the FEBS Journal

Multicomponent venom of the spider *Cupiennius salei*: a bioanalytical investigation applying different strategies

Christian Trachsel[1], Doreen Siegemund[1], Urs Kämpfer[1], Lukas S. Kopp[1], Claudia Bühr[1], Jonas Grossmann[2], Christoph Lüthi[3], Monica Cunningham[3], Wolfgang Nentwig[3], Lucia Kuhn-Nentwig[3], Stefan Schürch[1] and Johann Schaller[1]

1 Department of Chemistry and Biochemistry, University of Bern, Switzerland
2 Functional Genomics Center Zürich, University of Zürich/ETH Zürich, Switzerland
3 Institute of Ecology and Evolution, University of Bern, Switzerland

Keywords
Cupiennius salei; inhibitor cystine knot; mass spectrometry; short cationic peptides; spider venom analysis

Correspondence
J. Schaller, Department of Chemistry and Biochemistry, University of Bern, CH-3012 Bern, Switzerland
Fax: +41 31 631 34 21
Tel: +41 31 631 43 52
E-mail: johann.schaller@ibc.unibe.ch

(Received 26 March 2012, revised 14 May 2012, accepted 29 May 2012)

doi:10.1111/j.1742-4658.2012.08650.x

The multicomponent venom of the spider *Cupiennius salei* was separated by three different chromatographic strategies to facilitate subsequent analysis of peptidic venom components by tandem mass spectrometry (MALDI-TOF-MS and ESI-MS), Edman degradation and amino acid analysis: (a) desalting of the crude venom by RP-HPLC only, (b) chromatographic separation of the crude venom into 42 fractions by RP-HPLC, and (c) multidimensional purification of the crude venom by size exclusion and cation exchange chromatography and RP-HPLC. A total of 286 components were identified in the venom of *C. salei* by mass spectrometry and the sequence of 49 new peptides was determined *de novo* by Edman degradation and tandem mass spectrometry; 30 were C-terminally amidated. The novel peptides were assigned to two main groups: (a) short cationic peptides and (b) Cys-containing peptides with the inhibitor cystine knot motif. Bioinformatics revealed a limited number of substantial similarities, namely with the peptides CpTx1 from the spider *Cheiracantium punctorium* and U3-ctenitoxin-Asp1a from the South American fishing spider (*Ancylometes* sp.) and with sequences from a *Lycosa singoriensis* venom gland transcriptome analysis. The results clearly indicate that the quality of the data is strongly dependent on the chosen separation strategy. The combination of orthogonal analytical methods efficiently excludes alkali ion and matrix adducts, provides indispensable information for an unambiguous identification of isomasses, and results in the most comprehensive repertoire of peptides identified in the venom of *C. salei* so far.

Introduction

In the last few decades, improved bioanalytical methods have significantly enhanced the ability to analyse complex biological samples in detail. Concomitant with such studies, it was realized that venomous secretions of animals are an enormous source of bioactive compounds with a broad range of pharmacological properties [1–3]. A lot of work has been done in the last few years to investigate the pool of biologically active compounds of venomous animals, among them snakes, scorpions and cone snails [4–6].

As a result of soft ionization techniques, such as MALDI or ESI, with a concomitant improvement of resolution and sensitivity of mass analysers, MS has become the major analytical technique for screening

Abbreviations
AAA, amino acid analysis; CHCA, α-cyano-hydroxycinnamic acid; CsTx, *Cupiennius salei* toxin; EST, expressed sequence tag; ICK, inhibitor cystine knot; PTM, post-translational modification; SA, sinapinic acid; SCPs, short cationic peptides; TFA, trifluoroacetic acid.

large (natural) compound libraries [7–11]. Since the introduction of expressed sequence tag (EST) analysis [12], it has been applied and improved in many ways [13]. It is therefore not surprising that a combination of EST analysis and MS of transcribed peptides and proteins has become a fast and robust analytical protocol for the investigation of venoms [14,15].

A major drawback of this approach is the limited information about post-translational modifications (PTMs). The occurrence of PTMs like disulfide bridges, amidations, phosphorylations or glycosylations alters the mass of a peptide/protein, thereby making it difficult to assign MS data to polypeptides predicted from cDNA by algorithms and database information [13]. In particular, proteolytic cleavage of peptides and proteins into disulfide linked two-chain molecules [16] cannot be predicted to date with sufficient reliability. In contrast to these high throughput approaches stands the 'classical' proteomic/peptidomic strategy of isolating and analysing peptides.

In the present study, the venom of the spider *Cupiennius salei* was investigated. The venom is a complex multicomponent mixture, which can roughly be divided into three major groups: (a) salts, free amino acids and other low molecular mass components, (b) peptides in the mass range 1–10 kDa and (c) proteins in the mass range > 10 kDa [17,18]. The aim was to explore the peptidic venom composition of *C. salei* by bioanalytical methods.

The venom was separated by different multi-step purification protocols and analysed in detail by MS, tandem MS, Edman degradation and amino acid analysis (AAA). The combination of these orthogonally oriented analytical methods resulted in the most comprehensive repertoire of peptides identified in the venom of *C. salei* so far. The novel peptides were assigned to two main groups: (a) short cationic peptides (SCPs) and (b) Cys-containing peptides with the inhibitor cystine knot (ICK) motif. Identified peptides were compared with the UniProtKB database using the BLASTP algorithm (BLASTP 2.2.25) and with REGEX in combination with Unix EGREP. The results obtained give insight into the general composition of the *C. salei* venom in comparison with other spider species and increase the knowledge about the venom of *C. salei*.

Results

Peptide separation and mass spectrometric analysis

The complex multicomponent venom of the spider *C. salei* was fractionated by different chromatographic approaches to facilitate subsequent analysis. The purification of peptidic venom components was addressed by three strategies differing in their number and type of chromatographic steps (Fig. 1).

Strategy 1 is based on desalting of the venom by RP-HPLC only. The resulting peptide mixture was subjected to MALDI-TOF-MS analysis and the corresponding spectra are shown in Fig. 2. Peak assignment yielded 71 components using α-cyano-hydroxycinnamic acid (CHCA) as matrix and 74 components using sinapinic acid (SA) as matrix. Pronounced differences between the two applied matrices were observed regarding detected components, intensity of signals and distribution of mass numbers. The combined data sets finally consisted of 112 detected venom components (Table S1).

Strategy 2 is based on chromatographic separation of crude venom by RP-HPLC yielding 42 fractions mainly eluting between 20 and 80 min (Fig. 3). Fractions were analysed by MALDI-TOF-MS (using the matrices CHCA and SA) and by high resolution ESI-MS. Fractions containing a large number of components were rechromatographed and again subjected to MS analysis.

The substantial effect of α-cyclodextrin as a co-matrix in MALDI-TOF-MS analysis is demonstrated in Fig. 4. Venom fraction 30 (Fig. 3) was analysed without (Fig. 4A) and with (Fig. 4B) 10 mM α-cyclodextrin. The addition of co-matrix substantially reduced adduct-related peaks resulting in a spectrum of less complexity with increased ion intensity of actual compounds.

The total number of venom components detected in the 42 fractions by MALDI-TOF-MS was found to be 243 (Fig. 5 and Table S2). In all, 74 (22 + 52) of the 112 components detected in the desalted venom were confirmed and 169 components were observed additionally. High resolution ESI-MS analysis of the 42 fractions yielded 171 components (Table S3) and allowed the confirmation of 166 (52 + 114) of the 243 components found by MALDI-TOF-MS of the fractionated venom.

Strategy 3 is based on separation of the venom by an established, multidimensional purification protocol [19] with slight modifications (Fig. 1) involving size exclusion and cation exchange chromatography and RP-HPLC. With separation strategy 3, no additional venom components were identified upon MS analysis in comparison with separation strategy 2.

The results of MS analysis are summarized in Fig. 5. The number of components found by each method (separation strategy as well as ionization mode) is given in the corresponding circle, representing the total of 286 venom components. The majority of peptides

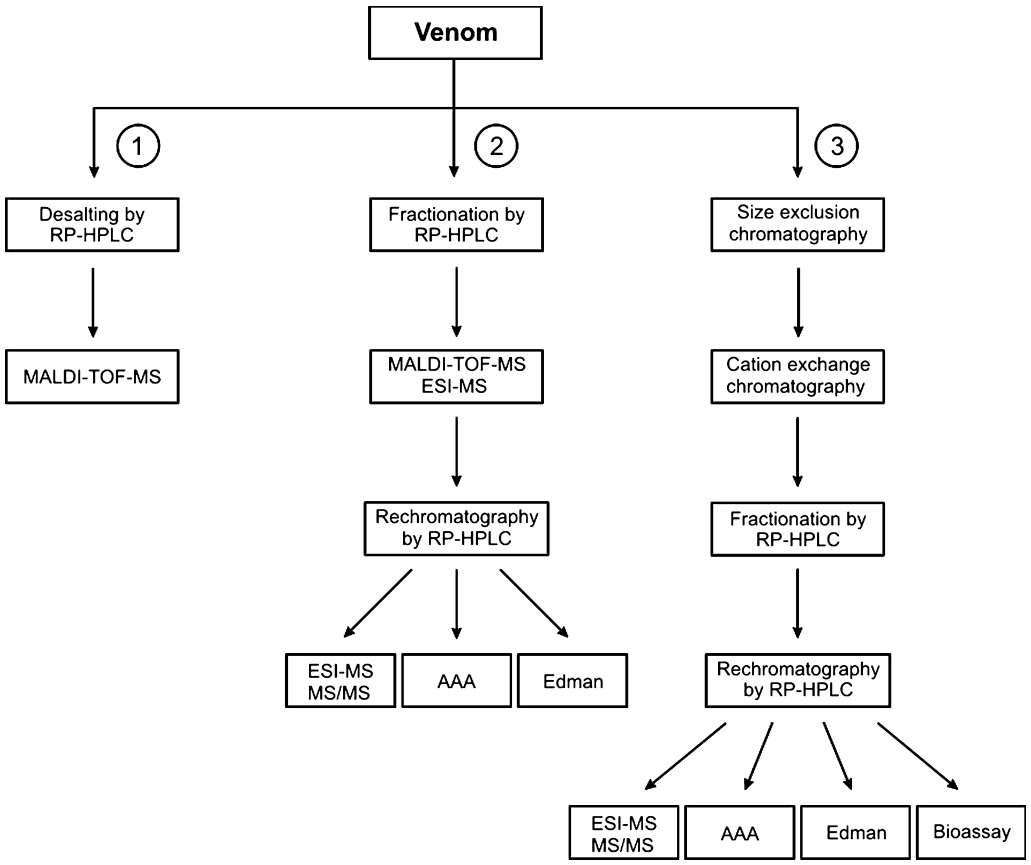

Fig. 1. Schematic representation of the applied strategies for the chromatographic separation and characterization of the venom of *Cupiennius salei*. Strategy 1: Desalting by RP-HPLC and characterization by MALDI-TOF-MS. Strategy 2: Fractionation by RP-HPLC, screening by MALDI-TOF-MS and ESI-MS, rechromatography and characterization by high resolution ESI-MS and tandem MS, Edman degradation and AAA. Strategy 3: Chromatographic separation by size exclusion, cation exchange and RP-HPLC and characterization by high resolution ESI-MS and tandem MS, Edman degradation, AAA and bioassays.

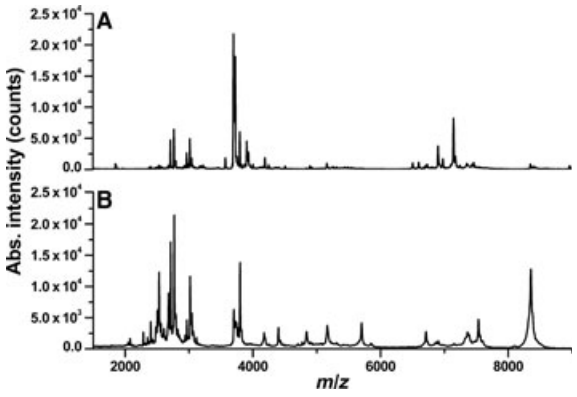

Fig. 2. MALDI-TOF mass spectra of the venom of *Cupiennius salei*. Desalted, unfractionated venom of *C. salei* was analysed by MALDI-TOF-MS using (A) SA and (B) CHCA as matrices, and 10 mM α-cyclodextrin as co-matrix.

(52%) were found in the mass range from 2000 to 4000 Da, and a second less pronounced group in the mass range from 6000 to 9000 Da (28%). Only 52 components (18%) were detectable by both MALDI-TOF-MS and ESI-MS in the desalted and in the fractionated venom. Compared with the 112 components detected by MALDI-TOF-MS of the desalted venom, MALDI-TOF-MS of fractionated venom yielded 166 (52 + 114) additional components. Similar results were obtained by ESI-MS of the fractionated venom, which gave 119 (5 + 114) additional components (Fig. 5).

One particular challenge for the generation of a robust venom peptidome based on MS is the presence of isomasses or near-isomasses [20], as exemplified by the detailed analysis of fractions 34 and 36 (Fig. 3). ESI-MS analyses of these fractions resulted in two

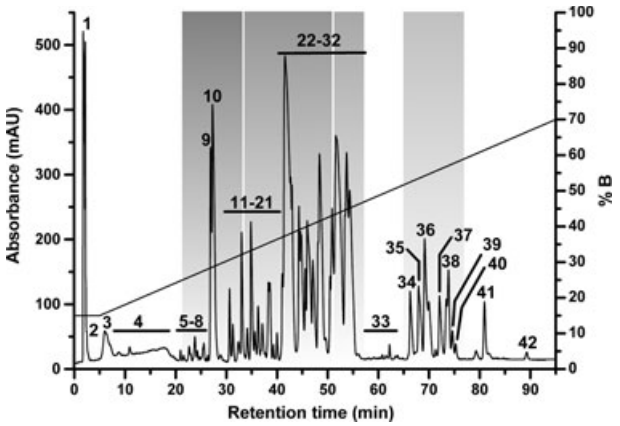

Fig. 3. RP-HPLC chromatogram of 1 μL of crude venom of *Cupiennius salei*. Separation: solvent A [0.1% (v/v) TFA in water] and 15% solvent B [0.1% (v/v) TFA, 80% (v/v) acetonitrile in water] was delivered at a flow rate of 0.2 mL·min^{-1} for 5 min, followed by a 90 min linear gradient to 70% solvent B on a Reprosil-PUR 300 phenyl column (5 μm, 2.0 × 100 mm). The numbers correspond to the collected fractions. The gradient is indicated by a solid line. The range of elution of the various peptide types (enhancer peptides, ICK toxins, cupiennin 1 and cupiennin 2) is highlighted in shades of grey.

masses differing by 22 mDa only (fraction 34, 3699.202 Da; fraction 36, 3699.224 Da; Fig. 6). AAA gave evidence for the presence of two different components with varying Gly, Ala and Val contents (Table S4). These results were confirmed by Edman degradation. The sequences differ in three amino acids at positions 1 (Ala/Gly), 14 (Ala/Gly) and 16 (Ala/Val) (Fig. 6). In addition, sequence analysis of the components in fraction 36 showed two different residues in equimolar ratio at position 32, namely Glu and Gln, and gave evidence for an altered C-terminus (Gln/Lys). Further purification yielded fractions 36A and 36B and AAA confirmed the expected composition (Table S4). MS/MS experiments revealed an amidation of the C-termini in both peptides and confirmed the expected sequences. The monoisotopic masses of the two peptides in fraction 36 are 3698.211 Da and 3699.224 Da, differing by 1.013 Da, and the monoisotopic mass of the peptide in fraction 34 is 3699.202 Da (Fig. 6). Consequently, the data give clear evidence for the presence of three different peptides and incomplete separation with double detection of a single peptide is excluded. All three peptides (cupiennin 2a, cupiennin 2d and cupiennin 2e) are members of the cupiennin 2 family [17].

Peptide characterization

Characterization of purified peptides by Edman degradation, AAA and high resolution tandem MS resulted in 59 complete sequences (Figs 7 and 8), of which 37 carry a C-terminal amidation; 49 peptides represent so far unassigned novel sequences determined by a combination of Edman degradation and MS/MS experiments. Their biochemical properties are compiled in Tables S5 and S6. The 59 obtained peptide sequences were assigned to two main groups. The first group consists of short, basic Cys-free peptides in the mass range 2–4 kDa termed SCPs which display a high sequence variety. For two SCP subgroups (cupiennins 1 and 2) a broad cytolytic activity has been experimentally proved [18,21,22], explaining their insecticidal properties.

All Cys-containing peptides were allocated to the second group. Among the characterized Cys-containing peptides, 15 peptides have their Cys residues arranged in the ICK motif, while *C. salei* toxin 16 (CsTx-16) (single interchain disulfide bridge) and CsTx-20 exhibit so far unassigned disulfide bridge patterns (Fig. 7). Peptides with the ICK motif can be subdivided into two structurally different subgroups exhibiting diverse biological activities: first, the single-chain peptides like CsTx-1, which act neurotoxically [23], and second, the

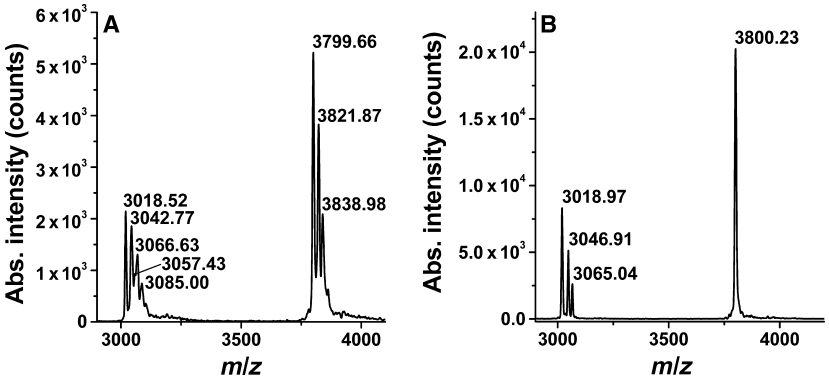

Fig. 4. MALDI-TOF mass spectrum of venom fraction 30 (Fig. 3) of *Cupiennius salei*. Fraction 30 was analysed with (A) CHCA only and (B) CHCA in the presence of 10 mM α-cyclodextrin as co-matrix.

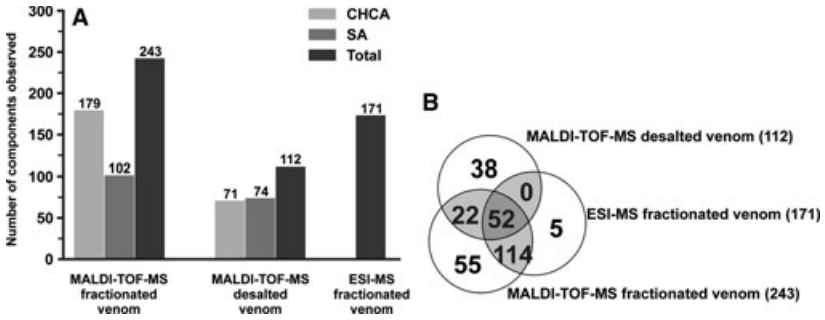

Fig. 5. Compilation of the MS data of the venom of *Cupiennius salei*. (A) Histogram representation of MALDI-TOF-MS and ESI-MS analysis of fractionated and desalted venom. For MALDI-TOF-MS, each sample was analysed with the matrices CHCA (light grey) and SA (dark grey) in the presence of the co-matrix α-cyclodextrin. The combined MALDI-TOF-MS and the ESI-MS data set is given as black bars. (B) Combination of all MS data of desalted and fractionated venom. The number of components found by each method is given in the corresponding circles. A total of 286 venom components were identified.

```
                    ▼           ▼ ▼             ▼ ▼
Fr 34 : Cu 2e  AFGTILKALAKIAAKAVKKLATKPGATYMLKQNLE-NH2  3699.202 Da
Fr 36A: Cu 2d  GFGTILKALAKIAGKVVKKLATKPGATYMLKQNLQ-NH2  3698.211 Da
Fr 36B: Cu 2a  GFGTILKALAKIAGKVVKKLATKPGATYMLKENLK-NH2  3699.224 Da
```

Fig. 6. Sequence alignment of peptides in fractions 34 and 36. Alignment of the sequences of peptides found in fractions 34, 36A and 36B with their monoisotopic masses (sequence variations indicated with triangles). The peptides are members of the cupiennin 2 family.

Fig. 7. Classification of CsTx toxin-like structures (Cys-containing peptides). The peptides are divided into two subgroups: single-chain peptides and two-chain peptides, both containing the ICK motif. CsTx-16 (single interchain disulfide bridge) and CsTx-20 exhibit so far unassigned disulfide bridge patterns. Cys residues are in bold type and shaded in grey. An asterisk indicates C-terminal amidation.

two-chain peptides like CsTx-13, which enhance the activity of neurotoxins [16,24].

Bioinformatics

Identified peptide sequences were compared against the UniProtKB/Swiss-Prot database using the BLASTP algorithm (BLASTP 2.2.25). Results indicate a high similarity of CsTx peptides to the Cys-containing, two domain peptides CpTx1 (D5GSJ8) from the spider *Cheiracantium punctorium* [25] and U3-ctenitoxin-Asp1a (P84001), a peptide fragment from the South American fishing spider (*Ancylometes* sp.) with a hitherto unknown disulfide bridge pattern. Additionally, 113 sequences were found to be derived from a *Lycosa singoriensis* venom gland transcriptome analysis [26]. Comparison of CsTx-1, CsTx-9 and CsTx-10 with toxin-like structures from the recently published EST-

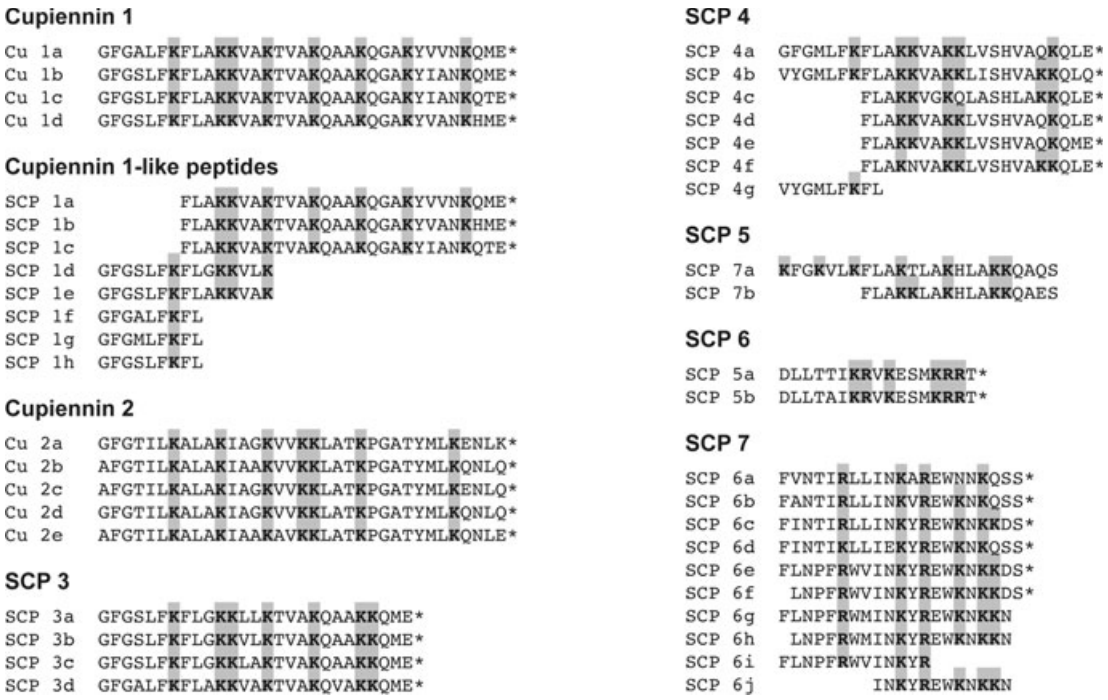

Fig. 8. Classification of short cationic peptides (SCPs). SCPs are classified as cupiennin 1, cupiennin 1-like, cupiennin 2 and SCPs 3–7. Type-defining Lys and Arg residues are in bold type and shaded in grey. An asterisk indicates C-terminal amidation.

tag library shows identities of 47.3% between CsTx-1 and B6DCP3, 40.6% between CsTx-9 and B6DCV0, and 57.1% between CsTx-10 and B6DCV0. Strikingly, the two-chain peptide CsTx-13 exhibits an identity of 57.7% with B6DCU1.

For the SCPs standard BLASTP searches did not result in any conclusive findings. As an alternative to BLASTP search, REGEX in combination with Unix EGREP was applied. In this approach, any user-predefined pattern can be searched against a FASTA database. For the Cys-containing peptides five different patterns were used for identification. No additional findings to BLASTP were observed for the conservative patterns 1–3. However, the broad patterns 4 and 5 were able to identify 63 and 115 not closely related additional structures, respectively. Among these additional, only distantly related peptides were 31 from the spider *Phoneutria* ssp. and four from *Ctenus ornatus* (both Ctenidae).

Whilst standard BLASTP searches did not result in any conclusive results for SCPs, searching with patterns 6 and 7 resulted in 73 and 101 sequences respectively from different species, among them 43 peptides from spiders. Of these 43 peptides, 20 derive from spiders which are known to possess venom components which act in a cytolytic fashion, namely two peptides of *Hogna carolinensis* [27,28], one of *Lycosa singoriensis* [29], 16 of *Lachesana tarabaevi* [30] and one of *Oxyopes takobius* [31,32].

Bioinformatics did not reveal significant matches of *C. salei* venom peptides with peptides from the phylogenetically closely related spider *Phoneutria* spp. (both Ctenidae). However, our data indicate some similarities among components of *C. salei* (Ctenidae) and *Lycosa singoriensis* (Lycosidae).

Discussion

The three applied separation strategies have advantages and disadvantages regarding the aim of the intended investigation.

Strategy 1 involving direct MALDI-TOF-MS analysis of desalted venom by RP-HPLC only is a fast and straightforward approach for the generation of a snapshot of the actual venom composition. Such an outline of the venom is useful in combination with sequence information for comparative studies of venomous animals of different species, gender or age [33–36]. However, since the peptides are not separated, ESI-MS analysis or alternative analytical techniques such as Edman degradation or AAA are not applicable.

In comparison with strategy 1, strategy 2 involving RP-HPLC separation of the crude venom into 42

fractions reduced the sample complexity, thereby enabling a detailed MS analysis. However, the achieved separation was insufficient for unambiguous *de novo* peptide sequencing by Edman degradation and/or MS/MS experiments or for investigations of biological activity.

Strategy 2 can be seen as a compromise between the fast and straightforward first strategy, which delivers a snapshot of the venom composition, and the much more time- and material-consuming third strategy (Fig. 1), which allows a thorough and detailed investigation of peptidic venom components.

Strategy 3 involving size exclusion and cation exchange chromatography and RP-HPLC provides a robust and reproducible separation of single venom components with high separation efficiency. Due to its time- and material-consuming nature, this strategy is unsuitable for an MS based high throughput screening of venom components. However, the high purity and the large amounts of isolated peptides allow an in-depth analysis of selected peptides by means of amino acid sequence, PTMs and biological function.

Some peptides isolated by strategy 3 showed an increased tendency to form oxidation products of Met residues with each additional chromatographic step [37]. In addition, one has to keep in mind the risk of material loss during multi-step separation protocols. In particular, low abundant or labile components might get lost during purification [9,38].

Given the complexity of a multicomponent mixture like spider venom, different analytical problems arise. Analytes differing by small mass increments might not be distinguishably resolved if analysed in a mixture, thus decreasing the number of observed components. Unambiguous assignment of such components requires additional analytical methods, such as AAA, sequence analysis by Edman degradation and/or MS/MS experiments. The detailed analysis of fractions 34 and 36 (Fig. 3) gives clear evidence for the necessity of such an approach, resulting in three clearly different peptide sequences, namely cupiennin 2a, cupiennin 2d and cupiennin 2e (Fig. 6).

Additionally, complex mixtures exhibit an increased tendency for peptide–peptide suppression effects. For example, rechromatography of fraction 36 (Fig. 3) resulted in two peptides (3698.211 and 3699.224 Da) differing by 1.013 Da. Both peptides were easily detected with high accuracy if separated, but only one peptide (3699.224 Da) was observed in the mixture.

Thus, fractionation was found to be essential for a comprehensive characterization of the venom of *C. salei* by MS; however, the distribution of the masses was similar for all approaches. Furthermore, the data indicate that thorough sample preparation and peptide separation is essential for a successful proteomic approach [34]. The quality of the obtained data is strongly dependent on the chosen separation strategy and separation efficiency.

An interesting observation was made regarding the elution behaviour of the different peptide subgroups. Whilst the ICK peptides, enhancer peptides, cupiennins 1 and cupiennins 2 elute in well-defined regions (Fig. 3), other SCPs were found throughout all fractions, independently of the chosen separation strategy. Only the application of size exclusion and cation exchange chromatography, followed by repeated RP-HPLC, enabled for instance the separation of the enhancer peptides CsTx-8 and CsTx-13 (Fig. 3, fractions 9 and 10). This reproducible elution behaviour could be indicative for hitherto undiscovered interactions of venom components and stands in good agreement with the previously proposed venom enhancement hypothesis [24].

Currently, no data of toxin-like structures for *Lycosa singoriensis* on the peptide level are available. Therefore, the potential processing of the single-chain cDNA template(s) into a two-chain peptide remains unclear. This finding clearly demonstrates the problem of the application of a single analytical approach. On one hand, the EST approach generates many more toxin-like sequences compared to *de novo* sequencing on the protein level. On the other hand, the EST approach lacks the information of post-translational processing, which is unambiguously determined in the classical proteomic approach.

In a recent review Kuhn-Nentwig *et al.* [39] have pointed out that the general knowledge on spider venom is rather limited (< 2000 identified and well characterized components). The potential of venom components in drug discovery seems to be tremendous once solid knowledge about these large peptide libraries is at hand (Editorial by Escoubas and King [40]).

The molecular diversity of spider venoms has been described in detail in several recent reviews [39,41,42]. Peptides in spider venom (1–10 kDa) can be divided into two main groups: (a) linear peptides without disulfide bridges, mainly cationic peptides, and (b) disulfide-containing peptides, often with an ICK motif (see Vassilevski *et al.* [41]). The present study clearly shows that the venom of *C. salei* contains peptides from both groups (Figs 7 and 8) which can act synergistically [17,24]. Escoubas *et al.* [14] have shown that the venoms of the Australian funnel-web spiders contain several hundred peptides with a bimodal mass distribution with the majority of the peptides in the range

3000–5000 Da (> 70%) and the second group between 6500 and 8500 Da. The mass distribution in the venom of *C. salei* is bimodal as well with the majority of peptides in the range 2000–4000 Da (52%) and the second group between 6000 and 9000 Da (28%). In the Australian funnel-web spiders this distribution is made up of smaller and larger disulfide-containing peptides; in *C. salei* the first group is composed of small cationic peptides and the second group mainly consists of ICK peptides.

A conservative calculation by Escoubas [42] assumes approximately 500 peptides per spider venom, which is significantly higher than the 286 components identified in the present study. In contrast, a review by Liang [43] summarized spider venom analysis applying different strategies, resulting in a range of 100 up to 1000 peptides per spider venom. Taking into account the variety of applied analytical strategies and the extreme complexity and variability of spider venom composition, it is difficult to judge if in all studies the venom components were recorded and assigned under the same stringent criteria.

In view of the difficulties in finding significant similarities with other spider venoms and the fact that the present study identified 286 components by MS but only 59 peptide sequences were determined, the data obtained do not currently provide a solid basis for a detailed comparison of the venom of *C. salei* with other spider venoms.

Conclusions

The peptidic venom components of the spider *C. salei* were separated by three different strategies and analysed in detail by MS and tandem MS, Edman degradation and AAA. MS data give evidence for the presence of 286 venom components and the amino acid sequence of 59 peptides was determined, of which 37 were C-terminally amidated. The combined data sets represent the most comprehensive *C. salei* venom repertoire identified so far.

The results clearly indicate that the quality of the data depends strongly on the chosen separation strategy. Furthermore, the combined application of orthogonal techniques, such as AAA, Edman degradation and tandem MS, is required to enable an unambiguous peptide assignment and to solve the problems of adduct formation, analyte suppression and isomasses.

Bioinformatic results demonstrate that conserved residues forming patterns are the key issue to a conclusive assignment of peptides to families and groups. Surprisingly, no substantial similarity among Cys-containing peptides of *C. salei* and *Phoneutria* ssp. was observed. However, bioinformatic data indicate a close proximity between *C. salei* and *Lycosa singoriensis*.

The results obtained in this study are a significant increase of knowledge regarding the venom composition of *C. salei* and provide the basis for further investigation of biological functions of venom peptides as well as for the characterization of the complete peptidome based on EST-tag library elucidation.

Material and methods

Materials

HPLC grade acetonitrile was obtained from Acros (Geel, Belgium) and neat trifluoroacetic acid (TFA) for HPLC solvents was from Supelco (Buchs, Switzerland). Protectan protection gas was obtained from Tetenal (Berlin, Germany). Enzymes for proteolytic degradation were of sequencing grade: chymotrypsin, endoproteinase Arg-C and endoproteinase Asp-N from Roche (Basel, Switzerland), endoproteinase Glu-C from Sigma (Buchs, Switzerland), TPCK-modified trypsin from Promega (Dübendorf, Switzerland) and DigesTip immobilized chymotrypsin from ProteoGen Bio (Pontedera, Italy). The reagents and solvents used for sequence analysis were purchased from Applied Biosystems (Bleiswijk, The Netherlands). All other chemicals were of analytical grade from Merck (Darmstadt, Germany), Sigma or Fluka (Buchs, Switzerland).

Venom collection

The venom was obtained from laboratory-bred adult spiders of *C. salei* and was collected as described previously [19]. Crude venom was separated from gland cell components and other impurities by centrifugation, split into aliquots, and frozen at −20 °C for immediate use or at −80 °C for long-term storage. MS analysis was performed with venom obtained from 44 female animals (siblings from a single cocoon). For peptide characterization, pooled venom from different spider generations was used.

Peptide purification

Venom was separated according to the scheme given in Fig. 1 either as described previously [17,19] with slight modifications or by RP-HPLC on a Hewlett-Packard liquid chromatograph 1090 (Agilent Technologies, Waldbronn, Germany). The obtained fractions were screened with MALDI-TOF-MS and, if required, subjected to further purification by RP-HPLC on an HP1090 equipped with a Reprosil-PUR 300 phenyl column, a Reprosil-PUR 120 C18 column (both 5 μm, 2.0 × 100 mm, Dr Maisch, Ammerbuch, Germany) or a Nucleosil 100-5 C18 Nautilus (5 μm, 4.6 × 250 mm, Macherey-Nagel, Oensingen,

Switzerland). Alternatively, a Solvent Delivery System 140B equipped with an absorbance detector 759A (Applied Biosystems, Foster City, CA, USA) was used with an Aquapore RP 300 C18 column (7 μm, 2.0 × 100 mm, Applied Biosystems). A binary solvent system [solvent A, 0.1% (v/v) TFA in water; solvent B, 0.1% (v/v) TFA, 80% (v/v) acetonitrile in water] was used with gradients adjusted to the corresponding separation. Fractions were collected manually and dried in a SpeedVac concentrator (Savant, Farmingdale, NY, USA). All dried samples were dissolved in 0.5% (v/v) formic acid, lyophilized three times and finally dried and stored at −20 °C until further use.

MALDI-TOF-MS

MALDI-TOF-MS analyses were carried out on a Voyager-DE Elite MALDI-TOF mass spectrometer (Applied Biosystems, Foster City, CA, USA) using the matrices CHCA dissolved in 50% (v/v) acetonitrile, 0.5% (v/v) formic acid in water (5–8 $mg{\cdot}mL^{-1}$), and SA dissolved in 30% (v/v) acetonitrile, 0.5% (v/v) formic acid in water (7–10 $mg{\cdot}mL^{-1}$). To suppress alkali- and matrix-dependent adducts, 10 mM α-cyclodextrin was added as a co-matrix to both matrices [44]. Sample solutions [in 0.5% (v/v) formic acid] were mixed with the corresponding matrix in a 1 : 4 ratio (v/v) and then spotted on a gold coated sample plate. Spectra were obtained in the positive ion mode using the linear and reflector mode (25 kV, 50 shots per spectrum). DATA EXPLORER™ (Applied Biosystems) was used to analyse the obtained MALDI-TOF-MS data.

ESI-MS

High resolution ESI-MS measurements were performed on an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) equipped with a nanoelectrospray ion source. Samples were dissolved in a mixture of acetonitrile, water and formic acid (50 : 50 : 0.1, v/v/v) and data were acquired in the positive full scan Fourier transform mode.

For MS/MS experiments (collision induced dissociation), the ions of highest intensity were isolated as precursor ions ($[M+nH]^{n+}$, n = 2–5) and activated with a normalized collision energy in the range 20–40 relative collision energy. Spectra were processed using the Thermo Electron MS data system XCALIBUR (Qual Browser, Thermo Fisher Scientific, San Jose, CA, USA).

Mass data evaluation

Masses were considered valid if they were detected by either high resolution ESI-MS (Orbitrap), by MALDI-TOF-MS with both matrices (CHCA and SA, with/without co-matrix α-cyclodextrin) or by both methods, ESI-MS and MALDI-TOF-MS.

High resolution ESI-MS enables an unambiguous detection of sodium (+21.98 Da) and potassium (+37.96 Da) adducts as well as oxidation products (+15.99 Da). The addition of α-cyclodextrin as co-matrix in MALDI-TOF-MS reduces the adduct formation by sodium and potassium and by the matrix SA (+207 Da). Comparison of the spectra with and without α-cyclodextrin facilitates the assignment of adduct formation. Masses identified as adducts were removed from the mass list.

Masses were considered 'identical' within a threshold of ±3 ppm for high resolution ESI-MS (Orbitrap) and within ±2 Da for low resolution MALDI-TOF-MS.

Amino acid analysis

Samples were hydrolysed in the gas phase with 6 M hydrochloric acid containing 0.1% phenol (v/v) for 22 h at 115 °C under N_2 vacuum. The liberated amino acids were reacted with phenylisothiocyanate and the resulting phenylthiocarbamoyl amino acids were analysed by RP-HPLC on a NovaPak C18 column (3.9 × 150 mm, 4 μm; Waters, Milford, MA, USA) on a Dionex HPLC system (Dionex, Olten, Switzerland) as described previously [45]. Disulfide-containing peptides were identified by the presence of diphenylthiocarbamoyl-cystine.

Reduction and alkylation

Cys-containing peptides were reduced and alkylated prior to analysis or proteolytic degradation. Peptides were dissolved in 100 mM Tris buffer, pH 7.8, containing 6 M guanidine hydrochloride. A 10-fold molar excess of dithiothreitol over the amount of disulfide bridges was added and the peptides were incubated for 1 h at 37 °C under protection gas. A 10-fold molar excess of iodoacetamide over the total amount of sulfhydryl groups was added and the reaction was continued for another hour at 37 °C in the dark. Peptides were desalted by RP-HPLC.

Edman degradation

N-terminal sequence analysis was carried out on a Procise 492 cLC protein sequencer (Applied Biosystems). The released amino acids were analysed online by RP-HPLC. In the case of disulfide-linked peptides, the diphenylthiohydantoin-cystine was detected in the corresponding cycle as a double peak in the vicinity of Tyr [46,47].

Proteolytic degradation

Longer peptides (> 25 amino acids) were subjected to proteolytic degradation with chymotrypsin, endoproteinase Arg-C, endoproteinase Asp-N, endoproteinase Glu-C or trypsin prior to Edman degradation. Digestions were carried

out according to the instructions of the manufacturer and the generated peptides were separated by RP-HPLC.

Bioinformatics

Peptide sequences were retrieved from UniProtKB in plain text format and the FASTA file was formatted with each amino acid sequence on a single line. Regular expression patterns (REGEX) were used in combination with Unix EGREP to identify sequences in the FASTA file which match the predefined amino acid patterns derived from *C. salei* peptides. Used patterns were as follows: pattern 1, Cx{6}Cx{6}CCx{8}CxCx{13}CxC; pattern 2, Cx{6}Cx{6}CCx{3,8}CxCx{13}CxC; pattern 3, Cx{6}Cx{6}CCx{8}CxCx{3,13}CxC; pattern 4, Cx{6}Cx{6}CCx{3,8}CxCx{3,13}CxC; pattern 5, Cx{3,6}Cx{3,6}CCx{3,8}CxCx{3,13}CxC; pattern 6, Kx{3}Kx{3}Kx{3}K; pattern 7, [K|R]x{3}[K|R]x{3}[K|R]x{3}[K|R]. Patterns 1–5 were used for searching Cys-containing peptides and patterns 6 and 7 for SCPs.

Acknowledgements

This work was supported by the Swiss National Science Foundation grant 206021-121304 to Stefan Schürch and by Swiss National Science Foundation grants 310030-127500 and 31003A-113681 to Wolfgang Nentwig.

Conflicts of interest

None declared.

References

1 Lewis RJ & Garcia ML (2003) Therapeutic potential of venom peptides. *Nat Rev Drug Discovery* **2**, 790–802.
2 Biass D, Dutertre S, Gerbault A, Menou JL, Offord R, Favreau P & Stöcklin R (2009) Comparative proteomic study of the venom of the piscivorous cone snail Conus consors. *J Proteomics* **72**, 210–218.
3 Vetter I, Davis JL, Rash LD, Anangi R, Mobli M, Alewood PF, Lewis RJ & King GF (2011) Venomics: a new paradigm for natural products-based drug discovery. *Amino Acids* **40**, 15–28.
4 Rodriguez de la Vega RC & Possani LD (2005) Overview of scorpion toxins specific for Na+ channels and related peptides: biodiversity, structure-function relationships and evolution. *Toxicon* **46**, 831–844.
5 Calvete JJ, Sanz L, Angulo Y, Lomonte B & Gutierrez JM (2009) Venoms, venomics, antivenomics. *FEBS Lett* **583**, 1736–1743.
6 Olivera BM & Teichert RW (2007) Diversity of the neurotoxic Conus peptides: a model for concerted pharmacological discovery. *Mol Interv* **7**, 251–260.
7 Favreau P, Menin L, Michalet S, Perret F, Cheneval O, Stöcklin M, Bulet P & Stöcklin R (2006) Mass spectrometry strategies for venom mapping and peptide sequencing from crude venoms: case applications with single arthropod specimen. *Toxicon* **47**, 676–687.
8 Escoubas P, Quinton L & Nicholson GM (2008) Venomics: unravelling the complexity of animal venoms with mass spectrometry. *J Mass Spectrom* **43**, 279–295.
9 Quinton L, Le Caer JP, Vinh J, Gilles N & Chamot-Rooke J (2006) Fourier transform mass spectrometry: a powerful tool for toxin analysis. *Toxicon* **47**, 715–726.
10 Legros C, Celerier ML, Henry M & Guette C (2004) Nanospray analysis of the venom of the tarantula *Theraphosa leblondi*: a powerful method for direct venom mass fingerprinting and toxin sequencing. *Rapid Commun Mass Spectrom* **18**, 1024–1032.
11 Shliapnikov IM, Kozlov SA, Fedorov AA & Grishin EV (2010) Comparison of polypeptide compositions from individual *Agelena orientalis* spider venoms. *Bioorg Khim* **36**, 81–88.
12 Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF *et al.* (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651–1656.
13 Nagaraj SH, Gasser RB & Ranganathan S (2007) A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief Bioinform* **8**, 6–21.
14 Escoubas P, Sollod B & King GF (2006) Venom landscapes: mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. *Toxicon* **47**, 650–663.
15 Diego-Garcia E, Peigneur S, Waelkens E, Debaveye S & Tytgat J (2010) Venom components from *Citharischius crawshayi* spider (Family Theraphosidae): exploring transcriptome, venomics, and function. *Cell Mol Life Sci* **67**, 2799–2813.
16 Wullschleger B, Kuhn-Nentwig L, Tromp J, Kämpfer U, Schaller J, Schürch S & Nentwig W (2004) CSTX-13, a highly synergistically acting two-chain neurotoxic enhancer in the venom of the spider *Cupiennius salei* (Ctenidae). *Proc Natl Acad Sci USA* **101**, 11251–11256.
17 Kuhn-Nentwig L, Schaller J & Nentwig W (2004) Biochemistry, toxicology and ecology of the venom of the spider *Cupiennius salei* (Ctenidae). *Toxicon* **43**, 543–553.
18 Kuhn-Nentwig L, Willems J, Seebeck T, Shalaby T, Kaiser M & Nentwig W (2011) Cupiennin 1a exhibits a remarkably broad, non-stereospecific cytolytic activity on bacteria, protozoan parasites, insects, and human cancer cells. *Amino Acids* **40**, 69–76.
19 Kuhn-Nentwig L, Schaller J & Nentwig W (1994) Purification of toxic peptides and the amino acid sequence of CSTX-1 from the multicomponent venom

of *Cupiennius salei* (Araneae: Ctenidae). *Toxicon* **32**, 287–302.
20 Pimenta AM, Stöcklin R, Favreau P, Bougis PE & Martin-Eauclaire MF (2001) Moving pieces in a proteomic puzzle: mass fingerprinting of toxic fractions from the venom of *Tityus serrulatus* (Scorpiones, Buthidae). *Rapid Commun Mass Spectrom* **15**, 1562–1572.
21 Kuhn-Nentwig L, Müller J, Schaller J, Walz A, Dathe M & Nentwig W (2002) Cupiennin 1, a new family of highly basic antimicrobial peptides in the venom of the spider *Cupiennius salei* (Ctenidae). *J Biol Chem* **277**, 11208–11216.
22 Pukala TL, Boland MP, Gehman JD, Kuhn-Nentwig L, Separovic F & Bowie JH (2007) Solution structure and interaction of cupiennin 1a, a spider venom peptide, with phospholipid bilayers. *Biochemistry* **46**, 3576–3585.
23 Kubista H, Mafra RA, Chong Y, Nicholson GM, Beirao PS, Cruz JS, Boehm S, Nentwig W & Kuhn-Nentwig L (2007) CSTX-1, a toxin from the venom of the hunting spider *Cupiennius salei*, is a selective blocker of L-type calcium channels in mammalian neurons. *Neuropharmacology* **52**, 1650–1662.
24 Wullschleger B, Nentwig W & Kuhn-Nentwig L (2005) Spider venom: enhancement of venom efficacy mediated by different synergistic strategies in *Cupiennius salei*. *J Exp Biol* **208**, 2115–2121.
25 Vassilevski AA, Fedorova IM, Maleeva EE, Korolkova YV, Efimova SS, Samsonova OV, Schagina LV, Feofanov AV, Magazanik LG & Grishin EV (2010) Novel class of spider toxin: active principle from the yellow sac spider *Cheiracanthium punctorium* venom is a unique two-domain polypeptide. *J Biol Chem* **285**, 32293–32302.
26 Zhang Y, Chen J, Tang X, Wang F, Jiang L, Xiong X, Wang M, Rong M, Liu Z & Liang S (2010) Transcriptome analysis of the venom glands of the Chinese wolf spider *Lycosa singoriensis*. *Zoology (Jena)* **113**, 10–18.
27 Yan Y & Adams ME (1998) Lycotoxins, Antimicrobial peptides from venom of the wolf spider *Lycosa carolinensis*. *J Biol Chem* **273**, 2059–2066.
28 Budnik BA, Olsen JV, Egorov TA, Anisimova VE, Galkina TG, Musolyamov AK, Grishin EV & Zubarev RA (2004) *De novo* sequencing of antimicrobial peptides isolated from the venom glands of the wolf spider *Lycosa singoriensis*. *J Mass Spectrom* **39**, 193–201.
29 Xu K, Ji Y & Qu X (1989) Purification and characterization of an antibacterial peptide from venom of *Lycosa singoriensis*. *Acta Zool Sinica* **35**, 300–305.
30 Kozlov SA, Vassilevski AA, Feofanov AV, Surovoy AY, Karpunin DV & Grishin EV (2006) Latarcins, antimicrobial and cytolytic peptides from the venom of the spider *Lachesana tarabaevi* (Zodariidae) that exemplify biomolecular diversity. *J Biol Chem* **281**, 20983–20992.
31 Corzo G, Villegas E, Gomez-Lagunas F, Possani LD, Belokoneva OS & Nakajima T (2002) Oxyopinins, large amphipathic peptides isolated from the venom of the wolf spider *Oxyopes kitabensis* with cytolytic properties and positive insecticidal cooperativity with spider neurotoxins. *J Biol Chem* **277**, 23627–23637.
32 Belokoneva OS, Villegas E, Corzo G, Dai L & Nakajima T (2003) The hemolytic activity of six arachnid cationic peptides is affected by the phosphatidylcholine-to-sphingomyelin ratio in lipid bilayers. *Biochim Biophys Acta* **1617**, 22–30.
33 Escoubas P, Chamot-Rooke J, Stöcklin R, Whiteley BJ, Corzo G, Genet R & Nakajima T (1999) A comparison of matrix-assisted laser desorption/ionization time-of-flight and liquid chromatography electrospray ionization mass spectrometry methods for the analysis of crude tarantula venoms in the Pterinochilus group. *Rapid Commun Mass Spectrom* **13**, 1861–1868.
34 Guette C, Legros C, Tournois G, Goyffon M & Celerier ML (2006) Peptide profiling by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry of the *Lasiodora parahybana* tarantula venom gland. *Toxicon* **47**, 640–649.
35 Cornett DS, Reyzer ML, Chaurand P & Caprioli RM (2007) MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat Methods* **4**, 828–833.
36 Pimenta DC, Prezoto BC, Konno K, Melo RL, Furtado MF, Camargo AC & Serrano SM (2007) Mass spectrometric analysis of the individual variability of *Bothrops jararaca* venom peptide fraction. Evidence for sex-based variation among the bradykinin-potentiating peptides. *Rapid Commun Mass Spectrom* **21**, 1034–1042.
37 Kuhn-Nentwig L, Schaller J, Kämpfer U, Imboden H, Malli H & Nentwig W (2000) A lyisne rich C-terminal tail is directly involved in the toxicity of CSTX-1, neurotoxic peptide from the venom of the spider *Cupiennius salei*. *Archiv Insect Biochem Physiol* **44**, 101–111.
38 Taggi AE, Meinwald J & Schroeder FC (2004) A new approach to natural products discovery exemplified by the identification of sulfated nucleosides in spider venom. *J Am Chem Soc* **126**, 10364–10369.
39 Kuhn-Nentwig L, Stöcklin R & Nentwig W (2011) Venom composition and strategies in spiders. *Adv Insect Physiol* **40**, 1–86.
40 Escoubas P & King GF (2009) Venomics as a drug discovery platform. *Exp Rev Proteomics* **6**, 221–224.
41 Vassilevski AA, Kozlov SA & Grishin EV (2009) Molecular diversity of spider venom. *Biochemistry (Moscow)* **74**, 1505–1534.

42 Escoubas P (2006) Molecular diversification in spider venoms: a web of combinatorial peptide libraries. *Mol Divers* **10**, 545–554.
43 Liang S (2008) Proteome and peptidome profiling of spider venoms. *Exp Rev Proteomics* **5**, 731–746.
44 Yamaguchi S, Fujita T, Fujino T & Korenaga T (2008) Suppression of matrix-related ions using cyclodextrin in MALDI mass spectrometry. *Anal Sci* **24**, 1497–1500.
45 Trachsel C, Kämpfer U, Bechtold R, Schaller J & Schürch S (2009) Elucidation of the disulfide bridge pattern of the recombinant human growth and differentiation factor 5 dimer and the interchain Cys/Ala mutant monomer. *Anal Biochem* **390**, 103–108.
46 Marti T, Rösselet SJ, Titani K & Walsh KA (1987) Identification of disulfide-bridged substructures within human von Willebrand factor. *Biochemistry* **26**, 8099–8109.
47 Lengweiler S, Schaller J & Rickli EE (1996) Identification of disulfide bonds in the ninth component (C9) of human complement. *FEBS Lett* **380**, 8–12.

Supporting information

The following supplementary material is available:
Table S1. Compilation of the average masses in the desalted, unfractionated venom of *C. salei* detected by MALDI-TOF-MS using CHCA and SA as matrices and α-cyclodextrin as co-matrix. A total of 112 venom components were assigned.
Table S2. Compilation of the average masses in the fractionated venom of *C. salei* detected by MALDI-TOF-MS using CHCA and SA as matrices and α-cyclodextrin as co-matrix, recorded in linear mode (asterisk indicates reflector mode; underlined masses only detectable by MALDI-TOF-MS). A total of 243 venom components were assigned.
Table S3. Compilation of the monoisotopic masses in the fractionated venom of *C. salei* detected by high resolution ESI-MS (underlined masses only detectable by high resolution ESI-MS). A total of 171 venom components were assigned.
Table S4. Amino acid analysis of fractions 34, 36A and 36B (mol amino acid per mol peptide); values from sequence in parentheses.
Table S5. Biochemical properties of fully characterized SCPs. The values for the pI and the net charge are calculated based on the amino acid sequence. Observed mass number of SCP u2 is low resolution.
Table S6. Biochemical properties of fully characterized CsTx toxin-like structures (Cys-containing). The values for the pI and the net charge are calculated based on the amino acid sequence. The mass numbers given for two-chain peptides are for Cys as carboxyamidomethylated Cys.

This supplementary material can be found in the online version of this article.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be reorganized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.