Solution structure of Cn5, a crustacean toxin found in the venom of the scorpions Centruroides noxius and Centruroides suffusus suffusus

ARTICLE *in* BIOCHIMICA ET BIOPHYSICA ACTA · AUGUST 2009

Impact Factor: 4.66 · DOI: 10.1016/j.bbapap.2009.07.006 · Source: PubMed

Biochimica et Biophysica Acta xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Rapid Report

Solution structure of Cn5, a crustacean toxin found in the venom of the scorpions *Centruroides noxius* and *Centruroides suffusus suffusus*☆,☆☆

Gerardo Corzo [a], Ada Prochnicka-Chalufour [b], Blanca I. García [a], Lourival D. Possani [a], Muriel Delepierre [b,*]

[a] Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, UNAM, Apartado Postal 510-3, Cuernavaca Morelos, 62210, Mexico

[b] Institut Pasteur, Unité de RMN des Biomolécules, CNRS URA 2185, 25 rue du Dr. Roux, 75724, PARIS Cedex 15, France

ARTICLE INFO

Article history:
Received 29 April 2009
Received in revised form 30 June 2009
Accepted 13 July 2009
Available online xxxx

Keywords:
Amino acid sequence
Centruroides noxius
Centruroides suffusus suffusus
Disulfide bridge
NMR three-dimensional structure
Scorpion toxin

ABSTRACT

The crustacean toxin Cn5 from *Centruroides noxius Hoffmann* and peptide Css39.8 from *Centruroides suffusus suffusus* scorpion venoms are identical peptides, as confirmed by amino acid sequence of purified toxins and by DNA sequencing of the two respective cloned genes. Therefore in this communication they will be simply named Cn5. Cn5 is a 66 amino acid long peptide with four disulfide bridges, formed between pairs of cysteines: C1–C8, C2–C5, C3–C6, and C4–C7 (the numbers indicate the relative positions of the cysteine residues in the primary structure). This peptide is non-toxic to mammals but deadly to arthropods (LD_{50} 28.5 mg/g body weight of crayfish). Its three-dimensional structure was determined by NMR using a total of 965 meaningful distance constraints derived from the volume integration of the 2D NOESY spectra. The Cn5 structure displays a mixed α/β fold stabilized by four disulfide bridges, with a kink induced by a *cis*-proline in its C-terminal part. Cn5 electrostatic surface is compared to that of Cn2 toxin toxic to mammals. The local differences produced by additional or substituted residues that would influence toxin selectivity towards mammalian or crustacean Na^+ channels are discussed.

1. Introduction

Scorpion toxins have been shown to affect ion permeability of excitable cells. At the present time, the best-studied scorpion toxins are those that recognize Na^+- and K^+-channels. Despite the fact that their primary structures can be quite different, there is a constant structural motif conserved among these families of proteins. All known Na^+-channel specific toxins are composed of 59–70 amino acid residues, stabilized, with few exceptions, by four disulfide bridges whereas those active on the K^+-channels are about 23–39 amino acid residues long and stabilized by three or four disulfide bridges (reviews in: [1,2]). Scorpion toxins have a highly conserved scaffold formed by an α-helix and a three stranded β-sheet structures. For the Na^+-channel specific toxins, two distinct protruding regions: B and J loops [3,4] are among some of the important features that differentiate between the mammalian alpha-scorpion toxins, which bind to the site 3 of the Na^+-channels, and the beta-scorpion toxins that bind to the site 4 [5,6]. Similarly, differences in the C-terminal part of the Na^+-channel specific toxins are thought to be at the origin of the remarkable species specificity of these toxins, capable of recognizing either mammalian or insect tissues [7]. The insect toxin I, from *Androctonus australis* Hector scorpion venom, shows a distinct arrangement of two non conserved disulfide bridges, situated out of the structural scaffold [8]. A proline rich segment at the C-terminal part of the insect specific toxins has been proposed to be responsible for the selectivity of these peptides towards distinct tissues [9]. From this short overview of the literature it is evident that more information is needed on the three-dimensional structure of these peptides, as well as on the charge distribution, and the spatial arrangement of certain residues capable of interacting specifically with a complementary surface on the receptor molecule. It is assumed that particular features of the 3D-structure of these peptides should provide them with the capacity of distinguishing between different classes of ion channels, or even more importantly, confer them a special property for discriminating sub-types of ion channels [10], with a remarkable variation in their affinities. Since our research group is interested in studying the structure–function relationships of scorpion toxins [11–13], here we report the solution structure of a crustacean specific toxin Css39.8 from the scorpion *Centruroides suffusus suffusus* (abbreviated *C. s. suffusus*), whose primary structure was shown to be identical to that of Cn5, purified originally from the venom of the Mexican scorpion *Centruroides noxius*. It contains 66 amino acid residues and its

Abbreviations: TFA, trifluoroacetic acid; ESI, electrospray ionization; LD_{50}, lethal dosis 50%; *C. s. suffusus*, *Centruroides suffusus suffusus*

☆ The protein and nucleotide sequence reported in this paper for Css39.8 has been submitted to the UniProt Knowledgebase and to the EMBL nucleotide database under the accession numbers P45663 and AM981271, respectively.

☆☆ The atomic coordinates of the family of ten conformers of Cn5 have been deposited in the Protein Data Bank as entry 2KJA and the proton chemical shifts have been deposited in the BioMagResBank under the accession number BMRB101189.

* Corresponding author.
E-mail address: murield@pasteur.fr (M. Delepierre).

1570-9639/$ – see front matter © 2009 Published by Elsevier B.V.
doi:10.1016/j.bbapap.2009.07.006

structure is stabilized by four disulfide bridges [14]. The present communication discusses several aspects of Css39.8 and/or Cn5 (here simply called Cn5) from the genus *Centruroides*: 1) the isolation, primary structure and disulfide bridge determination, 2) two sequences of cDNAs obtained from mRNAs of the two scorpions under study, assumed to correspond to Cn5, and 3) the ^{1}H-NMR three-dimensional solution structure, which permitted inspection of structural features of this crustacean toxin in comparison with those of mammalian specific toxins and of another crustacean specific toxin purified by our group from the venom of the scorpion *Centruroides limpidus limpidus* [12].

2. Materials and methods

2.1. Separation procedures

Since Cn5 from *C. noxius* was originally described by García et al. [14], this article will report only the procedure used for purification of Css39.8. The crude venom (10 mg) of *C. s. suffusus* was re-suspended in 0.1% aqueous TFA, and the insoluble material was removed by centrifugation at 14,000 *g* for 5 min. The supernatant was used directly for high performance liquid chromatography (HPLC) separation, using a Waters system (NJ, USA). The diluted venom was fractionated using a reverse-phase semipreparative C_{18} column ($5C_{18}$MS, 10×250 mm Nacalai Tesque Japan) equilibrated in 0.1% TFA, and eluted with a linear gradient of acetonitrile from 0 to 60% in 0.1% TFA, run for 60 min at a flow rate of 2 ml/min. Effluent absorbance was monitored at 280 nm. Fractions were collected and dried on a Speed Vac, model SC110 apparatus (Savant Instruments, Inc. Farmingdale, NY, USA). The fractions containing peptides with molecular masses similar to other known scorpion toxins were further purified by cation-exchange chromatography on a TSK-gel sulfopropyl column (SP-5PW, 4.6×75 mm, Tosoh, Japan), as originally described for other venoms [15]. The fractions were diluted to 200 µl with 20 mM ammonium acetate in 1 M acetic acid pH 2.9 and were separated using a linear gradient of 1 M ammonium acetate in 1 M acetic acid pH 5.9, in 50 min (1 ml/min). Absorbance was monitored at 280 nm. A final purification step was performed using a C_{18} reverse-phase column (4.6×250 mm, Nacalai Tesque, Japan) equilibrated in 0.1% TFA, and eluted with a linear gradient of acetonitrile from 20 to 60% in 0.1% TFA, run for 60 min at a flow rate of 1 ml/min. Effluent absorbance was monitored at 280 nm.

2.2. Bio-assays

Lethality tests were performed in three distinct species of animals, following approved protocols by the Animal Welfare Commission of the Institute of Biotechnology. Mice (strain CD1) were used to test chromatographic fractions. Bio-assays with insects and crustaceans were performed using crickets (*Achaeta* spp.) and a freshwater crayfish (*Cambarellus montezumae* spp.), respectively, as previously described [12]. Cn5 is toxic to crustaceans and insects, but it is not toxic to mice at doses up to 570 µg/mouse of 20 g body weight when injected intraperitoneally. Since the intoxication symptoms observed with Css39.8 are stronger for crayfish, its LD_{50} was determined. Three groups of 6 individuals were injected with different quantities of native toxin Css39.8 ranging from 25 µg to 35 µg. A value of 28.5 µg/g body weight of crayfish was estimated from the dose response-curve. The minimum amount of peptide required to kill crickets is more than 150 µg of Css39.8 per g body weight of insect.

2.3. Primary structure determination of Css39.8

The amino acid sequence of toxin Css39.8 was obtained by direct sequencing samples of native toxin, reduced and alkylated toxin (RC-toxin) and fragments generated by enzymatic cleavage of RC-toxin, using techniques already described by our group elsewhere [15,16]. The corresponding peptides were purified by HPLC and sequenced, using a Beckman LF3000 Sequencer (Palo Alto, CA, USA). When needed, amino acid analysis was performed on the fragments or on whole toxin. Mass spectrometry analysis was conducted using a LCQ^{Duo} spectrometer from Finnigan (San Jose, CA, USA), as described by Olamendi-Portugal et al. [16].

2.4. cDNA library construction and gene cloning

The clone (CngtII) coding for Cn5 was isolated from a cDNA library constructed from venomous glands of *C. noxius* Hoffmann scorpions, as described [17]. Briefly, to determine the nucleotidic sequence of the DNA coding for Cn5, a pair of oligonucleotides were synthesized: one (5′-ATGAAAGAAGGTTATCTGGTAAAC-3′) corresponding to the most N-terminal region of the mature peptide designed from the sequence encoding the first seven amino acids of toxin Cn5 (corresponding to bases at positions 1–21), but containing an additional methionine codon (ATG; not present in CngtII sequence) preceding the first residue encoded by the edited cDNA sequence, and the second oligonucleotide was designed according to the last seven residues (5′-TTAGCTGCAAGATTTATTAGGAAG-3′ positions 178–201 complementary) encoded by the same toxin (toxin Cn5). This pair of oligonucleotides was used to amplify the region encoding the mature toxin (residues 1–66) by means of PCR (for details, see Becerril et al. [17]).

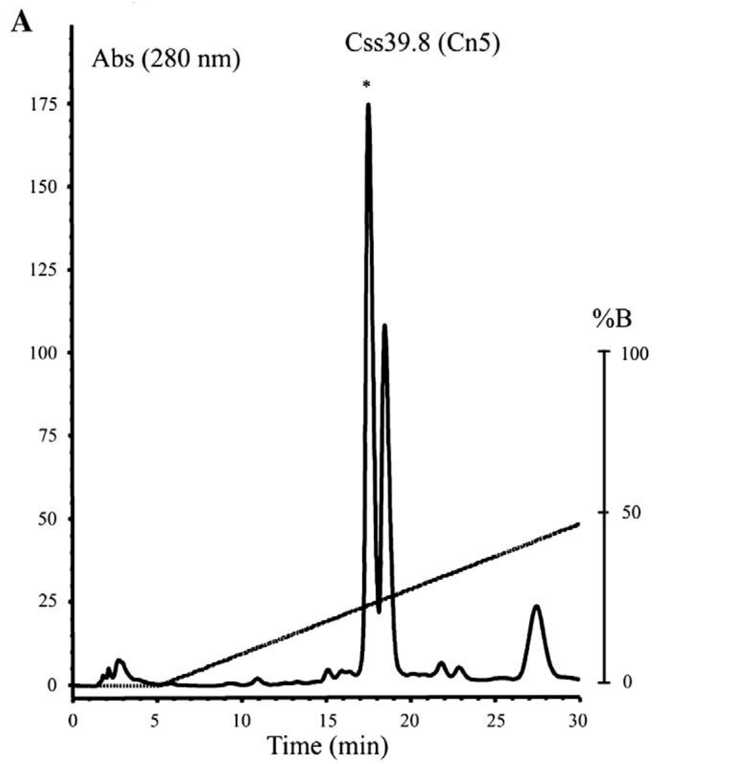
For cloning the cDNA corresponding to the nucleotidic sequence of Css39.8, the total RNA was isolated from a single venomous gland (telson) using the Total RNA Isolation System (Promega, USA) and used for constructing a full-length cDNA phagemid library using the SMART cDNA Library Construction Kit (CLONTECH Lab., Inc., USA). Based on the information obtained from Cn5 clone sequencing, a specific oligonucleotide was designed and used for the PCR reaction using as a template cDNA material from the library. The PCR reaction was performed in 1× Vent DNA polymerase buffer, 200 mM dNTPs, and 0.25 mM forward primer (5′-ATA AAG AGG GCT ATC TGG-3′); 0.25 mM reverse CDS3′primer (5′-ATT CTA GAG GCC GAG GCG GCC GAC ATG-3′) and 2 U of Vent DNA polymerase in a final volume of 50 ml in a Perkin Elmer 9600. The reactions were incubated at 94 °C for 5 min and 50 °C for 7 min before Vent polymerase was added. The mixtures were then incubated at 72 °C for 1 min for one cycle. After this initial cycle, the mixtures were incubated at 94 °C for 30 s, 50 °C for 40 s and 72 °C for 30 s per 35 cycles, followed by a final 7 min step at 72 °C. PCR products were purified using a Centricon 100 column (Amicon) following the manufacturer's instructions and then ligated into the *pKS*+EcoRV digested plasmid. The ligation reaction was used to transform competent *Escherichia coli* DH5-α cells. Positive clones were sequenced from both ends using the Thermo sequenase radio-labeled terminator cycle sequencing kit (Amersham). The complete clone was obtained using as a template the DNA extracted from the cDNA-SMART library. For the PRC amplification the 5′-adaptor primer provided by the cDNA-SMART library kit (5′-GGC CAT TAC GGC CGG GNN-3′) and a specific reverse primer were used (5′-AAT ATT TCT GTT GGT GGA CAA TAA AAA GCC-3′). The final cloning and sequencing were obtained as previously described [17].

2.5. ^{1}H NMR experiments

NMR experiments were recorded on an Inova 600 (Varian Inc., Palo Alto) spectrometer. The spectrometer was equipped with a cryogenically cooled 5 mm triple resonance ^{1}H{^{13}C/^{15}N} probe. Spectra were recorded, processed and analyzed using Vnmr 6.1C (Varian), and NMRView 5.2 [18]. The lyophilized toxin, 2.6 mg, was dissolved in 150 µl of H_2O/D_2O, 9:1 (v/v), or in 99.96% D_2O (Euriso-Top, Saclay, France) representing the required volume for 3 mm Shighemi NMR tubes (Shigemi Inc, Alison Park, United States). The solution pH was adjusted at 4.5 by adding small volumes of dilute NaOH. Experiments

Please cite this article as: G. Corzo, et al., Solution structure of Cn5, a crustacean toxin found in the venom of the scorpions *Centruroides noxius* and *Centruroides suffusus suffusus*, Biochim. Biophys. Acta (2009), doi:10.1016/j.bbapap.2009.07.006

A



B

Peptide	Sequence
Css39.8	KEGYLVNKSTGCKYGCLLLGKNEGCDKECKAKNQGGSYGYCYAFGCWCEGLPESTPTYPLPNKSCS
	:---- RC-Direct --------------------->
	:---- Glu-C 2 ------------->:-------- Glu-C 3 ------>:-- Glu-C 1->

Fig. 1. Cation-exchange HPLC chromatogram of the crustacean fraction and amino acid sequence of the pure peptide. (A) The protein fraction number 39, from a previous reverse-phase HPLC chromatogram [32], was re-purified under ion-exchange chromatography as described in Materials and methods. The asterisk shows the time of elution of the crustacean toxin Css39.8. (B) The amino acid sequence of Css39.8 was elucidated by a combination of direct Edman degradation of a reduced and alkylated sample of the native peptide and by protein cleavage using protease Glu-C. The first 38 amino acids were identified as indicated (RC-direct). Another sample of reduced and alkylated peptide Css39.8 was enzymatically cleaved and separated by HPLC. From the 5 peptides obtained (data not shown) three were sufficient to construct and overlapping sequence, as indicated by Glu-C1 to Glu-C3.All Glu-C fragments were sequenced by Edman degradation, and their molecular masses were verified by ESI mass spectrometry (data not shown).

were run at 303 K or 313 K with a 6982 Hz sweep width. Spectra were calibrated on the residual water signal at 4.73 ppm at 303 K and at 4.63 ppm at 313 K relative to 2.2-dimethylsilapentane-5-sulphonate (DSS) as external reference. The 2D proton NMR spectra were collected in the phase sensitive mode [19] with 2 K data points in the t_2 dimension and 400–512 t_1 increments, with typically 16–32 scans per increment. Zero-filling was applied prior to Fourier transformation and data were processed with shifted sine bell window functions in both dimensions. The 2D NMR experiments used for the identification of spin systems have the following pulse sequences; (a) total correlation spectroscopy or Clean TOCSY [20] recorded using a MLEV-17 pulse scheme with 100 ms and 60 ms isotropic mixing period [21,22], (b) two dimensional Double-Quantum Spectroscopy [23,24], and NOESY [19] spectra. NOESY spectra were acquired at 313 K in D_2O with a mixing time of 200 ms, and in H_2O with mixing times of 80, 120 and 200 ms. The strong signal from the

```
           1/1                                     31/11                                   61/21
Cn5                                                                                         ct tcg aag ATG AAC TCG TTG TTG ATG ATC
Css39.8    aag cag tgg tat caa cgc aga gtg gcc att acg gcc ggg gat tct aaa ctc gtt tag ata gct tcg aag ATG AAC TCG TTG TTG ATG ATC
                                                                                                        M   N   S   L   L   M   I
           91/31                                   121/41                                  151/51
Cn5        ACT GCT TGT TTG TTC CTG ATC GGA ACA GTG TGG GCA AAA GAA GGT TAT CTG GTA AAC AAG AGC ACG GGC TGC AAA TAC GGT TGC CTG CTA
Css39.8    ACT GCT TGT TTT TTC CTG ATC GGA ACA GTG TGG GCA AAA GAA GGT TAT CTG GTA AAC AAG AGC ACG GGC TGC AAA TAC GGT TGC CTG CTA
            T   A   C   F   F   L   I   G   T   V   W   A   K   E   G   Y   L   V   N   K   S   T   G   C   K   Y   G   C   L   L
           181/61                                  211/71                                  241/81
Cn5        TTG GGG AAA AAC GAA GGC TGC GAT AAG GAA TGC AAA GCG AAG AAC CAA GGA GGT AGT TAC GGC TAT TGC TAC GCT TTT GGG TGC TGG TGC
Css39.8    TTG GGA AAG AAC GAA GGC TGC GAT AAG GAA TGC AAA GCG AAG AAC CAA GGA GGT AGT TAC GGC TAT TGC TAC GCT TTT GGG TGC TGG TGC
            L   G   K   N   E   G   C   D   K   E   C   K   A   K   N   Q   G   G   S   Y   G   Y   C   Y   A   F   G   C   W   C
           271/91                                  301/101                                 331/111
Cn5        GAA GGT TTG CCC GAA AGT ACA CCG ACT TAT CCC CTT CCT AAT AAA TCT TGC AGC AAA AAA TAA tgg caa cga ctt tta ttg ttt acc aac
Css39.8    GAA GGT TTG CCC GAA AGT ACA CCG ACT TAT CCT CTT CCT AAT AAA TCA TGC AGC AAA AAA TAA ttg caa cga ctt ttt att gtc cac caa
            E   G   L   P   E   S   T   P   T   Y   P   L   P   N   K   S   C   S   K   K   *
           361/121
Cn5        aga aat att gta acg ctt gtt aat tgc agt taa atg
Css39.8    cag aaa tat t
```

Fig. 2. The cDNA and deduced amino acid sequences of Css39.8 and Cn5 (L05060). Amino acids are indicated by one-letter symbols. The cDNA encodes a mature peptide of 68 residues. The Lys–Lys pattern (residues 67–68) of the mature peptide seems to be required for post-translational processing; however, the C-terminal of both toxins is not amidated. The stop codon is shown by an asterisk.

Please cite this article as: G. Corzo, et al., Solution structure of Cn5, a crustacean toxin found in the venom of the scorpions *Centruroides noxius* and *Centruroides suffusus suffusus*, Biochim. Biophys. Acta (2009), doi:10.1016/j.bbapap.2009.07.006

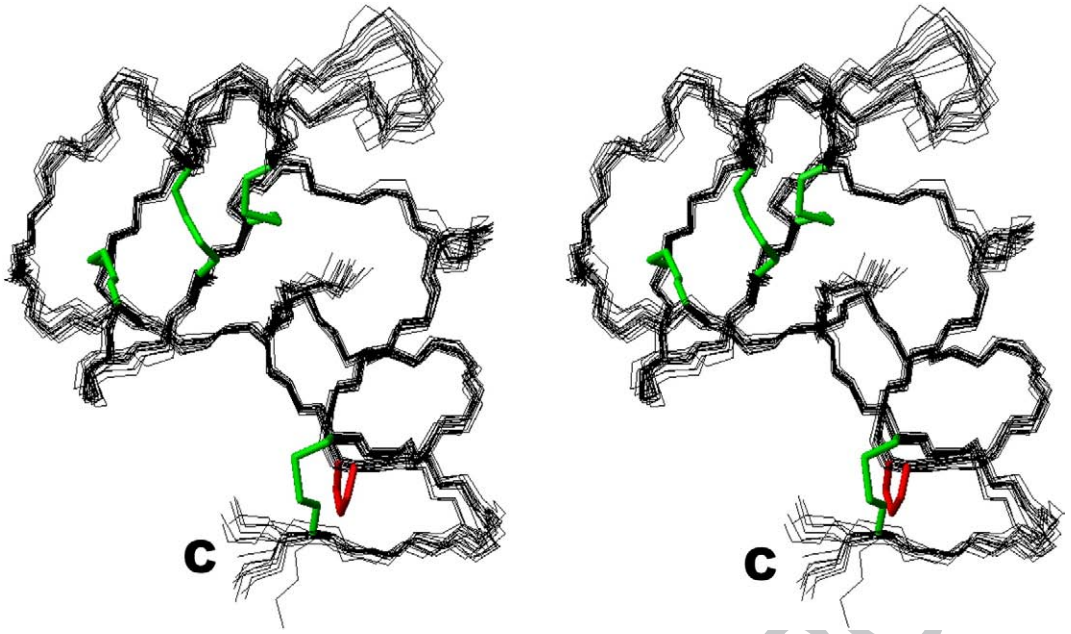


Fig. 3. Stereo-view of backbone superposition of the 15 selected conformers of Cn5. The disulfide bonds are indicated in green and the side chain of the *cis*-proline at the position 59 is shown in red. Letter C denotes the C-terminal. (Figure prepared using MOLMOL, [29].) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

solvent (HDO and H_2O) was suppressed either by presaturation during the recycle delay or by means of double pulsed field gradient spin echo (DPFGSE) in the NOESY experiment [25,26]. Hydrogen exchange rates were evaluated by freeze drying the Cn5 sample from H_2O and dissolving it in D_2O. One-dimensional experiments were carried out, at 303 K, 1 h after D_2O addition and then TOCSY and NOESY experiments were run allowing the identification of the slowest exchanging amide protons.

Assignment of signals to peptide protons was achieved by the standard method developed by Wüthrich [27].

2.6. Structure calculation

Structures were calculated from data on the native Cn5 peptide obtained at 35 °C using NOE-derived distance constraints and hydrogen bonds. Distance constraints were obtained from NOESY spectra recorded in H_2O with a mixing time of 200 ms at 35 °C and analyzed using VNMR (VariaInc) and NMRview. Additional distance constraints, in particular those involving the aromatic protons and the Hα protons were obtained from a NOESY spectrum recorded in D_2O at 35 °C with 200 ms mixing time.

Structure calculation was combined with automatic NOE cross-peak assignment using ARIA version 2.2 and CNS version 1.2 [28]. Four disulfide bridges (residues 12–65, 16–41, 25–46, and 29–48) based originally on sequence comparison and supported by NOE interactions observed between Hα and Hβ protons or between Hβ protons of paired cysteines were used in calculation. The residue Pro59 was in *cis* conformation, as indicated by the presence of NOE interactions between the Hα proton of the proline with that of the preceding residue Thr58.

Hydrogen bonds were set for the only very slowly exchanging amide protons of Asn7, Gly11, Asp26, Lys27, Glu28, Cys29, Lys30, Ser37, Cys41, Tyr42, Cys46, Trp47, Cys48, Gln49, and Thr55. Several cycles of ARIA were performed using standard protocols and by varying the chemical shift tolerance between 0.03 and 0.02 ppm. Rejected restraints, assignments and violations were analyzed after each cycle. Finally, 200 conformers were calculated with ARIA2 and refined in water. Fifteen structures with the lowest restraint energy values were used for statistical analysis. The structures were visualized and analyzed with MOLMOL [29], and their quality was assessed using PROCHECK [30] and WHATCHECK [31].

3. Results

3.1. Purification and amino acid sequence

For purification of Css39.8 a sample of soluble venom from *C. s. suffusus* was initially separated by HPLC. This separation gave many

Table 1
Structural statistics for the ensemble of 15 Cn5 conformers.

Parameter	Value
Constraints	
Unambiguous	710
Intraresidue	314
Sequential	144
Medium range	57
Long range	195
Ambiguous	255
Intraresidue	69.2
Sequential	51.5
Medium range	31.6
Long range	102.7
Total	965
Intraresidue	383.2
Sequential	195.6
Medium range	88.6
Long range	298.3
Residual distance constraint violations	
>=0.5 Å	5.27 +/− 0.66
>=0.3 Å	15.00 +/− 1.76
RMS from NOEs (Å)	0.10 +/− 0.003
Energies (kcal/mol)	
Total	−1721 +/− 57
Van der Waals	−182 +/− 15
Electrostatic	−2474 +/− 61
Mean pairwise RMSD (Å) for all residues	
Backbone atoms	0.69 +/− 0.14
Heavy atoms	1.22 +/− 0.17
Ensemble Ramachandran plot; % residues in	
Most favoured region	76.4 +/− 2.1
Additionally allowed regions	23.6 +/− 1.6

Please cite this article as: G. Corzo, et al., Solution structure of Cn5, a crustacean toxin found in the venom of the scorpions *Centruroides noxius* and *Centruroides suffusus suffusus*, Biochim. Biophys. Acta (2009), doi:10.1016/j.bbapap.2009.07.006

Table 2
Amino acid sequences of mammalian and crustacean neurotoxins from the genus *Centruroides*.

Peptide	Sequence	
	Crustacean toxins	
Cn5/ Css39.8	KEGYLVNKST GCKYGCLLLG KNEGCDKECK AKNQGGSYGY CYAFGCWCEG LPESTPTYPL PNKSCS	P45663
Cll toxin 1	KEGYLVNKST GCKYGCFWLG KNENCDKECK AKNQGGSYGY CYSFACWCEG LPESTPTYPL PNKSC	P45667
CsE-v1	KEGYLVKKSD GCKYDCFWLG KNEHCDTECK AKNQGGSYGY CYAFACWCEG LPESTPTYPL PNKSC[a]	P01492
	Mammalian toxins	
CssII	KEGYLVSKST GCKYECLKLG DNDYCLRECK QQYGKSSGGY CYAFACWCTH LYEQAVVWPL PNKTCN[a]	P08900
Cn2	**KEGYLVDKNT GCKYECLKLG DNDYCLRECK QQGYKGAGGY CYAFACWCTH LYEQAIVWPL PNKRCS**[a]	P01495
	1——10 ——20 ——30 ——40 ——50 ——60 ——66	

[a] Means amidated C-terminus.

distinct sub-fractions (see Corzo et al. [32]). The one eluting at time 39 min was not toxic to mice, but was toxic to insects and crustaceans. Since the product was not homogeneous, this fraction was further fractionated by ion-exchange HPLC chromatography as shown in Fig. 1A. The most abundant component, toxic to crustaceans, was finally re-applied to the analytical HPLC column (data not shown) providing a completely homogeneous peptide identical in mass to that of toxin Cn5 from *C. noxius* [14]. This peptide was shown to be toxic to crustaceans and insects, but non-toxic to mice when injected intraperitoneally up to 570 μg/mouse of 20 g body. The homogeneity of Css39.8 was verified by direct Edman degradation (only one amino acid per cycle of the sequencer) and mass spectrometry. A single component with molecular mass of 7136.0 a.m.u. was determined. Automatic N-terminal sequencing of a reduced and alkylated sample of the peptide allowed the identification of the first 38 amino acid residues (see RC-direct in Fig. 1B). Additionally, three overlapping peptides were sequenced after cleavage with endoprotease Glu-C and HPLC separation (data not shown) giving an unequivocal sequence as indicated in Fig. 1B. The sequence obtained from analysis conducted with the peptide was confirmed by the nucleotide sequence as described below. This component corresponds to 6–7% of the soluble venom. In order to identify the nucleotide sequence of the Css39.8 clone, a search of positive colonies was conducted using appropriate oligonucleotides, as discussed in Materials and methods. The cDNA was cloned and its messenger was shown to code for a peptide of 87 amino acid residues from which 19 residues are part of the signal peptide, and 68 residues correspond to the mature protein. The last two residues are cleaved without leaving a C-terminal amidation (Fig. 2).

3.2. NMR assignments

The sequence specific assignment of Cn5 was achieved according to a standard procedure [27]. All spin systems except that of Gly39 were identified. First, the spin systems of all amino acid residues were identified via their through-bond connectivity observed in COSY, TOCSY and double quantum experiments. The unique spin system Phe44 was used as starting point for the sequential assignment together with the easily assigned threonine and glycine spin systems. The sequential Hα proton NOE connectivity of Tyr58 with the Hα proton of Pro59, indicates that this Tyr-Pro amide bond is in *cis* conformation whereas sequential amide or Hα proton NOE connectivities of X residue with the Hδ protons of ProX+1, indicate that the Xaa-Pro amide bond is in *trans* conformation for all the remaining X-Pro fragments. Thr10 hydroxyl group was clearly observable in the

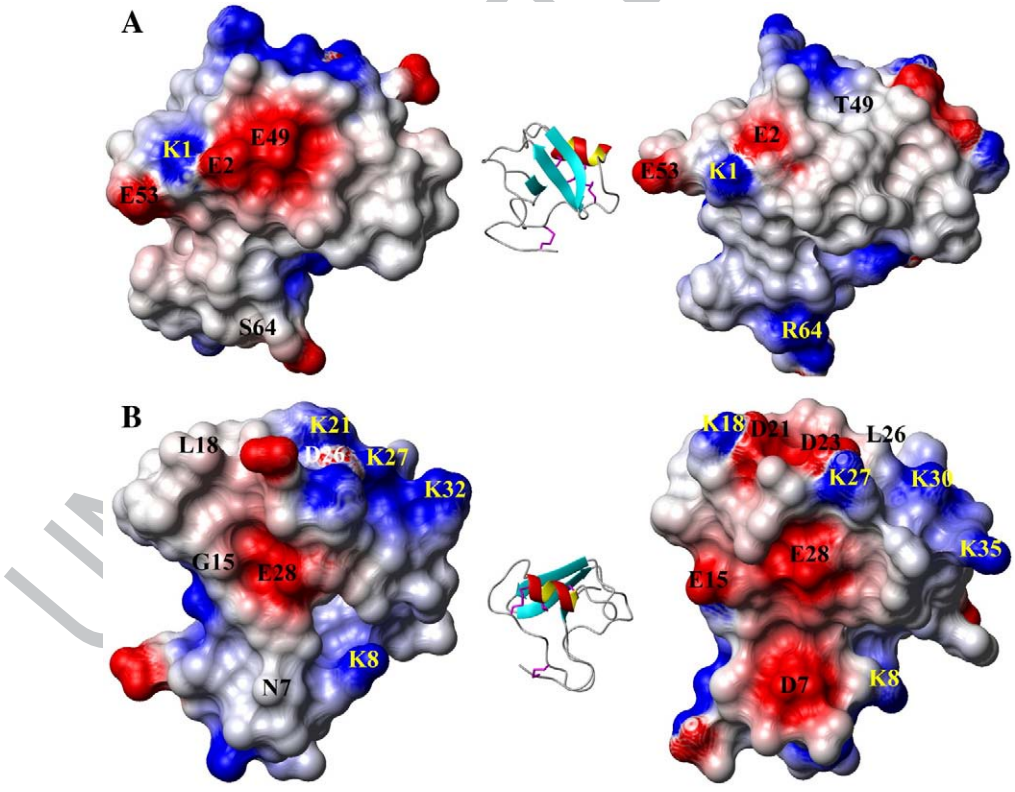


Fig. 4. Comparison of the electrostatic surfaces of Cn5 (left) and Cn2 (right). Cn5 ribbon representation insets with disulfide bonds indicated in magenta show the orientation of the molecules. (A) The side of the β-sheet facing the solvent. (B) The α-helix side. (Figure prepared using MOLMOL [29].) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: G. Corzo, et al., Solution structure of Cn5, a crustacean toxin found in the venom of the scorpions *Centruroides noxius* and *Centruroides suffusus suffusus*, Biochim. Biophys. Acta (2009), doi:10.1016/j.bbapap.2009.07.006

spectra obtained in H_2O indicating that this hydroxyl group is either involved in a hydrogen bond or is not accessible to solvent.

3.3. 3D structure of Cn5

The structures were calculated from data on the peptide in natural abundance using a total of 965 meaningful distance constraints derived from the volume integration of the 2D NOESY spectra. The structure of Cn5 is shown in Fig. 3. Cn5 displays the α/β scaffold characteristic of scorpion toxins. The structure is stabilized by four disulfide bridges. A *cis*-proline at position 59 induces a kink in its C-terminal part. The α-helix spans the region of residues Asn22–Cys29; the anti-parallel β-sheet is composed of three strands formed respectively by residues Gly3–Tyr4, Tyr38–Tyr42, and Gly45–Glu49. Furthermore, direct NMR data suggest that the turn following the first strand, and the turn running between the α-helix and the second strand can be characterized as type I β-turns, whereas a type II β-turn connects the second and the third strands. Similar features were also found in the Cll toxin 1[12]. The mean pairwise RMSD over all residues are 0.96 Å for backbone atoms and 1.50 Å for the heavy atoms. The values calculated over the secondary structure elements are 0.45 Å and 1.09 Å, respectively. The details of the constraints used in calculations and the structural characteristics of the family of 15 conformers are summarized in Table 1.

4. Discussion

This article reports structural aspects of the crustacean toxin (Cn5) from *C. noxius* Hoffmann, and of an identical peptide Css39.8 from *C. s. suffusus* venom, here simply named: Cn5. This work is aimed at obtaining structural information on this new class of crustacean specific toxins in order to compare it with other known scorpion toxins specific to mammals. Several questions were raised in this context. Which are the most important structural features of these short peptides that are relevant to recognition of ion channels? Which are the substitutions and charge density variations of all these structurally similar peptides that enable them to distinguish between different membrane-bound proteins, such as Na^+, K^+, Cl^- or Ca^{2+}-channels? More important yet, how is it that very similar structures like Cn5 and Cn2 [11,33] from *C. noxius* and CssII from *C. s. suffusus* venom [34], have such exquisite tissue specificity? Cn2 and CssII [33,34] are highly toxic to mammals, but non-toxic to crustacean, whereas Cn5 is toxic to crustacean, and non-toxic to mammals.

In this section we will briefly discuss some of the relevant structural features of Cn5.

4.1. Purification and species specificity

The first interesting observation is the species specificity of these peptides. The LD_{50} of Cn5 in crayfish is about 28.5 μg/g weight of animal. It is not absolutely specific for crustacean. Crickets injected with doses of about 40 μg/g of cricket weight, became paralysed, but 30 min after injection a complete recovery was observed. It was necessary to inject over 150 μg/g in order to kill crickets, although both animals are arthropods. The values reported here are comparable to those described earlier by Babin et al. [35], for the crickets assayed with toxin variants 1, 2, and 3 of the scorpion *Centruroides sculpturatus*. However, Cn5 was not toxic to mammals when doses up to 570 μg of protein/20 g mouse body weight were injected intraperitoneally. Cn5 induced a paralysis of lower limbs at 1 μg/20 g mouse body weight if injected intracranially, but after 20–30 min injection a complete recovery was observed. Toxin Cn2 from *C. noxius* [33] and CssII from *C. s. suffusus* [34] are typical mammalian toxins, [33,34] whose LD_{50} in mouse is of the order of 0.4–1.3 μg/20 g weight. Thus, they are extremely toxic to mammals but non-toxic to crustaceans at doses up to 25 μg/crayfish (Serrano, S. and Possani, L.D., private communication). It is thought that modifications in the total charge density of Cn5 and certainly some amino acid substitutions are responsible for the specific channel recognition and the specificity of these toxins. These modifications must account for the different intoxication symptoms observed in experimental animals (mice, crayfishes and crickets) when injected with these peptides, and will be discussed further.

4.2. Gene organization

Since most of scorpion toxin genes contain introns, it is worth mentioning that what we are reporting here is the product of the corresponding cDNAs obtained from mRNAs directly purified from the venomous glands of both species of scorpions. Thus, strictly speaking, we do not know if both species have identical genes at the level of the real DNA present in both species. Sequences of both, precursor and mature forms of Cn5 and Css39.8, as well as other scorpion peptide toxins indicate that the N-terminal of peptide toxins were generated by cleavage at Thr/Val/Trp/Ala endoproteolytic sites at the precursor polypeptide toxins (Fig. 2). It has been observed that the N-termini of peptide toxins from scorpions of the *Centruroides* genus were generated by cleavage of similar motives [36]. Moreover, the N-termini of peptide toxins from scorpions of Old World such LqhIT2 and LqqIT2 were generated by cleavage at similar motifs, such as Leu/Val/Asn/Ala [37]. Concerning the C-terminal amino acids of most scorpion toxins it is anticipated that they will undergo post-translational modification when a C-terminal glycine alone or flanked by a single or dibasic endoproteolytic site is located at the most C-terminal region [36,38]. This is a common characteristic for C-terminal amidation of the last residue of the peptide. The cDNA sequences of the mature forms of Cn5 and Css39.8 show a stop codon after a dibasic residue without a glycine. The mass spectrometry results as well as the Edman degradation show that the mature peptides Cn5 and Css39.8 are composed of 66 residues with no C-terminal amidation. However, the mature peptides Cn5 and Css39.8 suffer post-translational modification, because the dibasic residues found in the cloned genes are eliminated, leaving a non amidated serine residue as the last residue of the mature peptide. Therefore, post-translational modifications of scorpion peptides can occur without causing chemical modification such as the C-terminal amidation.

Polymorphic genes are common in scorpion venom glands [37]. Therefore, it is interesting to point out that five single-base substitutions (point mutations) occur between the precursor and mature forms of Cn5 and Css39.8 genes at the third position of the codons. There are three transitions (nucleotide positions 186, 189 and 316) and two transversions (nucleotide positions 102 and 316) (Fig. 2). These five single-base substitutions result in five silent mutations because they cause no change in the residue composition neither of the precursor nor of the mature scorpion toxins. Silent mutations in toxins such Cn5 and Css39.8 from two different species cannot be detected without sequencing the genes.

4.3. Comparison of crustacean versus mammalian toxins

It is worth noting that other toxic-peptides, also specific for crustaceans, but non-toxic to mammals are present in the subfractions of Fig. 1A. These crustacean toxins form a group of peptides that under cation-exchange chromatography elute earlier than the mammalian toxins, therefore confirming that the less charged peptides are usually toxic to arthropods whereas the most basic ones are toxic to mammals (Table 2).

In the case of Cn5, the disulfide bridges are in the expected positions, as based on our previous experience with crustacean toxin 1 from *C. limpidus limpidus* [12]. This was the first crustacean toxin for

Please cite this article as: G. Corzo, et al., Solution structure of Cn5, a crustacean toxin found in the venom of the scorpions *Centruroides noxius* and *Centruroides suffusus suffusus*, Biochim. Biophys. Acta (2009), doi:10.1016/j.bbapap.2009.07.006

which the complete covalent and three-dimensional structure was determined. Both Cn5 and Cll toxin 1 have the disulfide pairs situated in equivalent positions.

Comparing the different toxins from *C. noxius* (Cn5 and Cn2) it is worth mentioning that both are 66 residues long; they share 59% of sequence identity and display three conservative substitutions. The most important sequence variability is concentrated in the loop 31–38, which connects the α-helix with the second strand of the β-sheet and in the C-terminal part of the toxin (residues 49–57).

Most of the hydrophobic residues are conserved in the crustacean (Cn5, Cll1 and CsE-v1) and mammalian (Cn2 and CssII) toxins. Gly3, Leu5, Val6, Gly11, Leu51, Leu60 and the eight Cys residues define the hydrophobic core of both groups of toxins. A hydrophobic patch displayed on the β-sheet side facing the solvent is formed by conserved Tyr residues at positions 4, 40, 42, Ala43 and Trp47. Val57 reinforces the hydrophobic character of this patch in Cn2 and CssII, whereas a Thr residue is substituted at the corresponding position of Cn5, Cll1 and CsE-v1. The residues of Leu/Phe17, Leu19, Ala43, and Phe44 form a second hydrophobic patch situated on an edge of the molecule; atoms of the residues at positions 24 and 18 (Gly24 and Leu18 in Cn5, and Tyr24 and the aliphatic chain of Lys18 in Cn2 and CssII) also contribute to this surface.

Differences in the content of charged residues (six negatively and eight positively charged residues in Cn5 versus seven negatively and nine positively charged in Cn2) are reflected in the distribution of electrostatic charges on the surfaces of the toxins.

The electrostatic surface corresponding to the external side of the β-sheet is defined by three conserved residues Glu53, Lys1, and Glu2 as well as by Glu/Thr49. This latter Glu residue, not conserved in Cn2, contributes to a strongly negative patch in Cn5. A C-terminal substitution Ser64Arg results in a positive spot on the corresponding surface of Cn2, also visible on this side of the toxin (Fig. 4A).

The electrostatic features of the opposite side of the toxins, formed by the residues belonging to the α-helix and to the adjacent loops are represented on the upper part of the Fig. 4B. A positive patch defined in Cn5 by three lysine residues at positions 27, 30, and 32 finds its Cn2 equivalent formed by Arg27, Lys30 and Lys35; this is due to a shifted first residue of the α-helix that starts at position 22 in Cn5 and 23 in Cn2. The substitutions Leu18Lys and Lys21Asp (and, to a lesser extent that of Asp26Leu, Asp26 being not very much exposed in Cn5) determine the different character of the corresponding electrostatic surfaces in Cn5 and Cn2 toxins. Two N-terminal substitutions, namely Asn7Asp and Gly15Glu are responsible for the differences of the electrostatic surfaces of Cn5 and Cn2.

4.4. Specificity for crustaceans

The selectivity towards mammalian or crustacean Na^+ channels is certainly due to the presence of additional or substituted amino acid residues producing local differences capable of inducing or preventing the interaction with the specific receptor molecule. Sequence comparison indicates that the « decision » if a particular toxin of the genus *Centruroides* will be specific towards mammals or arthropods is influenced by the character of few residues which are clustered in two regions of the molecules, the first corresponding to the loop located between the end of the α-helix and the strand 2 of the β-sheet (residues 31 to 38), and the second following the strand 3 (residues 49 to 57) (Table 2). It is foreseen that comparative mutational analysis of Cn5 and their mammalian counterpart (either Cn2 or CssII) should be performed in order to corroborate our results. For this purpose various point mutations should be constructed and assayed. However, an occurrence of a more simple or unique modification such as those reported by Karbat et al. [39] for toxin Bj-xtrIT should not be discarded *a priori*. These authors showed that the substitution of single amino acid residues (Glu15) affects toxin specificity [39]. More experiments should be conducted in order to clarify the different structural aspects shown here, if we want to fully understand the preference of Cn5 specificity for crustaceans.

Acknowledgements

This work was supported in part by grants from *Dirección General de Asuntos del Personal Académico* (DGAPA-UNAM) IN226006 to GC and IN227507 to LDP, CONACyT 49773/24968 to GC and Institute Pasteur and CNRS URA 2185 to MD. The authors express their gratitude to Fernando Zamudio, Timoteo Olamendi-Portugal and Cipriano Balderas for technical assistance. The 600 MHz NMR spectrometer was funded by the Region Ile de France and the Institut Pasteur.

References

[1] R.C. Rodríguez de la Vega, L.D. Possani, Overview of scorpion toxins specific for Na+ channels and related peptides: biodiversity, structure–function relationships and evolution, Toxicon 46 (2005) 831–844.
[2] L.D. Possani, R.C. Rodríguez de la Vega, in: A.J. Kastin (Ed.), The Handbook of Biologically Active Peptides, Academic Press, San Diego, CA, USA, 2006, pp. 339–354.
[3] J.C. Fontecilla-Camps, C. Habersetzer-Rochat, H. Rochat, Orthorhombic crystals and three-dimensional structure of the potent toxin II from the scorpion *Androctonus australis* Hector, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 7443–7447.
[4] J.C. Fontecilla-Camps, R.J. Almassy, F.L. Suddath, D.D. Watt, C.E. Bugg, Three-dimensional structure of a protein from scorpion venom: a new structural class of neurotoxins, Proc. Natl. Acad. Sci. U. S. A. 77 (1980) 6496–6500.
[5] W.A. Catteral, S. Cestèle, V. Yarov-Yarovoy, F.H. Yu, K. Konoki, T. Scheuer, Voltage-gated ion channels and gating modifier toxins, Toxicon 49 (2007) 124–141.
[6] E. Jover, F. Couraud, H. Rochat, Two types of scorpion neurotoxins characterized by their binding to two separated receptor sites on rat brain synaptosomes, Biochem. Biophys. Res. Comm. 95 (1980) 1607–1614.
[7] I. Karbat, F. Frolow, O. Froy, N. Gilles, L. Cohen, M. Turkov, D. Gordon, M. Gurevitz, Molecular basis of the high insecticidal potency of scorpion {alpha}-toxins, J. Biol. Chem. 279 (2004) 31679–31686.
[8] H. Darbon, C. Weber, W. Braun, Two-dimensional 1H nuclear magnetic resonance study of AaH IT, an anti-insect toxin from the scorpion *Androctonus australis* Hector. Sequential resonance assignments and folding of the polypeptide chain, Biochemistry 30 (1991) 1836–1845.
[9] E.P. Loret, F. Sampieri, A. Roussel, C. Granier, H. Rochat, Conformational flexibility of a scorpion toxin active on mammals and insects: a circular dichroism study, Proteins 8 (1990) 164–172.
[10] E. Schiavon, T. Sacco, R. Cassulini, G. Gurrola, F. Tempia, L.D. Possani, E. Wanke, Resurgent current and voltage sensor trapping enhanced activation by a beta-scorpion toxin solely in Nav1.6 channel. Significance in mice Purkinje neurons, J. Biol. Chem. 281 (2006) 20326–20337.
[11] A. Pintar, L.D. Possani, M. Delepierre, Solution structure of toxin 2 from *Centruroides noxius* Hoffmann, a beta scorpion neurotoxin acting on sodium channels, J. Mol. Biol. 287 (1999) 359–367.
[12] F. Lebreton, M. Delepierre, A.N. Ramírez, C. Balderas, P. L.D., Primary and NMR three-dimensional structure determination of a novel crustacean toxin from the venom of the scorpion *Centruroides limpidus limpidus* Karsch, Biochemistry 33 (1994) 11135–11149.
[13] A. Prochnicka-Chalufour, G. Corzo, H. Satake, M.F. Martin-Eauclaire, A.R. Murgia, G. Prestipino, G. D'Suze, L.D. Possani, M. Delepierre, Solution structure of discrepin, a new K+-channel blocking peptide from the alpha-KTx15 subfamily, Biochemistry 45 (2006) 1795–1804.
[14] C. García, B. Becerril, B. Selisko, M. Delepierre, P. L.D., Isolation, characterization and comparison of a novel crustacean toxin with a mammalian toxin from the venom of the scorpion *Centruroides noxius* Hoffmann, Comp. Biochem. Physiol. B Biochem. Mol. Biol. 116 (1997) 315–322.
[15] G. Corzo, E. Villegas, L. Hon-Seok, T. Nakajima, Detection and purification of insecticidal peptides from scorpions and spiders: a rapid method for their isolation from their crude venoms using MALDI-TOF-MS analysis, Prot. Pep. Lett. 8 (2001) 375–383.
[16] T. Olamendi-Portugal, C.V.F. Batista, R. Restano-Cassulini, V. Pando, O. Villa-Hernández, A. Zavaleta-Martínez-Vargas, M.C. Salas-Arruz, R. Rodríguez de la Vega, B. Becerril, L.D. Possani, Proteomic analysis of the venom from the fish eating coral snake *Micrurus surinamensis*: novel toxins their function and phylogeny, Proteomics 8 (2008) 1919–1932.
[17] B. Becerril, A. Vázquez, C. García, M. Corona, F. Bolivar, L.D. Possani, Cloning and characterization of cDNAs that code for Na(+)-channel-blocking toxins of the scorpion *Centruroides noxius* Hoffmann, Gene 30 (1993) 165–171.
[18] B.A. Johnson, R.A. Blevins, NMRView: a computer program for the visualisation and analysis of NMR data, J. Biomol. NMR 4 (1994) 603–614.
[19] D.J. States, R.A. Haberkorn, D.J. Ruben, A two dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants, J. Magn. Reson. 48 (1982) 286–292.
[20] C. Griesinger, G. Otting, K. Wüthrich, R.R. Ernst, Clean Tocsy for 1H spin system identification in macromolecules, J. Am. Chem. Soc. 110 (1988) 7870–7872.
[21] M.H. Levitt, R. Freeman, T. Frenkiel, Broadband heteronuclear decoupling, J. Magn. Reson. 47 (1982) 328–330.

Please cite this article as: G. Corzo, et al., Solution structure of Cn5, a crustacean toxin found in the venom of the scorpions *Centruroides noxius* and *Centruroides suffusus suffusus*, Biochim. Biophys. Acta (2009), doi:10.1016/j.bbapap.2009.07.006

[22] A. Bax, D.G. Davis, MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy, J. Magn. Reson. 65 (1985) 355–360.
[23] U. Piantini, O.W. Sørensen, R.R. Ernst, Multiple quantum filters for elucidating NMR coupling networks, J. Am. Chem. Soc. 104 (1982) 6800–6801.
[24] M. Rance, O.W. Sørensen, G. Bodenhausen, G. Wagner, R.R. Ernst, K. Wüthrich, Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering Biochem, Biophys. Res. Comm. 117 (1983) 479–485.
[25] C. Dalvit, Efficient multiple-solvent suppression for the study of the interactions of organic solvents with biomolecules, J. Biomol. NMR 11 (1998) 437–444.
[26] T.L. Hwang, A.J. Shaka, Water suppression that works. Excitation sculpting using arbitrary waveforms and pulsed field gradients, J. Magn. Reson. A112 (1995) 275–279.
[27] K. Wüthrich, (1986), John Wiley & Sons, New York.
[28] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T.E. Malliavin, M. Nilges, ARIA2: automated NOE assignment and data integration in NMR structure calculation, Bioinformatics 23 (2007) 381–382.
[29] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: a program for display and analysis of macromolecular structures, J. Mol. Graphics 14 (1996) 51–55.
[30] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[31] R.W. Hooft, G. Vriend, C. Sander, E.E. Abola, Errors in protein structures, Nature 381 (1996) 272.
[32] G. Corzo, F. Papp, Z. Varga, O. Barraza, P.G. Espino, R.C. Rodríguez de la Vega, R. Gaspar, G. Panyi, L.D. Possani, A selective blocker of Kv1.2 and Kv1.3 potassium channels from the soluble venom of the scorpion *Centruroides suffusus suffusus*, Biochem. Pharm. 76 (2008) 1142–1154.
[33] F. Zamudio, R. Saavedra, B.M. Martin, G. Gurrola-Briones, P. Herion, L.D. Possani, Aminoacid sequence and immunological characterization with monoclonal antibodies of two toxins from the venom of the scorpion *Centruroides noxius* Hoffmann, Eur. J. Biochem. 204 (1992) 281–292.
[34] M.F. Martin, L.G. Garcia y Perez, M. el Ayeb, C. Kopeyan, G. Bechis, E. Jover, H. Rochat, Purification and chemical and biological characterizations of seven toxins from the Mexican scorpion, *Centruroides suffusus suffusus*, J. Biol. Chem. 262 (1987) 4452–4459.
[35] D.R. Babin, D.D. Watt, S.M. Goos, R.V. Mlejnek, Amino acid sequences of neurotoxic protein variants from the venom of *Centruroides sculpturatus* Ewing, Arch. Biochem. Biophys. 164 (1974) 694–706.
[36] M. Corona, N.A. Valdez-Cruz, E. Merino, M. Zurita, L.D. Possani, Genes and peptides from the scorpion *Centruroides sculpturatus* Ewing, that recognize Na+-channels, Toxicon 39 (2001) 1893–1898.
[37] T.I. Zaki, J.E. Maruniak, Three polymorphic genes encoding a depressant toxin from the Egyptian scorpion *Leiurus quinquestriatus quinquestriatus*, Toxicon 41 (2003) 109–113.
[38] G. Estrada, B.I. Garcia, E. Schiavon, E. Ortiz, S. Cestele, E. Wanke, P. L.D., G. Corzo, Four disulfide-bridged scorpion beta neurotoxin CssII: heterologous expression and proper folding in vitro, Biochim. Biophys. Acta 1770 (2007) 1161–1168.
[39] I. Karbat, L. Cohen, N. Gilles, D. Gordon, M. Gurevitz, Conversion of a scorpion toxin agonist into an antagonist highlights an acidic residue involved in voltage sensor trapping during activation of neuronal Na+ channels, FASEB J. 18 (2004) 683–689.

Please cite this article as: G. Corzo, et al., Solution structure of Cn5, a crustacean toxin found in the venom of the scorpions *Centruroides noxius* and *Centruroides suffusus suffusus*, Biochim. Biophys. Acta (2009), doi:10.1016/j.bbapap.2009.07.006