Biochimica et Biophysica Acta 1864 (2016) 1444–1454

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Research article

Structural characterization of the major ampullate silk spidroin-2 protein produced by the spider *Nephila clavipes*

José Roberto Aparecido dos Santos-Pinto [a,b], Helen Andrade Arcuri [a], Gert Lubec [b,*], Mario Sergio Palma [a,**]

[a] Center of the Study of Social Insects, Department of Biology, Institute of Biosciences of Rio Claro, São Paulo State University, Rio Claro, SP 13500, Brazil
[b] Department of Pharmaceutical Chemistry, University of Vienna, Vienna 1090, Austria

ARTICLE INFO

Article history:
Received 1 December 2015
Received in revised form 4 May 2016
Accepted 17 May 2016
Available online 18 May 2016

Keywords:
Silk proteins
Nephila clavipes
Mass spectrometry
Post-translational modification
Phosphorylation

ABSTRACT

Major ampullate spidroin-2 (MaSp2) is one of the most important spider silk protein, but up to now no information is available regarding the post-translational modifications (PTMs) of this protein. A gel-based mass spectrometry strategy using collision-induced dissociation (CID) and electron-transfer dissociation (ETD) fragmentation methods was used to sequence *Nephila clavipes* MaSp2 (including the N- and C-terminal non-repetitive domains, and the great part of the central core), and to assign a series of post-translational modifications (PTMs) on to the MaSp2 sequence. Two forms of this protein were identified, with different levels of phosphorylation along their sequences. These findings provide a basis for understanding mechanoelastic properties and can support the future design of recombinant spider silk proteins for biotechnological applications.

1. Introduction

Spiders of *Nephila* genus, have a set of abdominal glands composed of up to seven different types: major ampullate, minor ampullate, flagelliform, tubuliform, aciniform, pyriform and aggregate glands [1]. Different types of silk fibres are produced for various task-specific applications, such as prey capture, egg protection and as a lifeline to escape from predators [2,3]; and specialized secretions used for fibre lubrication and protection produced by aggregate gland [4].

Spider silk is one of the strongest biomaterials and exhibits unique mechanical properties due to an intrinsic assembly of large silk proteins, the spidroins. Major ampullate silk is one of six different types of silk fibres spun by the spiders of the genus *Nephila*, and is a nanostructured composite material containing two structural proteins [5], designated major ampullate spidroin-1 (MaSp1) and major ampullate spidroin-2 (MaSp2) [6–8], which are encoded by several genetic loci [9–11]. The MaSp2 of *Latrodectus hesperus*, *Latrodectus geometricus*, *Nephila madagascariensis*, *Nephila senegalensis and Nephila clavipes* spiders, for example, has been currently reported in several studies based on data obtained from genetic engineering and recombinant DNA technology [12–17].

The mechanical properties of major ampullate silk fibres depend on its highly repetitive backbone region, which has a structure reminiscent of a block copolymer, composed of alternating glycine-rich blocks, followed by an alanine-rich block [10]. MaSp1 contains GGX, $(GA)_n$, and $(A)_n$ motifs, whereas MaSp2 contains $(A)_n$ and GPGXX motifs. The tensile strength is apparently provided by the $(A)_n$ and/or $(GA)_n$ repeats, which form crystalline intra- and intermolecular β-sheet structures in the fibre, whereas the elasticity is dependent on intervening glycine-rich repeats (GGX and GPGXX motifs) [18,19]. The glycine-rich segments are postulated to form different structures, such as β-spirals and coil structures [20,21].

MaSp1 is present in large quantities in silk fibres, being uniformly found throughout the fibre core, whereas MaSp2 is heterogeneously distributed along silk fibres, is clustered in certain core areas, and is absent from the periphery of the fibre core [22,23]. Both spidroins are known to encode repetitive short amino acid sequence motifs that predominantly contain alanine and glycine. However, MaSp1 contains few proline residues, whereas MaSp2 is a proline-rich protein [18,24]. Liu et al. [25,26] studies demonstrated that proline-containing GPGXX motifs contribute to the elasticity of spider silks. Rauscher et al. [18] identified proline as the primary determinant of the elastin-like properties of MaSp2, whereas MaSp1 is proposed to affect tensile strength. The mixture of both proteins in the fibre creates an exceptional combination of tensile strength and extensibility; significant academic interest has been devoted to understanding the properties of these biomaterials at

* Correspondence to: M. Sergio Palma, CEIS-IBRC — UNESP, Av. 24 A, n° 1515, Bela Vista, Rio Claro, SP CEP 13506-900, Brazil.
** Correspondence to: G. Lubec, Department of Pharmaceutical Chemistry, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.
E-mail addresses: gert.lubec@univie.ac.at (G. Lubec), mspalma@rc.unesp.br (M.S. Palma).

http://dx.doi.org/10.1016/j.bbapap.2016.05.007
1570-9639/

the molecular level and their potential uses for industrial applications (e.g., biomedical/textile industries).

Considering the mechanical and physicochemical properties of spidroins, spider silks are well suited for many applications and have been produced using recombinant DNA technology [27–31], however few/ no information about PTMs is available. The structural assignments of N- and C-terminal domains were made using recombinant proteins. The demonstration that these domains were present in mature spidroins was made by using specific antibodies, coupled to mass spectrometry analysis [29]. Only recently was reported that the N- and C-terminal domains, as well as the central core, are part of the same protein [23].

The present study presents an extensive proteomic analysis, which led to the structural characterization of MaSp2. Previously, we reported a proteomic study on the characterization of MaSp1 [32], in which eight reliable phosphorylation sites were identified on the protein produced by *Nephila clavipes*, four were identified on the protein produced by *Nephila madagascariensis*, and two were identified on the silk protein expressed by *Nephila edulis*. The findings reported here might be valuable for understanding the physicochemical properties of the silk proteins, and in the design of recombinant spider silk proteins. In the present study, a gel-based mass spectrometry strategy involving collision-induced dissociation (CID) and electron-transfer dissociation (ETD) fragmentation methods was used to sequence and to assign the presence and location of the PTMs on spidroin-2.

2. Material and methods

2.1. Major ampullate silk samples

The silk of major ampullate gland constitutes the frame of the web, is easily identified to be collected separated from the remaining parts of the web. Thus, *N. clavipes* major ampullate silk (frame) was collected separately from radii and spiral of the web at the University of São Paulo State campus at Rio Claro, SP, Southeast Brazil, and processed as previously described by Santos-Pinto et al. [32]. Approximately 20 mg (dry weight) of major ampullate silk from 15 orb-webs were dissolved in 2 mL of saturated lithium thiocyanate (38 M LiSCN hydrate — Sigma, Deisenhofen, Germany) at 25 °C for 2 h under continuous shaking. The samples were centrifuged at 14,000 ×*g* for 20 min, and the supernatants were transferred to an Ultrafree-4 centrifugal filter unit (Millipore, Bedford, USA) after adding 2 mL of urea buffer [7 M urea in 20 mM Tris containing 2 M thiourea, 4% (*w*/*v*) CHAPS, 10 mM 1,4-dithioerythritol, 1 mM EDTA, 1 mM PMSF, 0.2% (*v*/*v*) protease inhibitor cocktail (Complete™ from Roche Diagnostics, Mannheim, Germany) and 0.2% (*v*/*v*) phosphatase inhibitor cocktail (Calbiochem, San Diego, USA)]. The sample was centrifuged at 4500 ×*g* until the volume was reduced to 0.5 mL. Subsequently, the LiSCN was removed and the buffer was dialysed against urea buffer prior to 2-DE analysis. The protein concentration was determined using the Bradford assay [33].

2.2. Scanning electron microscopy

To examine the structural characteristics of the major ampullate silk fibre surface, a short length (approximately 5 mm) was placed on scanning electron microscopy stubs. Subsequently, the stubs were coated with gold in a Cressington sputter coater using a planetary drive and studied using a Vega II scanning electron microscope (Tascan, Brno, Czech Republic) at 7 kV with a working distance of approximately 8 mm.

2.3. Two-dimensional gel electrophoresis (2-DE)

Samples (200 μg) of major ampullate silk protein were subjected to rehydration on 18-cm IPG strips and pH 7–10 nonlinear gradient strips. Isoelectric focusing (IEF) began at 200 V and was gradually increased to 8000 V (approximately 150,000. Vh). The IPG strips were incubated in equilibration buffer [50 mM Tris–HCl, pH 8.8, 6 M urea, 30% (*v*/*v*) glycerol, 2% (*w*/*v*) SDS containing 1% (*w*/*v*) DTT] for 15 min and then in equilibration buffer containing 4% (*w*/*v*) iodoacetamide for 15 min. The second dimension was run in SDS-PAGE gels (7–10% gradient) at 50 V (which was held constant overnight) and then at a constant 200 V for an additional 4 h at 10 °C. The gels were stained with *Coomassie Brilliant Blue R-250* (CBB) and were scanned and digitised for documentation.

2.4. In-gel digestion

Gel pieces were destained twice for 30 min at 25 °C with 10 mM ammonium bicarbonate/50% (*v*/*v*) acetonitrile, dehydrated in acetonitrile, dried, and subsequently treated with the following four proteolytic

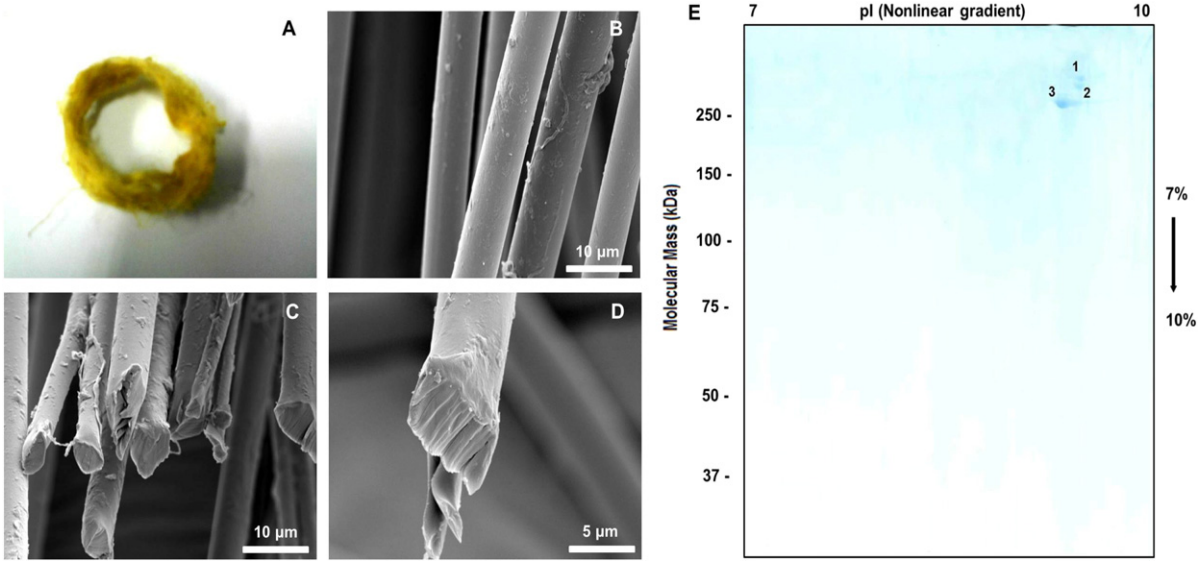


Fig 1. Scanning electron microscopy analysis and proteomic analysis of the major ampullate silk produced by the spider *N. clavipes*. (A) Fibres that form the frame of web. (B–D) Ultrastructure of the major ampullate silk fibres surface at various locations on the frame of web. (E) Representative 2-DE profile of the major ampullate silk proteins of *N. clavipes* stained with *Coomassie Colloidal Blue*.

Table 1
Proteolytic fragments of MaSp2 from the major ampullate silk produced by *N. clavipes*, obtained by digestions with trypsin, chymotrypsin, pepsin, and protease 10. The peptides are presented according to the protein domain to which they belong to (N-terminal domain, central repetitive domain, and C-terminal domain), as well are presented the *m/z* value of the respective molecular ion, charge state, spectral counting during mass spectrometric analysis, and normalized value of spectral counting.

Peptide sequence	*m/z* (charge)	Spectral counting	Normalized spectral counting (integer number)
Non-repetitive N-terminal domain			
SDTATADAFIQNFL	757.30 (+2)	5	1.0 (1)
AAVSGSGAF	766.33 (+1)	8	1.6 (2)
ARSNKSSQHKLQ	692.35 (+2)	6	1.2 (1)
SNKSSQHK	421.76 (+3)	7	1.4 (1)
LQALNMAFASSMAEIAAVEQGGMSMAVK	974.05 (+3)	5	1.0 (1)
QALNMAF	405.68 (+2)	6	1.2 (1)
TNAIVDGLNSAFYMTTGAANPQFVNEMR	1022.08 (+3)	6	1.2 (1)
YMTTGAANPQF	608.77 (+2)	8	1.6 (2)
MTTGAANPQFVNEMRSL	933.88 (+2)	5	1.0 (1)
AAAPSGYGPSQQGPSGSAAAAAAAAPSGYGPR	934.46 (+3)	6	1.2 (1)
APSGYGPSQQGPSGSAAAAAAAAPSGYGPRQQAPSGPAAAAA	1227.33 (+3)	5	1.0 (1)
GPRQQAPSGPAAAAAAAGAGGYGPAQRGPSVY	979.14 (+3)	7	1.4 (1)
QQAPSGPAAAAAAAGAGGYGPAQR	708.08 (+3)	6	1.2 (1)
YGPAQRGPSVYGPSGPGAAAAAAAGAGPGG	847.52 (+3)	6	1.2 (1)
GPSVYGPSGPGAAAAAAAGAGPGGYGPGQQGP	1001.16 (+3)	5	1.0 (1)
Central repetitive domain composed of 6 modules			
GGYGPGQQGPGGY	844.23 (+2)	320	64.0 (64)
GYGPGQQGPGGYGPGQQGPS	923.37 (+2)	120	24.0 (24)
GYGPGQQGPGGYGPGQQGPSGPG	1028.92 (+2)	90	18.0 (18)
GYGPGQQGPGGYGPGQQGPSGPGS	1072.41 (+2)	90	18.0 (18)
YGPGQQGPGGYGPGQQGPS	894.88 (+2)	120	24.0 (24)
YGPGQQGPGGYGPGQQGPSGPGS	1043.96 (+2)	90	18.0 (18)
YGPGQQGPGGYGPGQQGPSGPGSA	1079.49 (+2)	90	18.0 (18)
GPGQQGPGGY	819.96 (+1)	450	90.0 (90)
QQGPGGYGPGQQGPSGPG	813.310 (+2)	150	30.0 (30)
QQGPGGYGPGQQGPSGPGS	856.90 (+2)	150	30.0 (30)
QGPGGYGPGQQGPSGP	720.80 (+2)	150	30.0 (30)
GYGPGQQGPSGPGS	623.27 (+2)	150	30.0 (30)
GYGPGQQGPSGPGSAAAAAAAAA	983.89 (+2)	60	12.0 (12)
YGPGQQGPSGPG	1101.43 (+1)	180	36.0 (36)
YGPGQQGPSGPGS	1188.54 (+1)	180	36.0 (36)
GPGQQGPSGPGSAAAAAAAAAAGPGGY	846.70 (+3)	30	6.0 (6)
GPGQQGPSGPGSAAAAAAAAAAGPGGYGPGQQGPGGYGPGQQGPGRY	999.51 (+4)	30	6.0 (6)
SGPGSAAAAAAAAAAGPGGY	813.42 (+2)	30	6.0 (6)
AAAGPGGYGPGQQ	565.77 (+2)	120	24.0 (24)
GPGQQGPGGYGPGQQGPGRY	737.64 (+3)	30	6.0 (6)
GYGPGQQGPGRYGPGQQGPSGPG	1158.45 (+2)	30	6.0 (6)
GPGQQGPGRYGPGQQGPSGPGSAAAAAAGSGQQGPGGYGPRQQGPGGY	1499.02 (+3)	30	6.0 (6)
QGPGRYGP	911.36 (+1)	30	6.0 (6)
QGPGRYGPGQQGPSGPGSAAAA	709.96 (+3)	30	6.0 (6)
YGPGQQGPSGPGSAAAAAAGSGQQGPGGY	888.35 (+3)	30	6.0 (6)
GPGQQGPSGPGSAAAAAAGSGQQGPGGY	836.57 (+3)	30	6.0 (6)
GPGQQGPSGPGSAAAAAAGSGQQGPGGYGPRQQGPGGY	918.04 (+4)	30	6.0 (6)
AAAAAAGSGQQGP	568.75 (+2)	30	6.0 (6)
SGQQGPGGYGPRQQGPGGYGQGQQGPSGPG	960.07 (+3)	30	6.0 (6)
GPGGYGQGQQGPS	635.75 (+2)	30	6.0 (6)
GQGQQGPSGPGSAAAASAAASAESGQQGPGGY	768.52 (+4)	30	6.0 (6)
PGSAAAASAAA	924.41 (+1)	30	6.0 (6)
ESGQQGPGGYGPGQQGPGG	897.88 (+2)	30	6.0 (6)
GQQGPGGYGPGQQGPGGY	554.61 (+3)	150	30.0 (30)
GPGQQGPGGYGPGQQGPGGY	598.66 (+3)	120	24.0 (24)
QGPGGYGPGQQGPGGYGPGQQGPSGP	780.39 (+3)	60	12.0 (12)
GPGQQGPSGPGSAAAAAAAASGPGQQGPGGYGPGQQGPGGY	787.21 (+4)	60	12.0 (12)
AAAASGPGQQGPGGYGPGQQGP	955.33 (+2)	60	12.0 (12)
ASGPGQQGPGGYGPGQQGPG	877.35 (+2)	60	12.0 (12)
ASGPGQQGPGGYGPGQQGPGG	905.86 (+2)	60	12.0 (12)
ASGPGQQGPGGYGPGQQGPGGYGPGQQGPSGPG	965.13 (+3)	30	6.0 (6)
ASGPGQQGPGGYGPGQQGPGGYGPGQQGPSGPGSA	1017.78 (+3)	30	6.0 (6)
SGPGQQGPGGYGPGQQGPG	841.82 (+2)	60	12.0 (12)
SGPGQQGPGGYGPGQQGPGGYGPGQQGPSGPG	941.44 (+3)	30	6.0 (6)
SGPGQQGPGGYGPGQQGPGGYGPGQQGPSGPGSA	994.05 (+3)	30	6.0 (6)
GPGGYGPGQQGLSGPGSAAAAAAA	1025.89 (+2)	30	6.0 (6)
GPGQQGLSGPGSAAAAAAAGPGQQGPGGY	841.77 (+3)	30	6.0 (6)
SGPGSAAAAAAAGPGQQGPGGY	869.35 (+2)	30	6.0 (6)
AAGPGQQGPGGYGPGQQGPSGPG	989.92 (+2)	60	12.0 (12)
AAGPGQQGPGGYGPGQQGPSGPGS	1033.38 (+2)	60	12.0 (12)
GPGQQGPSGPGSAAAAAAAAAGPGGY	1000.26 (+2)	30	6.0 (6)
GPGQQGPSGPGSAAAAAAAAAGPGGYGPGQQGPGGY	780.94 (+4)	30	6.0 (6)
SGPGSAAAAAAAAAGPGGYGPGQQGPGGYGPGQQGPS	1054.51 (+3)	30	6.0 (6)
GPGQQGPGGYGPGQQGPSGAGSAAAAAAAGPGQQGL	1017.91 (+3)	30	6.0 (6)
GPGQQGPGGYGPGQQGPSGAGSAAAAAAAGPGQQGLGGY	841.27 (+4)	30	6.0 (6)

Table 1 (*continued*)

Peptide sequence	*m/z* (charge)	Spectral counting	Normalized spectral counting (integer number)
Central repetitive domain composed of 6 modules			
GPGQQGPSGAGSAAAAAAAGPGQQGL	554.03 (+4)	30	6.0 (6)
GPGQQGPSGAGSAAAAAAAGPGQQGLGGY	1479.18 (+2)	30	6.0 (6)
QGLGGYGPGQQGPGGYGPGQQGPGGYGPGSASAAA	827.05 (+4)	30	6.0 (6)
GGYGPGQQGPGGYGPGQQGPGGY	668.06 (+3)	60	12.0 (12)
GPGQQGPGGYGPGSASAAAAAAGPGQQGPGGY	749.23 (+4)	30	6.0 (6)
GPGSASAAAAAAGPGQQGPGGY	874.80 (+2)	30	6.0 (6)
GSASAAAAAAGPGQQGPGGYGPGQQGPSGPGSASAAAAAAA	1130.77 (+3)	30	6.0 (6)
GPGGYGPGQQGPSGPGSASAA	636.92 (+3)	30	6.0 (6)
GYGPGQQGPSGPGSASAAAAAAAA	986.38 (+2)	30	6.0 (6)
GPGQQGPSGPGSASAAAAAAAAGPGGY	827.13 (+3)	30	6.0 (6)
GPGQQGPSGPGSASAAAAAAAAGPGGYGPGQQGPGGY	806.97 (+4)	30	6.0 (6)
PGQQGPSGPGSASAAAAAA	816.80 (+2)	60	12.0 (12)
GPGSASAAAAAAAAGPGGY	777.92 (+2)	60	12.0 (12)
SASAAAAAAAAGPGGYGPGQQGPGGYAPGQQGPS	988.73 (+3)	60	12.0 (12)
SAAAAAAAAG	811.39 (+1)	60	12.0 (12)
QGPGGYAPGQQGPSGPG	756.28 (+2)	60	12.0 (12)
QGPGGYAPGQQGPSGPGS	799.84 (+2)	60	12.0 (12)
QGPGGYAPGQQGPSGPGSA	835.34 (+2)	60	12.0 (12)
APGQQGPSGPGSASAAAAAAAAGPGGYGPGQQGPGGY	814.96 (+4)	30	6.0 (6)
QGPGGYAPGQQGPSGPGSAAAAAAAAA	773.99 (+3)	30	6.0 (6)
APGQQGPSGPGSAAAAAAAAAGPGGY	548.76 (+4)	30	6.0 (6)
Non-repetitive C-terminal domain			
GYGPAQQGPAGYGPGSAVAASA	1007.53 (+2)	5	1.0 (1)
GPAQQGPAGYGPGSAVAASAGAGSAGY	838.98 (+3)	5	1.0 (1)
GPGSAVAASAGAGSAGY	451.87 (+3)	6	1.2 (1)
GPGSQASAAASRL	583.65 (+2)	8	1.6 (2)
RLASPDSGARVASAVSNLVSSGPTS	902.77 (+3)	5	1.0 (1)
ASPDSGARVASAVSNLVSSGPTSSAAL	820.48 (+3)	5	1.0 (1)
ASPDSGARVASAVSNL	940.34 (+2)	7	1.4 (1)
DSGARVASAVSNL	663.90 (+2)	5	1.0 (1)
VASAVSNLVSSGPTSSAAL	859.38 (+2)	5	1.0 (1)
VSSGPTSSAAL	606.79 (+2)	6	1.2 (1)
VSSGPTSSAALSSVISNAVSQIGASNPGL	1313.73 (+2)	5	1.0 (1)
SGPTSSAAL	870.36 (+1)	7	1.4 (1)
SSVISNAVSQIGASNPGLSGCDVLIQAL	926.59 (+3)	5	1.0 (1)
VSQIGASNPGLSGCDVL	837.36 (+2)	6	1.2 (1)
QIGASNPGLS	943.38 (+1)	6	1.2 (1)
IQALLEIVSACVTIL	749.91 (+2)	5	1.0 (1)
EIVSACVTILSSSSIGQVNY	1036.30 (+2)	5	1.0 (1)
SACVTILSSSSIGQVN	621.57 (+3)	6	1.2 (1)
SSSSIGQVNY	540.42 (+2)	7	1.4 (1)
SSSSIGQVNYGAASQF	801.90 (+2)	7	1.4 (1)
SSIGQVNYGAASQFAQVVGQ	1086.05 (+2)	5	1.0 (1)
QVNYGAASQFAQVVGQSVLSA	762.06 (+3)	5	1.0 (1)
GAASQFAQVVGQSVL	743.31 (+2)	7	1.4 (1)
AQVVGQSVL	450.80 (+2)	6	1.2 (1)
AQVVGQSVLSAF	407.21 (+3)	5	1.0 (1)

enzymes: 40 ng/mL of trypsin (Promega, Madison, USA) in 5 mM octyl β-D-glucopyranoside (OGP) and 10 mM ammonium bicarbonate pH 7.9 at 37 °C for 18 h; 50 ng/mL of chymotrypsin (Roche Diagnostics) in 5 mM OGP and 25 mM ammonium bicarbonate pH 7.9 at 30 °C for 2 h; pepsin (Sigma, Deisenhofen, Germany) in 0.1 M HCl pH 1.0 at 37 °C for 4 h; and proteinase 10 (syn.: Thermolysin) in 25 mM Tris–HCl pH 7.5 at 50 °C for 2 h. Peptide extraction was performed using 0.5% (*v*/*v*) formic acid and 0.5% (*v*/*v*) formic acid in 30% (*v*/*v*) acetonitrile. The extracted peptides were then pooled for nanoLC-ESI-CID/ETD-MS^n analysis.

2.5. NanoLC-ESI-CID/ETD-MS^n

The HPLC instrument used in the analysis was an Ultimate 3000 system (Dionex, Sunnyvale, CA, USA) equipped with a PepMap100 C-18 trap column (300 mm × 5 mm) and a PepMap100 C-18 analytic column (75 mm × 150 mm). The gradient (A — 0.1% formic acid in water, B — 0.08% formic acid in acetonitrile) was 4–30% B from 0 min to 105 min, 80% B from 105 min to 110 min, and 4% B from 110 min to 125 min. An HCT ultra ETD II instrument (Bruker Daltonics, Bremen, Germany) was used to record peptide spectra over a mass range of *m/z* 350–3500, and MS/MS spectra were acquired in information-dependent data acquisition mode over a mass range of *m/z* 100–3500. MS spectra were recorded repeatedly and then three data-dependent CID MS/MS spectra and three ETD MS/MS spectra were generated from the three precursor ions of the highest intensity. An active exclusion of 0.4 min was used after two spectra to detect low-abundance peptides. The voltage between the ion spray tip and the spray shield was set to 1500 V. The nitrogen gas used for drying was heated to 150 °C, and the flow rate used was 10 L/min. The collision energy was set automatically according to the mass and charge state of the peptides that were chosen for fragmentation. Multiple charged peptides were selected for analysis in the MS/MS experiments due to their good fragmentation characteristics. The MS/MS spectra were interpreted and peak lists were generated using DataAnalysis 4.0 (Bruker Daltonics). The combined use of CID and ETD generated spectral data, which were analysed using a MASCOT protein engine search and Modiro®.

2.6. Protein identification and post-translational modifications

Searches were conducted using MASCOT 2.2.06 (Matrix Science, London, UK) against the latest NCBI database (http://blast.ncbi.nlm.

nih.gov, on 08/18/2014) for protein identification; for this purpose, we selected all 57,321 entries contained in the taxa Araneae (spider). The databanks mentioned above were appended with common external contaminants from cRAP, a maintained list of contaminants, laboratory proteins and protein standards provided through the Global Proteome Machine Organization (http://www.thegpm.org/crap/index.html). The search parameters were set as follows: enzyme selected as trypsin or alternative enzymes such as chymotrypsin, pepsin and proteinase 10, two maximum missing cleavage sites allowed, 0.2 Da peptide mass tolerance for MS and 0.2 Da tolerance for MS/MS spectra; carbamidomethyl (C) was specified as a fixed modification, while methionine oxidation and phosphorylation (of Y, T, and S) were specified in MASCOT as variable modifications. After protein identification, an error-tolerant search was performed to detect non-specific cleavage and unassigned modifications [34]. Proteins identified after database search were subjected to additional filtering using Scaffold 4.3.2 (Proteome Software Inc., Portland, OR) to validate peptide identification and to obtain a false discovery rate (FDR) of <1%; FDR was calculated from the forward and decoy matches by requiring significant matches to at least 2 distinct sequences. According to a Local FDR algorithm implemented into Scaffold, the peptide probability was set to a minimum of 90%, whereas the protein probability was set at 95%. PTM searches were performed using Modiro™ — PTM Explorer 1.1 software (Protagen AG, Dortmund, Germany) with the following parameters: enzyme selected as trypsin or alternative enzymes such as chymotrypsin, pepsin and proteinase 10, two maximum missing cleavage sites allowed, 0.2 Da peptide mass tolerance for MS, 0.2 Da tolerance for MS/MS spectra, and carbamidomethyl (C) and methionine oxidation were specified as modifications 1 and 2, respectively. The modification data returned were checked to obtain confirmed PTM identification by Modiro software. Searches for unknown mass shifts, amino acid substitution, and calculation of significance were selected in advanced PTM-explorer search strategies. The Modiro software is complementary to the MASCOT software, using already identified sequences and has the advantage that also unknown mass shifts can be handled [35]. A list of 172 common modifications including phosphorylation was selected and applied to virtually cleaved and fragmented peptides that were compared with experimentally obtained MS/MS spectra. Protein identification was first listed according to the spectral view by Modiro. Subsequently, each identified peptide considered significant based on the ion-charge status of the peptide, *b*- and *y*-ion fragmentation qualities, ion score (>200), and significance score (>80), which are components set of Modiro software, were manually assigned in the respective spectrum. The gel-based proteomic approach used was based on that described by Kang et al. [36].

2.7. Quantification of the proteolytic peptides

The quantification of the proteolytic peptides was performed using the spectral counting of the data generated during the NanoLC-ESI-MS analysis as previously done in other studies [37–41] of the individual digests of *N. clavipes* MaSp2 with trypsin, chymotrypsin, pepsin, and proteinase 10. The extracted ion chromatograms of each peptide were manually inspected for spectral counting of each peptide. The results were normalized dividing the individual counting of each peptide by the lowest counting observed amongst all the peptides.

2.8. Western blotting

To verify the phosphoserine modifications of the major ampullate spidroin proteins, 20 μg of major ampullate silk protein extract was loaded onto 1D-SDS-PAGE gel; after electrophoresis, the proteins were transferred onto PVDF membrane (Millipore) at 21 °C using a semi-dry Bio-Rad transfer system. The membrane was blocked by incubation for 1 h in T-TBS containing 0.1% (*v*/*v*) Tween 20 and 5% BSA (bovine serine albumin). After washing, the membrane was incubated with diluted primary anti-phosphoserine antibodies (1:2000, Abcam) at 4 °C overnight. The membrane was then washed three times by gentle agitation in T-TBS (containing 0.1% (*v*/*v*) Tween 20) and incubated with HRP-coupled anti-mouse IgG secondary antibodies (1:10,000, Abcam). As an experimental control, the major ampullate spidroin proteins were treated with phosphatase (calf intestine alkaline phosphatase, New England Biolabs) prior to resolving on 1D-SDS-PAGE gels. The membrane was developed with the Amersham ECL-plus western blotting detection system (GE Healthcare).

3. Results

Considering that Masp2 is one of the major ampullate silk proteins produced by *N. clavipes* spider for use in the construction of the frame and radii of orb webs, and as a dragline to escape from predators, an experimental approach was developed combining 2-DE with multiple rounds of proteolytic in-gel digestion, followed by mass spectrometry analysis for the study of MaSp2.

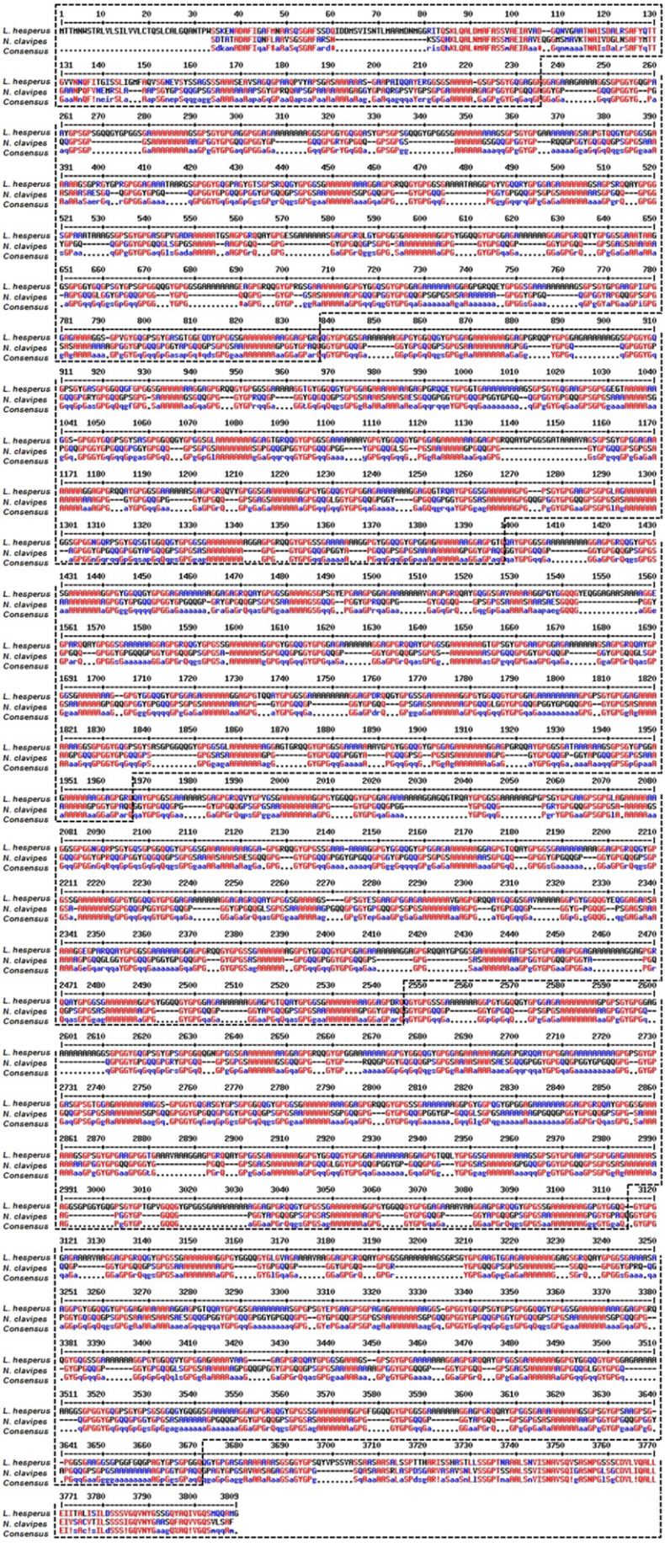
Scanning electron microscopy analysis of major ampullate silk revealed the ultrastructure of the silk fibres at different frame positions, showing that the fibres are constituted by compact arrays of several layers of silk (Fig. 1A–D). The solubilization of silk in lithium thiocyanate represented a key step in dissolving the silk proteins prior to separation by 2-DE. The electrophoretic profile of major ampullate silk from *N. clavipes* (Fig. 1E) revealed three spidroin spots with molecular weights above 250 kDa, which were labelled 1, 2 and 3. Proteomic analysis of these three spots identified MaSp1 (spots 1, 2, and 3) (GenBank ID accession number P19837) and MaSp2 (spots 2 and 3) (GenBank ID accession numbers P46804 and B5SYS7).

The objective of this manuscript is to describe the proteomic characterization of MaSp2; therefore, the sequence assignment was focused on spots 2 and 3, as shown in Tables S1 and S2. Because the sequence assignment of MaSp1 has been reported previously [32], the proteomic identification and sequence assignment of this protein are not presented here.

The spidroins obtained from spots 2 and 3 were cleaved using proteases and subjected to tandem mass spectrometry analysis. This enabled the detection and sequence assignment of 157 proteolytic peptides sequence related to the sequence of MaSp2: 26 tryptic peptides, 11 peptic peptides, 68 chymotryptic peptides and 52 proteolytic peptides were detected after digestion with proteinase 10. The Table S3 shows the list of proteolytic peptides generated by the digestion of MaSp2 with trypsin, chymotrypsin, pepsin, and proteinase 10, presented according to the protein domain to which they belong to (N-terminal domain, central repetitive domain, and C-terminal domain). This table also presents the *m*/*z* values of the respective molecular ions, charge state, spectral counting during mass spectrometric analysis, normalized value of spectral counting, and the position of occurrence of each peptide along the sequence of *N. clavipes* MaSp2. Considering that this table is very large to be included in the body of the manuscript, its summarized version is presented in Table 1. In this table are reported 116 non-redundant peptides.

When the sequence of these peptides was aligned with the sequence of MaSp2 from *L. hesperus*, it was observed that they were aligned all over the most of sequence of the protein (Fig. 2), revealing an identity of 49.7% of *N. clavipes* MaSp2 in relation to *L. hesperus* protein. Meanwhile, the alignment with *Nephila madagascariensis* MaSp2 (Fig. S1) revealed an identity of 60%; however, the absence of the sequence of C-terminal region of *N. madagascariensis* did not permit to observe the alignment in this region. There are several

Fig 2. Alignment of sequence of all the proteolytic fragments obtained for *N. clavipes* MASp2 with the sequence of the protein from *L. hesperus*. Red — high sequence conservation; Blue — low sequence conservation.


L. hesperus
N. clavipes
Consensus
NR-NTD
CRD Module 1
CRD Module 2
CRD Module 3
CRD Module 4
CRD Module 5
CRD Module 6
NR-CTD

gaps in the sequence of *N. clavipes* protein when compared to *L. hesperus* (Fig. 2) which may be due to missed peptides lost during sample analysis, or even these sequences are absent from *N. clavipes* MaSp2. When the comparison is made with *N madagascariensis* protein (Fig. S1) it is possible to observe that in the central repetitive region of *N. clavipes* protein contains many stretches of sequences which seems to be absent from *N. madagascariensis* protein.

A detailed observation of these alignments reveals that several peptides, mainly from the central repetitive region occurred repeatedly many times at different locations of the sequence (Table S3 shows the exact locations of each peptide). As a consequence it is possible to identify six modules of the central repetitive domain in tandem, between the N- and C-terminal domains (Fig. 2 and S1).

There are peptides whose the sequence occurred many times all over *N. clavipes* MaSp2, while other peptides occurred a single time, as shown in Tables 1 and S3; thus, it was necessary to verify if the concentration of each peptide detected in the Total Ion Chromatogram of the digests of MaSp2 was proportional to the number of times that each sequence was repeated along the complete of the silk protein. Technically this task is difficult to be performed since it would require the isolation of each peptide for individual quantification; however, we used successfully a semi-quantitative method based on spectral counting during the LC–MS. The results are shown in Tables 1 and S3.

High-sequence coverage of MaSp2 enabled the detection of a series of PTMs. The combined use of CID and ETD generated spectral data, which were analysed using a MASCOT protein engine search and Modiro®, and the results permitted the assignment of a series modifications in the MaSp2 sequence. Tables, S1 and S2, present detailed proteomic data describing all of the modifications assignments. Modifications such as oxidation, deamidation and methylation might have occurred due to various artefacts arising from sample preparation or analytical procedures.

Phosphorylation was the major PTM assigned onto the MaSp2 sequence. Considering that there is insufficient space to present all of the spectra used to assign the position of each phosphorylation site, we present two representative spectra that were used for this purpose. Thus, Fig. 3A shows the CID spectrum of the chymotryptic peptide S*GPGSAAAAAAAAAAGPGGY (200–219) used to assign the phosphorylation on *S200; meanwhile, Fig. 3B shows the CID spectrum of the proteolytic peptide SSIGQVNY*GAAS*QFAQVVGQ (3121–3140), which was obtained by digesting spidroin-2 with proteinase 10 and used to assign the phosphorylation sites at *Y3128 and *S3132. The mass difference of 18 Da between the theoretical and experimental data shown in Fig. 3A and B, indicates the potential loss of water normally observed for all the series of *y*-type of fragment-ions; this fact does not change the peptides sequence assignment. Table 2 shows a map of the phosphorylation sites observed on spidroin-2 present in the spots 2 and 3 (Fig.1E), highlighting the position of these modifications along the sequence, and 36 phosphorylation sites were observed on residues S200, Y239, S294, S423, S561, S673, Y712, S896, S1034, S1146, Y1185, S1240, S1369, S1507, S1619, Y1658, S1713, S1842, S1980, S2092, Y2131, S2186, S2315, S2453, S2565, Y2604, S2659, S2788, S2926, T3074, S3120, S3121, S3122, Y3128 and S3132 of the MaSp2 protein. The two types of fragmentation, CID (*b*- and *y*-ions) and ETD (c and z-ions), were complementary to each other, showing the mass difference of 80 Da (HPO_3) representative of phosphorylation sites [42–44]. Fig. 4 is showing the sequence of *N. clavipes* MaSp2, indicating the N- and C-terminal regions, as well six modules of the repetitive central domain with the assignment of the positions of phosphorylation for the MaSp2.

In addition to the mass spectral data shown in Fig.3A and B, some other mass spectral assignments of phosphorylation sites are also shown in Figs. S2–S8. Thus, the numbering sequence obtained in these locations assignment does not match the position of these sequences in the accession numbers by which the protein has been identified in databases (Tables S2 and S3); in the present manuscript we are using a numbering based on the proteins used for alignment of the sequence of MaSp2, while in proteins DBs the numbering system refers to each partial sequences. In order to confirm the phosphorylation of *N. clavipes* MaSp2 the solution of the protein was submitted to SDS-PAGE, and then to immunoblotting, using IgG anti-phosphoserine for staining. An intense staining was observed for the protein without any treatment, while the treatment of the protein with alkaline phosphatase removed completely the staining (Fig. S9A and S9B).

The potential sequences of the N- and C-terminal domains from *N. clavipes* MaSp2 were aligned with the sequences of the same domains from MaSp2 from other web-spiders (Fig. 5A and B), revealing to be relatively well conserved, mainly in relation to the non-repetitive C-terminal domain.

4. Discussion

Silk fibres from *N. clavipes* are nanostructured composite materials that are predominantly composed of two structural proteins: MaSp1 and MaSp2 [5,45,46]. Most of the structural assignments for the N- and C-terminal domains were determined by isolating mRNA from silk glands and expressing recombinant protein [29]. Thus, it was unclear whether these domains were present in mature spider silk proteins. However, the use of specific antibodies associated with mass spectrometry analysis permitted to demonstrate that the N- and C-terminal domains are part of the same protein, together with the central core [23,47–49]. It is known that spider silk proteins generally consist of a large repetitive core sequence that is flanked by non-repetitive amino- and carboxy-terminal domains [47]. Therefore, the aim of the present initiative was to characterize the structure of MaSp2 (including the non-repetitive N- and C-terminal domains and the central β-hairpin core) using a proteomic approach. For this purpose, 2-D gel electrophoresis and multiple proteolytic *in-gel* protein digests followed by mass spectrometry analysis using collision-induced dissociation (CID) and electron-transfer dissociation (ETD) for fragmentation were used to generate high-sequence coverage of MaSp2 from the major ampullate silk expressed by *N. clavipes*.

The results of 2-DE (Fig. 1E) suggest that spot 1 contains only MaSp1, whereas spots 2 and 3 should contain hetero-oligomers of both proteins; however, the apparent MW value observed for the spots 2 and 3 (~270 kDa) suggest that the cross-linking between both proteins was reduced under the denaturing conditions of 2-DE described in material and methods (in the present study it was not investigated the chemical nature of these crosslinks). Thus, MaSp1 and MaSp2 presented electrophoretic mobility of large monomeric proteins, not well resolved from each other by the gel system (constituted of SDS-PAGE under polyacrylamide gradient from 7 to 10%), making both spots partially overlapped to each other. Despite the differences in MW between MaSp1 and MaSp2, the occurrence of extensive post-translational modifications of these proteins possibly masked these differences, which were not enough for discriminating both proteins based on their individual MW and pI values.

The protein databases GenBank (http://www.ncbi.nlm.nih.gov/genbank/submit) and Uniprot (http://www.uniprot.org/help/uniprotkb) contain several entries for different fragments of the protein that were virtually translated from DNA sequences individually, and deposited by different authors. Thus, the strategy for sequencing MaSp2 was to cover as much of the sequence as possible, based on the proteolytic fragments in spots 2 and 3, and then overlapping the peptide sequences using the existing sequence information in GenBank and Uniprot to obtain the highest possible coverage. Digestion of the proteins from spots 2 and 3 with four proteases resulted in 157 different proteolytic peptides with MaSp2-related sequences. Although the number of peptides was large, they were identified by only two accession numbers: B5SYS7 (a large fragment containing the sequence of the N-terminal domain) and P46804 (a very large fragment containing the sequence of the central and C-terminal domains).

A careful observation of the alignment of the proteolytic peptides of *N. clavipes* MaSp2 with the sequence of *L. hesperus* protein (Fig. 2)

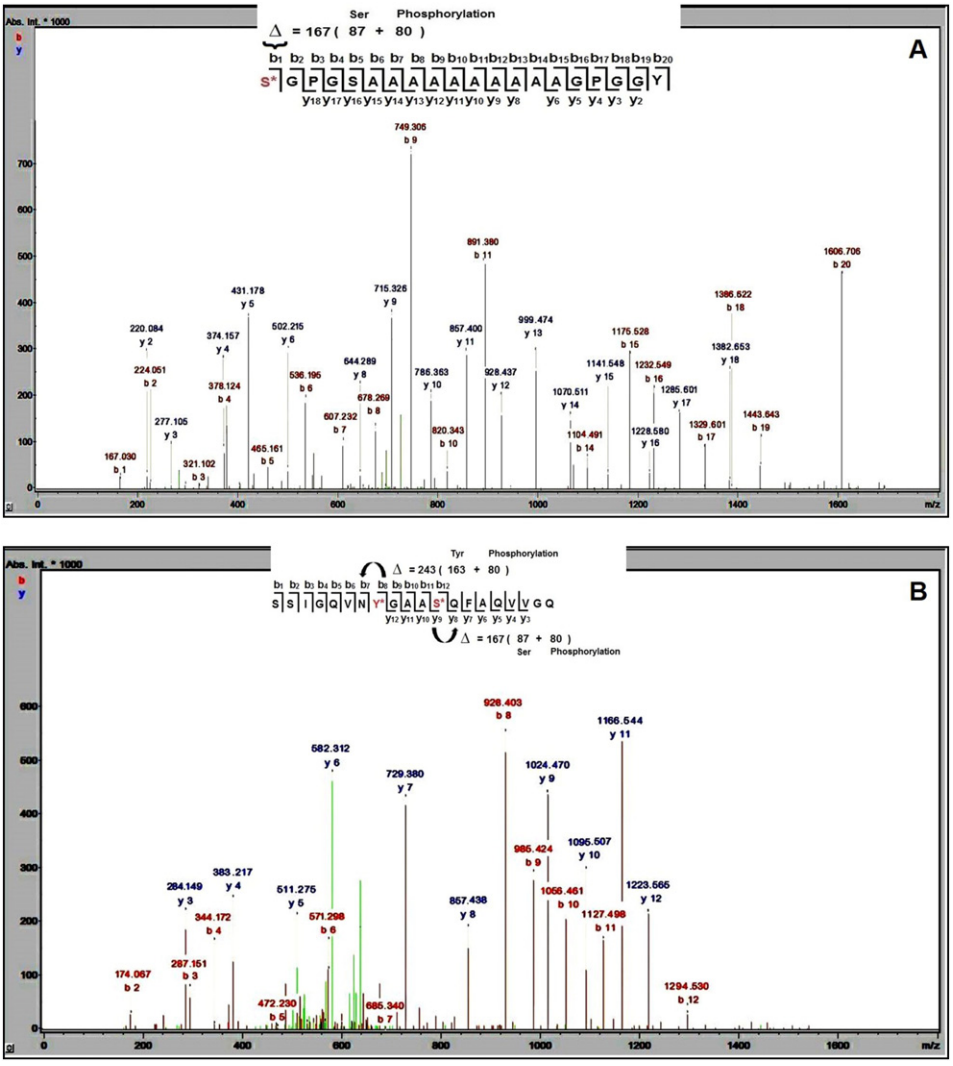


Fig 3. Representative mass spectra of *N. clavipes* MaSp2 (A) CID spectrum of the chymotryptic peptide S*GPGSAAAAAAAAAAGPGGY (200–219), which was obtained by selecting the *m/z* 813.424 $[M + 2H]^{2+}$ as a precursor ion and showing the phosphorylation site at *S200. (B) CID spectrum of the proteolytic peptide SSIGQVNY*GAAS*QFAQVVGQ (3121–3140), which was generated by the hydrolysis of spidroin-2 with proteinase 10 and selecting the *m/z* 1086.050 $[M + 2H]^{2+}$ as a precursor ion. The phosphorylation sites at *Y3128 and *S3132 are shown.

reveals that some of these peptides presented a single occurrence in MaSp2 sequence, such as the peptide "QALNMAF" which occurs only between the positions 35 and 41. Meanwhile, other peptides such as "GGYGPGQQGPGGY" occur 64 times at different positions of the MaSp2 (Tables 1 and S3). In order to verify if the concentrations of each one of the 116 non-redundant proteolytic peptides in the chromatograms were proportional to their occurrence in MaSp2 sequence, the peptides were semi-quantified using the spectral counting strategy. This is

Table 2
Mapping of the phosphorylation sites observed on *N. clavipes* MaSp2. All the positions of these modifications in the sequences of protein from spots 2 and 3 are shown in Fig. 4.

PTMs	Enzyme	Fragmentation method	Comments
S200, S673, S1146, S1619, S2092, S2565	Chymotrypsin	CID	Table S1 (*spot* 3)
Y239, Y712, Y1185, Y1658, Y2131, Y2604	Proteinase 10	CID	Table S1 (*spot 3*)
S294, S767, S1240, S1713, S2186, S2659	Proteinase 10	CID	Table S1 (*spot* 2)
S423, S896, S1369, S1842, S2315, S2788	Chymotrypsin	CID	Table S1 (*spot* 2)
S561, S1034, S1507, S1980, S2453, S2926	Proteinase 10	CID	Table S1 (*spot* 2)
T3074	Proteinase 10	CID	Table S1 (*spot* 2)
S3120	Proteinase 10	CID/ETD	Table S1 (*spot* 2)
S3121	Proteinase 10	CID/ETD	Table S1 (*spot* 2)
S3122	Proteinase 10	CID/ETD	Table S1 (*spot* 2)
Y3128	Proteinase 10	CID	Table S1 (*spots 2* and 3)
S3132	Proteinase 10	CID/ETD	Table S1 (*spots* 2 and 3)

Fig 4. Potential sequence of *N. clavipes* MaSp2 , showing the Non-Repetitive N-terminal Domain (NR-NTD), the six modules of the Central Repetitive Domain (CRD), and the Non-Repetitive C-Terminal Domain (NR-CTD). The sites of phosphorylation are assigned in yellow and marked with red "p" letters; the last two phosphorylation positions in the NR-CTD (Y3128 and S3132) were common to the proteins from spots 2 and 3.

important to ensure that the amount of each peptide in the chromatogram is reflecting its occurrence all over the sequence of the MaSp2; the completeness of the proteolytic digestions of *N. clavipes* MaSp2 makes its sequencing reliable. The results of these analyses are shown in Tables 1 and S3. The integer value of the normalized spectral counting of each peptide seems to be reflecting directly the number of occurrences of each specific peptide along the *N. clavipes* MaSp2. Thus, the results of MaSp2 sequencing are reliable, and apparently the proteolytic digestions used for generating the peptides (tools of sequencing) achieved the completeness.

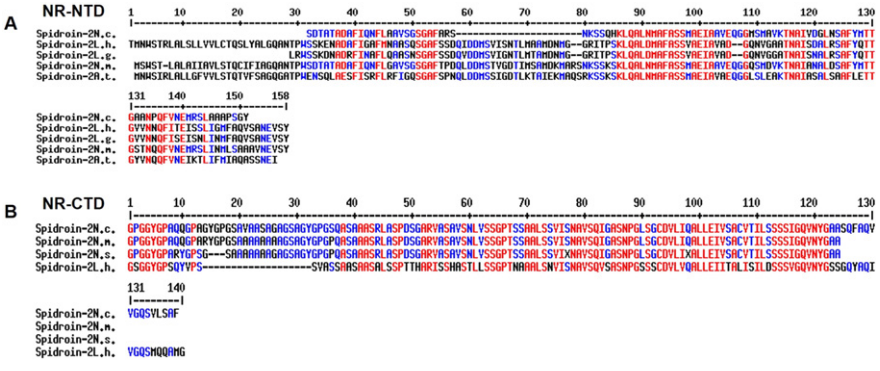


Fig. 5. Alignment of the sequences of known spidroin-2 homologues found in the silk of various species of spiders and comparison of the NR-NTD (non-repetitive N-terminal domain) and NR-CTD (non-repetitive C-terminal domain). (A) NR-NTD**:** N. c. — *Nephila clavipes*; L. h. — *Latrodectus hesperus*; L. g. — *Latrodectus geometricus*; N. m. — *Nephila madagascariensis*; and A. t. — *Argiope trifasciata*. (B) NR-CTD: N. c. — *Nephila clavipes*; N. m. — *Nephila madagascariensis*; N. s. — *Nephila senegalensis*; and L. h. — *Latrodectus hesperus*. Red — high sequence conservation; and Blue — low sequence conservation.

The MW calculated for the sequence of *N. clavipes* MaSp2 as shown Fig. 2, is 269 kDa; considering that the apparent MW values observed for protein spots in 2-DE gel profile are >250 kDa (Fig. 1E), and that the error in MW determination in the electrophoresis protocol is about 5%, both results are consistent to each other, even taking into account that may be missing some parts of the sequence of *N. clavipes* MaSp2.

High-sequence coverage of MaSp2 protein was revealed and a series of PTMs in their sequences were observed, such as phosphorylation, deamidation, hydroxylation, methylation and oxidation. Up to now all the mechanical properties of the spidroin have been characterized without any consideration about the existence of modifications in the sequence of spidroins. Modifications such as oxidations, deamidations and methylations might be artefacts arising from sample preparation or analytical procedures, although deamidations reportedly occur due to the presence of deamidases and/or deiminases in biological extracts [50].

Phosphorylation was the major PTM assigned onto the MaSp2 sequence. The introduction of phosphate groups into a protein backbone may cause significant conformational changes that affect the activity of the protein and/or its interactions with other proteins [43]. Importantly, proteomic studies have revealed the presence of such PTMs on some known silk proteins (i.e., the heavy chain of fibroin, the light chain of fibroins, and the P25 protein) including the fibroin from *Bombyx mori* [44,51,52] and the major ampullate silk spidroin-1 protein, from the spiders *N. clavipes*, *N. madagascariensis* and *N. edulis* [32]. In a previous study, eight phosphorylation sites were observed on *N. clavipes* MaSp1, four on the MaSp produced by *N. madagascariensis* and two on the MaSp1 produced by *N. edulis* [32]. It was postulated that phosphate residues might play important roles in the storage of silk proteins at high concentrations in the silk gland, as well as in the self-assembly of the proteins during the fibre-spinning process [48,53]. It is particularly noteworthy that in vivo, the transition from the soluble protein to solid fibres involves a combination of chemical and mechanical stimuli (such as ion exchange, water extraction and shear forces) [13,54,55]. Winkler et al. [56] reported that it is possible to control the assembly of β-sheet, as well as to enhance protein solubility through the enzymatic phosphorylation/dephosphorylation of spidroin-inspired recombinant short proteins (ca. 25 kDa). In the present investigation, careful observation of the location of the phosphorylation sites on MaSp2 reveals that all these PTMs are mostly located within the GPGXX and $(A)_n$ structural motifs, that are responsible for the silk's mechanical properties [13,28,57].

A careful observation of Table 2 reveals that MaSp2 from the spots 2 and 3 present different number of phosphorylation sites, when compared to each other, suggesting the occurrence of two forms of this protein in *N. clavipes* web. Apparently, the MaSp2 from the spot 2 present a higher number of phosphorylation sites than the protein from spot 3, making the former to present higher MW and pI values than the protein from spot 3 as shown in Fig. 1E. The role of each form of MaSp2 is still unknown; however, considering that each type of silk fibre is produced for specific purposes by the spider, the MaSp1 and MaSp2 are used to build the frame, the radii and the dragline [8]. Thus, it may be speculated that the two different forms of MaSp2 may be differentially used to crosslink with MaSp1 for tailoring the web frame and radii (in our preparation there was no dragline).

In fact no phosphorylation was observed in the N-terminal domain, while a total of thirty phosphorylation sites were observed in each repetitive module of the central domain, and six phosphorylation sites were observed in the C-terminal domain (Fig. 4); the MaSp2 sequence from spot 2 presented 24 phosphorylation sites, meanwhile the protein from spot 3 presents 14 of phosphorylation sites. It is important to emphasize that two of these sites at the NR-CTD were common between both forms of MaSp2 (Y3128 and S3132).

The phosphorylation sites were initially assigned by using CID fragmentations in mass spectrometric analysis; no phosphorylation was detected with the use of ETD in the present study. Thus, in order to make these assignments reliable the phosphorylation of *N. clavipes* MaSp2 was confirmed by immunoblotting (at least in relation to the residues of serine), in the absence and presence of treatment with alkaline phosphatase (Fig. S9). The biological significance of the multiple phosphorylation sites observed on the spidroins produced by orb-weaving spiders is still unknown. However, data obtained from the literature suggest that these PTMs might be related to the conformation of the GPGXX and $(A)_n$ domains that are responsible for the silk's mechanoelastic properties.

Comparison of the sequence alignments of the N- and C-terminal non-repetitive domains of *N. clavipes* MaSp2 with the same domains from four other species of orb-weaving spiders (Fig. 5 A–B), revealed that these domains are highly conserved, indicating that they might be involved in important function(s) related to the role of MaSp2 in the web. It has been suggested that the initiation of assembly and fibre elongation depend on this central core, which might require the occurrence of PTMs to exert its biological actions [29,45]. Most phosphorylation sites are located in the central core and apparently determine the conformation of this domain, which may dictates the silk's mechanical properties.

5. Conclusions

Spider silk has great potential as a biomaterial; to exploit this potential, a better understanding of the structure-function relationship of the spidroins is required, Therefore, chemical information, such as knowledge of the PTMs normally present on the proteins, may provide a basis for understanding the mechanoelastic properties of silk for use biotechnological and biomedical applications. In the present study a gel-based mass spectrometry strategy using CID and ETD fragmentation methods in mass spectrometric analysis were used to sequence *N. clavipes* MaSp2. The N- and C-terminal non-repetitive domains were sequenced, as well as great part of the central core. In addition to this, two different forms of MaSp2 were identified, being one of them with twenty four phosphorylation sites along its sequence, while another form presented 14 phosphorylation sites (two sites were common between both forms). The sequencing data reported herein might be relevant for developing new approaches for the synthetic or recombinant production of novel spider web-based polymers.

Conflict of interest

The authors declare no conflict of interests.

Transparency document

The Transparency document is associated with this article can be found, in online version.

Acknowledgments

This work was supported by grants from FAPESP (Proc. 2010/19051-6, Proc. 2011/51684-1 and Proc. 2013/26451-9), CNPq and the Gert Lubec Proteomics Laboratory at the University of Vienna. M.S.P. is a researcher from the National Research Council of Brazil-CNPq; G.L. is a researcher from the Gert Lubec Proteomics Laboratory at the University of Vienna, Vienna, Áustria and J.R.A.S.P. is a Post-Doc Research fellow from FAPESP at São Paulo State University — UNESP, Rio Claro, Brazil.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbapap.2016.05.007.

References

[1] M.E. Rousseau, T. Lefèvre, M. Pézolet, Conformation and orientation of proteins in various types of silk fibers produced by *Nephila clavipes* spiders, Biomacromolecules 10 (2009) 2945–2953.
[2] F. Vollrath, Strength and structure of spiders' silks, J. Biotechnol. 74 (2000) 67–83.
[3] M. Heim, L. Römer, T. Scheibel, Hierarchical structures made of proteins. The complex architecture of spider webs and their constituent silk proteins, Chem Soc Rev. 39 (2010) 156–164.
[4] J. Kovoor, Comparative structure and histochemistry of silk producing organs in arachnids, in: W. Nentwig (Ed.), Ecophysiology of Spiders, Springer, Berlin 1987, pp. 159–186.
[5] J.G. Hardy, T.R. Scheibel, Composite materials based on silk proteins, Prog. Polym. Sci. 35 (2010) 1093–1115.
[6] M. Xu, R.V. Lewis, Structure of a protein superfiber: spider dragline silk, Proc. Natl. Acad. Sci. U. S. A. 87 (1990) 7120–7124.
[7] A. Sponner, B. Schlott, F. Vollrath, E. Unger, Characterization of the protein components of *Nephila clavipes* dragline silk, Biochemistry 44 (2005) 4727–4736.
[8] L. Römer, T. Scheibel, The elaborate structure of spider silk: structure and function of a natural high performance fiber, Prion 2 (2008) 154–161.
[9] A. Rising, J. Johansson, G. Larson, E. Bongcam-Rudloff, W. Engström, G. Hjälm, Major ampullate spidroins from *Euprosthenops australis*: multiplicity at protein, mRNA and gene levels, Insect Mol. Biol. 16 (2007) 551–561.
[10] A. Sponner, W. Vater, S. Monajembashi, E. Unger, F. Grosse, K. Weisshart, Composition and hierarchical organization of a spider silk, PLoS One 2 (10) (2007), e998.
[11] W.A. Gaines, W.R. Jr, Marcotte, identification and characterization of multiple spidroin 1 genes encoding major ampullate silk proteins in *Nephila clavipes*, Insect Mol. Biol. 17 (2008) 465–474.
[12] N.A. Ayoub, J.E. Garb, R.M. Tinghitella, M.A. Collin, C.Y. Hayashi, Blueprint for a high-performance biomaterial: full-length spider dragline silk genes, PLoS One 2 (2007) e514-e514.
[13] N.A. Ayoub, C.Y. Hayashi, Multiple recombining loci encode MaSp1, the primary constituent of dragline silk, in widow spiders (*Latrodectus*: Theridiidae), Mol. Biol. Evol. 25 (2) (2008) 277–289.
[14] M.B. Hinman, R.V. Lewis, Isolation of a clone encoding a second dragline silk fibroin *Nephila clavipes* dragline silk is a two-protein fiber, J. Biol. Chem. 267 (1992) 19320–19324.
[15] A. Chinali, W. Vater, B. Rudakoff, A. Sponner, E. Unger, F. Grosse, K.H. Guehrs, K. Weisshart, Containment of extended length polymorphisms in silk proteins, J. Mol. Evol. 70 (2010) 325–338.
[16] D. Motriuk-Smith, A. Smith, C.Y. Hayashi, R.V. Lewis, Analysis of the conserved N-terminal domains in major ampullate spider silk proteins, Biomacromolecules 6 (2005) 3152–3159.
[17] J. Gatesy, C. Hayashi, D. Motriuk, J. Woods, R. Lewis, Extreme diversity, conservation, and convergence of spider silk fibroin sequences, Science 291 (2001) 2603–2605.
[18] S. Rauscher, S. Baud, M. Miao, F.W. Keeley, R. Pomes, Proline and glycine control protein self-organization into elastomeric or amyloid fibrils, Structure 14 (2006) 1667–1676.
[19] L. Eisoldt, J.G. Hardy, M. Heim, T. Scheibel, The role of salt and shear on the storage and assembly of spider silk proteins, J. Struct. Biol. 170 (2010) 413–419.
[20] Y. Liu, Z. Shao, F. Vollrath, Relationships between supercontraction and mechanical properties of spider silk, Nat. Mater. 4 (2005) 901–905.
[21] T. Lefévre, F. Paquet-Mercier, J.F. Rioux-Dubé, M. Pézolet, Structure of silk by Raman spectromicroscopy: from the spinning glands to the fibers, Biopolymers 97 (2011) 322–336.
[22] D. Motriuk-Smith, A. Smith, C.Y. Hayashi, R.V. Lewis, Analysis of conserved N-terminal domains in major ampullate spider silk proteins, Biomacromolecules 6 (2005) 3152–3259.
[23] G. Askarieh, M. Hedhammar, K. Nordling, A. Saenz, C. Casals, A. Rising, J. Johansson, S.D. Knight, Self-assembly of spider silk proteins is controlled by a pH-sensitive relay, Nature 465 (2010) 236–238.
[24] M. Humenik, A.M. Smith, T.R. Scheibel, Recombinant spider silks — biopolymers with potential for future applications, Polymers 3 (2011) 640–661.
[25] Y. Liu, A. Sponner, D. Porter, F. Vollrath, Proline and processing of spider silks, Biomacromolecules 9 (1) (2008) 116–121.
[26] Y. Liu, Z. Shao, F. Vollrath, Elasticity of spider silks, Biomacromolecules 9 (7) (2008) 1782–1786.
[27] J.G. Hardy, L.M. Römer, T.R. Scheibel, Polymeric materials based on silk proteins, Polymers 49 (2008) 4309–4327.
[28] A. Rising, M. Widhe, J. Johansson, M. Hedhammar, Spider silk proteins: recent advances in recombinant production, structure–function relationships and biomedical applications, Cell. Mol. Life Sci. 68 (2011) 169–184.
[29] L. Eisoldt, A.M. Smith, T. Scheibel, Decoding the secrets of spider silk, Mater. Today 14 (2011) 80–86.
[30] S.W. Cranford, A. Tarakanova, N.M. Pugno, M.J. Buehler, Nonlinear material behavior of spider silk yields robust webs, Nature 482 (2012) 72–76.
[31] U. Slotta, N. Mougin, L. Römer, A.H. Leimer, Synthetic spider silk proteins and threads, Soc. Biol. Engin. (2012) 43–49.
[32] J.R.A. dos Santos-Pinto, G. Lamprecht, W.Q. Chen, S. Heo, J.G. Hardy, H. Priewalder, T.R. Scheibel, M.S. Palma, G. Lubec, Structure and post-translational modifications of the web silk protein spidroin-1 from *Nephila* spiders, J. Proteome 105 (2014) 174–185.
[33] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye biding, Anal. Biochem. 72 (1976) 248–254.
[34] D.M. Creasy, J.S. Cottrell, Error tolerant searching of uninterpreted tandem mass spectrometry data, Proteomics 2 (2002) 1426–1434.
[35] T. Nuwal, S. Heo, G. Lubec, E. Bunchner, Mass spectrometric analysis of synapsins in *Drosophila melanogaster* and identification of novel phosphorylation sites, J. Proteome Res. 10 (2011) 541–550.
[36] S.U. Kang, K. Fuchs, W. Sieghart, A. Pollak, E. Csaszar, G. Lubec, Gel-based mass spectrometric analysis of a strongly hydrophobic $GABA_A$-receptor subunit containing four transmembrane domains, Nat. Protoc. 4 (2009) 1093–1102.
[37] L. Arike, L. Peil, Spectral counting label-free proteomics, Methods Mol. Biol. 1156 (2014) 213–222.
[38] M.E. Feltcher, H.P. Gunawardena, K.E. Zulauf, S. Malik, J.E. Griffin, C.M. Sassetti, X. Chen, M. Braunstein, Label-free quantitative proteomics reveals a role for the *Mycobacterium tuberculosis* SecA2 pathway in exporting solute binding proteins and Mce transporters to the cell wall, Mol. Cell. Proteomics 14 (6) (2015) 1501–1516.
[39] C. Ramus, A. Hovasse, M. Marcellin, A.M. Hesse, E. Mouton-Barbosa, D. Bouyssié, S. Vaca, C. Carapito, K. Chaoui, C. Bruley, J. Garin, S. Cianférani, M. Ferro, A. Van Dorssaeler, O. Burlet-Schiltz, C. Schaeffer, Y. Couté, A. Gonzalez de Peredo, Benchmarking quantitative label-free LC–MS data processing workflows using a complex spiked proteomic standard dataset, J. Proteome 132 (2016) 51–62.
[40] S. Zhang, X. Cao, Y. He, S. Hartson, H. Jiang, Semi-quantitative analysis of changes in the plasma peptidome of Manduca sexta larvae and their correlation with the transcriptome variations upon immune challenge, Insect Biochem. Mol. Biol. 47 (2014) 46–54.
[41] J.R. Dos Santos-Pinto, A.M. Garcia, H.A. Arcuri, F.G. Esteves, H.C. Salles, G. Lubec, M.S. Palma, Silkomics: insight into the silk spinning process of spiders, J. Proteome Res. 15 (4) (2016) 1179–1193.
[42] M. Mann, O.N. Jensen, Proteomic analysis of post-translational modifications, Nat. Biotechnol. 21 (2003) 255–261.
[43] C.E. Eyers, S.J. Gaskell, Mass Spectrometry to Identify Post-Translational Modifications, Wiley Encyclopedia Chem. Biol., 2008, http://dx.doi.org/10.1002/9780470048672.wecb469.
[44] W.Q. Chen, H. Priewalder, J.P.P. John, G. Lubec, Silk cocoon of *Bombyx mori*: Proteins and posttranslational modifications — heavy phosphorylation and evidence for lysine-mediated cross links, Proteomics 10 (2010) 369–379.
[45] L. Eisoldt, C. Thamm, T. Scheibel, The role of terminal domains during storage and assembly of spider silk proteins, Biopolymers 97 (2011) 355–361.
[46] M. Heim, D. Keerl, T. Scheibel, Spider silk: from soluble protein to extraordinary fiber, Angew. Chem. Int. Ed. 48 (2009) 3584–3596.
[47] A. Rising, G. Hjalm, W. Engstrom, J. Johansson, N-terminal non-repetitive domain common to dragline, flagelliform, and cylindriform spider silk proteins, Biomacromolecules 7 (2006) 3120–3124.
[48] C.L. Mattina, R. Reza, X. Hu, A.M. Falick, K. Vasanthavada, S. McNary, R. Yee, C.A. Vierra, Spider minor ampullate silk proteins are constituents of prey wrapping silk in the cob weaver *Latrodectus hesperus*, Biochemistry 47 (2008) 4692–4700.
[49] J. Garb, N. Ayoub, C. Hayashi, Untangling spider silk evolution with spidroin terminal domains, BMC Evol. Biol. 10 (2010) 243.
[50] K. Kubota, T. Yoneyama-Takazawa, K. Ichikawa, Determination of sites citrullinated by peptidylarginine deiminase using ^{18}O stable isotope labeling and mass spectrometry, Rapid Commun. Mass Spectrom. 19 (2005) 683–688.
[51] P. Zhang, K. Yamamoto, Y. Aso, Y. Banno, D. Sakano, Y. Wang, H. Fujii, Proteomic studies of isoforms of the P25 component of *Bombyx mori* fibroin, Biosci. Biotechnol. Biochem. 69 (2005) 2086–2093.
[52] P. Zhang, Y. Aso, K. Yamamoto, Y. Banno, Y. Wang, K. Tsuchida, Y. Kawaguchi, H. Fujii, Proteome analysis of silk gland proteins from the silkworm, *Bombyx mori*, Proteomics 6 (2006) 2586–2599.
[53] C.A. Michal, A.H. Simmons, B.G. Chew, D.B. Zax, L.W. Jelinski, Presence of phosphorus in *Nephila clavipes* dragline silk, Biophys. J. 70 (1996) 489–493.
[54] D.P. Knight, F. Vollrath, Changes in element composition along the spinning duct in a Nephila spider, Naturwissenschaften 88 (2001) 179–182.
[55] D. Huemmerich, C.W. Helsen, S. Quedzuweit, J. Oschmann, R. Rudolph, T. Scheibel, Primary structure elements of spider dragline silks and their contribution to protein solubility, Biochemistry 43 (2004) 13,604–13,612.
[56] S. Winkler, D. Wilson, D.L. Kaplan, Controlling beta-sheet assembly in genetically engineered silk by enzymatic phosphorylation/dephosphorylation, Biochemistry 39 (2000) 12,739–12,746.
[57] T. Lefèvre, M.E. Rousseau, M. Pézolet, Protein secondary structure and orientation in silk as revealed by Raman spectromicroscopy, Biophys J. 92 (2007) 2885–2895.