Primary and NMR Three-Dimensional Structure Determination of a Novel Crustacean Toxin from the Venom of the Scorpion *Centruroides limpidus limpidus* Karsch†

F. Lebreton and M. Delepierre*

Institut Pasteur, 25 rue du Dr. Roux, 75724, Paris Cedex 15, France

A. N. Ramírez, C. Balderas, and L. D. Possani

Department of Biochemistry, Biotechnology Institute, National Autonomous University of Mexico, Av. Universidad, 2001 Cuernavaca, Mexico 62271

Received April 4, 1994; Revised Manuscript Received June 30, 1994®

ABSTRACT: A crustacean-specific toxin from the Mexican scorpion *Centruroides limpidus limpidus* was purified, and its primary sequence was determined, including disulfide bonds. This toxin has 66 amino acid residues and is stabilized by four disulfide bridges (Cys12–Cys65, Cys16–Cys41, Cys25–Cys46, and Cys29–Cys48). A detailed nuclear magnetic resonance structure of this protein was obtained using a combination of two-dimensional proton NMR experiments. The NMR parameters that gave 69 dihedral restraints and 418 distance constraints were used in molecular dynamics calculations in order to determine the solution conformation of the toxin. It is composed of a short α-helix and a three-stranded antiparallel β-sheet. Although the regular secondary structure of this crustacean toxin is common to the structural motif of other scorpion toxins, detailed conformational analysis was performed in order to highlight structural features that might be responsible for the differential modulation of the toxin on sodium channels of distinct tissues: mammalian versus crustacean.

Scorpion toxins have been shown to affect ion permeability of excitable cells (Catterall, 1980). Three main distinct families of proteins have been described, specific for Na^+ (Zlotkin *et al.*, 1978; Rochat *et al.*, 1979), K^+ (Carbone et al., 1982; Possani et al., 1982; Miller *et al.*, 1985), and Ca^{2+} channels (Valdivia *et al.*, 1992). Recently, a new peptide that affects Cl^- permeability was also reported (DeBin *et al.*, 1993). The most thoroughly studied proteins are those that recognize Na^+ and K^+ channels (Possani, 1984; Lazdunski *et al.*, 1986; Meves *et al.*, 1986; Moczydlowski *et al.*, 1988; Blaustein *et al.*, 1991). Despite the fact that their primary structures can be quite different, there is a constant structural motif conserved between these two families of proteins: an α-helix and three strands of β-sheet structure, stabilized by either four (Na^+ channel) or three (K^+ channel) disulfide bridges (Kopeyan *et al.*, 1974; Sugg *et al.*, 1990). Among the conserved amino acids of the primary structures are the cysteine residues and two peptide stretches with the sequences Cys-X-X-X-Cys and Cys-X-Cys (Kobayashi *et al.*, 1991; Menez *et al.*, 1992), where X is a variable amino acid. These segments of the primary structure seem to be responsible for the formation of the constant motif underlying the main structural core of the scorpion toxins. However, the acceptor molecules (ion channels) to which these proteins bind are widely variable (Catterall, 1991). Whereas some toxins are specific for mammals, insects, or crustaceans (Zlotkin *et al.*, 1991), others are not species restricted and can affect both mammals and insects (Zlotkin *et al.*, 1978; Possani, 1984; Loret *et al.*, 1991; Zlotkin *et al.*, 1991). Certain toxins (α-scorpion toxins) which bind to site 3 of the mammalian Na^+ channels, modify the closing mechanism of the channel (Catterall, 1980; Jover *et al.*, 1980; Catterall, 1991); others, which bind to site 4 (β-scorpion toxins), modify the activation mechanism of the channel (Catterall, 1980; Jover *et al.*, 1980; Kirsch *et al.*, 1989; Catterall, 1991). The same toxin may bind excitable cells and nonexcitable cells with different affinities (Vaca *et al.*, 1993). Toxins of very similar amino acid sequence may bind to the same tissue with widely variable affinities (Bablito *et al.*, 1986; De Lima *et al.*, 1986). Thus, the species specificity, the type of channel affected, and the binding affinities of the scorpion toxins seem to depend on variations of the amino acid sequence, which can modify the spatial arrangement and charge distribution, thus influencing the fine tuning of the molecular mechanism of action.

Several comparative reports in the literature discuss similarities and differences among the sequences of various scorpion toxins and postulate possible structural characteristics that might be involved in the determination of the function and the species specificity (Fontecilla-Camps *et al.*, 1980; Possani, 1984; Meves *et al.*, 1986; Loret *et al.*, 1991; Menez *et al.*, 1992). In our opinion, however, for comparative purposes it is of the utmost importance to obtain information on the three-dimensional structure of such proteins, before trying to correlate structures with known functions.

In the literature, 3D[1] structures for several mammalian toxins, determined by X-ray and/or NMR methods have been reported (Fontecilla-Camps *et al.*, 1988; Pashkov *et al.*, 1988; Martin *et al.*, 1990; Bontems *et al.*, 1991; Johnson & Sugg, 1992; Mikou *et al.*, 1992; Zhao *et al.*, 1992). Other publications give structural information of neurotoxic peptides

† The work in Mexico was partially financed by grants from the Howard Hughes Medical Institute, No. 75191-527104; DGAPA—National University of Mexico, No. IN205893; and CONACYT, Mexican Council for Science and Technology, No. 18-N9105 to L.D.P. The work in Paris was supported by funds from the Pasteur Institute and the National Center of Scientific Research.

* Address correspondence to this author at Institut Pasteur, Laboratoire de Résonance Magnétique Nucléaire, 28 rue du Dr. Roux, 75015 Paris, France.

® Abstract published in *Advance ACS Abstracts,* August 15, 1994.

0006-2960/94/0433-11135$04.50/0

belonging to the insect toxin group (Fontecilla-Camps *et al.*, 1981; Almassy *et al.*, 1983; Arseniev *et al.*, 1984; Nettesheim *et al.*, 1989; Fontecilla-Camps, 1989; Darbon *et al.*, 1991).

The toxin isolated from the venom of the scorpion *Centruroides limpidus limpidus* displays a remarkable specificity for Na^+ channels of crustaceans and is the first example of this class for which the structure is determined. The first part of this paper concerns the toxin purification, primary structure determination, and disulfide pair arrangements. The second part deals with the toxin ^1H NMR assignments, the secondary structure determination, and the refinement of the global structure. Additional structural comparisons are performed with other known toxins and are discussed.

MATERIAL AND METHODS

Isolation and Primary Structure Determination of the Centruroides limpidus limpidus Toxin

Source of Venom. Venom was collected and prepared in the laboratory, by electrical stimulation of anesthetized animals, as described (Dent *et al.*, 1980).

Chemicals and Reagents. Only analytical grade chemicals were used throughout this work. Reagents for sequencing were all obtained from MilliGen/Biosearch, Division of Millipore (Burlington, MA). Acetonitrile and trifluoroacetic acid were from Pierce (Rockford, IL). Trypsin and chymotrypsin were sequence grade enzymes from Boehringer Mannhein (Mannhein, Germany), while protease V8 from *Staphylococcus aureus* was a gift from Dr. Brian Martin (NIMH, Bethesda, MD). Sephadex G-50 medium was from Pharmacia Fine Chemicals (Uppsala, Sweden). (Carboxymethyl)cellulose (CM-32) was from Whatman (Clifton, NJ). Reagents for reduction and alkylation were from Sigma (St. Louis, MO). Water deionized and double distilled, over quartz, was used for purification and chemical characterization.

Separation Procedure. Separation of toxins was performed by a series of chromatographic steps, similar to previously reported protocols (Ramírez *et al.*, 1988). Briefly, the first step consisted of Sephadex G-50 (medium) gel filtration, followed by (carboxymethyl)cellulose (CM-cellulose) column chromatography and high-performance liquid chromatography (HPLC). Technical information is given in the figure captions and in the Results. Homogeneity of the samples was judged by polyacrylamide gel electrophoresis in two systems (Reisfeld *et al.*, 1962; Laemmli, 1970), by the HPLC profile and by direct sequencing of the proteins. For calculation of recoveries from chromatographic steps, the samples were read at 280 nm. Columns were run at room temperature (298 K).

Bioassays. Venom and fractions were assayed in three animal models: albino mouse (strain CD1); "acociles" (*Cambarellus montezumae* spp.), a freshwater crayfish (crustacean) native of the state of Morelos, Mexico; and insects (crickets). Mice were inoculated intraperitoneally; crickets and crustaceans were dorsally inoculated between the third and fourth most distal segments of the postabdomen. Appropriate controls were conducted using water, saline, or various amounts of the buffers used for separation. HPLC samples were systematically dried (Speed-Vac) to eliminate solvents and trifluoroacetic acid before they were dissolved in water or appropriate buffers for assay. Three designations were used to quantitate the effects: "Nontoxic" means that the animal showed no symptoms of intoxication and behaved in a way similar to those inoculated with saline or buffer alone (controls); "Toxic" means the animal showed one or more of several symptoms—restlessness (excitability), convulsions, partial paralysis, total limb paralysis—but did not die with the dosage used; "Lethal" means the animal showed the symptoms described and died within 20 h after inoculation.

Sequence Determination. Purified toxin was sequenced using a ProSequencer Model 6600 apparatus from MilliGen/Biosearch (Division of Millipore), which automatically performs the Edman degradation of samples either adsorbed into Immobilon-P or covalently attached to Sequelon-DITC or Sequelon-AA membranes, following protocols described by the company. Two types of samples were used: (i) native toxin as well as reduced and carboxymethylated toxin (1 nmol of each loaded) for direct sequencing and (ii) isolated peptides obtained by HPLC fractionation of toxin after cleavage with enzymes (also 1 nmol of each). Trypsin and *Staphylococcus aureus* protease V8 digestions were performed with reduced and carboxymethylated toxin (100 μg each time) in 100 mM ammonium bicarbonate buffer, pH 7.8, for 4 h at 310 K. Chymotryptic digestion was performed with native toxin (1 mg) in the same buffer, pH 7.8, overnight at 310 K.

Disulfide Bridge Determination. The disulfide bridges were assigned by sequencing heterodimeric and polymeric peptides obtained after mild hydrolysis of native toxin and HPLC separation of the corresponding fragments, by a procedure similar to that described by Sugg *et al.* (1990). For these purposes, trypsin and chymotrypsin were used under the experimental conditions described above. Amounts of 1 mg of toxin were used for each protocol. The entire process was repeated twice, independently. When needed, a second HPLC separation of partially clean fragments was performed in a C18 microbore column, using a Millipore system, equipped with a Waters 996 photodiode array detector.

NMR Spectroscopy

Sample Preparation. The lyophilized toxin was dissolved in H_2O/D_2O, 9:1 (v/v), or in 99.96% D_2O (Euriso-Top). The toxin concentration was 1 mM in both H_2O, and D_2O, and the solution pH was adjusted to 3.5 by adding drops of dilute ^2HCl or NaO^2H.

^1H NMR Experiments. ^1H NMR experiments at 303 or 318 K were run at 500 MHz on a Varian Unity spectrometer equipped with a Sun Sparc 2 computer. Data processing was carried out with the VNMR 4.1 program. The sweep width was 5999.7 Hz in 90% H_2O/10% D_2O or 5424.5 Hz in D_2O. Spectra were referred to the water signal at 4.73 ppm at 303 K and at 4.58 ppm at 318 K relative to 3-trimethylsilyl[2,2,3,3-$^2H_4$]propionate (TMSP) as external reference. Quadrature detection was employed in all experiments with the carrier frequency always maintained at the solvent resonance. The 2D ^1H NMR spectra were recorded in the phase-sensitive mode (States *et al.*, 1982) with 2K data points in the t_2 dimension and 512 t_1 increments. For P.COSY experiments 640 t_1 experiments were collected with 4K data points. Zero-filling was applied prior to Fourier transformation, and data were processed with shifted sine bell window functions in both dimensions. The strong signal from the solvent (HDO and H_2O) was suppressed by selective saturation during the recycling delay and, for NOESY, during the mixing period.

Identification of spin systems was obtained through analysis and comparison of two-dimensional P.COSY (Marion & Bax,

[1] Abbreviations: CII, *Centruroides limpidus limpidus* Karsch; CsE, *Centruroides sculpturatus* Ewing; AaH, *Androctonus australis* Hector; NMR, nuclear magnetic resonance; 2D, two dimensional; 3D, three dimensional; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; COSY, correlated spectroscopy; TOCSY, total correlated spectroscopy; rms, root mean square; AMX, spin system of three different protons.

1988), clean TOCSY (Bax & Davis, 1985; Griesinger *et al.*, 1988), DQ (two-dimensional double-quantum spectroscopy; Boyd *et al.*, 1983), and NOESY (Kumar *et al.*, 1980) spectra. TOCSY spectra were recorded using a MLEV-17 pulse scheme with an 80-ms isotropic mixing period. NOESY spectra were acquired at 303 K in D_2O with a mixing time of 100 ms and in H_2O with mixing times of 60, 100, 150, 200, and 300 ms. NOE buildup showed that the spin-diffusion effects were important only for mixing times greater than 200 ms. Thus, peak volumes were measured on NOESY spectra run with a mixing time of 100 ms. A post-acquisition treatment to suppress the residual water signal in NOESY spectra recorded in 90% H_2O/10% D_2O was applied (Sodano & Delepierre, 1993).

Proton–deuterium exchange of the amide groups of *Centruroides limpidus limpidus* scorpion toxin was measured on a sample lyophilized from H_2O and redissolved in pure D_2O. One-dimensional ^1H spectra were collected sequentially for several hours to follow the residual NH signals at 303 K. Clean TOCSY, NOESY, and P.COSY experiments were then recorded at 303 K to identify the locations of the slowly exchanging amide protons.

NH–CαH and CαH–CβH coupling constants were obtained from P.COSY experiments collected respectively in H_2O and D_2O. In these experiments, the digital resolution after zero filling along ω_1 was 0.7 Hz.

Determination of Dihedral and Distance Constraints. Distance constraints as well as ϕ and χ angle constraints were used in the molecular-modeling calculations. Sixty-nine angle constraints were derived from $J_{NH-C\alpha H}$ and $J_{C\alpha H-C\beta H}$, respectively. Thirty-one ϕ backbone torsion angles were restrained within a variable interval centered on −120° depending on the value of $J_{NH-C\alpha H}$ (Pardi *et al.*, 1984): for $J > 9$ Hz, $\phi = -120° \pm 35°$; for $J > 8$ Hz, $\phi = -120° \pm 55°$. For $J < 5$ Hz, the ϕ backbone torsion angles were constrained between −90° and −40°, and for $J < 7$ Hz, $\phi = -80° \pm 25°$. The range of χ angle constraints were determined from a combined use of $J_{C\alpha H-C\beta H}$ and NOE information between Hα/Hβ and NH/Hβ (Zuiderweg *et al.*, 1985; Wagner *et al.*, 1987; Basus, 1989). Using this procedure, the stereospecific assignments of a few β protons were possible, and 14 χ angle conformations could be defined as g^+, t, and g^-, which permitted definition of χ angle constraints.

Four hundred and eighteen distance constraints were calculated after integration and calibration of NOE cross peaks. The calibration was based on the standard distance of 0.22 nm for the $d_{\alpha N}$ cross peaks in β-strands (Wüthrich, 1986). In a first approximation, the NOE intensity between a pair of protons was considered to be directly proportional to the inverse sixth power of the distance separating the protons. The NOEs were divided into three classes, weak, medium and strong, with a lower limit of 0.19 nm and with upper limits of 0.26, 0.33, and 0.5 nm for intraresidual and interresidual distance constraints. Pseudoatom corrections were used for non-stereospecifically assigned protons, especially for methyl protons.

Constraints were added to define the four disulfide bridges. The covalent distance between two sulfur atoms, Sγ–Sγ, was restrained between 0.19 and 0.25 nm. Additional constraints were derived from the presence of hydrogen bonds given by the amide-exchange proton data and by several NOEs characteristic of secondary structures. For the slowest hydrogen to exchange, a distance restraint of 0.19–0.25 nm was used between hydrogen and oxygen. All of this information concerning the geometry of the backbone gave 25 distance constraints.

Protocol for Structure Calculations

Structure calculations of the *Centruroides limpidus limpidus* scorpion toxin were carried out using simulated annealing (SA; Clore *et al.*, 1986). The SA program belongs to NMRchitect, the NMR-toolbox from Biosym Technologies (San Diego). The structures were displayed and analyzed on a Silicon Graphics Personal IRIS 4D/30 workstation, using the Insight II (version 2.2) and Discover (version 2.8) packages. All of the dynamical calculations were based on the constant valence force field parameter set (CVFF) (Daubert-Osguthorpe *et al.*, 1988). Several rounds of SA calculations and reexamination of the restraint file were used in order to complete the assignments of the NOESY cross peaks. The data set of 512 experimental constraints included 148 interresidue and 255 intraresidue distance constraints, 25 generic distances, and 55 ϕ and 14 χ dihedral constraints. Constraints were built with an energetic term using skewed biharmonic functions for the lower and upper boundaries for the distances and the dihedral angles.

The starting structure of Cll toxin 1 was built using the available coordinates of CsE-v3 (Almassy *et al.*, 1983) as a template. Fifty identical simulated annealing protocols were performed using a restrained dynamic algorithm as described in Nilges *et al.* (1988). Starting from a random array of atomic coordinates, this refinement protocol began with 1-ps step heating to 1000 K and a 30-ps propagation step at 1000 K. The NOE restraint force constant K_{NOE} was increased gradually from 0.01 to 10 kcal mol^{-1} Å^{-2}, and the nonbonded term was left at 0.001 kcal mol^{-1} Å^{-4} during this phase. During the folding phase of 50 ps, K_{NOE} was increased to its full value (20 kcal mol^{-1} Å^{-2}), while the internal force constants were scaled up in small steps until they reached 450 kcal mol^{-1} Å^{-2} or 450 kcal mol^{-1} rad^{-2} and the nonbonded force constant reached 0.25 kcal mol^{-1} Å^{-4}. The last refinement stage consisted of 10 ps of restrained dynamics during which the temperature was cooled to 300 K. (At the end of this phase, the nonbonded force constant reached 1 kcal mol^{-1} Å^{-4}.) These 50 resulting structures were then energy minimized and selected by the following criteria: no distance violation greater than 0.1 nm, no dihedral violation greater than 30°, and no chirality violation. The remaining structures showed an identical global fold. The most energetically favorable structure was used as a starting structure for a restrained dynamic calculation consisting of several stages. This modeling protocol started with restrained dynamics at 300 K for 5 ps (with K_{NOE} increasing from 10 to 60 kcal mol^{-1} Å^{-2}). The temperature was then decreased to 50 K in 50 K steps of 2 ps (with K_{NOE} decreasing, from 60 to 10 kcal mol^{-1} Å^{-2}), and the resulting structures were minimized. Among the 20 resulting structures, the best were selected using the criteria that they present the minimal number of distance violations larger than 0.05 nm, the minimal number of angular violations larger than 5°, and acceptable ϕ/ψ angles in the Ramachandran map. These structures were used to back-calculate the theoretical matrix of NOESY cross-peak intensities using the IRMA program (iterative relaxation matrix approach; Boelens *et al.*, 1989), included in the NMRchitect package. The matrices of theoretical and experimental NOE cross-peak volumes were merged to calculate a new rate matrix and a set of distances which were used to optimize the restraints file. These updated distances served as new distances for restrained dynamics to generate new structures which were analyzed, classified and used as starting structure for successive IRMA/SA cycles. We used a correlation time of 3 ns (confirmed by Stocke's equation; Boere & Kidd, 1983) and two mixing times of 100 and 300 ms for the IRMA cycles.

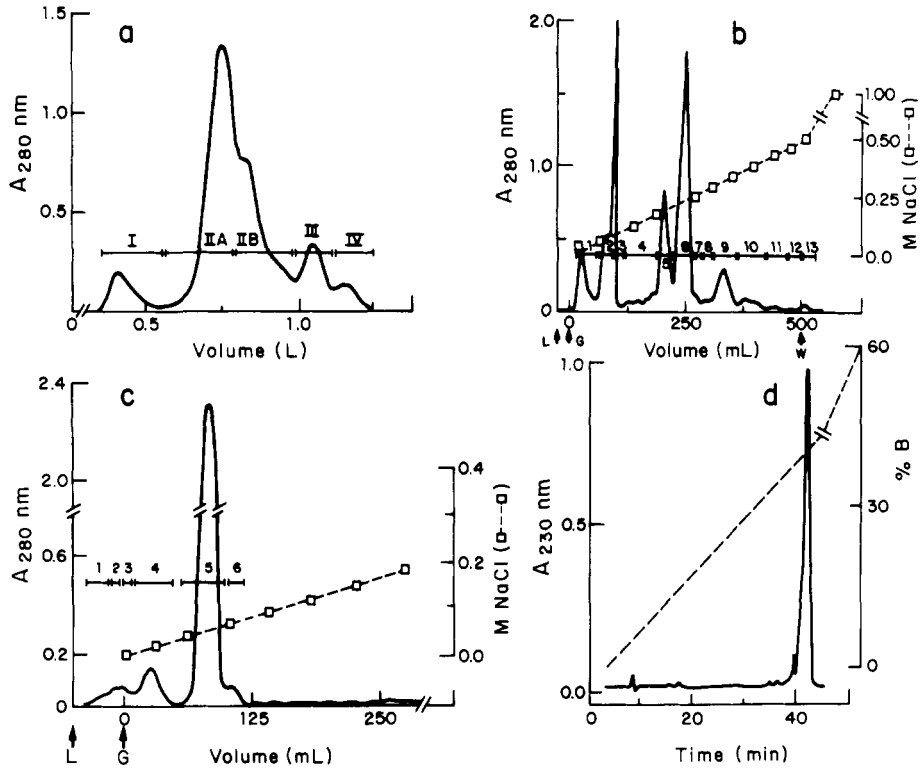

FIGURE 1: Purification of toxin 1 by chromatography. (a) Soluble venom from *C. limpidus l.* (301 mg in 8 mL) was applied to a Sephadex G-50 (medium) column (size: 3 × 200 cm) and eluted with 20 mM ammonium acetate buffer, pH 4.7, at flow rate of 30 mL/h; 6.5-mL fractions were collected. Material absorbing at 280 nm was pooled in five subfractions, as indicated by the horizontal bars (I–IV). Recoveries were 6.1 (I), 36.0 (IIA), 17.1(IIB), 9.4 (III), and 3.8% (IV), respectively, with 82.4% overall recovery. (b) Fraction IIB (135 mg, from various Sephadex G-50 columns) was applied to a CM-cellulose column (0.9 × 30 cm) equilibrated and run in the presence of 20 mM ammonium acetate buffer, pH 4.7, at a flow rate of 30 mL/h; 2.5 mL was collected from each tube. Elution was obtained with a linear gradient, by mixing 250 mL of buffer alone with 250 mL of buffer containing 0.5 M NaCl (dashed lines). A solution with 1 M NaCl was used for washing at the end of the gradient. Fractions were pooled as indicated by the horizontal bars (1–13). Overall recovery was 89%, from which fraction 6 corresponded to 33%. L means loading the sample; G, starting the gradient; W, washing with 1 M NaCl, dissolved in the same buffer. (c) Fraction IIB-6 (41.5 mg), from previous columns (panel b), was separated in identical manner as in panel b, except for the buffer (50 mM sodium phosphate, pH 6.0) and the salt gradient, which went up to 0.3 M NaCl. Six subfractions were separated as indicated by the horizontal bars, from which component 5 was about 80% of the material recovered, with overall recoveries of 97.5% in this column. (d) A sample (9.5 mg) of fraction IIB-6-5 was further separated on a preparative C4 reverse-phase column (Vydac, Hisperia, CA), using a Waters 600E liquid chromatograph equipped with a Waters Model 481 variable wavelength detector. A linear gradient was used from solution A (0.12% trifluoracetic acid in water) to 60% solution B (0.10% trifluoroacetic acid in acetonitrile), run for 60 min. The main component was pure toxin 1.

RESULTS

Purification and Primary Structure Determination

Figure 1 summarizes the results of chromatographic steps used to obtain the pure toxin. In the Sephadex column (Figure 1a) four main components were separated. Because component II is lethal to the animals of our bioassays and was poorly resolved, we divided component II into component IIA, which contained most of the mammal-specific toxins, and IIB, which was rich in crustacean toxins. Overall recovery of this column was about 82%, from which component IIB amounts to 27% of the material recovered. This fraction separated into 13 subfractions on a CM-cellulose column (Figure 1b), the last one usually by washing with 1 M NaCl. Both columns were equilibrated and run in the presence of 20 mM ammonium acetate buffer, pH 4.7. Recovery from the CM-cellulose column was 89%, from which subfraction IIB-6 was 33% of the total material recovered. This subfraction was lethal to acociles and was further applied to a CM-cellulose column equilibrated and run with 50 mM phosphate buffer, at pH 6.0 (Figure 1c). Component 5 of this column was lethal to acociles and still contained a small contaminant (about 5.5%) that was eliminated by HPLC, using a C4 reverse-phase column (Figure 1d). Recovery from the CM-cellulose column at pH 6.0 was almost quantitative (97.5%), from which the lethal fraction IIB-6-5 (IIB comes from Sephadex; 6, from CM-cellulose at pH 4.7; and 5, from CM-cellulose at pH 6.0) was about 80% of the material recovered. In this manner, the pure toxin corresponds to about 6.7% of the material recovered. Considering that the overall recovery from the various steps of separation is on the order of 70% (assuming quantitative recovery from the HPLC), this component is about 9.5% of all venom. This protein was shown to be homogeneous by polyacrylamide gel electrophoresis (data not shown), using acidic conditions (Reisfeld *et al.*, 1962) and in the presence of SDS (Laemmli, 1970). Direct amino acid sequencing of native protein gave only one amino acid residue in every step, from which we concluded that this toxin was pure, and we propose to call it crustacean toxin 1 from *Centruroides limpidus limpidus*. The bioassays showed that this protein was not toxic to mouse (using up to 500 μg of protein per 20 grams of mouse body-weight) and barely toxic to insects (using up to 30 μg per cricket), but very effective against crustaceans. Thus, it displays a remarkable specificity for crustaceans.

The total amino acid sequence was determined, as shown in Figure 2. Direct Edman degradation of native toxin (D) gave an unequivocal sequence up to residue Asn24. A reduced and alkylated sample of toxin (DA) was sequenced up to Glu28. Further amino acid sequence information was determined from fragments obtained by enzymatic cleavage: tryptic peptide (T2) confirmed the sequence from Lys27 (site of trypsin

FIGURE 2: Amino acid sequence determination of toxin 1. The ProSequencer Model 6600 from MilliGen/Biosearch (Millipore Co.) was used for the complete determination of the primary structure of toxin 1. Sequencing native toxin gave the first 24 amino acid residues (labeled --D--). A reduced and alkylated sample gave unequivocal results for amino acids 1–28 (labeled --DA--), confirming the first 24 residues. Tryptic peptides (..T1.. and ==T2==) gave residues at positions 16–30 and 28–46, respectively. Protease V8 hydrolyzed toxin-produced peptides (.-V1.- and .=V2.=) from which amino acids at positions 29–49 and 50–66 were identified. Finally a chymotryptic fragment (-=CT-=) was sequenced and overlapped amino acids at positions 48–58. Toxin and peptides were used at about 1 nmol of sample per application. Three types of membranes were used, according to protocols of MilliGen (Millipore): Immobilon-P (for adsorbed peptides) and Sequelon-AA and/or Sequelon-DITC (for covalently attached peptides). Enzymatic hydrolyses provided many distinct peptides after HPLC (data not shown), and most of them were placed into the sequencer. However, we are reporting here only those necessary for construction of an unequivocal primary structure.

action), starting with Glu28 to Cys46; tryptic peptide T1 overlapped sequences from Tyr14 to Lys30; protease-V8 fragment (labeled V1) revealed the sequence from residue Cys29 to Glu49, while peptide V2 continued sequences from Gly50 to Ser66. The linkage between the two last segments of the molecule was confirmed by a chymotryptic fragment (CT), which gave the sequence from Cys48 to Tyr58. Amino acid composition of samples of native toxin (data not shown) confirmed that the last residue (Ser66) was present in trace amounts only. All of the peptide fragments sequenced were obtained after HPLC separation (data not shown) of reduced and carboxymethylated toxin, subjected to the various treatment with specific enzymes. The cysteine residues were identified as carboxymethylated derivatives. Sequence determination of peptide fragments containing the C-terminus showed that they were covalently attached to Sequelon-AA membranes. Practically all of the sequences reported in Figure 2 were re-confirmed during analysis of additional peptides prepared for disulfide bridge determination. For this purpose, several samples of native toxin were hydrolyzed with different enzymes, separated by HPLC (data not shown), and placed in the Sequencer. Table 1 shows a summary of the most significant results, which led to the determination of the following disulfide bridges: Cys12–Cys65, Cys16–Cys41, Cys25–Cys46, and Cys29–Cys48, for crustacean toxin 1 of *C. limpidus l.* The three first pairs were directly sequenced as single heterodimers, with unequivocal interpretation. The last one was obtained as a trimeric peptide, whose amino acid sequences confirmed the assignment of Cys25–Cys46, and gave results consistent with the fourth pair: Cys29–Cys48.

NMR Three-Dimensionnal Structure Determination

NMR signal assignments of the *Centruroides limpidus limpidus* toxin 1 were obtained by the standard two-stage procedure at 303 K (Wüthrich, 1986).

Identification of Spin Systems. Unique amino acid types in the primary sequence of Cll toxin 1 were identified from their characteristic chemical shift patterns in the TOCSY spectra in H_2O and were used as starting points for the sequential assignment. From the amide protons, most of the spin systems were identified via relayed TOCSY connectivities combined with P.COSY connectivities observed in H_2O (Figure 3). Side-chain proton assignments were made on the basis of the TOCSY patterns of the α, β, γ, δ, and ϵ protons and via direct coupling in the P.COSY spectra obtained in H_2O and especially in D_2O. Double-quantum experiments were performed and analyzed when coupled peaks had only small chemical shift differences, such as for glycine Hα or for some Hβ protons. The DQ spectrum in H_2O was also used to complete lysine side-chain assignment and in particular the NH_ϵ–H_δ connectivities. Aromatic residues were distinguished from other AMX systems on the basis of the NOE observed between certain aromatic ring protons and the Hβ protons. NOE effects in H_2O were also used to assign the amino protons of the glutamine and asparagine residues with the exception of Asn7. Intraresidue assignments were confirmed in the NOESY spectra for all residues but Lys1 and Cys29 which showed only very weak intraresidue connectivities in the NOESY spectra.

Sequential Assignments. The NOESY spectra run at two mixing times, 100 (Figure 4a) and 300 ms, and two

Table 1: Amino Acid Sequence of Heterodimeric Peptides Corresponding to the Disulfide Bridges of Toxin 1[a]

retention time HPLC	amino acid sequence	corresponding disulfide
CT-22.21	1 2 3 4 5 6 7 Val-Asn-Lys-Ser-Thr-Gly-CYS(Lys-Tyr) Lys-Ser (Xxx-Ser)	Cys12–Cys65
T-39.20 and CT-28.00	1 2 3 4 5 6 7 8 9 Asn-Gln-Gly-Gly-Ser-Tyr-Gly-Tyr-CYS(Tyr) Tyr-Gly (Tyr-Xxx-Tyr)	Cys16–Cys41
CT-36.65 and T-34.19	1 2 3 4 5 6 7 8 9 10 Trp-Leu-Gly-Lys-Asn-Glu-Asn-CYS-Asp-Lys Ser-Phe-Ala (Xxx-Trp)	Cys25–Cys46
CT-34.74 and T-30.00	1 2 3 4 5 6 7 8 9 10 11 12 Leu-Gly-Lys-Asn-Glu-Asn-CYS (Asp-Lys-Glu-Xxx-Lys) Ser-Phe-Ala-Xxx-Trp Xxx-Glu-Gly-Leu-Pro-Glu-Ser-Thr	Cys29–Cys48 and Cys25–Cys46

1 10 20 30 40 50 60

Toxin 1 KEGYLVNKSTGCKYGCFWLGKNENCDKECKAKNQGGSYGYCYSFACWCEGLPESTPTYPLPNKSCS

[a] Amino acid sequences were obtained by microsequencing heteropolymeric peptides cleaved from native toxin by treatment with chymotrypsin (CT) or trypsin (T), or a combination of both, and separated by HPLC at the times indicated in minutes. Numbers above the amino acids are the positions in the sequence; Xxx represents a blank position corresponding to a half-Cys residue (cysteine), while CYS indicates the position of a cystine residue (usually in our machine we can identify a component that absorbs at 330 nm which corresponds to this amino acid). Amino acids in parentheses were seen in trace amounts or were surmised on the basis of the specificity of the enzyme that cleaved the toxin and on its sequence shown immediately above (one-letter code; cysteinyl residues are in bold).

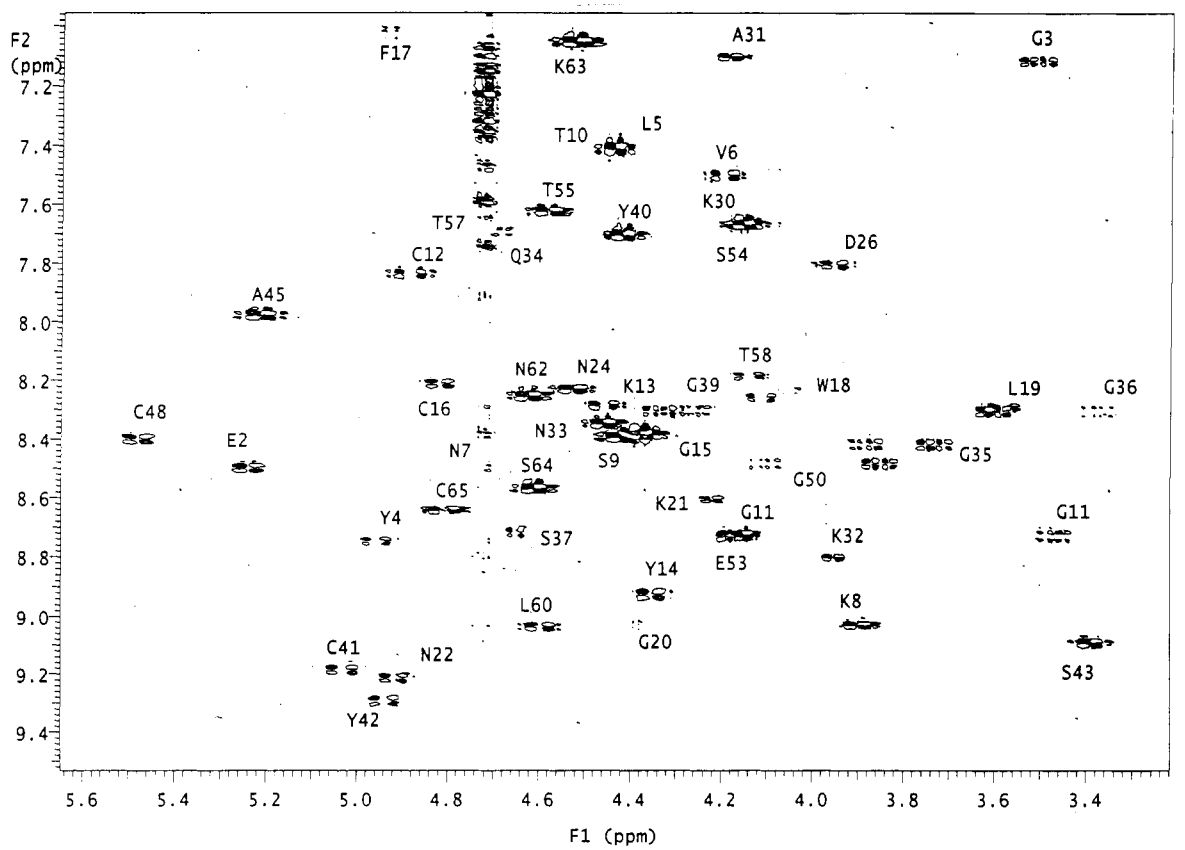

FIGURE 3: Fingerprint for the *Centruroides limpidus limpidus* toxin 1 identifying the HN/Hα P.COSY cross peaks at 303 K. Fifty assigned HN/Hα cross peaks are labeled. Few missing cross peaks were observed in the TOCSY spectra and in spectra at 318 K.

temperatures, 303 and 318 K, were analyzed to perform the sequential assignment (Wüthrich, 1986).

Starting with Ala31, the assignment proceeded in the C-terminal direction up to Tyr38 since none of the sequential connectivity to Gly39 was detected. However, observation of cross peaks between Tyr38 and Tyr40 and between Gly39 and Tyr40 allowed the identification of these residues. Starting with Ala45 and going in both directions, C- and N-terminal,

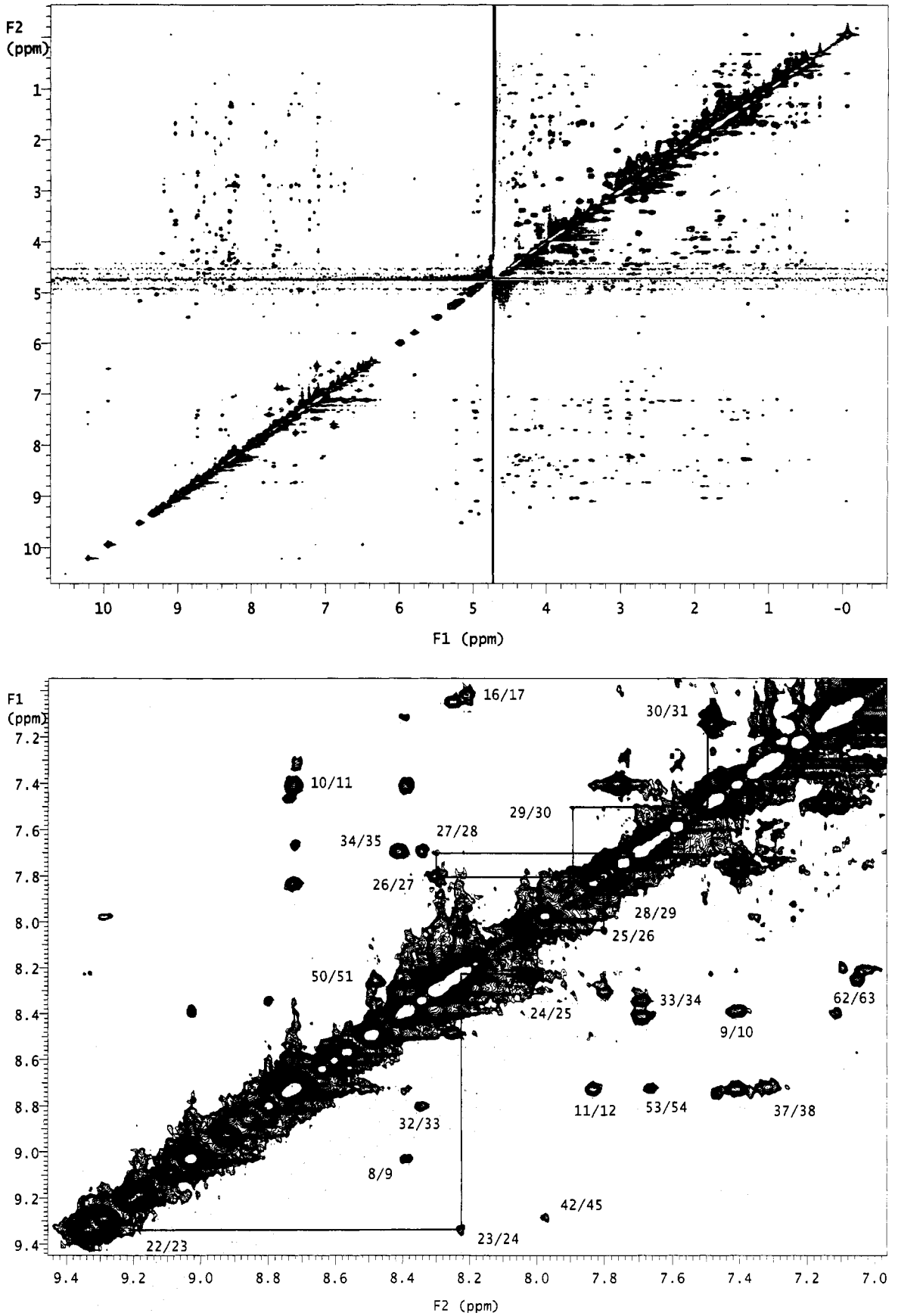

FIGURE 4: (a) Contour plot of the 100-ms NOESY spectrum of the *Centruroides limpidus limpidus* toxin 1 at 303 K. (b) Contour plot of the amide proton region of a 100-ms NOESY spectrum of Cll toxin 1. The sequential d_{NN} connectivities for residues 22–31 are shown.

via $d_{\alpha N}$, d_{NN} but also $d_{\beta N}$, connectivities, it was possible to assign unambiguously the Ala31 to Ser66 fragment. No sequential connectivity was found for Ser43 and Phe44, but NOEs between residues Tyr42 and Phe44 and between Phe44 and Ala45 were observed. In the C-terminal region, there are several prolines, and their positions in the sequence were characterized through either $H_{\delta i}/NH_{i+1}$ or $H_{\alpha i-1}/H_{\delta i}$ connectivities or both. The very strong NOE between the Pro59

Table 2: ^{1}H Chemical Shifts for the Toxin from the Venom of the Scorpion *Centruroides limpidus limpidus* (pH = 3.5, 303 K)[a]

residue	NH	αH–α′H	βH–β′H	γH–γ′H	δH–δ′H	others		residue	NH	αH–α′H	βH–β′H	γH–γ′H	δH–δ′H	others	
Lys1	nd	4.88	nd	nd	nd	nd		Asn33	8.34	4.46	2.93–2.77			γNH_2	6.89–7.64
Glu2	8.50	5.24	3.06–2.23	2.47–2.03				Gln34	7.69	4.67	2.5–2.31	2.03–2.10		δNH_2	6.46–7.12
Gly3	7.12	3.50–1.66						Gly35	8.41	3.74–3.89					
Tyr4	8.75	4.96	2.89–3.15			2,6H	7.47	Gly36	8.30	4.34–3.38					
						3,5H	7.35	Ser37	8.72	4.65	3.92–3.84				
Leu5	7.41	4.44	1.63–1.73	1.53	0.89–0.65			Tyr38	7.31	4.98	3.41–2.58			2,6H	6.72
Val6	7.50	4.20	1.48	0.51–0.63										3,5H	7.16
Asn7	8.50	4.72	2.48–2.67			γNH_2	nd	Gly39	8.30	4.27–3.60					
Lys8	9.03	3.90	1.88	1.76–1.65	1.45	ϵCH_2	3.05	Tyr40	7.70	4.42	2.85–3.20			2,6H	6.75
						ϵNH_2	7.65							3,5H	6.65
Ser9	8.40	4.42	3.96–3.98					Cys41	9.19	5.04	2.75–3.82				
Thr10	7.40	4.46	4.51	1.08				Tyr42	9.30	4.95	2.57–2.9			2,6H	6.83
Gly11	8.72	4.17–3.47												3,5H	6.38
Cys12	7.83	4.89	2.69–3.4					Ser43	9.09	3.39	3.35–4.02				
Lys13	8.29	4.46	1.66–1.28	nd	nd	nd		Phe44	6.00	3.76	3.54–3.21			2,6H	7.22
Tyr14	8.92	4.35	2.67–2.91			2,6H	6.93							3,5H	7.32
						3,5H	6.55							4H	7.10
Gly15	8.37	3.88–4.38						Ala45	7.98	5.21	1.30				
Cys16	8.21	4.82	2.69–3.19					Cys46	8.70	5.16	2.65–nd				
Phe17	7.03	4.93	3.4–3.2			2,6H	7.09	Trp47	9.50	4.61	2.77–nd			NH	9.93
						3,5H	7.23							2H	6.50
						nd								4H	5.80
Trp18	8.23	4.05	2.90–nd			NH	10.21							5H	6.62
						2H	7.35							6H	7.07
						4H	7.75							7H	7.145
						5H	7.29	Cys48	8.40	5.48	2.48–nd				
						6H	7.32	Glu49	8.86	4.75[b]	2.67–nd	2.28–2.10			
						7H	7.58	Gly50	8.48	4.10–3.85					
Leu19	8.30	3.58	1.65–nd	1.33	–0.05–0.68			Leu51	8.26	4.12	1.35	1.17	0.32–0.73		
Gly20	9.02	3.66–4.40						Pro52		4.59	2.35–2.02	2.76	3.82		
Lys21	8.61	4.22	1.87–nd	1.58–1.76	1.58	ϵCH_2	3.07	Glu53	8.72	4.16	2.71–2.63	nd			
						ϵNH_2	7.55	Ser54	7.67	4.16	4.07–3.79				
Asn22	9.21	4.92	2.89–2.47			γNH_2	7.15–7.48	Thr55	7.62	4.58	4.08	1.53			
Glu23	9.34	4.31	2.13–2.19	2.19				Pro56		4.22	1.87–2.26	2.27–2.05	3.71–3.95		
Asn24	8.23	4.53	3.20–2.83			γNH_2	6.75–6.55	Thr57	7.70	4.68	4.38	0.57			
Cys25	8.04	4.42	2.66–3.20					Tyr58	8.18	4.14	2.66–2.91			2,6H	7.10
Asp26	7.81	3.97	1.62–2.16	2.05–1.87										3,5H	6.90
Lys27	8.31	3.81	1.99–1.87	1.45–1.5	1.70	ϵCH_2	3.01	Pro59		3.65	1.09–1.66	1.72	3.4–3.57		
						ϵNH_2	7.53	Leu60	9.03	4.59	1.89	1.69	0.95–0.89		
Glu28	7.70	3.96	2.32–2.08	2.27–2.02				Pro61		4.09	1.53–2.00	1.73–1.88	3.91		
Cys29	7.90	nd	2.69–nd					Asn62	8.25	4.63	2.83–2.90			γNH_2	6.90–7.59
Lys30	7.50	4.19	1.95–1.84	1.12–1.02	1.63–1.72	ϵCH_2	3.13–2.68	Lys63	7.06	4.52	1.60	1.21	1.55	ϵCH_2	2.94
						ϵNH_2	7.53							ϵNH_2	7.45
Ala31	7.10	4.18	1.55					Ser64	8.56	4.60	3.8–3.8				
Lys32	8.80	3.96	1.89–1.85	1.53	1.71	ϵCH_2	3.00	Cys65	8.64	4.81	3.10–3.21				
						ϵNH_2	7.52	Ser66	8.38	4.17	3.11				

[a] All data were obtained in H_2O/D_2O, 9:1 (v/v), and in D_2O. Chemical shifts are in ppm relative to the protons of H_2O (4.73 ppm downfield from TMSP at 303 K). The protein concentration was 1.0 mM. nd = chemical shift not determined. [b] This chemical shift was found at 318 K.

Hα and the Tyr58 Hα indicates that Pro59 adopts a *cis* conformation. The connectivities presented by the three other proline residues are those observed for *trans* conformation.

In the N-terminal region, sequential assignment was achieved by the analysis of the characteristic through-space connectivities. The Cll toxin 1 assignments are presented in Table 2.

Secondary and Tertiary Structures. In addition to sequential assignment information, NOE effects involving NH, Hα, and Hβ protons provide information about secondary structure. Thus, the presence of NN connectivities from residue 22 to residue 31, together with three $H\alpha_i/H\beta_{i+3}$ and four $H\alpha_i/NH_{i+3}$ connectivities, suggests an α-helical conformation for this fragment. The presence of these typical medium-range cross peaks is important because the d_{NN} connectivities are weaker than expected for a regular α-helix (Figure 4b). The presence of the helix was confirmed both by the small values of the $^3J_{NH\alpha}$ coupling constants, with the exception of Asn22, and by the slow rate of exchange observed in D_2O for some of the amide protons in this region. Furthemore, the observed $d_{\alpha N(i,i+4)}$ allowed us to classify this helix as an α-helix.

Strong $d_{\alpha N}$ sequential connectivities were observed for residues 2–5, 36–41, and 47–51. The last two stretches are connected by the 42–45 fragment which displayed several connectivities characteristic of a tight turn. In addition, a number of long-range d_{NN} [3:48; 38:49], $d_{\alpha N}$ [4:48; 2:50], and $d_{\alpha\alpha}$ [4:47; 41:46; 39:48] connectivities indicated that the structure is a triple-stranded antiparallel β-sheet. This triple-stranded antiparallel β-sheet is consistent with the slow rate of exchange for the amide protons expected to be involved in hydrogen bonds between the strands. There was an alternating pattern of slowly and rapidly exchangeable amide protons observed for strands 2–5 and 36–41, typical of external strands, while all of the amide protons of the central strand 47–51 exchanged slowly with the exception of Gly50. The large $^3J_{NH\alpha}$ coupling constant measured for most of the residues in the β-sheet confirmed the schematic structure represented in Figure 5.

In the NH/NH region of the NOESY spectrum shown in Figure 4b, the strong d_{NN} connectivities for segments 8–12, 16–17, 31–35, 50–51, 53–54, and 62–63 are indicative of the presence of several turns or bends. Some of the NMR data are summarized in Figure 6, as the sequential and medium-range connectivities, the backbone vicinal coupling constant data, and the slowly exchanging amide hydrogens.

In order to analyze further the overall structure of toxin 1, the deviations from the random-coil position of the Hα and NH chemical shifts were analyzed using the chemical shift index method (Wishart *et al.*, 1992). The Hα and NH

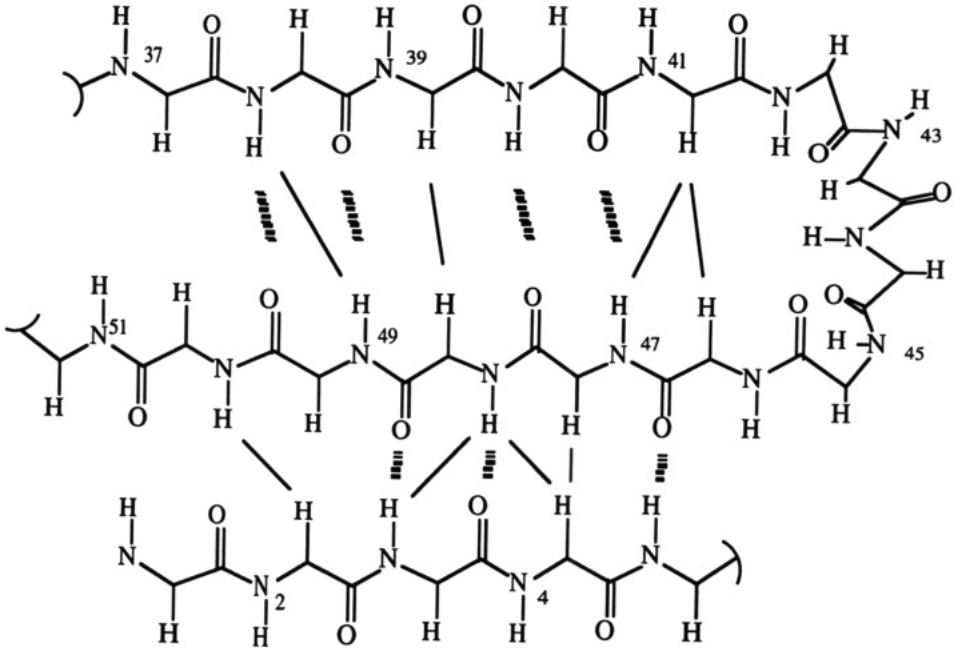

FIGURE 5: Schematic representation of the β-sheet secondary structure seen in the *Centruroides limpidus limpidus* toxin 1. The solid lines show the NOE interstrand connectivities, and the broken lines represent the hydrogen bonds deduced from NH exchange.

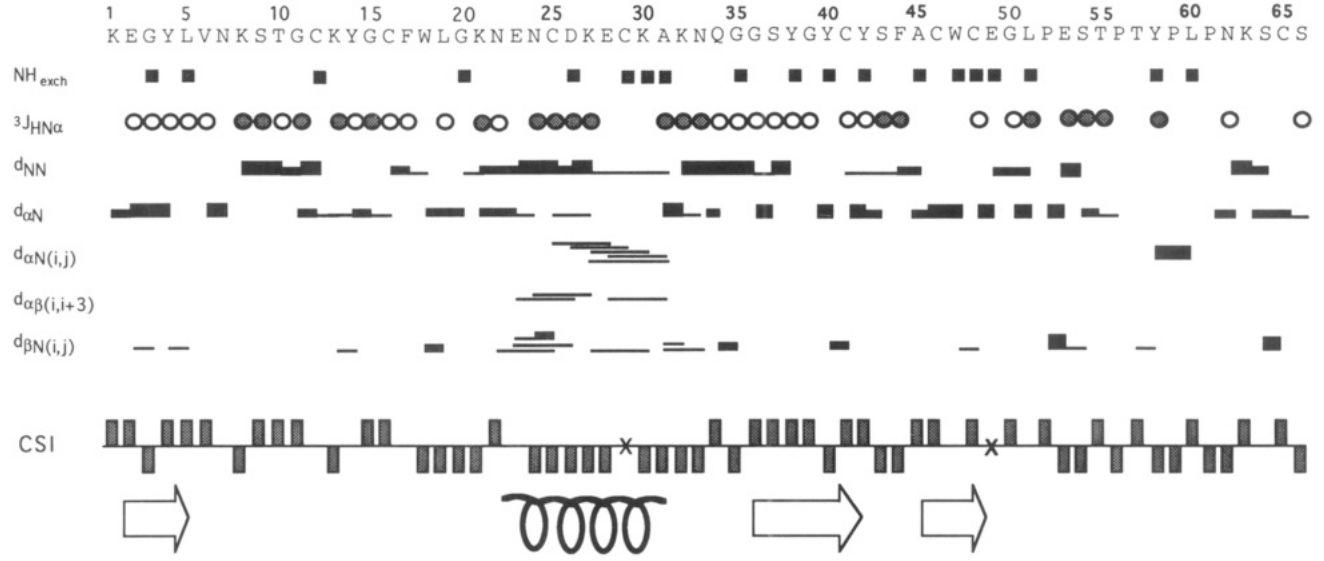

FIGURE 6: Amino acid sequence of CII toxin 1 and summary of the NOE connectivities used in the sequential assignment procedure and the secondary structure determination. Data are summed from NOESY spectra obtained at 303 K, pH 3.5, using a mixing time of 100 ms. For proline, sequential NOEs involving $C_\delta H$ instead of NH are indicated. $^3J_{HN\alpha}$ coupling constants greater than 8 Hz are represented with open circles, while filled circles correspond to coupling constants lower than 7 Hz. The filled squares indicate slowly exchanging amide protons. The chemical shift indices for the Hα protons are indicated at the bottom: no rectangle represents a zero; a rectangle above or below the line represents +1 or −1, respectively; and X indicates insufficient information (Wishart et al., 1992). The locations of the secondary structures (α-helix and β-strand) found in CII toxin 1 are also indicated.

chemical shifts present upfield shifts with respect to the random-coil values in a helical conformation and downfield shifts in a β-strand extended conformation. The results are in reasonably good agreement for Hα protons with what is expected from other NMR parameters (Figure 6). The α-helical structure is observed between residues 24 and 31, while β-strand is seen for residues 36 to 39. However, results for the two other strands, 1–4 and 46–50, are not clear, although the NH secondary shifts observed for residues 46–50 are positive as expected for a β-strand conformation. The presence of numerous aromatic residues in the sequence, especially Tyr4, Tyr38, Tyr40, Tyr42, Phe44, and Trp47, could explain the results obtained with the chemical shift index method. Several NOEs were observed between the ring protons of several aromatic residues, showing that the aromatic residues were close in space and formed a hydrophobic core. The three-dimensional arrays of the rings are confirmed by the ring current effects inducing shifts in the resonances of the protons of these aromatic residues and their closest neighbors (Table 2).

Assignments of the proton resonances of the toxin were compared with those of CsE-v3 (Nettesheim *et al.*, 1989). Figure 7 presents the chemical shift differences for both Hα and amide protons between CsE-v3 and CII toxin 1. This diagram shows that their chemical shifts are very similar and that the largest differences are observed in the region where residues are distinct, that is, mainly in the N-terminal region. The fact that the resonances of both toxins have small deviations shows that they have common elements of secondary structure.

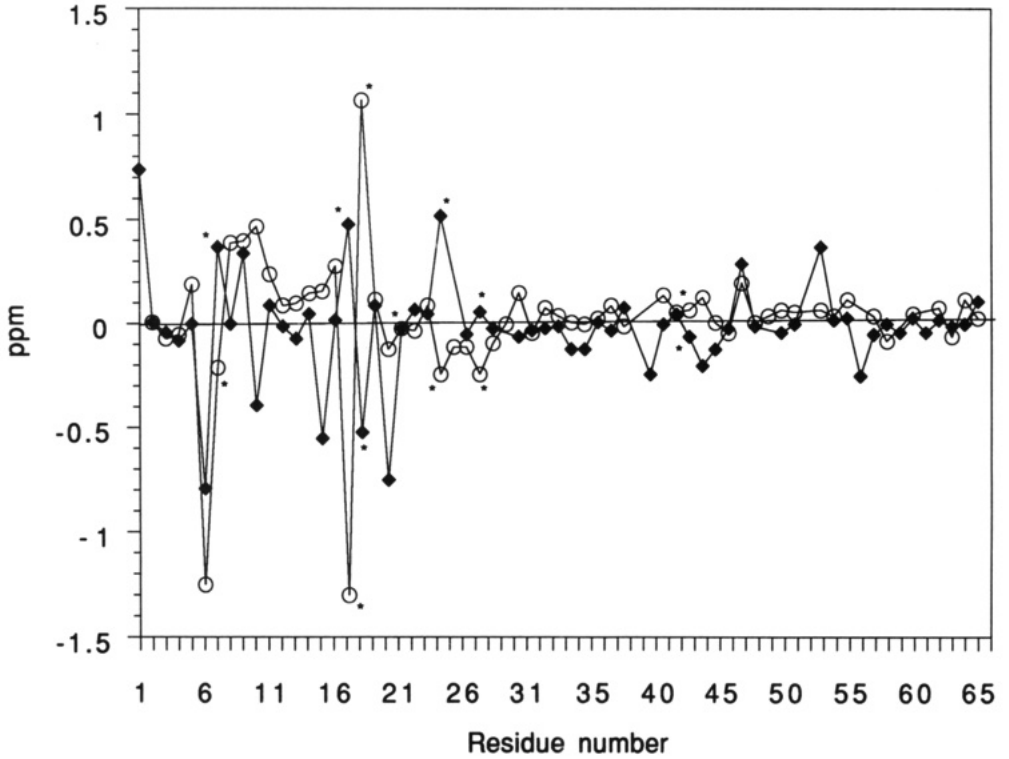

FIGURE 7: Representation of CαH (filled diamonds) and NH (open circles) chemical shift differences between Cll toxin 1 and CsE-v3. Sequence numbering corresponds to the amino acids of Cll toxin 1. Asterisks indicate residues that differ between the two toxins.

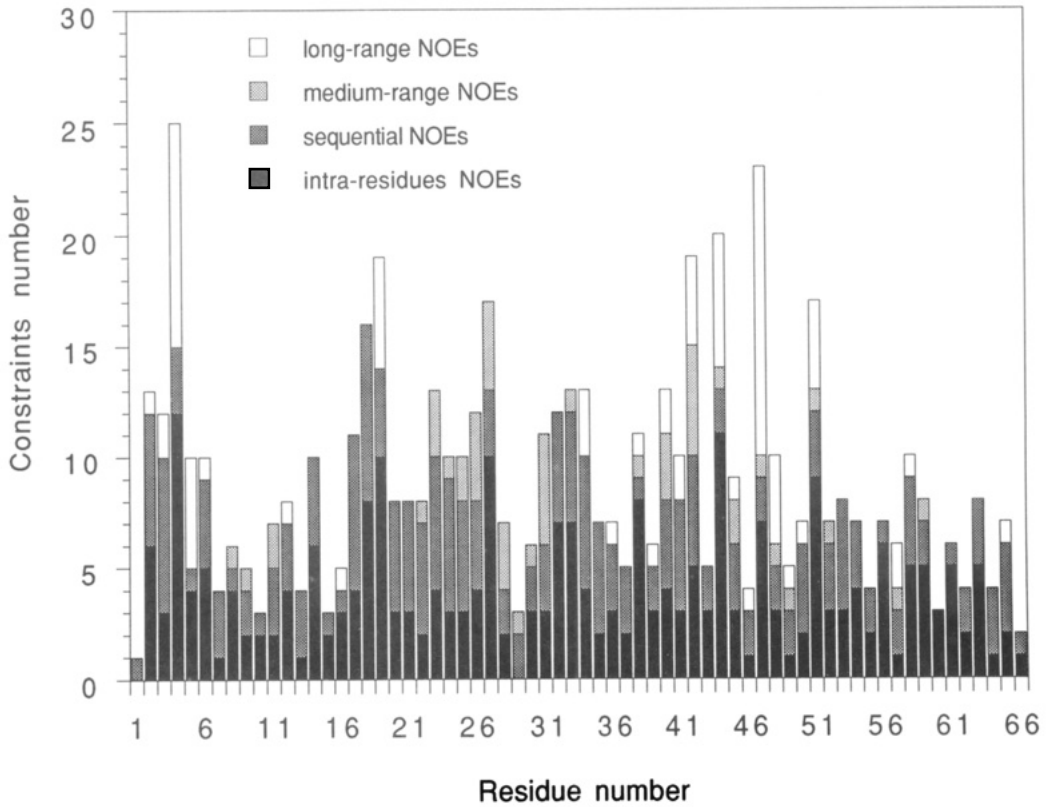

FIGURE 8: Sequence distribution of the NOE constraints used in the calculation of the structure of the *Centruroides limpidus limpidus* toxin 1. The number of NOE constraints is plotted versus the position of the amino acid residue.

Disulfide Bridge Analysis. It was possible to confirm some of the disulfide bridges from the use of the observed NOEs between the protons of the cysteines. Thus, the observation of a cross peak between Cys12 NH and Cys65 Hβ in the 100-ms NOESY spectrum allowed identification of the disulfide bridge Cys12–Cys65, which is the unusual disulfide pair shown by the AaH IT anti-insect toxin (Darbon *et al.*, 1991). Using a longer mixing time (300 ms), very weak NOE cross peaks could be observed between Cys16–Phe44 and Glu28–Cys48, suggesting two other disulfide bridges at positions 16–41 and 29–48, although these connectivities might also be due to a spin-diffusion effect.

Structural Calculations. The simulated annealing (SA) method using dihedral and distance constraints was employed to refine the structures and to identify a family of conformations consistent with the experimental data. The starting

FIGURE 9: (a) The 10 best structures resulting from the molecular dynamics calculations. (b) CII toxin 1 global fold of one of the 10 final structures.

structure of CII toxin 1 was modeled using the known high-resolution structure of CsE-v3 (Almassy et al., 1983). The scorpion toxins CII toxin 1 and CsE-v3 present 88% primary sequence identity and the same topology for their four disulfide bridges. In addition, the elements of secondary structure found in CII toxin 1 by NMR are present in the structure of the scorpion toxin CsE-v3 obtained by X-ray diffraction. The distribution of the NOE constraints per residue used in the constraints file is represented in Figure 8. Four IRMA cycles were necessary to obtain the smallest and therefore the desired values of the R-factors, with $R_1 = 0.918$ (Gonzales *et al.*, 1991). This procedure gave a family of 10 structures which have no distance violations greater than 0.05 nm and no ϕ angle violations greater than 5° (Figure 9a). In addition, analysis of Ramachandran plots for the final structures after each cycle shows that all of the ϕ/ψ angles for non-glycine residues lie within the allowed region.

The overall agreement among the different conformations can be summarized by an rms deviation value. The average of the rms deviations of all pairwise combinations of the 10 structures was 0.046 nm for the backbone and 0.13 nm for all of the atoms. The averages of the rms deviations for pairwise combinations of the 10 structures with a calculated average structure are 0.031 and 0.085 nm for the backbone and all the atoms superposition, respectively. Figure 10 represents the backbone rms deviation per residue calculated for all of the final structures, both between them and for the calculated average structure. This diagram confirms that some parts of the molecule are really well defined, especially the double-stranded β-sheet from residue 36 to 48.

Thus, the structure is characterized by an α-helix involving residues 23–31 and a triple-stranded antiparallel β-sheet made of residues 2–5, 36–42, and 45–49 (Figure 9b). It is possible to describe the 42–45 turn as a type II β-turn, considering the values of the dihedral angles ϕ and ψ in this region as well as the strong d_{NN} and $d_{\alpha N}$ connectivities between residues 44 and 45 and between residues 42 and 43, respectively. A medium-range NOE between residues Tyr42 and Phe44 and the slow rate of exchange of the amide proton of Ala45 further characterize this type II β-turn with the (i, i+3) hydrogen bond.

Some other turns or bends were not so well characterized, but the dihedral angles $\phi(i+1, i+2)$ and $\psi(i+1, i+2)$ given by the final structures and the NOE connectivities are enough to confirm their presence in the molecule. There is a bend at the 15–18 position, and for the 8–11 and 32–35 fragments, strong d_{NN} and small $d_{\alpha N}$ were observed, which indicates a type I β-turn. The (i, i+3) hydrogen bond corresponding to this β-turn was found for the 32–35 segment. The 52–55 and 60–63 turns seem to be a type I β-turn even though the ϕ and ψ angles and NMR data were not entirely consistent with this type of turn. Indeed, a medium-range NOE connectivity was observed between Pro52 Hβ and Ser54 NH, and a H-bond between Thr55 NH and the oxygen of the carbonyl of Pro52 is present in the resulting structures. The NOE connectivities showed that, among the four prolines in the small C-terminal loop of the protein, three of them are in *trans* conformation, and one, Pro59, is in *cis* conformation. This *cis*-proline defines a type VIb turn, described with strong $d\alpha\alpha(i+1, i+2)$ and $d\alpha N(i+1, i+3)$ connectivities. The presence of a *cis*-proline has been reported already in the structure determination of the CsE-v3 toxin (Nettesheim *et al.*, 1989). However, all of the proline residues were described with a *trans* conformation in the mammalian neurotoxin M9, which is also rich in proline residues in its C-terminal region (Pashkov *et al.*, 1988).

The conformations of the disulfide bridges appear to be well defined. The expected values around ±90° ± 10° are obtained for χ_s, corresponding to right- or left-handed disulfide conformations. χ_{1i} and χ_{1j} give reasonable values around −60°, 60°, and 180°, corresponding to the staggered orientations about the Cα–Cβ bond (Srinivasan *et al.*, 1990). The χ angle value was determined from the NMR data for only one

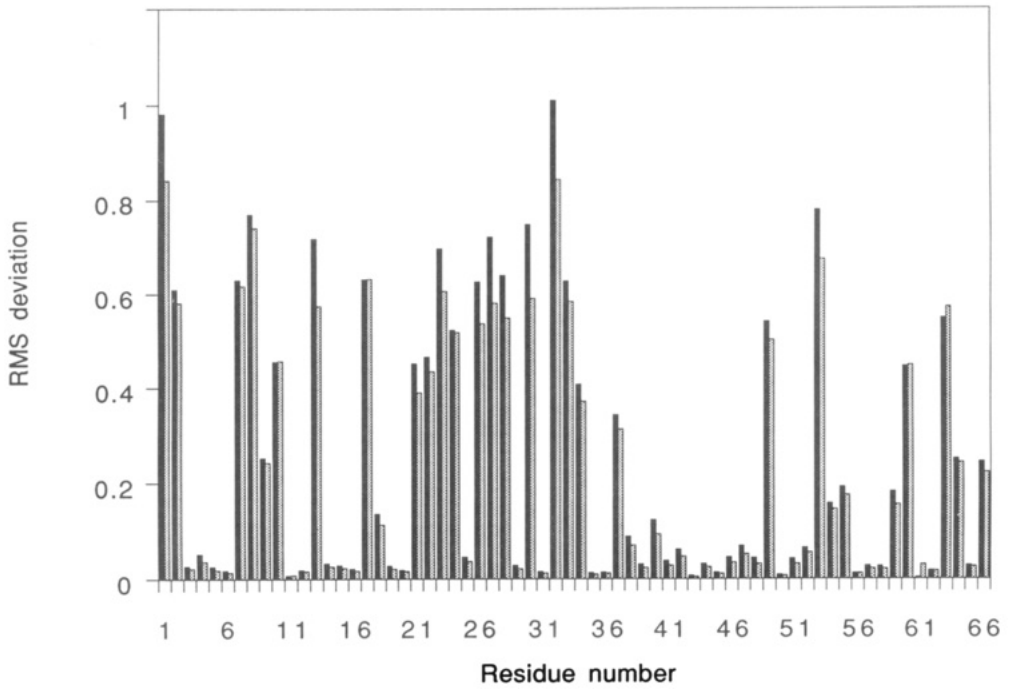

FIGURE 10: Representation of the average backbone rms deviations per residue of all pairwise combinations of the ten structures (solid bars) and for pairwise combinations of the ten structures with a calculated average structure (stippled bars).

disulfide bridge, 12–65, and was thus introduced as a restraint in the calculations.

DISCUSSION

Purification and Primary Structure Determination. The isolation procedures and bioassays used to determine the species specificity of this toxin were straightforward. No further discussion is needed, other than a brief comment on Cll toxin 1. An important observation is that toxin 1 constitutes about 9.5% of the total venom of *C. limpidus l.* The biological meaning of the high content of this toxin is certainly linked to the feeding habits of this scorpion. The mammal-directed toxins are in much smaller quantities (Alagon *et al.*, 1988). Scorpions probably need a high content of toxin for appropriation of their prey rather than for defense against mammals, their main predators.

A considerable degree of difficulty was encountered in the determination of the disulfide bridges (Table 1). In many instances, heteropolymeric forms were obtained (data not shown) that contained more than one disulfide bridge. They were eluted in clusters, making it impossible to elucidate the correct disulfide pairing. The native toxin was considerably resistent to mild enzymatic cleavages, contrary to the reduced and carboxymethylated one, which was easily cleaved out by the enzymes used in this work (Figure 2). Comparison of the primary sequences of several known toxins (Figure 11) shows that, in spite of the differences in their primary sequences, the eight half-cysteine residues in Cll toxin 1 are located in the same positions in the related toxins (Narayanan *et al.*, 1992; Loret *et al.*, 1991).

Secondary and Tertiary Structure Determination. The three-dimensional structure of the crustacean toxin 1 from the venom of the scorpion *Centruroides limpidus limpidus* was determined and shown to be very similar to the global structure of all other scorpion toxins studied so far, that is, a highly structured region containing an α-helix and a triple-stranded antiparallel β-sheet (Figure 9b).

The Cartesian coordinates of the X-ray crystal structures of two scorpion toxins are available. These are two mammalian toxins, an α-scorpion toxin, *Androctonus australis* Hector toxin II (AaH II) (Fontecilla-Camps *et al.*, 1988), and *Centruroides sculpturatus* Ewing variant 3 (CsE-v3), a weak β-toxin (Fontecilla-Camps *et al.*, 1981; Almassy *et al.*, 1983; Nettesheim *et al.*, 1989; Zhao *et al.*, 1992). Since the long toxins share identical elements of secondary structure, these conformational motifs were superimposed in Cll toxin 1, in AHa II, and in CsE-v3, for comparative purposes, especially focused on possible species selectivity features. The study of their spatial organization shows that the main structural differences are localized in the loops 6–15 and 40–43 (38–43 for AHa II, which has additional residues in this region like most mammalian toxins) and in the C-terminus. These structural comparisons were further analyzed by studying the differences between the ϕ and ψ dihedral angles in Cll toxin 1 and CsE-v3 (data not shown). This plot confirmed that large differences appear in the fragment from residue 7 to residue 15 and in the C-terminal region but also showed differences in the fragment 20–23. The coordinates of AaH IT (Darbon *et al.*, 1991) are not available, so it is not possible to compare its structure with that of Cll toxin 1, CsE-v3, or AHa II.

Numerous NOE connectivities between aromatic protons were observed for Cll toxin 1, in particular between Tyr4, Tyr42, Trp47, Tyr40, Tyr38, and Tyr58, which form an aromatic cluster as described in Figure 12. The relative positions of the aromatic residues are well resolved and are in favor of an identical spatial arrangement as observed in AaH IT (Darbon et al., 1991), AaH II (Fontecilla-Camps, 1989), and CsE-v3 (Almassy *et al.*, 1983; Krishna *et al.*, 1989). In particular, the aromatic rings Trp47 and Tyr4 present an orthogonal arrangement, as has been reported recently in the solution structure of CsE-v3 (Lee *et al.*, 1994). This region, which has been called "the conserved hydrophobic surface", not only has a function in stabilizing the tertiary structure of the scorpion neurotoxins but is also thought to be involved in the toxin–receptor interaction (Fontecilla-Camps, 1989). It has been proposed that the toxin specificity was due to changes in the positioning of loops coming from this highly conserved core structure. It is interesting to note that, in the case of Cll toxin 1, another small aromatic cluster composed of Tyr14,

FIGURE 11: Amino acid sequences of scorpion toxins. AaH IT (from *Androctonus australis* Hector) is a long insect toxin; Cll toxin 1 (from *Centruroides limpidus limpidus*) is a long crustacean toxin; CsE-v3 (from *Centruroides sculpturatus* Ewing), AaH II, AaH III, and Be M9 (from *Buthus eupeus*) are long mammalian toxins acting on Na^+ channels; ChTX (charybdotoxin from *Leiurus quinquestriatus hebraeus*), IbTX (iberiotoxin from *Buthus eupeus*), and ScyTX (scyllatoxin or leiurotoxin from *Leiurus quinquestriatus hebraeus*) are small mammalian toxins acting on K^+ channels; I_5A (from *Buthus eupeus*) is a short insectotoxin. The gaps were introduced in order to enhance similarities (taken from Bontems et al., 1991).

Phe44, Phe17, and Trp18 is also present, lying parallel to the one described above (Figure 12). This interesting feature was further investigated, in particular the position of Trp18, which seems to point out from the surface of the molecule. This was confirmed by the determination that a high percentage of its surface is accessible to solvent, which is rather unusual for a hydrophobic residue (its surface accessibility to the solvent is eight times more than that of Trp47). In addition, sequential alignment of several toxins showed that a hydrophilic residue is generally present at this position in mammal-directed toxins, usually a Lys or an Arg, in contrast to crustacean or insect toxins, which present a hydrophobic residue such as Trp but quite often a Leu at this position. The presence of a tryptophan at position 18 might be an important factor for the Cll toxin 1 specificity in the interaction with the Na^+ channel of the crustaceans.

It has been suggested that not only hydrophobic interactions but also charged molecules may contribute to the channel recognition in the toxins. Thus, it is interesting to analyze their primary sequences in terms of charge, hydrophobic residue, and hydrophilic residue positions. As expected, most of the hydrophilic residues are pointing out of the surface of the *Centruroides limpidus limpidus* scorpion toxin 1. This is also the case for the eight lysine residues in the molecule. The surface accessibility to the solvent was calculated for each, allowing them to be ranked from the most accessible to the less accessible: Lys32 > Lys1 > Lys63 > Lys30 > Lys21 > Lys27 > Lys13 > Lys8. Many lysine residues, such as Lys32, Lys30, Lys27, and Lys8, form a cluster of positive charges at the opposite side of the core. These lysine residues may also play a role in the specific interaction with the sodium channel. Another positively charged residue, Arg60, was shown to be essential in the biological activity of the mammalian neurotoxins (Pashkov *et al.*, 1988). A hydrophilic residue is also present at this position in other mammalian toxins like AaH II but is replaced by the hydrophobic residue Tyr58 in Cll toxin 1, confirming that this residue may be important in the specificity of the interaction with the sodium channel.

Another important feature of the scorpion toxin family which might be involved in structure–function relationships is the presence of the disulfide bridges. The four pairs 12–65, 16–41, 25–46, and 29–48 have been characterized for Cll toxin 1. Unlike most of the proteins with disulfide linkages, it has been reported recently that the disulfide bridges in the scorpion toxin family present topological isomeric forms with the same topological chirality (Mao, 1993). The cystine-stabilized α-helical motif (CSH) is an important feature which seems to be correlated with the common ion channel blocking activity of the scorpion toxins (Kobyashi *et al.*, 1991) (Figure 11). A surprising convergence of the secondary structure elements is observed in the scorpion toxin family, suggesting that even for the higher molecular weight scorpion toxin, only the CSH motif may act as a binding anchor. The selectivity toward a mammalian, insect, or crustacean Na^+ channel might be due to the presence of additional or substituted amino acid residues producing local differences capable of inducing or preventing the interaction with the specific acceptor molecule (ion channel).

Despite the presence of the disulfide bridges, it is important to note the presence of flexible regions in the Cll toxin 1. The large backbone rms deviations per residue shown in Figure 10 for some regions of the protein can be simply explained by the fact that the structure is not well defined in these regions either as a result of insufficient NMR data or a greater flexibility of the fragments. Indeed, large rms deviation values were obtained for residues 7–10 and 32–34, which confirms the difficulty of characterizing the type of turn formed by these amino acids. However, concerning the segment 21–30, the rms deviation for most of the residues is greater than expected for residues involved in known secondary structure. A rather dynamic conformation of the helix, which might be

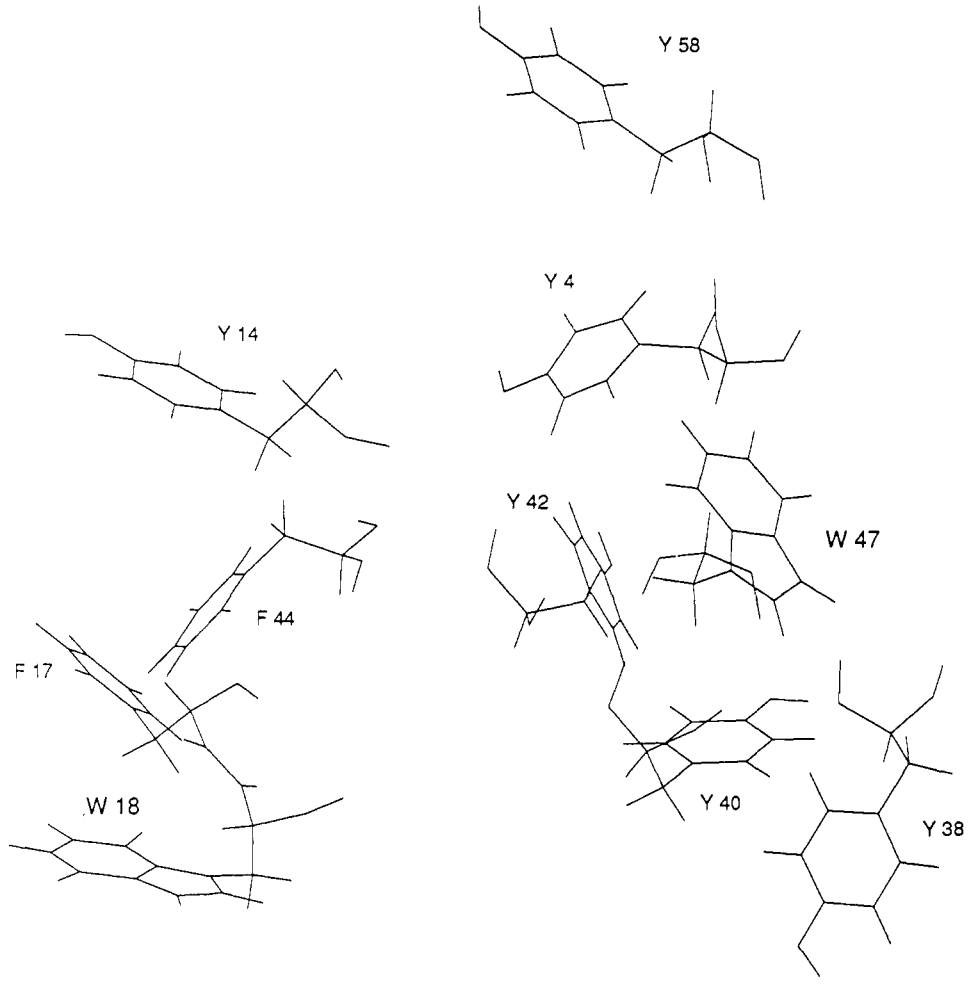

FIGURE 12: Spatial arrangement of the aromatic residues in Cll toxin 1.

partially destabilized under the experimental conditions, might explain these observed values. At pH 3.5, both Glu23 and Asp26, located near the helix N-terminus, are protonated. On the basis of the helix dipole model with charged group effects as described by Nettesheim (Nettesheim *et al.*, 1989) the protonation of the two acidic groups could destabilize this region of the helix. Unfortunately experiments at pH values above the pK_a of the two acidic groups could not be achieved because the samples precipitated. The flexibility of the helical region 23–30 is also confirmed by its rather weak d_{NN} connectivities and the difficulty of assigning the Cys29 residue. In the description of the crystal structure of CsE-v3 refined at 0.12 nm, it has been reported that the two disulfide bridges between Cys12 and Cys65 and between Cys29 and Cys48 are very flexible (Zhao *et al.*, 1992). This seems to be the case for the Cll toxin 1 as far as the disulfide bridge between Cys29 and Cys48 is concerned. It would be interesting to investigate the internal dynamics of the molecule which seem to be important in Cll toxin 1.

From the study presented here, it emerges that some common elements in the scorpion toxins found also in Cll toxin 1 are necessary for the interaction with the Na^+ channel, i.e, the secondary structures, the disulfide bridges, and the hydrophobic core. However, these are not sufficient to explain the neurotoxic activity and the species specificity which might be due to the presence of peculiar residues in this toxin 1, such as Trp18, Tyr58, and a few lysine residues, making it a crustacean toxin rather than a mammalian one.

Therefore, it would be interesting to compare and analyze in detail other structures from new scorpion toxins presenting similar primary sequences but different neurotoxic activities. These comparisons could be the starting point to mutations and bioassays that could precisely define the mode of action of these scorpion toxins.

ACKNOWLEDGMENT

We thank Dr. Rose for helpful discussions; Dr. Fontecilla-Camps and Dr. Bontems for giving us the coordinates of toxin II from *A. australis* and charybdotoxin, respectively; and Dr. Bentley for stylistic revision. The technical assistance of Fernando Zamudio and Timoteo Olamendi during purification and sequencing is kindly appreciated.

REFERENCES

Alagon, A. C., Guzman, H. S., Martin, B. M., Ramirez, A. N., Carbone, E., & Possani, L. D. (1988) *Comp. Biochem. Physiol. 89B*, 153.

Almassy, R. J., Fontecilla-Camps, J. C., Suddath, F. L., & Bugg, C. E. (1983) *J. Mol. Biol. 170*, 497.

Arseniev, S. A., Kondakov, V. I., Maiorov, V. N., & Bystrov, V. F. (1984) *FEBS Lett. 165*, 1.

Bablito, J., Jover, E., & Couraud, F. (1986) *J. Neurochem. 46*, 1763.

Basus, V. L. (1989) *Methods Enzymol. 177*, 132.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355.

Blaustein, M. P., Rogowski, R. S., Schneider, M. J., & Krueger, B. K. (1991) *Mol. Pharmacol. 40*, 932.

Boelens, R., Koning, T. M. G., Van der Marel, G. A., Van Boom, H., & Kaptein, R. (1989) *J. Magn. Reson. 82*, 290.

Boere, R. T., & Kidd, R. G. (1983) *Annu. Rep. NMR Spectrosc. 12*, 319.

Bontems, F. (1992) Ph.D Thesis, Université de Paris Sud, centre

d'Orsay.
Bontems, F., Roumestand, C., Boyot, P., Gilquin, B., Doljansky, Y., Menez, M., & Toma, F. (1991) *Eur. J. Biochem. 196*, 19.
Boyd, J., Dobson, C. M., & Redfield, C. (1983) *J. Magn. Reson. 55*, 170–176.
Carbone, E., Wanke, E., Prestipino, G., Possani, L. D., & Maelicke, A. (1982) *Nature 296*, 90.
Catterall, W. A. (1980) *Annu. Rev. Pharmacol. Toxicol. 20*, 15.
Catterall, W. A. (1991) *Curr. Opin. Neurobiol. 1*, 5.
Clore, G. M., Nilges, M., Sukumaran, D. K., Brunger, A. T., Karplus, M., Easthope, P., & Havel, T. F. (1986) *EMBO J. 5*, 2729.
Darbon, H., Weber, C., & Braun, W. (1991) *Biochemistry 30*, 1836.
Daubert-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., & Hagler, A. T. (1988) *Proteins: Struct., Funct., Genet. 4*, 31.
DeBin, J. A., Maggio, J. E., & Strichartz, G. R. (1993) *Am. J. Physiol. 264*, 361.
De Lima, M. E., Martin, M. F., Diniz, C. R., & Rochat, H. (1986) *Biochem. Biophys. Res. Commun. 139*, 296.
Dent, M. A. R., Possani, L. D., Ramirez, G. A., & Fletcher, P. L., Jr. (1980) *Toxicon 18*, 343.
Fontecilla-Camps, J. C. (1989) *J. Mol. Evol. 29*, 63.
Fontecilla-Camps, J. C., Almassy, R. J., Suddath, F. L., Watt, D. D., & Bugg, C. E. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 6496.
Fontecilla-Camps, J. C., Almassy, R. J., Suddath, F. L., Watt, D. D., & Bugg, C. E. (1981) *Trends Biochem. Sci. 6*, 291.
Fontecilla-Camps, J. C., Habersetzer-Rochat, C., & Rochat, H. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7443.
Gonzales, C., Rullmann, J. A. C., Bonvin, A. M. J. J., Boelens, R., & Kaptein, R. (1991) *J. Magn. Reson. 91*, 659.
Griesinger, C., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870.
Johnson, B. A., & Sugg, E. E. (1992) *Biochemistry 31*, 8151.
Jover, E., Couraud, F., & Rochat, H. (1980) *Biochem. Biophys. Res. Commun. 95*, 1607.
Kirsch, G. E., Skattebol, A., Possani, L. D., & Brown, A. M. (1989) *J. Gen. Physiol. 93*, 67.
Kobyashi, Y., Takashima, H., Tamaoki, H., Kiogoku, Y., Lambert, P., Kuroda, H., Chino, N., Watanabe, T. X., Kimura, T., Sakakibara, S., & Moroder, L. (1991) *Biopolymers 31*, 1213.
Kopeyan, C., Martinez, G., Lissitzky, S., Miranda, F., & Rochat, H. (1974) *Eur. J. Biochem. 47*, 483.
Krishna, N. R., Nettesheim, D. G., Klevit, R. E., Drobny, G., Watt, D. D., & Bugg, C. E. (1989) *Biochemistry 28*, 1556.
Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 64*, 2229.
Laemmli, U. K. (1970) *Nature 227*, 680.
Lazdunski, M., Frelin, Ch., Barhanin, J., Lombet, A., Meiri, H., Pauron, D., Romey, G., Schmid, A., Schweitz, H., Vigne, P., & Vijverberg, H. P. M. (1986) in *Tetrodotoxin, Saxitoxin and The Molecular Biology of The Sodium Channels* (Kao, C. Y., & Levinson, S. R., Eds.) Vol. 479, pp 204–220, New York Academy of Sciences, New York.
Lee, W., Moore, C. H., Watt, D. D., & Krishna, N. R. (1994) *Eur. J. Biochem. 218*, 8.
Loret, E. P., Martin-Eauclaire, M., Mansuelle, P., Sampieri, F., Granier, C., & Rochat, H. (1991) *Biochemistry 30*, 633.
Mao, B. (1993) *Protein Sci.* 2, 1057.
Marion, D., & Bax, A. (1988) *J. Magn. Reson. 80*, 528.
Martin, J. C., Zhang, W., Tartar, A., Ladzunski, M., & Borremans, F. A. M. (1990) *FEBS Lett. 260*, 249.
Menez, A. (1993) *Sci. Am. 190*, 34.
Menez, A., Bontems, F., Roumestand, C., Gilquin, B., & Toma, F. (1992) *Proc.—R. Soc. Edinburgh 99B*, 83.
Meves, H., Marc, S., & Watt, D. (1986) in *Tetrodotoxin, Saxitoxin and the molecular biology of the sodium channels* (Kao, C. Y., & Levinson, S. R., Eds.) Vol. 479, pp 113–132, New York Academy of Sciences, New York.
Mikou, A., Laplante, S. R., Guittet, E., Lallemand, J. Y., Martin-Eau Claire, M.-F., & Rochat, H. (1992) *J. Biomol. NMR 2*, 57.
Miller, C., Moczydlowski, E., Latorre, R., & Phillips, M. (1985) *Nature 313*, 316.
Moczydlowski, E., Lucchesi, K., & Ravindran, A. (1988) *J. Membr. Biol. 105*, 95.
Narayanan, P., & Lala, K. (1992) *Life Sci. 50*, 683.
Nettesheim, D. G., Klevit, R. E., Drobny, G., Watt, D. D., & Krishna, N. R. (1989) *Biochemistry 28*, 1548.
Nilges, M., Clore, G. M., & Gronenborn, A. (1988) *FEBS Lett. 239*, 129.
Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741.
Pashkov, V. S., Maiorov, V. N., Bystrov, V. F., Hoang, A. N., Volkova, T. M., & Grishin, E. V. (1988) *Biophys. Chem. 31*, 121.
Possani, L. (1984) in *Handbook of Natural Toxins* (Tu, A. T., Ed.) Vol. 2, pp 513–550, Marcel Dekker, Inc., New York.
Possani, L. D., Martin, B. M., & Svendsen, I. (1982) *Carlsberg Res. Commun. 47*, 285.
Ramirez, A. N., Gurrola, G. B., Martin, B. M., & Possani, L. D. (1988) *Toxicon 26*, 772.
Reisfield, R. A., Lewwis, U. J., & Williams, D. E. (1962) *Nature 195*, 281.
Rochat, H., Bernard, P., & Couraud, F. (1979) in *Advances in Cytopharmacology* (Ceccarelli, B., & Clementi, F., Eds.) Vol. 3B, pp 325–334, Raven Press, New York.
Sodano, P., & Delepierre, M. (1993) *J. Magn. Reson. A 104*, 88.
Srinivasan, N., Sowdhamini, R., Ramakrishnan, C., & Balaram, P. (1990) *Int. J. Protein Res. 36*, 147.
States, D. J., Haberkorn, R. A., & Rubens, D. J. (1982) *J. Magn. Reson. 48*, 286.
Sugg, E. E., Garcia, M. L., Reuben, J. P., Patchett, A. A., & Kacsorowski, G. J. (1990) *J. Biol. Chem. 265*, 18745.
Vaca, L., Gurrola, G. B., Possani, L. D., & Kunze, D. L. (1993) *J. Membr. Biol. 134*, 123.
Valdivia, H. H., Kirby, M. S., Lederer, W. J., & Coronado, R. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 12185.
Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., & Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992) *Biochemistry 31*, 1647.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
Zhao, B., Carson, M., Ealick, S. E., & Bugg, C. E. (1992) *J. Mol. Biol. 227*, 239.
Zlotkin, E., Miranda, F., & Rochat, H. (1978) in *Handbook of Experimental Pharmacology, Vol. 48, Arthropod Venoms* (Bettini, S., Ed.) pp 317–369, Springer-Verlag, Berlin.
Zlotkin, E., Eitan, M., Bindokas, V. P., Adams, M. E., Moyer, M., Burkhart, W., & Fowler, E. (1991) *Biochemistry 30*, 4814.
Zuiderweg, E. R. P., Boelens, R., & Kaptein, R. (1985) *Biopolymers 24*, 601.