Identification of a precursor processing protease from the spider Cupiennius salei essential for venom neurotoxin maturation

Received for publication, August 8, 2017, and in revised form, December 13, 2017 Published, Papers in Press, December 21, 2017, DOI 10.1074/jbc.M117.810911

Nicolas Langenegger‡, Dominique Koua§, Stefan Schürch¶, Manfred Heller†, Wolfgang Nentwig‡, and Lucia Kuhn-Nentwig‡§†
From the ‡Institute of Ecology and Evolution, University of Bern, Balzertorstrasse 6, CH-3012 Bern, Switzerland, the §Institut National Polytechnique Félix Houphouët-Boigny, BP 1093 Yamoussoukro, Côte d’Ivoire, the ¶Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, CH-3012 Bern, Switzerland, and the †Department of Clinical Research, Proteomics and Mass Spectrometry Core Facility, University of Bern, Freiburgstrasse 15, CH-3010 Bern, Switzerland

This work was supported by Swiss National Science Foundation Grant 31003A_162564 (to W. N.). The authors declare that they have no conflicts of interest with the contents of this article.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EBI Data Bank with accession number(s)MF537412, MF537413, MF537414, MF537415, MF537416, MF537417, MF537418, MF537419, MF537420, MF537421, MF537422, MF537423, MF537424, MF537425, MF537426, MF537427, MF537428, MF537429, and MF537430. The mass spectrometry proteomics data are available from ProteomeXchange with the accession number PXD007126.

1 To whom correspondence should be addressed: Institute of Ecology and Evolution, University of Bern, Balzertorstrasse 6, CH-3012 Bern, Switzerland. Tel.: 41-31-631-45-32; E-mail: lucia.kuhn@iee.unibe.ch.

© 2018 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.

Spider venom neurotoxins and cytolytic peptides are expressed as elongated precursor peptides, which are post-translationally processed by proteases to yield the active mature peptides. The recognition motifs for these processing proteases, first published more than 10 years ago, include the processing quadruplet motif (PQM) and the inverted processing quadruplet motif (iPQM). However, the identification of the relevant proteases was still pending. Here we describe the purification of a neurotoxin precursor processing protease from the venom of the spider Cupiennius salei. The chymotrypsin-like serine protease is a 28-kDa heterodimer with optimum activity at venom’s pH of 6.0. We designed multiple synthetic peptides mimicking the predicted cleavage sites of neurotoxin precursors. Using these peptides as substrates, we confirm the biochemical activity of the protease in propeptide removal from neurotoxin precursors by cleavage C-terminal of the PQM. Furthermore, the PQM protease also cleaves the iPQM relevant for heterodimerization of a subgroup of neurotoxins. An involvement in the maturing of cytolytic peptides is very likely, due to high similarity of present protease recognition motifs. Finally, bioinformatics analysis, identifying sequences of homolog proteins from 18 spiders of 9 families, demonstrate the wide distribution and importance of the isolated enzyme for spiders. In summary, we establish the first example of a PQM protease, essential for maturing of spider venom neurotoxins. In the future, the here described protease may be established as a powerful tool for production strategies of recombinant toxic peptides, adapted to the maturing of spider venom toxicins.

The toxic action of spider venoms relies on a complex mixture of different components. The lead toxic components are mainly enzymes, as reported for Loxosceles (Sibariidae) (1), cytolytic peptides (e.g. in Lachesana tarabaevi (Zodariidae)) (2) or, in the majority of spider venom, neurotoxins (3). The activity of neurotoxins can thereby be enhanced by cytolytic peptides, if co-occurring in the venom (4, 5). Cytolytic and neurotoxic peptides are expressed in the venom gland as precursors, which undergo proteolytic processing to yield the mature active peptide (6). Neurotoxic peptide precursors are reported to commonly comprise a signal peptide followed by an acidic propeptide and a mature peptide. The last four amino acid residues of the propeptide have been demonstrated to form a conserved motif featuring an Arg residue at position –1 and at least one Glu residue at positions –2, –3, and/or –4. This motif is named the processing quadruplet motif (PQM),2 and it is supposed to act as recognition site for a propeptide processing protease (6), henceforth referred to as PQM protease. In precursors of cytolytic peptides, an additional processing motif, featuring an Arg residue at position 1 and at least one Glu residue within the three following amino acids was identified and named inverted PQM (iPQM) (7, 8). Despite the description of the PQM and iPQM as potential protease recognition sites and the widespread existence of similar precursor structures of spider venom toxicins, few is known about the proteases responsible for precursor maturation and there is still no experimental evidence for the involvement of a specific protease. Nevertheless, different venom proteases have been hypothesized to be involved in venom peptide maturation. Among them, a recently identified thiol protease from Lycosa singoriensis (Lycosidae) showing similarity to the cysteine protease cathepsin L (9) and a venom protein of Acanthoscurria geniculata (Theraphosidae) homologous to proprotein convertases (10). In addition, multi-

2 The abbreviations used are: PQM, processing quadruplet motif; DABCYL, 4-((4-(dimethylamino)phenyl)azo)benzoic acid; EDANS, 5–((2-aminoethyl)amino)naphthalene-1-sulfonic acid; SEC, size exclusion chromatography; CEX, cation exchange chromatography; RP-HPLC, reversed-phase HPLC; AEBSF, 4-(2-aminoethyl)benzenesulfonyl fluoride; iPQM, inverted processing quadruplet motif; PRIDE, proteomics identifications; HCD, higher-energy collisional dissociation; ETnCD, electron transfer/higher-energy collision dissociation; PAM, peptidylglycine α-amidating monooxygenase; AGC, automatic gain control.
ple metalloproteases and some serine proteases with unclear functions have been identified in spider venoms (3).

In conclusion, different proteins have been hypothesized to be involved in propeptide processing of spider venom peptides. Nevertheless, the involvement in the maturation process of venom neurotoxins has so far not been experimentally confirmed for any protease isolated from spider venom. Here, we report the isolation and characterization of a PQM processing protease from the venom of the Central American spider Cupiennius salei (Ctenidae). We demonstrate the specificity of the PQM protease for propeptide removal from spider venom neurotoxin precursors. Additionally, we provide evidence for the involvement of this protease in the generation of the two heterodimeric neurotoxins CsTx-13 (C. salei, P83919) (11) and /H9275-agatoxin-1A (Agelenopsis aperta, P15969) (12).

Results

Purification of the native protease

The results are based on processing a total of 11.5 ml of crude venom of C. salei, corresponding to the venom obtained from approximately 1150 spider milkings (13). We purified the protease by three successive chromatographic steps. All fractions of each chromatographic step were tested for proteolytic activity against two substrates mimicking a PQM and an iPQM, respectively (DABCYL-PFFENEQARKDDKN-Glu(EDANS)-NH2 and DABCYL-FYPSQRSETD-Glu(EDANS)-NH2). Proteolytically active fractions were pooled, analyzed by SDS-PAGE, and subjected to the next chromatographic step. First, size exclusion chromatography (SEC) of venom on a Sephadex G-75 column allowed the separation of the active fraction (Fig. 1A, peak 2) from the majority of peptides and proteins appearing at masses smaller than ~35 kDa in SDS-PAGE (Fig. 1D). Second, cation exchange chromatography (CEX) of size exclusion chromatography peak 2 on a MiniS PE column led to the enrichment of the proteolytic activity in a broad peak (Fig. 1B, peak 5). We observed a loss in total proteolytic activity of this peak compared with SEC peak 2 (Table 1). This may be due to denaturation of the protease leading to a reduced specific activity and/or due to the loss of protease material during purification. The assumed loss of protease is in accordance with our observations comparing SDS-PAGE lane SEC and lane CEX (Fig. 1D), where we noticed a reduction of the intensity of band a (protease) relative to the intensity of band b (contaminating peptides). Third, RP-HPLC on a butyl column enabled the final purification of the protease (Fig. 1C, peak 3). The contaminating peptides could be efficiently separated (Fig. 1C, peaks 1 and 2). Tandem mass spectrometry of chymotryptic fragments of the peptides in peaks 1 and 2 showed top hits for the neurotoxins CsTx-39, CsTx-243, CsTx-9, and CsTx-12 (Tables S1 and S2). The isolated enzyme appeared at a mass slightly below 30 kDa on SDS-PAGE (Fig. 1D, lane RP-HPLC), which is in good agreement with the measured mass of 28 kDa by MALDI-TOF-MS (Fig. 2A). Contamination of the purified enzyme could neither be detected by SDS-PAGE, nor by MALDI-TOF-MS with different matrices (Fig. 2A, Figs. S1 and S2). The specific activity of the pure enzyme was (6,364 ± 2,306) × 10^{-12} mol liter^{-1} s^{-1} µg^{-1}. This is 1,097-fold higher than the specific activity of the venom (Table 1). The venom concentration of the isolated protease can be estimated to 0.78 ± 0.2 µM by quantification of the isolated enzyme (amino acid analysis). However, it
is reasonable to assume that the venom concentration is underestimated due to loss of protease during the purification procedure.

**Structural characterization and cDNA sequence of the protease**

The isolated protease was identified to be heterodimeric by liquid chromatography-mass spectrometry (LC-MS) of the reduced and alkylated protein (Fig. 2B). We sequenced the light chain (Fig. 2B, retention time 12.06 min) by tandem mass spectrometry (Fig. S3 and Table S3) and found it to be C-terminally amidated and connected to the heavy chain by one disulfide bridge. The monoisotopic mass of the uncharged light chain was determined to be 1,621.935 Da (Cys carboxyamidomethylated), which is in good agreement with its theoretical mass of 1,621.934 Da (Fig. 2C). The heavy chain eluted from the HPLC column in a broad peak (Fig. 2B, retention time 18.42–19.60 min). The mass spectra measured between these retention times provide evidence for the presence of multiple isoforms of the heavy chain (Fig. S4). Nevertheless, the number of isoforms and their accurate masses cannot be assessed due to unknown methionine oxidation state and possible adducts.

By analysis of the venom gland transcriptome, we identified two transcripts giving rise to two isoforms of a monomeric protease precursor, both differing in three amino acid residues of the later heavy chain (Fig. 3). The more abundant isoform (relative read count approximately 90%) was named venom serine protease 1a1 (VSP1_a1), the less abundant (relative read count approximately 10%) was named venom serine protease 1a2 (VSP1_a2). The amino acid sequences of the heavy chains were validated by tandem mass spectrometry of peptides resulting from trypsin, chymotrypsin, or proteinase K digests. The obtained data were searched against our in-house protein database venomProteome.fasta as described below, and the results were mapped against the cDNA-derived amino acid sequence of the two isoforms of the protease as identified at transcriptome level. The sequence coverage was 100% in the region 36Ile–Thr280 with a false discovery rate of ≤1% (Fig. 3). Alignment of the amino acid sequence to the cDNA sequence shows that the 840-bp open reading frame codes for a signal peptide (Met–Gly18) followed by the light chain (Lys–Gln33), a two-amino acid spacer (34Gly–Arg35), and the heavy chain (36Ile–Thr280). Functional analysis (14) of the amino acid sequence predicts the protein belongs to the serine peptidase S1A subfamily with active sites His80, Asp130, Ser231, and substrate-binding sites Asp225, Ala251, and Gly253. The amino acid sequence differences of the identified isoforms F98Y, and D176N

**Table 1**

**Summary of the purification process of the PQM protease**

The data shown are from two independent purifications starting with 3 ml of crude venom each. The total activity was measured as degradation of the PQM substrate in triplicate for each purification. The data are given as mean ± S.D. The standard deviation of the specific activity was calculated according to the rules of error propagation.

<table>
<thead>
<tr>
<th>Step</th>
<th>Total activity</th>
<th>Protein content</th>
<th>Specific activity</th>
<th>Fold purification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venom</td>
<td>5.93 ± 1.40 × 10⁻¹ mol l⁻¹ s⁻¹ (µl venom eq.)⁻¹</td>
<td>101.750 ± 6.124 µg (µl venom eq.)⁻¹</td>
<td>5.8 ± 1.7 × 10⁻¹ µmol l⁻¹ s⁻¹</td>
<td>1</td>
</tr>
<tr>
<td>SEC</td>
<td>5.26 ± 0.53 × 10⁻¹</td>
<td>3.581 ± 0.110</td>
<td>146.8 ± 19.2 × 10⁻¹</td>
<td>25</td>
</tr>
<tr>
<td>CEX</td>
<td>0.85 ± 0.14 × 10⁻¹</td>
<td>0.053 ± 0.009</td>
<td>1,601.7 ± 527.8 × 10⁻¹</td>
<td>276</td>
</tr>
<tr>
<td>HPLC</td>
<td>1.40 ± 0.16 × 10⁻¹</td>
<td>0.022 ± 0.006</td>
<td>6,364.3 ± 2,305.6 × 10⁻¹</td>
<td>1,097</td>
</tr>
</tbody>
</table>

**Figure 2. Mass spectrometry of the PQM protease.** A, mass spectrum of the intact protein acquired by MALDI-TOF-MS. The matrix used was α-cyano-4-hydroxy-cinnamic acid. B, chromatogram of the reduced and alkylated isolated protease given as total ion current. The heterodimerity of the protease is shown by separation of the light chain (peak 1) and the heavy chain (peak 2) showing the enlargement of the region at 1621–1627 m/z. The mass spectrum was deconvoluted to neutral monoisotopic masses with Xcalibur version 2.2 and its plug-in Xtract.
were both validated on protein level. The isobaric amino acid change I100L cannot be detected by mass spectrometry.

The disulfide bridge pattern of the PQM protease was elucidated by selective proteolysis followed by LC-MS/MS/MS analysis under nonreductive conditions (Fig. 4, **A** and **B**, Figs. S5–S11 and Tables S4–S8). The identified disulfide bridge pattern is identical to one of the peptide isomerase of *A. aperta* (15), the protein with the highest sequence identity to the PQM protease of all proteins with published disulfide bridge patterns.

**Enzymatic activity characterization of the protease**

We measured the maximal activity of the isolated protease at pH 6. The enzyme was not active at buffer pH ≤ 4 or ≥ 8 (Fig. 5A). The divalent cations Co2+, Mn2+, Mg2+, Ni2+, and Ca2+ did not notably affect the activity of the isolated protease, whereas the presence of 5 mM Zn2+ in the assay buffer led to an activity decrease by a factor of 5 (Fig. 5B). A slight decrease in activity was observed in the presence of the chelating agent EDTA (10 mM), whereas the serine protease inhibitor AEBSF (4-(2-aminoethyl)benzenesulfonyl fluoride) (10 mM) caused a total loss of activity. The measured monovalent cations did not noticeably affect the activity of the isolated protease (Fig. 5B).

**Substrate specificity of the protease**

To assess the specificity of the isolated protease we: 1) analyzed known neurotoxic precursor sequences of *C. salei* for possible PQM protease processing sites, 2) synthesized peptides covering these sites, and 3) analyzed the digestion products of these peptides by LC-MS. The neurotoxins CsTx-1 and CsTx-9 exhibit the two most prominent PQMs in *C. salei* processing quadruplet motif processing protease

Figure 3. cDNA and amino acid sequence of the purified PQM protease. Shown are the cDNA and the amino acid sequence of the isolated protease main isoform. Non-silent mutations of the minor isoform are indicated below the main sequence. Amino acids in bold letters were validated on the protein level by tandem mass spectrometry of the reduced and alkylated chain light chain, or tandem mass spectrometry of peptides generated by trypsin, chymotrypsin, or protease K digests of the heavy chain. Amino acids in boxes are part of the catalytic triad. Cysteines are black boxed.
peptides (EQAR and EQVR, respectively), present in ~85% of all transcripts coding for neurotoxin precursors. We detected specific cleavage on the C-terminal side of these PQMs (Fig. 6, A and B). The involvement of the isolated protease in generating heterodimeric toxins, such as CsTx-13 (C. salei) or ω-agatoxin-1a (A. aperta), and the potential heterodimer CsTx-40 was investigated using appropriate substrates (Fig. 6C). The three substrates were all cut twice: 1) C-terminal of the PQM, and 2) C-terminal of Arg6. Arg6 is part of an iPQM in all three substrates and additionally part of a PQM in the substrate mimicking CsTx-40 (Fig. 6C). Cutting on the C-terminal side of the PQM of the heterodimeric precursors liberates the -chain of the toxins. The C-terminal amino acid of the -chains of the mature heterodimers corresponds to position 5 of the synthetic substrates. Nevertheless, cutting was observed C-terminal of Arg6 of the synthetic substrates. It is therefore tempting to assume that the C-terminal Arg from the liberated -chain is removed by a carboxypeptidase during maturation of heterodimeric neurotoxins.

Discussion

Isolation and characterization of spider venom enzymes can be challenging due to: 1) their often up to 1000-fold lower venom concentration compared with the main toxins (e.g. for C. salei: venom hyaluronidase, 1.3–9.3 μM (16) versus main neurotoxin CsTx-1, 1.4–3.3 mM (13)), and 2) potential loss of activity during the purification. The latter requires optimization of the purification strategy, which is only possible if venom is accessible in high quantities.

3 L. Kuhn-Nentwig, personal communication.
Structural characterization

We identified the isolated PQM protease to be a heterodimer. The heterodimeric structure is likely to be a result of proteolytic cleavage of a monomeric precursor C-terminal of Arg35. Nevertheless, the C-terminal amino acid of the light chain was identified to be Gln33. We propose that this is a result of exoproteolytical removal of Arg35 followed by peptidylglycine α-amidating monooxygenase (PAM)-mediated C-terminal amidation of Gln33, using Gly34 as amide donor (7). A PAM has been identified in our C. salei venom gland transcriptome.3

The biological function of the heterodimeric structure of the PQM protease is unclear. Some proteases are expressed aszymogens that are activated by limited proteolysis. Trypsin, for example, emerges from its precursor trypsinogen by activation through proteolytical removal of a six-amino acid propeptide (17). In analogy, cleavage of the monomeric precursor of the PQM protease into a light and a heavy chain may be required for the activation of the protease. Thus, the light chain might be comparable with the propeptide of trypsin, but stays connected to the activated protease due to the presence of an interchain disulfide bridge. Nevertheless, the structural advantage of heterodimerity of the PQM protease needs further investigation requiring an active protease obtained by recombinant expression.

Isoforms

We detected isoforms of the PQM protease heavy chain bytranscriptome analysis and validated the respective amino acid sequences by bottom-up proteomics. Isoforms of venom pro-
The isolated protease shows high amino acid sequence identity with two proteins isolated from spider venom: U21-ctenitoxin-Pn1a (76%) (P84033) a monomeric venom protease from Phoneutria nigriventer (20), and a heterodimeric venom peptidase isomerase (56%) (Q9TWH3/Q9TXD8) from A. aperta (15). Comparable identities (59–98%) were calculated for putative protease homologs from 18 spiders of nine families (if we regard Cupiennius as non-ctenidae, see below) identified in our in-house spider venom gland cDNA libraries of RTA clade spiders (Fig. S19). The high amino acid sequence conservation and the presence of the enzyme in the venom glands of all investigated spider families demonstrate the high biological relevance of the protease for spiders. Phylogenetic analysis (Fig. 7) shows a topology in good agreement with a recently published phylogeny (21, 22). Interestingly, the sequences of the two Cupiennius species cluster distantly from their currently assigned family Ctenidae and group close to Lycosidae and Pisauridae. Corresponding observations concerning ambiguous family affiliation of Cupiennius have recently been discussed in taxonomic studies (22, 23), indicating that the structure of venom components can provide relevant phylogenetic arguments.

Multiple sequence alignment of the homolog amino acid sequences shows conserved catalytic residues, substrate-binding sites, and high overall identities (Fig. S19). All shown sequences feature the same signal peptidase cleavage site (Gly18, prediction using SignalP 4.1 (24)), a conserved cysteine pattern, and conserved amino acid residues (35Arg–Val37) at the predicted proteolytic cleavage site for heterodimerization of the monomeric precursor. It is therefore reasonable to propose, that the homologs undergo all the same maturation mechanism resulting in a heterodimeric structure. Nevertheless, this is not in agreement with the description of U21-ctenitoxin-Pn1a of Phoneutria nigriventer (P84033) as a monomeric enzyme (20). However, the cDNA-derived amino acid sequence of another closely related species in the same genus, Phoneutria fera, exhibits the structural features as prerequisite for heterodimerization and a high sequence identity of 95% with U21-ctenitoxin-Pn1a (Fig. S19).

The light chain of the protease from C. salei is C-terminally amidated. This amidation seems not to be conserved, as some sequences lack Gly34, the supposed amide-donor for the amidation by a PAM. Interestingly, the peptide isomerase from A. aperta (Q9TWH3/Q9TXD8) exhibits the same amino acid residues in the predicted active center and substrate-binding sites, as well as the same disulfide bridge pattern as the PQM protease of C. salei. Functional analysis of the amino acid sequence predicts the enzyme to belong to the peptidase S1A subfamily. Therefore, it seems conceivable that the one-step purification procedure of the peptide isomerase from the venom of the funnel web spider (15) was not appropriate to purify the enzyme to homogeneity, resulting in a misinterpretation of the enzymatic function due to contamination of the purified protein by an isomerase. A comparable case has been reported for Tex31, a cysteine-rich venom protein belonging to the CAP superfamily (IPR014044 and IPR001283), with homologs in spider venom (10). Tex31, isolated from the venom of the cone snail Conus textile, was initially thought to proteolytically process conotoxin precursors (25). However, Qian et al. (26) could demonstrate that the reported proteolytic activity of Tex31 is rather caused by a contamination of Tex31 with an unknown protease than by the protein itself.

Specificity of the PQM protease

Spider venom prototoxin-processing sites have already been well investigated at the transcriptome level, resulting in the description of the protease recognition motifs PQM and iPQM (6, 7). In our study, we investigated the cutting specificity of the isolated protease on different synthetic substrates. We identi-
fied a specific cutting C-terminal of all Arg residues of PQMs. Furthermore, cutting on the C-terminal side of Arg residues of all iPQMs and PQMs was observed for synthetic substrates mimicking the processing sites of the heterodimeric neurotoxin precursors of CsTx-13, CsTx-40, and α-Aga-1a. In contrast, we could not observe processing of the iPQM present in close proximity to the PQM in the substrate mimicking the CsTx-9 propeptide–mature peptide sequence stretch. It seems conceivable that the PQM protease lacks the specificity for this iPQM, and/or that the short distance between the Arg residues of the PQM and the nearby iPQM may impede correct substrate recognition (three amino acids in the case of the CsTx-9 precursor and four to five amino acids in the case of the precursors of heterodimeric neurotoxins).

Not all predicted cutting sites were cleaved with the same efficiency. This is followed from the observed HPLC peak height ratios between digestion fragments and undigested substrates. It is well known that the turnover number of a protease is highly dependent on the substrate (27), which may explain the observed differences in cleavage efficiency.

The here described cleavage specificity of the isolated protease is congruent with the predicted specificity of a propeptide peptidase (7) and points toward an involvement of the PQM protease in generation of heterodimeric toxins from their monomeric precursors. Our data indicates that heterodimeric neurotoxins are generated by: 1) removal of the inter-chain spacer of the monomeric precursors by the PQM protease, followed by 2) removal of the C-terminal Arg of the α-chain of heterodimeric neurotoxins by a carboxypeptidase. This is in accordance with the supposed maturation process of toxin precursors based on transcriptomic studies (7). The number of reported heterodimeric neurotoxins is rather low in proportion to all described spider venom neurotoxins (28). This raises the question if heterodimerization of monomeric neurotoxins is of biological benefit. In C. salei venom, a family of heterodimeric neurotoxins (CsTx-8, -12, and -13) are present in millimolar concentrations (29). The LD50 values of these heterodimeric neurotoxins (CsTx-13: 16.3 pmol/mg of fly; CsTx-8: 6.3 pmol/mg of fly) are significantly higher than the ones of the main monomeric neurotoxins (CsTx-1: 0.35 pmol/mg of fly; CsTx-9: 3.12 pmol/mg of fly) (11, 29). Surprisingly, CsTx-13 enhances synergistically the paralytic activity of CsTx-1 and CsTx-9 toward Drosophila flies even in nontoxic concentrations (11).

Cytolytic venom peptide precursors have been reported to feature a comparable organization to neurotoxic precursors, but in some cases to give rise to more than one mature peptide. Multiple mature peptides on one precursor are separated by short spacers flanked by iPQMs and PQMs. These spacers are proteolytically removed during the maturation process and are comparable with the above-mentioned inter-chain spacers of heterodimeric neurotoxin precursors. In analogy to our findings for heterodimeric neurotoxins, maturation of some cytolytic precursors is thought to involve a carboxypeptidase (7). We assume that maturation of cytolytic precursors is comparable with the maturation of heterodimeric neurotoxins and involves the same proteases. Furthermore, we expect the PQM protease to occur in all spiders producing neurotoxic and/or cytolytic peptide precursors.

To the best of our knowledge, this is the first report about the isolation and characterization of the protease responsible for the removal of propeptides from spider venom neurotoxin precursors. In summary, our data confirm the proposed maturation mechanism of monomeric neurotoxins by proteolytic cleavage C-terminal of the PQM (6). Moreover, specificity studies imply analogous maturation mechanisms of heterodimeric neurotoxins and complex cytolytic precursors. We provide experimental evidence to support the described maturation mechanism of complex precursors by PQM and iPQM cleavage followed by exoproteolytical removal of the residual arginine of the iPQM (7). As consequence, the PQM protease may enable efficient production strategies for recombinant peptides and proteins, adapted to the maturing of complex cytolytic peptides as supposed already by Vassilevski et al. (30).

**Experimental procedures**

**Spider maintenance and venom collection**

C. salei (Ctenidae) were laboratory bred and milked as previously reported (13). For the purification and characterization of the PQM protease, venom of adult male and female spiders was used. For transcriptomic studies, Ancylostometes rufus, P. fera, Cupiennius getazi, Viridasius fasciatus, Zoropsis spinimana, Hogna radiata (Italy), and H. radiata (Spain) were laboratory bred. Eratigene atrica (formerly Tegenaria atrica), Pisaura mirabilis, Anypheana accentuata, Amaurobius ferox, and Alopecosa cuneata were collected in Switzerland, Lycosa hispanica in Spain, Geolycosa vultuosa in Hungary, Peucetia striata in Tansania, Lycosa praegrandis in Greece, and Alopecosa marikovskyi in Kasachstan. To reduce gene pool variation, spiders of the same species were collected from one population. Collection of spiders did not require specific permissions because we collected on public and/or private land of the authors without any protection status or because spiders were laboratory bred. None of the above-mentioned spiders belongs to a protected or endangered species.

**cDNA libraries of venom glands**

The cDNA venom gland library of C. salei is based on 454 sequencing and has been previously described (31). The libraries of other spiders were built using the following procedure: venom glands of adult spiders (see on top) were dissected in different time intervals after milking and stored in RNAlater (Qiagen). Total RNA was extracted using an in-house protocol combining phenol/chloroform extraction with the RNeasy mini kit (Qiagen). cDNA libraries were prepared using Illumina’s TruSeq-stranded mRNA prep kit. We used double barcoded and selected fragments with lengths between 300 and 600 bp for further sequencing (Pippin HT system, Sage Science). Each spider venom gland cDNA library was multiplexed (25% per lane) with other libraries (mostly genome) from non-arthropods and sequenced by the next generation sequencing platform of the University of Bern on an Illumina HiSeq3000 platform using 2 × 150 bp paired-end sequencing cycles. The resulting reads have been assembled using Trinity version 2.1.1 (32).
PQM protease cDNA sequences

The sequences of the ORFs encoding the two isoforms of the C. salei PQM protease (VSP1_a1 and VSP1_a2) were established from 189 contigs comprising 1960 reads. Sequences of homolog proteins from other spiders were obtained by a venom gland cDNA library search using a combination of Blast and Exonerate 2.2.0 (33) with the sequence of the C. salei PQM protease as template. Sequences were manually reviewed at read level.

Substrate design and enzymatic assays

The proteolytic activity was measured in a fluorescence resonance energy transfer (FRET)-based assay using two peptide-substrates carrying a PQM and an iPQM, respectively. The PQM substrate (DABCYL-PFFENEQARKDDKN-Glu(EDANS)-NH2, PQM underlined) was adapted to the propeptide of CsTx-1 (P81694) and to the mature peptide of CsTx-9 (P58604). CsTx-1, the most abundant neurotoxin in the venom of C. salei, comprises a cysteine at position 2 (13). To prevent any possible effect of a free cysteine residue on proteolytic processing of the synthetic substrate or on the activity test, we adapted the mature peptide mimicking region of the PQM substrate to CsTx-9 (the second most abundant neurotoxin in C. salei venom), whereas retaining the PQM of the main neurotoxin CsTx-1. The iPQM substrate (DABCYL-FYPQRSQETD-Glu(EDANS)-NH2, iPQM underlined) was adapted to C. salei’s main heterodimeric neurotoxin CsTx-13 (P83919). For the activity test, 1.2 µg of the substrates, (supplied by GeneCust, Luxembourg) were resolved in 130 µl of activity test buffer (200 mM NH₄OAc, pH 6.0, 100 mM KCl, 1 mM CaCl₂) and incubated for 30–120 min with 20 µl of each fraction of a purification step. Incubation was performed at room temperature in assay plates (96-well, flat bottom, non-binding surface, black polystyrene; Corning). End point fluorescence measurements were performed at 25 °C using a Synergy HT Microplate reader (BioTek) equipped with the following filters: 340/30 and 485/20 nm. The activity of a given fraction was calculated as percentage of substrate digested after incubation relative to the negative control (buffer only) and the positive control (1 µg of trypsin).

Isolation and purification

We purified the native protease by applying three successive steps of chromatographic separation, starting with a total of 2–3 ml of venom per batch. First, size exclusion chromatography of 100 µl of venom per separation diluted with 300 µl of running buffer (100 mM NH₄OAc, pH 5.0) was performed on a Superdex 75 HR (10 x 300 mm, GE Healthcare Life Sciences) column at a flow rate of 0.8 ml/min. Second, fractions exhibiting proteolytic activity were combined, diluted 1:1 (v/v) by addition of H₂O, and reseparated on a MiniS PE (4.6 x 250 mm, GE Healthcare Life Sciences) column equilibrated with 50 mM NH₄OAc, pH 5.0. Analytes were eluted by applying a gradient of 0–0.65 M NaCl in 21 min at a flow rate of 0.8 ml/min. Third, the eluted proteolytic active fraction was further separated by RP-HPLC on a Nucleosil 300–5, C4 (5 µm, 4.6 x 250 mm, Macherey-Nagel, Düren, Germany) column equilibrated with 0.05% (v/v) TFA in water. Proteins were eluted with 0.05% (v/v) TFA in acetonitrile applying a gradient of 0–60% acetonitrile in 42 min at a flow rate of 0.5 ml/min.

Processing quadruplet motif processing protease

Polyacrylamide gel electrophoresis

SDS-PAGE and silver staining was performed using a PhastGel System and PhastGel Gradient 8–25 (all GE Healthcare).

Amino acid analysis

Amino acid analysis was performed as previously reported (34).

Enzymatic characterization

We determined the effect of different cations, protease inhibitors, and pH on the activity of the isolated protease. First, 55 ng of the isolated protease was preincubated in the respective buffer in assay plates at 37 °C for 15 min. Second, the substrate (DABCYL-PFFENEQARKDDKN-Glu(EDANS)-NH₂) was added to a final concentration of 5 µM and the increase in fluorescence was immediately measured in intervals of 1 min during 15 min at 37 °C using a Synergy HT Microplate reader (BioTek) as described above. The following assay buffers were used: 0.2 M NH₄OAc, 0.2 M NaOAc, or 0.2 M KOAc, all pH 6.0, for monovalent cations; 0.2 M NH₄OAc, pH 6.0, and 5 mM of one of the following additives: CoCl₂, MnCl₂, ZnCl₂, NiCl₂, or CaCl₂ for divalent cations, 10 mM EDTA or AEBSF for inhibitors, or 0.2 M NH₄OAc, pH 3.5 to 8.5, by steps of 0.5 for pH. All assay buffers contained additionally 7.5% (v/v) acetonitrile. For each condition, a negative control (no protease) and a positive control (10 µg trypsin) were included. Trypsin was preincubulated with peptide prior to the addition of the respective buffer to ensure full digestion of the substrate at t = 0. The proteolytic activity was calculated as an increase in fluorescence per time relative to the respective negative and positive controls. All measurements were done in triplicate.

Substrate specificity

To investigate the function of the isolated protease in the maturation process of venom neurotoxins, we analyzed the digestion products of different substrates (all supplied by GeneCust, Luxembourg) after incubation with the isolated enzyme. Two substrates have been designed to mimic the propeptide–mature peptide cutting site of C. salei neurotoxins: PFENFENQRASKIP (CsTx-1) and PFLLAQVRKKDDKN (CsTx-9). Additionally, two substrates mimic potential processing sites within the mature chain region of C. salei venom neurotoxin precursors: WGLEWRNEEAERSPC (CsTx-40) and FPFLAREQVRKKDDKN (CsTx-9). SDS-PAGE and silver staining was performed using a PhastGel System and PhastGel Gradient 8–25 (all GE Healthcare).

Amino acid analysis was performed as previously reported (34).

Enzymatic characterization

We determined the effect of different cations, protease inhibitors, and pH on the activity of the isolated protease. First, 55 ng of the isolated protease was preincubated in the respective buffer in assay plates at 37 °C for 15 min. Second, the substrate (DABCYL-PFFENEQARKDDKN-Glu(EDANS)-NH₂) was added to a final concentration of 5 µM and the increase in fluorescence was immediately measured in intervals of 1 min during 15 min at 37 °C using a Synergy HT Microplate reader (BioTek) as described above. The following assay buffers were used: 0.2 M NH₄OAc, 0.2 M NaOAc, or 0.2 M KOAc, all pH 6.0, for monovalent cations; 0.2 M NH₄OAc, pH 6.0, and 5 mM of one of the following additives: CoCl₂, MnCl₂, ZnCl₂, NiCl₂, or CaCl₂ for divalent cations, 10 mM EDTA or AEBSF for inhibitors, or 0.2 M NH₄OAc, pH 3.5 to 8.5, by steps of 0.5 for pH. All assay buffers contained additionally 7.5% (v/v) acetonitrile. For each condition, a negative control (no protease) and a positive control (10 µg trypsin) were included. Trypsin was preincubulated with peptide prior to the addition of the respective buffer to ensure full digestion of the substrate at t = 0. The proteolytic activity was calculated as an increase in fluorescence per time relative to the respective negative and positive controls. All measurements were done in triplicate.

Substrate specificity

To investigate the function of the isolated protease in the maturation process of venom neurotoxins, we analyzed the digestion products of different substrates (all supplied by GeneCust, Luxembourg) after incubation with the isolated enzyme. Two substrates have been designed to mimic the propeptide–mature peptide cutting site of C. salei neurotoxins: PFENFENQRASKIP (CsTx-1) and PFLLAQVRKKDDKN (CsTx-9). Additionally, two substrates mimic potential processing sites within the mature chain region of C. salei venom neurotoxin precursors: WGLEWRNEEAERSPC (CsTx-40) and FPFLAREQVRKKDDKN (CsTx-9). SDS-PAGE and silver staining was performed using a PhastGel System and PhastGel Gradient 8–25 (all GE Healthcare).
Processing quadruplet motif processing protease

cetonitrile, 0.5% (v/v) formic acid in water, the analytes were eluted by applying a gradient of 2–80% acetonitrile, 0.5% (v/v) formic acid. The gradient length was 20 min, the flow rate during loading and isotropic wash was 0.15 ml/min, and during the gradient was 0.3 ml/min. The eluted sample compounds were identified by online mass spectrometry in full scan positive ion mode on a LTQ Orbitrap Velos mass spectrometer (ThermoFisher Scientific). The scan range was 110–2000 m/z, and the resolution was set to 100,000. Data were processed using Excalibur version 2.2. LC-MS data are available as Fig. S12–S18.

MALDI-TOF-MS

MALDI-TOF-MS was performed on an Autoflex III Smartbeam mass spectrometer (Bruker Daltonics). 1 μl of acidified sample was directly mixed with 1 μl of matrix solution (α-cyano-4-hydroxycinnamic acid, sinapic acid, or 2,6′-dihydroxyacetophenone) on a ground steel target plate and air dried. The spectra were acquired in linear positive mode between 10 and 100 kDa and analyzed using the Flex analysis software (Bruker Daltonics). Spectra are shown baseline subtracted.

Tandem mass spectrometry

For protein identification, dried protein fractions were reconstituted in Laemmli buffer containing 20 mM DTT and separated on a 12.5% SDS-PAGE. Protein bands were excised and prepared for mass spectrometry as described elsewhere (35). For enzymatic digests with proteases other than trypsin, the procedure was slightly modified. For chymotrypsin digestion, incubations were done at room temperature for 6 h. For proteinase-K digestion, no ProteaseMax (Promega) was used and the digestion was carried out at 37 °C for 30 min. Peptide extracts were analyzed by nano-LC-MS/MS as described elsewhere (36).

For the mass spectrometric characterization of the protease light–heavy chain dimer, 3 μg of HPLC purified and dried protein was reconstituted in 10 μl of 8 M urea, 50 mM Tris-HCl, pH 8.0, reduced with 10 mM DTT at 37 °C for 30 min, followed by alkylation with 50 mM iodoacetamide at 37 °C for 30 min in the dark. Excess iodoacetamide was quenched with DTT. The protein was diluted to a concentration of 0.05 μg/μl with 0.1% TFA and 2 μl of this solution was analyzed by nano-LC-MS/MS on an Ultimate3000 chromatograph connected to an Orbitrap Fusion Lumos mass spectrometer (ThermoFisher Scientific). Analytes were trapped on a Pepmap 100 Trap C18 column (300 μm × 5 mm, ThermoFisher Scientific) and separated by backflush over a C18 column (3 μm, 100 Å, 75 μm × 150 mm, Nikkyo Technos Co. Ltd., Japan) applying a gradient of 0 to 38% acetonitrile, 0.1% formic acid in 10 min at a flow rate of 300 nl min⁻¹. MS scans were recorded in data-dependent acquisition mode acquiring MS1 and MS2 spectra in the Orbitrap detector with a cycle time of 3 s and the following settings: scan range 400–2000 m/z; MS1 and MS2 resolution at 120,000 and 30,000; AGC of 400,000 and 50,000; and a maximum injection time of 50 ms for both, respectively; including 2+ and 3+ ions only without exclusion time and an isolation width of 1.6 m/z; HCD fragmentation with a relative collision energy of 30%. All fragment spectra peak list files in Mascot generic file format were created with ProteomeDiscoverer version 2.1 (ThermoFisher Scientific) keeping the six most intense fragment ions in a sliding window of 20 m/z.

Tandem mass spectrometry data analysis

LC-MS data were analyzed with Excalibur (ThermoFisher Scientific, version 2.2), fragment spectra were interpreted with SearchGUI (version 3.2.14) (37) and Peptide Shaker (version 1.16.5) (38). We used an in-house database (venomProteome fasta) containing: 1) all Araneae sequences in Swiss-Prot (as of 04/03/2017, n = 1490), 2) additionally identified peptides and proteins from our C. salei venom gland transcriptome, including the cDNA-derived amino acid sequences of the two isoforms of the PQM protease (n = 85), and 3) known contaminants (n = 218). The search engine was X! TANDEMD (version 2015.12.15.2) and the search parameters: maximum missed cleavages, 3 (for trypsin and chymotrypsin) and option “unspecific” (for proteinase K); precursor ions tolerance, 10 ppm; fragment ion tolerance, 0.01 Da; fixed modifications, carbamidomethylation of C; variable modifications, oxidation of M. The results were filtered by applying a false discovery rate of ≤1%. Data are available via ProteomeXchange with identifier PXD007126. Details of identified peptides and proteins are given in Tables S1, S2, and S9.

Elucidation of the disulfide bridge pattern

2.5 μg of the dried purified enzyme were reconstituted in 20 μl of 8 M urea, 50 mM Tris-HCl, pH 8.0, by strong vortexing for 30 s and sonication for 1 min in a water bath. The protein was digested by incubation for 6 h at 37 °C with 100 ng of trypsin after dilution of urea to 1.6 M with 20 mM Tris-HCl, 2 mM CaCl₂, pH 8.0. For the digestion with agarose-bound trypsin (Sigma), the enzyme was reconstituted in 20 μl of 50 mM Tris-HCl, pH 8.0, and incubated with immobilized enzyme for 22 h at 37 °C. After digestion, half of the samples were reduced by addition of DTT to a final concentration of 10 mM and incubation for 30 min at 50 °C. Reduced and native samples were acidified with TFA to a final concentration of 1% and 100 ng of digested enzyme were analyzed as described under “Tandem mass spectrometry” with a few changes. Only peptide ions with a charge state between 3 and 8 were subjected to HCD fragmentation and the three most intense fragments were subjected to another HCD fragmentation (MS3) using a normalized fragmentation energy of 35%. MS2 spectra were recorded in the orbitrap with resolution 15,000 and AGC of 50,000, whereas MS3 spectra were recorded in the ion trap with AGC of 10,000 and maximum injection times of 35 ms for both detectors, respectively. Additionally, the MS3 acquisition mode was repeated with a mixed MS2 fragmentation mode of ETD and HCD (ETHcD) (25%) for peptide ions of charge 4–8, whereas peptides with charge 2–3 were fragmented with HCD as above. Relevant spectra were annotated manually and are available as Fig. S5–S11.

Phylogenetic analysis

The amino acid sequence alignment was generated using Clustal Omega (39) and was manually reviewed (alignment...
available in Fig. S19). The best fitting substitution model was identified to be WAG+G+I by using MEGA 7.0.26 (40). The maximum likelihood analysis was conducted in RAxML 8.2.9 (41) using default parameters and 1000 random starting trees under the mentioned model. The clade support was obtained by rapid bootstrapping (42) executing 1000 rapid bootstrap inferences. The bootstrap support was drawn on the best-scoring tree. Calculations were performed on UBELEX, the HPC cluster at the University of Bern.

Database deposits

The nucleotide sequences of the two PQM protease isoforms of C. salei have been deposited in the GenBank™ database under accession numbers MF537412 and MF537413. Nucleotide sequences of homologous transcripts form other spiders are accessible under the following GenBank™ accession numbers: L. hispanica accession number MF537414, G. vultuosa accession number MF537415, A. rifus accession number MF537416, P. fera accession number MF537417, C. getazi accession number MF537418, P. striata accession number MF537419, V. fasciatus accession number MF537420, P. mirabilis accession number MF537421, E. atrica accession number MF537422, A. accentuata accession number MF537423, Z. spinimana accession number MF537424, H. radiata (Italy) accession number MF537425, L. praegrandis accession number MF537426, A. ferox accession number MF537427, A. marikovskyi accession number MF537428, H. radiata (Spain) accession number MF537429, and A. cuneata accession number MF537430. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (43) partner repository with the dataset identifier PXD007126.

Author contributions—N. L. conceived the study, designed and conducted most experiments, analyzed the results and transcriptomic data, and wrote the paper. L. K. N. conceived the study, analyzed the results and transcriptomic data, wrote the paper, bred and milked the spiders, and dissected the venom glands. D. K. contributed to transcriptomic data analysis. S. S. provided technical assistance for nano-LC-MS/MS, contributed to data analysis, and wrote the methods part for tandem mass spectrometry. W. N. conceived transcriptomic studies, and collected and identified spider species. All authors reviewed the results, corrected the manuscript, and approved its final version.

Acknowledgments—We are grateful to Benjamin Eggs, Christian Kropf, Miguel Richard, and André Leetz for help in spider sampling, Seraina Klopfstein for help with phylogenetic analyses, and Benoît Zuber for providing access to his fluorimeter.

References
Processing quadruplet motif processing protease


