Novel γ-Carboxyglutamic Acid-Containing Peptides from the Venom of Conus textile

Eva Czerwiec¹,², Dario E. Kalume³#, Peter Roepstorff³, Björn Hambe⁴, Bruce Furie¹,², Barbara C. Furie¹,² and Johan Stenflo¹,⁴.

¹ Marine Biological Laboratory, Woods Hole, Ma, 02543, USA
² Center for Hemostasis and Thrombosis Research, Beth Israel Deaconess Medical Center, and Harvard Medical School, Boston, MA 02215, USA
³ Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense University, Campusvej 55, DK-5230 M, Odense. Denmark
⁴Department of Clinical Chemistry, Lund University, University Hospital, Malmö, S-20502, Malmö, Sweden
#Present address: McKusick-Nathans Institute of Genetic Medicine and the Department of Biological Chemistry, Johns Hopkins University, School of Medicine, Baltimore, MD 21287, USA
*eTo whom correspondence should be sent.
Running title

Gla-containing peptides from the venom of *C. textile*

Corresponding author

Eva Czerwiec, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543, USA. Fax +1 508 540 6902.
e-mail: czerwiec@mbl.edu

1 Abbreviations

Gla, γ-carboxyglutamic acid; BrTrp: 6-L-bromotryptophan; Hyp: *trans*-4-hydroxyproline; γ-CRS, γ-carboxylation recognition site; CHAPS: 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate; PCR: polymerase chain reaction; RACE-PCR: rapid amplification of cDNA ends-PCR

2 Non-standard bases: M = A or C; I = deoxyinosine; R = A or G; S = C or G; W = A or T; Y = C or T

Key words

γ-carboxyglutamic acid, vitamin K, conotoxin, *Conus textile*, propeptide
Abstract

The cone snail is the only invertebrate system in which the vitamin K dependent carboxylase (or γ-carboxylase) and its product γ-carboxyglutamic acid (Gla)\(^1\) have been identified. It remains the sole source of structural information of invertebrate γ-carboxylase substrates. Four novel γ-carboxyglutamic acid (Gla)\(^1\) containing peptides were purified from the venom of *Conus textile* and characterized by biochemical methods and mass spectrometry. The peptides Gla(1)-TxVI, Gla(2)-TxVI/A, Gla(2)-TxVI/B and Gla(3)-TxVI each have 6 Cys residues and belong to the *O*-superfamily of conotoxins. All four conopeptides contain 4-*trans*-hydroxyproline and the unusual amino acid 6-L-bromotryptophan. Gla(2)-TxVI/A and Gla(2)-TxVI/B are isoforms with an amidated C-terminus that differ at positions +1 and +13. Three isoforms of Gla(3)-TxVI were observed that differ at position +7: Gla(3)-TxVI, Glu\(^7\)-Gla(3)-TxVI and Asp\(^7\)-Gla(3)-TxVI. The cDNAs encoding the precursors of the four peptides were cloned. The predicted signal sequences (amino acids –46 to –27) were nearly identical and highly hydrophobic. The predicted propeptide region (–20 to –1) that contains the γ-carboxylation recognition site (γ-CRS) is very similar in Gla(2)-TxVI/A, Gla(2)-TxVI/B and Gla(3)-TxVI, but is more divergent for Gla(1)-TxVI. Kinetic studies utilizing the *Conus* γ-carboxylase and synthetic peptide substrates localized the γ-CRS of Gla(1)-TxVI to the region –14 to –1 of the polypeptide precursor: the \(K_m\) was reduced from 1.8 mM for Gla (1)-TxVI lacking a propeptide to 24 µM when a 14-residue propeptide was attached to the substrate. Similarly, addition of an 18-residue propeptide to Gla(2)-TxVI/B reduced the \(K_m\) 10-fold.
Introduction

Venom from marine snails of the genus *Conus* contains a plethora of highly potent neurotoxins, many of which block voltage-gated and ligand-gated ion channels. The peptides are typically 12 to 30 amino acids in length and contain disulfide bonds and a wide variety of posttranslationally modified amino acids [1, 2]. Particularly abundant are 4-*trans*-hydroxyproline (Hyp), L-6-bromotryptophan and γ-carboxyglutamic acid (Gla) [3-6].

Gla is formed by γ-carboxylation of glutamyl residues, a reaction mediated by a vitamin K-dependent γ-glutamyl carboxylase located in the endoplasmic reticulum. The *Conus* carboxylase is a homologue of its vertebrate counterpart and is predicted to be an integral membrane protein with several transmembrane-spanning regions [7-10]. γ-Carboxylases from several vertebrates and the invertebrate *Conus textile* have been expressed and kinetically characterized [8, 11, 12].

The biosynthesis of Gla is a complex reaction that involves replacement of a proton on the γ-carbon of a Glu residue with a CO₂ molecule [13]. The γ-glutamyl carboxylase is the sole enzyme known to use vitamin K as a cofactor. Carboxylation of Glu in the nascent polypeptide chain requires the presence of a γ-carboxylation recognition site (γ-CRS) that typically resides within a 12- to 28-residue propeptide located immediately adjacent to the N-terminal signal peptide [7, 14-17]. The propeptide mediates binding of the substrate to the carboxylase and also activates the enzyme.

The discovery of γ-carboxylated conotoxins and, more recently, the cloning and characterization of the γ-carboxylase from cone snails and *Drosophila melanogaster* [14,19] has evoked fresh interest in the function of
vitamin K and the vitamin K-dependent carboxylase [8, 9, 18, 19]. New functions for the vitamin and Gla are anticipated; functions that may be phylogenetically older than blood coagulation and bone formation [19]. This has stimulated research aimed at identifying novel Gla-containing proteins and peptides from non-vertebrate sources. The only invertebrate peptides in which Gla has been identified to date and thus the only source of structural information of non-vertebrate carboxylase substrates are the conotoxins [6, 7, 16, 17, 19-24]. Comparison of the structure of vertebrate and invertebrate γ-carboxylase substrates provides information about possible alternate functions for this unique enzyme and the mechanistic properties of an ancestral carboxylation system.

In this paper we describe the purification and characterization of four novel Gla-containing conotoxins from C. textile. All of the peptides have six Cys residues, belong to the O-superfamiily of conotoxins and have uniquely spaced Glu residues in the mature peptide. The cDNAs encoding the predicted prepropeptide precursors were cloned and synthetic peptide substrates based on the precursor sequences were used as substrates in kinetic experiments that localize the γ-CRS in the propeptides.
Results

Sequence analysis and posttranslational modifications of Gla(2)-TxVI/A Gla(2)-TxVI/B and Gla(3)-TxVI. Peptides were purified by gel filtration and HPLC chromatography as described in Material and Methods. Edman degradation identified Gla at position 10 and hydroxyproline at position 12 in Gla(2)-TxVI/A and Gla(2)-TxVI/B 12 and showed that these peptides are isoforms that differ at positions 1 and 13 (Table S1). Amino acid sequence analysis of Gla(3)-TxVI yielded 26 residues and showed a microheterogeneity (Gla/Glu/Asp) at position 7 (Table S1). The UV spectrum of all peptides suggested the presence of a tryptophan residue but this residue was not identified during sequence analysis. The full sequence including posttranslational modifications of the peptides was obtained by additional mass spectrometry analysis (Table 1).

Positive ion linear mode MALDI mass spectra of native Gla(2)-TxVI/A and Gla(2)-TxVI/B showed main ion signals at \( m/z \) 2966.75 and 2979.70, respectively (Fig. S2). The discrepancy between the theoretical molecular masses (2836.81 for Gla(2)TxVI/A and 2849.81 Da for Gla(2)-TxVI/B) and the observed molecular masses can be explained by the presence of a bromotryptophan residue and an amidated C-terminus. These posttranslational modifications were confirmed by analysis of the respective fingerprints after enzymatic digestion. The isotopic distribution of the peak at \( m/z \) 901.18 indicates a bromine-containing peptide (Fig. 1A and B, inset). The peak at \( m/z \) 626.29 is consistent with amidation of the C-terminal fragment (DVVCS), as is the observed 14-Da mass increase (to \( m/z \) 640.31) following methyl-esterification of the fragment (Fig. S3). The presence of six cysteinyI residues was confirmed by observation of an average mass
increment of ~640.5 Da after pyridylethylation of the reduced peptides (Data not shown).

The MALDI mass spectrum of native Gla(3)-TxVI produced three main ion signals consistent with the presence of Gla, Glu and Asp at position +7 (Fig. 2A). The isotope distribution of the most intense peak obtained after enzymatic digestion and analysis by nano-ESI/MS corresponds to a bromine-containing peptide (Fig. 2B). In addition, the mass of this peptide is in agreement with the presence of 6-L-bromotryptophan in the C-terminal fragment (residues 17–27) (Fig. S4). MS as well as MS/MS of the C-terminal peptide showed that all three Gla(3)-TxVI isoforms have a free carboxyl group at the C-terminus.

Cloning of cDNAs encoding the Gla(1)-TxVI, Gla(2)-TxVI/B and Gla(3)-TxVI precursors. The isolated 580-bp cDNA encoding the Gla(1)-TxVI precursor includes the 5’ and 3’ untranslated regions and contains an open-reading-frame (ORF) of 228 bp. The ORF encodes the 30-residue mature peptide, which is preceded by a 46-amino acid prepropeptide that is absent in the secreted conotoxin (Fig. 3A). The cloned cDNA, though considerably longer, exactly matches a 342-bp conotoxin sequence deposited in GenBank (accession number AF215016.1).

We cloned cDNAs encoding the precursors to Gla(2)-TxVI/B and Gla-(3)-TxVI using 5’RACE- and 3’RACE-PCR with primers based on the 5’ and 3’ untranslated regions of Gla(1)-TxVI [25]. A 481-bp cDNA was obtained for Gla(2)-TxVI/B (Fig. 3C). It includes an ORF of 216 bp encoding a 72-residue precursor comprising the mature conotoxin and a 46-amino acid N-terminal prepropeptide. The precursor contains a C-terminal Gly residue, as would be expected for a peptide that undergoes
posttranslational α-amidation. We were unable to obtain a clone for the Gla(2)-TxVI/A isoform, but identified a 510-bp cDNA sequence in GenBank (accession number AF215024.1) that contains the ORF encoding prepro-Gla(2)-TxVI/A (Fig. 3B). Though we anticipated the possibility of isolating two cDNAs encoding the Gla(3)-TxVI isoforms we were only able to obtain a clone specifying Glu at position +7. The 520-bp cDNA contains an ORF encoding a 73-residue precursor comprising the 27-residue mature peptide and a 46-residue N-terminal prepropeptide (Fig. 3D). We also identified cDNA sequences in GenBank which encode the precursors to conotoxins that are nearly identical to the Glu7- and Asp7-containing isoforms of Gla(3)-TxVI (accession numbers AF215021.1 and AF215023.1). The amino acid sequences predicted from the cDNAs in GenBank differ from our sequence only at position –15, where we find Leu instead of Phe. This substitution probably would not lead to a major perturbation of the overall structure or properties of the precursor. Our results suggest that the mature conopeptides encoded by accession numbers AAG60449.1 and AAG60451.1 would also be γ-carboxylated.

In all cases, the deduced precursor sequences have a conserved hydrophobic N-terminal region that is predicted by the PSORTII algorithm to serve as a signal sequence [26]. The predicted cleavage site is located between residues 19 and 20 of the precursor forms. The remaining sequence that is located between the signal peptide and the mature peptide contains a region that bears a resemblance to the propeptide sequences of other Gla-containing peptides (see below).
The γ-carboxylation recognition site of Gla(1)-TxVI and Gla(2)-TxVI/B.
The predicted propeptide regions of the Gla(1)-TxVI, Gla(2)-TxVI/A, Gla(2)-TxVI/B and Gla(3)-TxVI precursors have features resembling propeptides from other conotoxins, which suggested they would positively modulate carboxylation of the mature peptide. We tested this hypothesis by performing γ-carboxylation experiments with peptide substrates that either lacked a propeptide or that contained at least part of the predicted propeptide (Table 2). A peptide comprising amino acids +1 to +18 of mature Gla(1)-TxVI (lacking any potential propeptide) was a poor substrate for the Conus γ-carboxylase, exhibiting a $K_m$ of around 1.8 mM. Addition of amino acids –8 to –1 (a strongly charged part of the precursor) decreased the $K_m$ about 3-fold, whereas addition of amino acids –14 to –1, which also included the mostly hydrophobic amino acids located between positions –14 and –8, decreased the $K_m$ 75-fold (to 24 µM). These results are similar to those obtained in our previous study with conotoxin ε-TxIX, where we found that the hydrophobic amino acids located in the propeptide region form an important structural element of the γ-carboxylation recognition site [16]. Similarly, a synthetic substrate based on amino acids +1 to +11 of mature Gla(2)-TxVI/B exhibited a $K_m$ of ~540 µM, whereas the $K_m$ was reduced ~10-fold by including amino acids –18 to –1 of the prepropeptide region (Table 3). Though in this case the decrease in $K_m$ was not as marked as that observed with the Gla(1)-TxVI substrates, it nevertheless clearly showed that the presence of a propeptide substantially enhances γ-carboxylation of the Gla(2)-TxVI/B substrate.
Discussion

The marine cone snail remains the sole invertebrate in which the vitamin K-dependent amino acid Gla has been identified. Although a homologue of the vitamin K-dependent carboxylase gene has been identified in another invertebrate and recently in a bacteria, no Gla-containing polypeptides have been isolated from these organisms [18, 27]. Thus, the Gla-containing conopeptides remain the only source of structural information for invertebrate γ-carboxylase substrates. Isolation of novel Gla-containing peptides and determination of the predicted precursor forms continues to provide information about structural features important for the γ-carboxylation system. The mechanistic properties of the invertebrate and vertebrate carboxylases are similar and the vertebrate and invertebrate carboxylase enzymes are able to carboxylate their respective substrates. However, while the bovine carboxylase does not efficiently carboxylate cone snail substrates, certain bovine substrates are carboxylated as efficiently by the cone snail enzyme as by the bovine enzyme [8, 16]

Our recent studies indicate that the cone snail enzyme may tolerate a greater degree of structural variability in its substrates than the bovine enzyme. Indeed, while the γ-carboxylation recognition site (γ-CRS) is located within an N-terminal propeptide in virtually all known substrates of the vertebrate γ-carboxylase, in cone snail substrates this recognition site can also be located in a C-terminal ‘postpeptide’ in the precursor [20]. Moreover, a rigorous consensus sequence for the cone snail γ-CRS has not yet been identified also suggesting less stringent amino acid sequence requirements for recognition by the cone snail carboxylase. In an effort to obtain more information on the structure of invertebrate carboxylase
substrates we purified four γ-carboxylated peptides from *Conus textile*, a species whose venom is particularly rich in Gla-containing peptides.

All four isolated conopeptides have 6 Cys residues arranged in the typical VI/VII scaffold and belong to the *O*-superfamily of conotoxins [28]. Gla(1)-TxVI and Gla(3)-TxVI contain a motif –γCCS– that is found in four other Gla-containing peptides, TxVIIA from *C. textile*, γ-PnVIIA from *C. pennaceus*, d7a from *C. delessertii* and as7a from *C. austini* [1, 21, 22, 29]. Conotoxins that contain this motif are grouped into a subfamily of the *O*-superfamily, designated as the γ-conotoxins. TxVIIA, γ-PnVIIA are both excitatory conotoxins that increase firing in mollusk neurons and it has been suggested that the presence of the γCCS motif is involved in their biological activity [1].

The predicted modular structure of the precursor forms of Gla(1)-TxVI, Gla(2)-TxVI/A, Gla(2)-TxVI/B and Gla(3)-TxVI is consistent with other γ-carboxylated conopeptides, where the mature peptide is preceded by a prepropeptide containing a highly conserved signal sequence (–46 to –27) and a more divergent propeptide (residues –20 to –1). The propeptide regions of the conotoxins reported here share structural and physicochemical properties with the pro- and postpeptides of other Gla-containing peptides from *Conus spp.* (Table 3). All four propeptides have a high Lys/Arg content and are strongly basic, as is typical for pro- and postpeptides of Gla-containing conotoxins [20]. In addition, the newly identified propeptides contain a putative consensus sequence found in the precursors of Gla-containing conotoxins but not in the precursors of non-carboxylated conotoxins (Table 3). This sequence involves one hydrophobic and two basic residues arranged in the motif Lys/Arg-X-X-J-X-X-X-
Lys/Arg, where $J$ is typically a hydrophobic amino acid and $X$ is any amino acid [20]. This consensus sequence is also found in the propeptide of the mammalian vitamin K-dependent proteins prothrombin and Factor IX (Table 3). Coincidentally, synthetic substrates based on the sequences of the precursor forms of prothrombin (proPT28) and Factor IX (proFIX28) are both low $K_m$ substrates for the cone snail carboxylase [8]. It is anticipated that additional structural parameters such as the $\alpha$-helicity of the propeptide and the position of certain residues relative to the $\alpha$-helix are likely to be important to confer substrate efficiency. In this context it is noteworthy that a charged amino acid is present close to the predicted $\alpha$-helical domain in several of the propeptides (Table 3). Unfortunetaly, lack of information on the three-dimensional structure of propeptide containing conotoxins has hampered identification of essential $\gamma$-carboxylase substrate features.

The presence of a vitamin K-dependent carboxylase and of $\gamma$-carboxyglutamic acid in phyla as disparate as Chordata and Mollusca suggests the existence of an ancestral carboxylation system with a purpose predating blood coagulation and bone formation. Because $\gamma$-carboxylation requires tight cellular control, carboxylase substrates must contain the structural information necessary for subcellular localization, substrate recognition and tight enzyme-substrate binding. The observation that cone snail propeptides do not contain sufficient structural information to drive efficient carboxylation by the mammalian system, yet certain mammalian propeptides contain sufficient structural information to drive carboxylation by the cone snail system suggests that the vitamin K dependent carboxylation has evolved towards a more tightly controlled process. Identification of overlapping structural elements between the vertebrate and
invertebrate substrates could identify the minimum requirements for an ancestral propeptide and this information could be used as a filter in the quest to identify novel Gla-containing proteins.
Materials and Methods

Materials. Live specimens of *C. textile* were obtained from Suva (Fiji) and frozen specimens of *C. textile* from Nha Trang (Vietnam). NaH$^{14}$CO$_3$ (55 mCi/mmol) was purchased from Amersham Life Sciences (Arlington Heights, IL), Sephadex G-50 Superfine and Superose 12 resins from Pharmacia (Piscataway, NJ), and Endoproteinase Asp-N and elastase from Boehringer Mannheim Biochemicals GmbH (Mannheim, Germany). 2,5-Dihydroxybenzoic acid was from Aldrich Chemical Company (Steinheim, Germany) and ammonia solution (25%) from Merck (Darmstadt, Germany). Ultra-pure Milli-Q water (Millipore, Bedford, MA, USA) was used in the preparation of all solutions for mass spectrometry. A marathon cDNA Amplification Kit, DNA polymerase, and PCR buffer were purchased from Clontech (Palo Alto, CA) and AmpliTaq Gold polymerase and buffer from Perkin Elmer (Branchburg, NJ). Primers were synthesized by Gibco BRL Life Technologies (Gaithersburg, MD). Qiaquick Gel Extraction Kits were obtained from Qiagen (Santa Clarita, CA) and a TA Cloning Kit and Micro Fasttrack kit from Invitrogen (Carlsbad, CA). Atomlight scintillation fluid was from Packard (Meriden, CT), vitamin K from Abbott Laboratories (North Chicago, IL), and DL-Dithiothreitol (DTT), FLEEL, L-phosphatidylcholine (type V-E) and CHAPS$^1$ from Sigma (St. Louis, MO). Spectra/Por dialysis tubing (6 Membrane MWCO 1000) was obtained from Spectrum Laboratories Inc. (Rancho Dominguez, CA). All other chemicals were of the highest grade commercially available.

Purification of Gla(1)-TxVI, Gla(2)-TxVIA, Gla(2)-TxVIB and Gla(3)-TxVI. Venom was extruded from the venom duct, taken up in water and
lyophilized. Lyophilized venom (200 mg from five snails) was extracted in 0.2 M ammonium acetate buffer, pH 7.5, and chromatographed on a Sephadex G-50 Superfine column (2.5 x 92 cm) as described previously [30, 31]. The $A_{280}$ and Gla content of column fractions were monitored (Fig. S1A). Purification and characterization of the Gla-containing material in peak 10 (i.e. Gla(1)-TxVI) was performed as described previously [32]. The material in the Gla-containing peaks in pools 12 (Gla(2)-TxVI/A), 13 (Gla(2)-TxVI/B), and 14 (Gla(3)-TxVI) was further purified by reversed-phase HPLC in 0.1% TFA on a HyChrom C18 column (Fig. S1B and C)(5 µm; 10 x 250 mm), elution being achieved with a linear gradient of acetonitrile (0–80%) at a flow rate of 2 mL/min. Peptide Gla(3)-TxVI was essentially homogenous after gel filtration and gave a single major peak during reversed-phase HPLC (data not shown).

**Amino acid analysis and sequencing.** Amino acid compositions were determined after acid hydrolysis, except for Gla, which was determined after alkaline hydrolysis as described [23,24]. Peptide sequencing was performed using a Perkin Elmer ABI Procise 494 sequencer (Foster City, CA). Gla was identified after methyl-esterification as described [33, 34].

**Mass spectrometry.** MALDI-TOF MS and Nano ESI-MS was performed on the same instruments and in the same conditions as described for Gla(1)-TxVI [32].

**Cloning of Gla(1)-TxVI, Gla(2)-TxVIB and Gla(3)-TxVI.** PCR was performed using the degenerate oligonucleotides DGR1 (5’-GGMATGTGGGGIGARTGYAAR-3’)2 based on amino acid residues 1–7
of Gla (1)-TxVI, and DGR2 (5’-CCACATCGTRSAISWGCCYTCRSA-3’) based on amino acid residues 23–31 of Gla (1)-TxVI. A C. textile Lambda ZAP II library was used as the template [16]. Sequence information obtained from the degenerate PCR experiment was used to design the gene-specific primers GSP1 (5’-CTCTGAGGGCGCCAAACATGTG-3’) and GSP2 (5’-CGACATGTTGGCAGCCTCAGAG-3’) in 5’ RACE and 3’ RACE PCR reactions that employed a C. textile RACE library as the template. Amplification parameters were as indicated by the manufacturer. The cDNAs encoding Gla(2)-TxVI/B and Gla(3)-TxVI were obtained by RACE-PCR using oligonucleotides complementary to the conserved 5’ untranslated (5’UNT) (5’-CTCTTGAAGCCTCTGAAGAGGAGAGTGG-3’) and 3’ untranslated (3’UNT) (5’-CTCCCTGACAGCTGCCTTCAGTCGACC-3’) regions of Gla(1)-TxVI

**Enzyme assays.** The amount of [¹⁴C]O₂ incorporated into exogenous peptide substrates was measured in reaction mixtures of 125 µL containing 222 µM reduced vitamin K, 0.72 mM NaH[¹⁴C]O₃ (5 mCi), 28 mM MOPS (pH 7.0), 500 mM NaCl, 0.16% (w/v) phosphatidylcholine, 0.16% (w/v) CHAPS, 0.8 M ammonium sulfate, 10 µL microsomal preparation and peptide substrate. Microsomal preparations of Sf21 insect cells expressing the cone snail γ-glutamyl carboxylase were prepared as described previously [8]. All of the assay components except carboxylase were prepared as a master mixture. The reaction was initiated by adding the enzyme to the assay mixtures. The amount of [¹⁴C]O₂ incorporated into the peptides over a period of 30 min was assayed in a scintillation counter [35]. Peptides were synthesized using
standard FMOC/NMP chemistry on an Applied Biosystems Model 430A peptide synthesizer [36]
Acknowledgments

This work was supported by grants K2001-03X-04487-27A and K2001-03GX-04487-27, 08647, 13147 from the Swedish Medical Research Council, the European Union Cono-Euro-Pain (QLK3-CT-2000-00204), the Swedish Foundation for Strategic Research, the Kock Foundation, the Påhlsson Foundation and the Foundation of University Hospital, Malmö. Work performed at the Marine Biological Laboratory was supported by the National Institutes of Health. We also thank Ingrid Dahlqvist for performing sequence and amino acid analyses and peptide synthesis and Margaret Jacobs for peptide synthesis.
References


**Table 1:** Amino acid sequences of conopeptides Gla(1)-TxVI, Gla(2)-TxVI/A, Gla(2)-TxVI/B and Gla(3)-TxVI* obtained by combined Edman degradation and mass spectrometry analysis. Posttranslational modifications are highlighted in bold. **W:** 6-L-Bromotryptophan, **γ:** γ-carboxyglutamic acid, **O:** 4-trans-hydroxyproline, **#:** amidated C-terminus.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gla(1)-TxVI</td>
<td>GMWGγCKDGLTTCLAOγCCSγDCγGSCTMWM</td>
</tr>
<tr>
<td>Gla(2)-TxVI/A</td>
<td>SCSDDWQYCγSOTDCCSWDCDVVCS#</td>
</tr>
<tr>
<td>Gla(2)-TxVI/B</td>
<td>NCSDDWQYCγSOSDCCSWDCDVVCS#</td>
</tr>
<tr>
<td>Gla(3)-TxVI*</td>
<td>LCODYTγOCSAHAYCCSWNCYNGHCTG</td>
</tr>
<tr>
<td>Glu7/Gla(3)-TxVI*</td>
<td>LCODYTγOCSAHAYCCSWNCYNGHCTG</td>
</tr>
<tr>
<td>Asp7/Gla(3)-TxVI*</td>
<td>LCODYTγOCSAHAYCCSWNCYNGHCTG</td>
</tr>
</tbody>
</table>

*Position 7 in Gla(3)TxVI displays a microheterogeneity with Gla, Glu and Asp occurring in a ratio of 1:1:2, respectively (see also Table S1)
Table 2: Kinetic parameters of synthetic substrates based upon the sequences of Gla(1)-TxVI and Gla(2)-TxVI/B and their predicted precursors.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>K_m (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gla(1)-TxVI/18</td>
<td>GMWGECKDGLTTCLAPSE</td>
<td>1800 ± 300</td>
</tr>
<tr>
<td>pro-Gla(1)-TxVI/26</td>
<td>KRKRAADDRGMWGECKDGLTTCLAPSE</td>
<td>550 ± 30</td>
</tr>
<tr>
<td>pro-Gla(1)-TxVI/32</td>
<td>NINFLLKRRKRAADDRGMWGECKDGLTTCLAPSE</td>
<td>24 ± 2</td>
</tr>
<tr>
<td>Gla(2)-TxVI/B/11</td>
<td>NCSDDWQYCES</td>
<td>540 ± 20</td>
</tr>
<tr>
<td>pro-Gla(2)-TxVI/B/29</td>
<td>KIDFLSKGKADAEKQRKRNCSDDWQYCES</td>
<td>51 ± 5</td>
</tr>
</tbody>
</table>

*a* The propeptide sequence is shaded. *b* K_m values were calculated by the Lineweaver-Burke method and are given as the mean ± 1 S.D.
Table 3: Comparison of propeptide and postpeptide amino acid sequence

<table>
<thead>
<tr>
<th>Conotoxin</th>
<th>Amino acid sequence</th>
<th>Position</th>
<th>pI</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gla(1)-TxVI</td>
<td>HSKENINFLKRRRAKAD-R</td>
<td>-1/-20</td>
<td>11.64</td>
<td>-</td>
</tr>
<tr>
<td>Gla(2)-TxVI/A</td>
<td>KKIDFLSKKTRDAEKQQR</td>
<td>-1/-20</td>
<td>10.69</td>
<td>-</td>
</tr>
<tr>
<td>Gla(2)-TxVI/B</td>
<td>KKIDFLSKKTRDAEKQQR</td>
<td>-1/-20</td>
<td>11.07</td>
<td>-</td>
</tr>
<tr>
<td>Gla(3)-TxVI</td>
<td>EKIKLLKKRTDAEKQQR</td>
<td>-1/-20</td>
<td>11.07</td>
<td>-</td>
</tr>
<tr>
<td>Gla-TxX</td>
<td>CRRLIMGR</td>
<td>+48/+57</td>
<td>12.81</td>
<td>[20]</td>
</tr>
<tr>
<td>Gla-TxXI</td>
<td>GRKAKLLEFFRQR</td>
<td>+32/+44</td>
<td>12.24</td>
<td>[20]</td>
</tr>
<tr>
<td>k-BtX</td>
<td>GRKSLQEFQR</td>
<td>+32/+44</td>
<td>12.24</td>
<td>[37]</td>
</tr>
<tr>
<td>PnVIIA</td>
<td>QQAKINFLSKRPSAEWR</td>
<td>-1/-20</td>
<td>12.52</td>
<td>[22]</td>
</tr>
<tr>
<td>TxVIIA</td>
<td>RKAEIFSETKRKNKQR</td>
<td>-1/-20</td>
<td>12.12</td>
<td>[21]</td>
</tr>
<tr>
<td>Tx9.1</td>
<td>DNRNLOKMPFLVLYMQQR</td>
<td>-1-20</td>
<td>12.11</td>
<td>[17]</td>
</tr>
<tr>
<td>Con-G</td>
<td>GKRDLTQKRILRKCNKA-R</td>
<td>-1/-20</td>
<td>12.53</td>
<td>[7]</td>
</tr>
<tr>
<td>Con-L</td>
<td>GNRDLTQKRILRKCNKA-R</td>
<td>-1/-20</td>
<td>12.53</td>
<td>[38]</td>
</tr>
<tr>
<td>Con-R</td>
<td>GNRDLTQKRILRKCNKA-R</td>
<td>-1/-20</td>
<td>12.53</td>
<td>[38]</td>
</tr>
<tr>
<td>Glacon-M</td>
<td>GDNFGRARKRRMKVL</td>
<td>-1/16</td>
<td>12.69</td>
<td>[23]</td>
</tr>
<tr>
<td>Mr5.2</td>
<td>PLASHVHKRTLQIL-RDKR</td>
<td>-1/-20</td>
<td>12.24</td>
<td>[39]</td>
</tr>
<tr>
<td>Mr5.3</td>
<td>PLASHVHKRTLQIL-RNK</td>
<td>-1/-20</td>
<td>12.81</td>
<td>[39]</td>
</tr>
<tr>
<td>γ-TxIX</td>
<td>PLSSLNLKRTITLNR</td>
<td>-1/-19</td>
<td>12.68</td>
<td>[16, 31]</td>
</tr>
<tr>
<td>Human II</td>
<td>HVFLAPQQAPSLORVVR</td>
<td>-1/-18</td>
<td>12.98</td>
<td>[35]</td>
</tr>
<tr>
<td>Human FIX</td>
<td>TVFLEDHENAK—TRNPKR</td>
<td>-1/-18</td>
<td>10.76</td>
<td>[40]</td>
</tr>
</tbody>
</table>

*a* amino acids forming the consensus sequence are boxed and their positions highlighted by an asterisk. Basic amino acids are shown bold. Shaded residues are those predicted to form an γ-helix by the program Nnpredict (http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html). The γ-CRS identified in propeptides of human prothrombin (factor II) and human factor IX is underlined.
Legends to the Figures

**Figure 1. Posttranslational modification of Gla(2)-TxVI/A and Gla(2)-TxVI/B.** Positive ion reflector mode MALDI mass spectra of an endoproteinase Asp-N digest of (A) pyridylethylated Gla(2)-TxVI/A and (B) Gla(2)-TxVI/B. The characteristic monoisotopic distribution of the peaks at \( m/z \) 901.18 and 901.21 (insets) suggests a bromotryptophan-containing peptide. Peptide alkali (Na\(^+\) and K\(^+\)) adducts are labeled with asterisks.

**Figure 2. Posttranslational modification of Gla(3)-TxVI.** (A) Positive ion linear mode MALDI mass spectrum of native conotoxin Gla(3)-TxVI. The three high-intensity peaks at \( m/z \) 3167.5, 3180.6 and 3225.0 correspond to three isoforms containing Asp, Glu and Gla, respectively. (B) Nano-ESI mass spectrum of an elastase digest of the reduced Gla(3)-TxVI peptide. The distinctive monoisotopic distribution (inset) of the C-terminal peptide (\( m/z \) 660.18) reveals it is a bromotryptophan-containing peptide. The doubly charged ions at \( m/z \) 935.32, 942.33 and 964.33 correspond to the N-terminal peptides of the three conotoxin isoforms containing Asp, Glu and Gla at position 7, respectively.

**Figure 3. The cDNA and deduced amino acid sequences of the precursors of (A) Gla(1)-TxVI, (B) Gla(2)-TxVI/A, (C) Gla(2)-TxVI/B and (D) Gla(3)-TxVI.** The open-reading-frames of the cDNA sequences are shown in uppercase typeface and untranslated regions in lowercase. The amino acid sequences of the mature conotoxins, as determined by Edman degradation and mass spectrometry, are shown.
in bold typeface and Glu residues that are posttranslationally modified to Gla are shown in parentheses. The signal peptide is underlined and the propeptide that contains the γ-CRS is shaded. * Sequence retrieved from GenBank (accession number AF215024.1) # amidated C-terminus.
Figure 1
Figure 2

![Image of figure 2 showing mass spectrometry data and relative abundance percentages.]
**Figure 3**

**A**  
Gla(1)-TxVI

```
agtcatctctctcagtctcctgtcagctgccttcagtcgaccctgccgtcatctcagcgcagacttggtaagaag
T Q A L V E R A G E N H S K E N I N F L
```

**B**  
Gla(2)-TxVI/A*

```
gtcatctctctcagtctcctgtcagctgccttcagtcgaccctgccgtcatctcagcgcagacttggtaagaag
T Q A L V E R A G E N H S K E N I N F L
```

---

**A**  
Gla(1)-TxVI

```
agtcatctctctcagtctcctgtcagctgccttcagtcgaccctgccgtcatctcagcgcagacttggtaagaag
```

**B**  
Gla(2)-TxVI/A*

```
gtcatctctctcagtctcctgtcagctgccttcagtcgaccctgccgtcatctcagcgcagacttggtaagaag
```

---

**A**  
Gla(1)-TxVI

```
agtcatctctctcagtctcctgtcagctgccttcagtcgaccctgccgtcatctcagcgcagacttggtaagaag
```

**B**  
Gla(2)-TxVI/A*

```
gtcatctctctcagtctcctgtcagctgccttcagtcgaccctgccgtcatctcagcgcagacttggtaagaag
```
### Table S1: Edman degradation of Gla(2)-TxVI/A, Gla(2)-TxVI/B and Gla(3)-TxVI

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Assigned residue</th>
<th>Yield (pmol)</th>
<th>Assigned residue</th>
<th>Yield (pmol)</th>
<th>Assigned residue</th>
<th>Yield (pmol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ser</td>
<td>25</td>
<td>Asn</td>
<td>25</td>
<td>Leu</td>
<td>530</td>
</tr>
<tr>
<td>2</td>
<td>Cys</td>
<td>-</td>
<td>Cys</td>
<td>-</td>
<td>Cys</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Ser</td>
<td>20</td>
<td>Ser</td>
<td>23</td>
<td>hPro</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>Asp</td>
<td>15</td>
<td>Asp</td>
<td>23</td>
<td>Asp</td>
<td>389</td>
</tr>
<tr>
<td>5</td>
<td>Asp</td>
<td>21</td>
<td>Asp</td>
<td>22</td>
<td>Tyr</td>
<td>510</td>
</tr>
<tr>
<td>6</td>
<td>Trp</td>
<td>5</td>
<td>Trp</td>
<td>5</td>
<td>Tyr</td>
<td>510</td>
</tr>
<tr>
<td>7</td>
<td>Gln</td>
<td>9</td>
<td>Gln</td>
<td>15</td>
<td>Asp##</td>
<td>113</td>
</tr>
<tr>
<td>8</td>
<td>Tyr</td>
<td>8</td>
<td>Tyr</td>
<td>8</td>
<td>hPro</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>Cys</td>
<td>-</td>
<td>Cys</td>
<td>-</td>
<td>Cys</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>Gla</td>
<td>-</td>
<td>Gla</td>
<td>-</td>
<td>Ser</td>
<td>145</td>
</tr>
<tr>
<td>11</td>
<td>Ser</td>
<td>4</td>
<td>Ser</td>
<td>8</td>
<td>His</td>
<td>165</td>
</tr>
<tr>
<td>12</td>
<td>hPro</td>
<td>-</td>
<td>hPro</td>
<td>-</td>
<td>Ala</td>
<td>213</td>
</tr>
<tr>
<td>13</td>
<td>Thr</td>
<td>3</td>
<td>Ser</td>
<td>8</td>
<td>His</td>
<td>128</td>
</tr>
<tr>
<td>14</td>
<td>Asp</td>
<td>5</td>
<td>Asp</td>
<td>7</td>
<td>Gla</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>Cys</td>
<td>-</td>
<td>Cys</td>
<td>-</td>
<td>Cys</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>Cys</td>
<td>-</td>
<td>Cys</td>
<td>-</td>
<td>Cys</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>Ser</td>
<td>3</td>
<td>Ser</td>
<td>4</td>
<td>Ser</td>
<td>65</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>Asp</td>
<td>4</td>
<td>Asp</td>
<td>4</td>
<td>Asn</td>
<td>105</td>
</tr>
<tr>
<td>20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>Asp</td>
<td>5</td>
<td>Asp</td>
<td>3</td>
<td>Tyr</td>
<td>99</td>
</tr>
<tr>
<td>22</td>
<td>Val</td>
<td>2</td>
<td>Val</td>
<td>2</td>
<td>Asn</td>
<td>81</td>
</tr>
<tr>
<td>23</td>
<td>Val</td>
<td>2</td>
<td>Gly</td>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Cys</td>
<td>-</td>
<td>His</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Thr</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# Reduced and alkylated peptides were analyzed. Gla was identified in separate runs after methylesterification (25).

## At this position Asp, Glu and Gla were found in ≈50%, 25% and 25% of relative abundance, respectively
Legends to Supplementary Figures

**Figure S1. Purification of conotoxins.** (A) Venom from *C. textile* was chromatographed on a Sephadex G-50 Superfine column. Gla(1)-TxVI was eluted in fraction pool 10 (P10), Gla(2)-TxVI/A in pool 12, Gla(2)-TxVI/B in pool 13 and Gla(3)-TxVI in pool 14. The vertical arrow denotes one column volume. (—) Absorbance at 280 nM; (–o–) Gla content. (B) Isolation of Gla(2)-TxVI/A (peak indicated by arrow) by reversed-phase HPLC on a C18 column (C) Isolation of Gla(2)-TxVI (peak indicated by arrow) on the same column.

**Figure S2. Positive ion reflector mode MALDI mass spectrum of Gla(2)-TxVI/A and Gla(2)-TxVI/B.** The observed monoisotopic molecular masses of (A) Gla(2)-TxVI/A (2966.75 Da) and (B) Gla(2)-TxVI/B (2979.70 Da) differ from the theoretical molecular masses - 2836.81 for Gla(2)-TxVI/A and 2849.81 for Gla(2)-TxVI/B. The discrepancy can be explained by the presence of a L-6-bromotryptophan and an amidated C terminus. Partial decarboxylation of the Gla residue present in both conotoxins is observed.

**Figure S3. Possttransalational modification of Gla(2)-TxVI/A: confirmation of C-terminal amidation.** After methyl-esterification of Gla(2)-TxVI/A, the C-terminal peptide (peak at m/z 626.3) exhibits a 14-Da mass increase consistent with methylation of the side chain carboxyl group of the N-terminal Asp residue confirming amidation of the C-terminus. Partial methylation of the internal peptide (residues 4–13) is observed.
Figure S4 Posttransalational modification Gla(3)-TxVI: confirmation of the presence of 6-L-bromotryptophan. Product ion mass spectrum of the doubly charged ion at m/z 660.18. The isotopic distribution of the b₂ ion (inset) indicates the presence of bromine. The MS/MS spectrum allows assignment of the sequence SW*NNCYNGHCTG, where W* is the bromotryptophan residue.
Figure S1
Figure S2

A

Intensity (arbitrary units)

MH\(^+\)
2966.75

-CO

m/z

2400 2800 3200 3600 4000

B

Intensity (arbitrary units)

MH\(^+\)
2979.70

-CO

m/z

2600 2800 3000 3200 3400 3600 3800
Figure S3