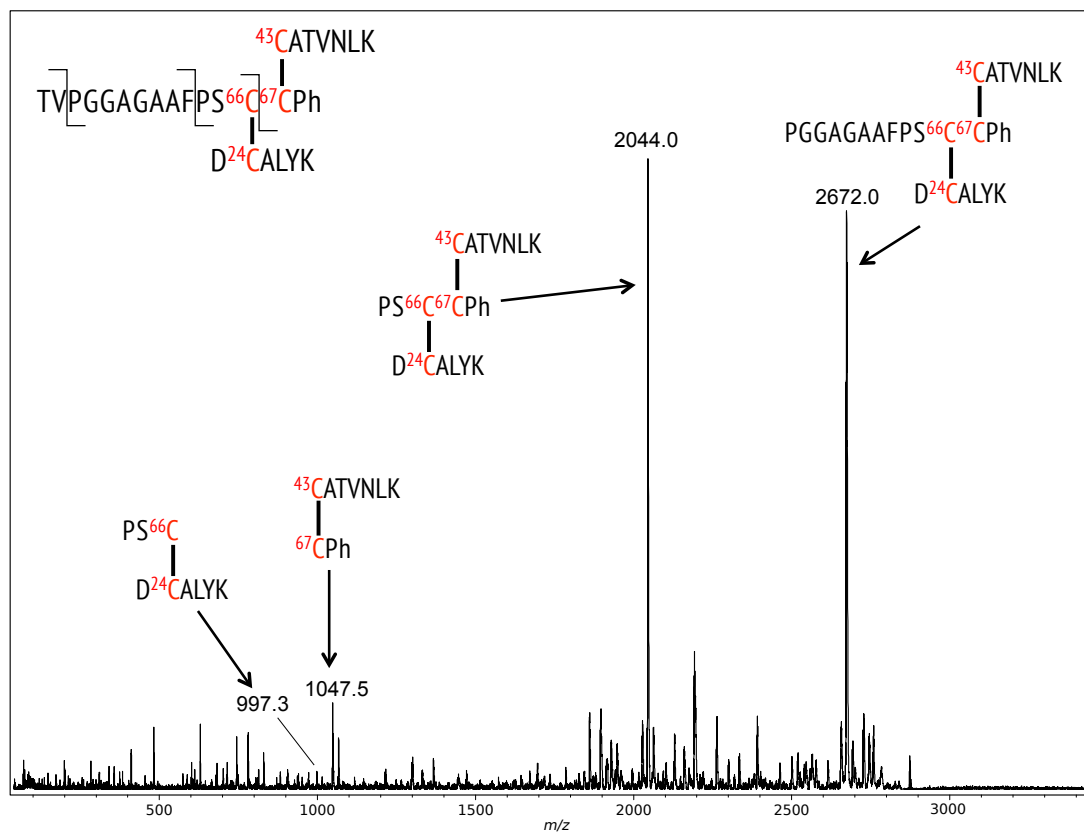
Fig. 8 MS/MS analysis of the peptide fragment containing two disulfide bonds. The letter "h" in the peptide sequence denotes homoserine lactone.

a)

FGES**C**IAGRFIVPLGQQVTDQRD**C**ALYK**C**VNYNKKFALETKR**C**ATVNLKSG**C**KTVPGGAGAAFPS**CC**PMVT**C**K

b)

YVEFQEAGS**C**VQDGQRYNDKDVWKPEP**C**RI**C**V**C**DTGTVL**C**DDII**C**EDVKD**C**LSPEIPFGE**CC**PI**C**PADLAAAA
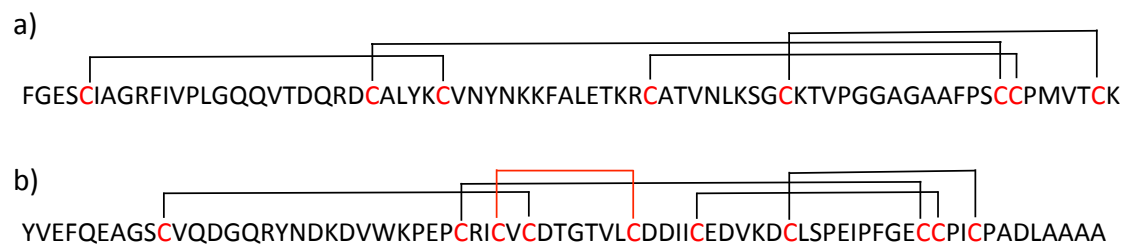
Fig. 9 Disulfide bonding pattern of La1 (a) and its comparison with that of chordin-like cysteine-rich repeat from collagen IIA (b). The disulfide bond shown in red is missing in La1.