Solution Structure of an Insect-Specific Neurotoxin from the New World Scorpion *Centruroides sculpturatus* Ewing[†,‡]

Michael J. Jablonsky,[§,||] Patricia L. Jackson,[§,⊥] and N. Rama Krishna*[,§,#]

Comprehensive Cancer Center and Department of Biochemistry and Molecular Genetics, The University of Alabama at Birmingham, Birmingham, Alabama 35294-2041

Received February 1, 2001; Revised Manuscript Received May 8, 2001

ABSTRACT: We report the high-resolution solution structure of the 6.3 kDa neurotoxic protein CsE-v5 from the scorpion *Centruroides sculpturatus* Ewing (CsE, range southwestern U.S.). This protein is the second example of an Old World-like neurotoxin isolated from the venom of this New World scorpion. However, unlike CsE-V, which is the first Old World-like toxin isolated and shows both anti-insect and anti-mammal activity, CsE-v5 shows high specificity for insect sodium channels. Sequence-specific proton NMR assignments and distance and angle constraints were obtained from 600 MHz 2D-NMR data. Distance geometry and dynamical simulated annealing refinements were performed to produce a final family of 20 structures without constraint violations, along with an energy-minimized average structure. The protein structure is well-defined (0.66 and 0.97 D rmsd for backbone and all heavy atoms, respectively) with a compact hydrophobic core and several extending loops. A large hydrophobic patch, containing four aromatic rings and other aliphatic residues, makes up a large area of one side of the protein. CsE-v5 shows secondary structural features characteristic of long-chain scorpion toxins: a two and a half-turn α-helix, a three-strand antiparallel β-sheet, and four β-turns. Among the proteins studied to date from the CsE venom, CsE-v5 is the most compact protein with nearly 50% of the amide protons having long exchange lifetimes, but CsE-v5 is unusual in that it has loop structures similar to both Old and New World toxins. Further, it also lacks prolines in its C-terminal 14 residues. It shows some important differences with respect to CsE-V not only in its primary sequence, but also in its electrostatic potential surface, especially around areas in register with residues 8, 9, 17, 18, 32, 43, and 57. The loss of anti-mammal activity in CsE-v5 and the differences in its anti-insect activity compared to that of other proteins such as CsE-V, v1, and v3 from this New World scorpion may be related to residue variations at these locations.

Scorpion venom typically consists of a number of toxic proteins responsible for the neurotoxic effects of the scorpion sting. Both long chain neurotoxins (MW ∼ 7 kDa), which bind to the sodium channels, and short chain toxins (MW ∼ 4 kDa), which bind to other channels such as K^+ and Cl^-, are typically present in the venom. The long chain scorpion neurotoxins can be divided into two groups, insect and mammalian toxins (*1*), but toxins acting on both groups (e.g., refs *2* and *3*) and toxins with novel properties have also been characterized. Two classes of mammalian toxins, α- and β-toxins, are distinguished on the basis of their effect on the sodium channel (*1*, *4*−*8*). The α-toxins slow the inactivation of the sodium ion permeability in the channel. In contrast, the β-toxins shift the voltage of activation toward more negative potentials, causing the membrane to fire spontaneously and repetitively. Insect toxins can be divided as either excitatory or depressant, depending on their mode of paralysis. The three-dimensional structures of some long chain scorpion toxins have been determined by NMR and crystallography (e.g., see references sited in refs *6*−*10*). All toxins have similar secondary structural elements and globular structures. These features include a hydrophobic core made of a two and a half-turn α-helix and a three-strand antiparallel β-sheet. The relative arrangement of the helix wrt β-sheet is stabilized by two disulfide linkages in the αβDB motif (*11*). Several loops extend from this core, and two additional disulfide linkages help anchor these loops. The depressant insect and mammalian toxins have conserved disulfide linkages. In general, the toxins from New World species have a shorter B loop and longer J and M loops than the Old World toxins (Figure 1). A hydrophobic surface is present on one face of the protein (referred to as the front surface), which is thought to be the channel binding site. Near this surface are several key residues which have been implicated in the activity and selectivity of the toxins (*6*, *12*).

[†] Partial support of this work by the NSF Grant MCB-9630775 and NCI Grant CA-13148 (NMR Shared Facility) is gratefully acknowledged.

[‡] This paper is warmly dedicated to Dr. Dean D. Watt, Emeritus Professor of Biomedical Science at Creighton University, Omaha, NE, for his lifetime contributions which have led to our current understanding of the biology, structures, and functions of neurotoxins from the New World scorpion species *Centruroides scupturatus* Ewing.

* To whom correspondence should be addressed. Phone: (205) 934-5695. Fax: (205) 934-6475. E-mail: NRKrishna@bmg.bhs.uab.edu.

[§] Comprehensive Cancer Center.

[||] Current address: Department of Chemistry, UAB. Phone: (205) 975-9942. Fax: (205) 934-2543. E-mail: jabo@uab.edu.

[⊥] Current address: Department of Physiology & Biophysics, UAB. Phone: (205) 934-3032. Fax: (205) 934-1445. E-mail: plj@uab.edu

[#] Department of Biochemistry and Molecular Genetics.

10.1021/bi010223h CCC: $20.00
Published on Web 06/20/2001

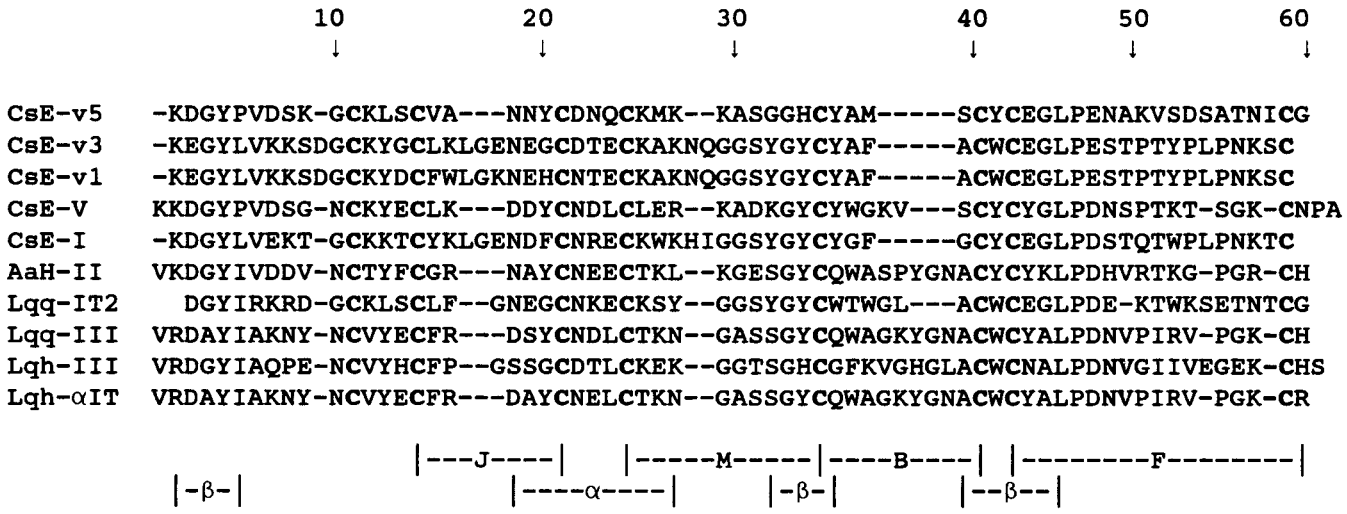

FIGURE 1: Amino acid sequences for selected neurotoxins: CsE-v5, -v3, -v1, -V, and -I (from ref *1*), AaH-II (*33*), LqqIT2 (*34*), Lqq-III (*35*), LqhIII (*36*), and LqhαIT (*37*). The numbering is with respect to CsE-v5. The presence of a terminal glycine-amide residue in CsE-v5 has been established in the current study. Sequences are aligned with respect to the cysteines. The four major loops (J, M, B, and F) and the approximate locations of the secondary structures are shown.

Insect toxins have been isolated from several species of Old World scorpions, and they show a broad spectrum of effects. Lqq-III[1] is both a powerful insect and mammalian α-toxin, binding to Na^+ channels in both groups (*3*). Lqh-III is an α-like toxin with unique physiological properties (*9*), and does not compete with AaH-II. AaH IT_4 is an insect toxin which also effects the binding to α- and *β*-sites of mammalian channels (*13*). Lqh-αIT is one of the most powerful insect toxins known; it slows inactivation of insect Na^+ channels (α-toxin), but has only a weak effect on mammalian channels (*14*, *15*). AaH-II is a powerful mammalian toxin, which has been used as a reference in competitive binding studies (*16*).

The Na^+-channel binding scorpion toxins have been broadly categorized into as many as 10 different groups based on structural features and functional features such as species specificity, effects on the sodium current, and binding properties (*8*). Thus, not surprisingly, competition measurements suggest that the mode of recognition of scorpion toxins by sodium channels is complex, e.g., the specificity and differences in insect versus mammal activity of α- and α-like scorpion toxins may involve structural differences in both the toxins and the homologous receptor sites on insect and mammalian sodium channels; different toxins from the same species may bind to different overlapping receptor sites on the same sodium channel or to different sodium channel subtypes; α-toxins from different scorpion species may bind to different receptor sites on insect and mammal sodium channels (*17*). Species specificity between the New World and the Old World toxins has also been reported (*18*).

Recently, much attention has been focused on insect toxins, with the goal of determining how these toxins interact with the insect and mammalian and sodium channels. As part of an ongoing study, we have determined the solution state structure of the neurotoxin CsE-v5 (also referred to as the variant-5 toxin) from the New World scorpion *Centruroides sculpturatus* Ewing (range Southwestern USA). It behaves similarly to other variants, CsE-v1 to -v4 and -v6, i.e., as a weak α-toxin with little toxicity toward mammals, but specifically toxic to insects (*1*). In fact, CsE-v5 is more toxic to insects than any of these variants (*1*). The toxicities of the individual toxins from CsE are CsE-v1 (>20, 12.0), CsE-v2 (>20, 21.5), CsE-v3 (>20, 255), CsE-v4 (>20, 32.0), CsE-v5 (>20, 2.5), CsE-v6 (8.0, 5.9), and CsE-V (1.6, 0.9), where the first and second numbers in parentheses stand for LD_{50} (mg/kg) and ED_{50} (mg/kg), respectively (*1*). All these toxins have been shown to prolong the inactivation of the sodium current (*5*). Further, variant toxins from CsE do not appear to compete with the classical α-toxin AaH-II (*18*). Thus CsE-v5 could be considered as an α-like toxin. The receptor site on the sodium channel for the α-like toxins is not yet known (*6*). Among these, interestingly, CsE-v5 has shortened J- and M-loops, which makes it an Old World-like toxin. However, it also shares a shorter B-loop with other New World toxins. In this respect, CsE-v5 is very similar to CsE-V, the first example of an Old World-like neurotoxin isolated from the New World scorpion *C. sculpturatus* (*2*). Both the CsE-v5 and CsE-V toxins are examples of evolutionary carry over of the structure quite prevalent in the Old World scorpion toxins, and provide an insight into the evolutionary relationship among these toxins (*2*). However, whereas CsE-V exhibits both high anti-mammal and high anti-insect activities, CsE-v5 is unusual in that it is highly specific for insects. It is also unique among the CsE toxins in that it has a long C-terminal loop (residues 47−60), which is devoid of any prolines whereas other toxins tend to have two or more prolines in this region. This absence of prolines portends some significant conformational differences in the C-terminal loop, compared to other scorpion toxins.

We have employed 600 MHz ^1H NMR spectroscopy and X-PLOR refinement to determine the solution structure of CsE-v5. Electrostatic potential surfaces for the two Old World-like toxins from CsE, viz., CsE-v5 and CsE-V, were

[1] Abbreviations: CsE, *Centruroides sculpturatus* Ewing; CsE-v5, variants-5 neurotoxin from CsE; CsE-V, toxin-V from CsE; AaH, *Androctonus australis* Hector; Lqq, *Leiurus quinquestriatus* quinquestriatus; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; DQF−COSY, double quantum filtered correlated spectroscopy.

generated with DELPHI (*19*) and displayed with GRASP (*20*). The backbone structure of CsE-v5 is compared to a few representative toxins to highlight the significant conformational differences and similarities with these toxins.

MATERIALS AND METHODS

CsE-v5 Purification and Sample Preparation. CsE-v5 (6.354 kDa) was isolated and purified by a three-step column chromatography procedure. First, the crude venom is separated on a carboxymethylcellullose column with a pH and ammonium acetate gradient to produce 12 toxin containing zones (*21*). The variant-5 is separated from zones 7−8 using two additional columns with ammonium acetate gradients: DEAE-Sephadex A-25 at pH 8.5 to remove a 7.286 kDa impurity and CM-Sephadex at pH 6.0 to remove a 7.136 kDa impurity. The resulting pure toxin is then lyophilized repeatedly to remove ammonium acetate. Identity and purity were confirmed by MALDI-TOF mass spectral analysis. The NMR sample was made at 1.0 mM and pH 4.0 with 10% D_2O, in a volume of 250 μL for use in a Shigemi microcell. After completion of the experiments in 90% H_2O, the sample was lyophilized to dryness and then brought up in 100% D_2O for the ^1H-^2H exchange experiment and additional experiments in D_2O.

NMR Spectroscopy. The NMR measurements were performed on a Bruker AM-600 spectrometer equipped with an Aspect 3000 computer. Data were collected at 293 and 303 K. NOESY, DQF-COSY, and TOCSY measurements were performed in pure absorption mode using time proportional phase increment and presaturation for water suppression. Mixing times of 100, 200, and 300 ms for NOESY in H_2O, and 200 ms and 400 ms in D_2O were employed. NOESY data in H_2O were collected using a jump-return read sequence. For assignment purposes, 70 ms TOCSY experiments were performed in both H_2O and D_2O. The ^1H-^2H exchange experiment was a 30 ms TOCSY with 32 scans and 256 t_1 points run immediately after dissolving the protein in D_2O. For all other NOESY and TOCSY experiments, 128 scans of 2K complex data points were collected for each of the 512 serial files. For the COSY experiments, the data size was 4K complex with 1024 t_1 points. Data were processed with FELIX using various window functions. COSY data were zero filled to 8K real in the observe dimension before Fourier transformation.

Experimental Constraints. The NOESY constraints were grouped as strong (1.8−2.7 Å), medium (1.8−3.3 Å), or weak (1.8−5.0 Å) based on peak intensities using 100 ms and 200 ms data. Additional NMR constraints of very strong category (1.8−2.3 Å) were also obtained from the 100 ms NOESY in H_2O (*22*). A total of 919 distance contraints were used: 339 intraresidue, 216 sequential, 96 medium range ($1 < |i - j| \leq 4$), 186 long range ($|i - j| > 4$), 70 hydrogen bond (2 per bond), and 12 disulfide bridge constraints. Thirty φ angle constraints ($-60 \pm 30°$ for $^3J_{HN\alpha} < 7$ Hz and $-120 \pm 40°$ for $^3J_{HN\alpha} > 8$ Hz), were obtained from the DQF-COSY in H_2O. Fifteen additional coupling constants for residues not giving rise to COSY peaks were determined from NOESY and TOCSY spectra (*23*). Twenty-eight stereospecific assignments for the AMX spin systems and the valines were made from $^3J_{\alpha\beta\beta'}$ coupling constants and the 100 ms NOESY spectra; 14 additional χ_1 constraints were obtained using STEREOSEARCH (*24*). The χ_2 constraints for the leucines were determined from nOe patterns (*25*). The χ_1 constraints were given as the optimum angle (180, 60, or $-60°$) $\pm$ 30°. In addition, one ω constraint (180 $\pm$ 5°) was applied for each proline, giving a total of 91 angular constraints. This gave a total of 1010 constraints or an average of 16.8 constraints/residue. Most residues had at least 20 constraints per residues, except at the C-terminus which showed fewer constraints (Figure 3). A much smaller constraint data set (resulting in a lower resolution structure) was previously used as a test data set for a semi-automated assignment program (*26*). Pseudo atoms were used for the nonstereospecifically assigned atoms (*27*). A subset of these restraints, less hydrogen bonds and some stereospecific nOe and angle constraints, was used in the first round of modeling calculations to produce an initial family of structures. By careful examination of the resulting structures and use of STEREOSEARCH, additional constraints including H-bonds, χ_1 angles, and stereospecific ring nOes were added for the final refinement calculations.

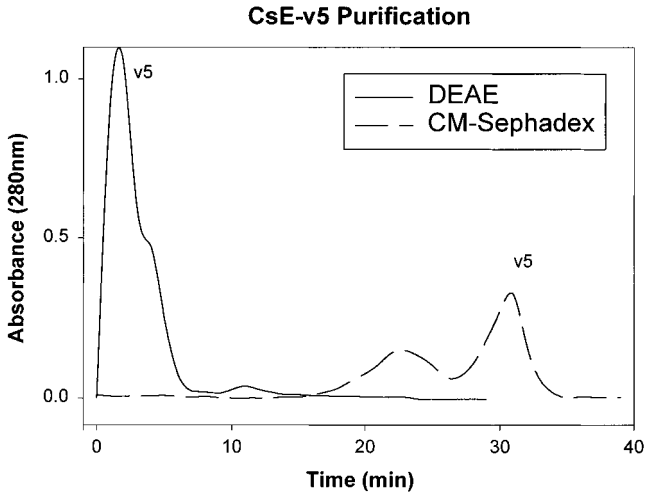

FIGURE 2: Chromatograms of the CsE Zone 7−8 for the isolation of CsE-v5. The peaks containing CsE-v5 are identified.

Molecular Modeling. The structure refinement calculations were performed with the XPLOR/QUANTA/CHARMm package (Molecular Simulations, Inc.) on an Origin 200 (Silicon Graphics, Inc.). The solution-state structures were generated using a hybrid method consisting of distance geometry combined with dynamic simulated annealing based on the work of Clore and Gronenborn and co-workers (*28*) with some minor modifications as previously described (*11*). Distance geometry substructures were first generated using only medium- and long-range (residues i, i+2 or greater) distance constraints and torsion angle constraints. Dynamical simulated annealing followed by energy minimization was then performed using a four-step protocol. This protocol was first run without hydrogen bond or stereospecific ring constraints. From the structures obtained, the H-bond and stereospecific ring constraints were determined. The protocol was repeated with these additional constraints to produce the final structures. Of this family of structures, only those with nOe distance violations less than 0.3 Å and dihedral angle violations less than 5°, designated as $\langle SA \rangle_k$, were considered acceptable, and included in the statistical analysis. The average structure, $\langle \overline{SA} \rangle_k$, was obtained by taking the averages of the atomic coordinates of these acceptable

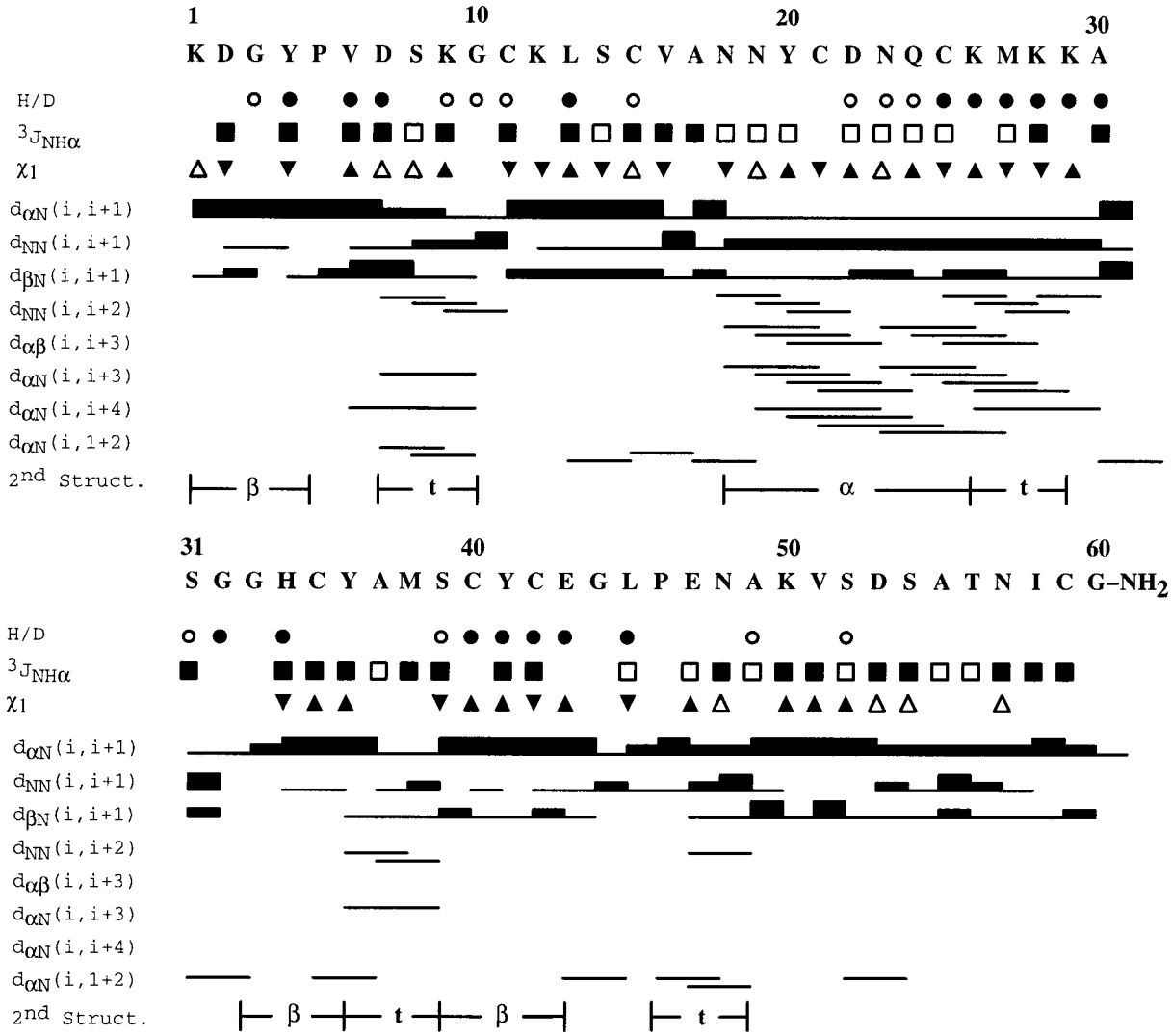

FIGURE 3: Summary of NMR data for CsE-v5. Sequential NOESY connectivities are shown as filled bars according to nOe intensity, while other connectivities are shown as lines. In the case of prolines, the contacts are to the δ hydrogens. Vicinal coupling constants: $^3J_{N\alpha} > 8$ Hz (■) and $^3J_{N\alpha} < 7$ Hz (□); torsion angles $\chi_1 = 180°$ (▲), 60° (▼), and −60° (△); slowing exchanging amide protons visble after 1 h (○). Visible after 48 h (●). Regions of α-helix and β -sheet are also shown.

structures. This structure was subjected to energy minimization first with and then without NMR derived constraints, to remove bond length and bond angle distortions, giving the final energy minimized average structure, $\langle\overline{SA}\rangle_{kr}$. Electrostatic potentials were generated using a nonlinear Poisson−Boltzmann treatment in DELPHI (*19*) and were mapped with GRASP (*20*). Coordinates for CsE-v3, Lqq-III, Lqh-III, Lqh-αIT, and AaH-II were obtained from the Protein Data Bank.

RESULTS

CsE-v5 Purification. The protein was purified from the crude venom using standard procedures (*1*, *21*). The elution profile from CM-cellulose of the whole venom resulted in about 12 separate zones of which zone 7−8 contains the CsE-v5 toxin. Seventy-four milligrams of zones 7/8 was used for the two-step purification. From this, about 20 mg of pure toxin were obtained after rechromatography (Figure 2). Mass spectral analysis gave a single peak at mass 6353.6.

Spin System and Sequence-Specific Assignments. The spin system and sequence specific assignments were made using standard procedures for unlabeled proteins (*27*) and agreed with the previously published sequence (*1*), except for the addition of a glycine-amide in our sequence. Using TOCSY data at 293 and 303 K, all protons were uniquely (not necessarily stereospecifically) assigned except for cases of overlap within the residue, e.g., the β and χ protons for Lys-9, Lys-12, Met-38, and Lys-50, and the β protons of Ser-31 and Cys-59. The two prolines, Pro-6 and Pro-46, were determined to be in the trans configuration based on the presence of strong Val-5/Pro-6 and Leu-45/Pro-46 $d_{(\alpha,\delta\delta')}(i,i+1)$ contacts. Strong nOe contacts between the Gly-60 α protons and an amide group indicated that the C-terminus was amidated. No evidence of isomerization (multiple peaks) was observed from either the prolines or the cysteines. The cis and trans primary amide protons were assigned according to the relative strength of the $H^{\beta}-H^{\delta}$ or $H^{\gamma}-H^{\epsilon}$ nOe contacts for Asn and Gln, respectively.

Structural Elements. Figure 3 gives a summary of sequential NOE connectivities, location of some secondary structural elements, and data on coupling constants and slowly exchanging amide protons. The NMR structures were calculated using an experimental data set consisting of 1010 constraints (16.8 constraints/residue). The distribution of

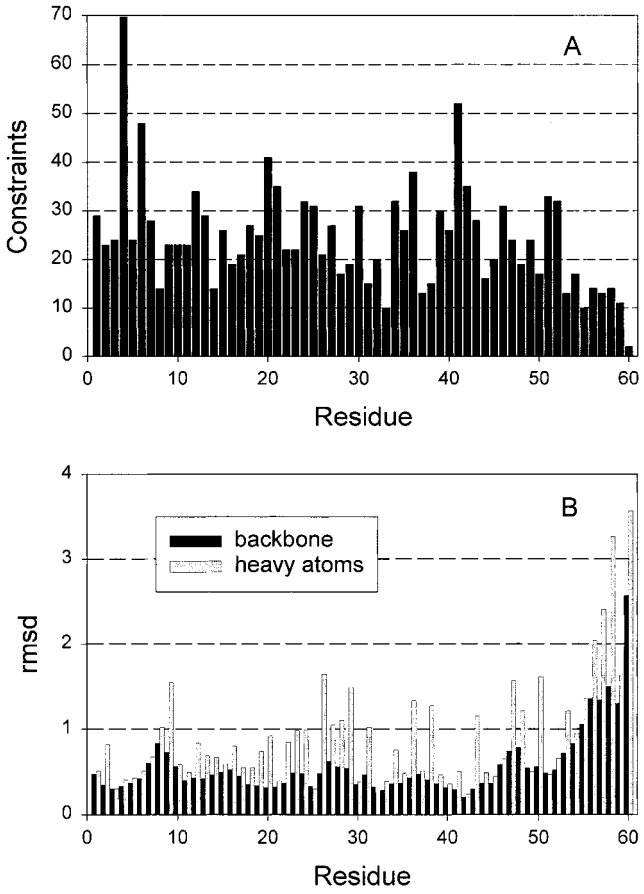

FIGURE 4: Constraints and atom rmsd per residue. (A) total constraints per residue; (B) backbone and heavy atom rmsd for the twenty $\langle SA \rangle_k$ structures with respect to the $\langle \overline{SA} \rangle_{kr}$ structure.

these constraints among the residues is shown in Figure 4. The general structure of CsE-v5 is a compact core consisting of an α-helix and a β-sheet, with the helix being held to the central strand of the sheet by two disulfide bridges in the $\alpha\beta$DB motif (*11*). Two loops, containing residues 7−12 and 53−58, extend from this core. The C-terminal loop region is comparatively not as well defined in the NMR structure as the rest of the regions in CsE-v5. This was due to the general lack of long-range nOes and the absence of hydrogen bond constraints due to a lack of slowly exchanging amides in this segment. Three secondary structural elements were observed in CsE-v5: a two and a half turn α-helix (18−26), a three-strand antiparallel β-sheet (1−5, 39−43,-33−36), and four β-turns (7−10, 26−29, and 46−49 Type I; 36−39, Type I′) (*29*). These features were indicated from the characteristic nOes, coupling constants, and hydrogens bonds (Figure 3). The secondary structures of CsE-v5 are very stable as shown by the observation of all expected nOes and long-lived exchangeable amide protons. In addition, many slowly exchanging amides were observed from residues involved in long-range hydrogen bonds, e.g., Leu-13 NH/Pro-5 CO, Ser-52 NH/Tyr-4, and Val-6/Lys-50. The hydrogen bonding in CsE-v5 was particularly stable, with many amide exchange peaks visible even after 48 h (Figure 3). All slowly exchanging amide protons had an acceptor oxygen atom within the acceptable hydrogen-bonded distance and angle except for Tyr-4 and Asn-18. A comparison of the spectra at 293 and 303 K confirmed that these slowly exchanging amides had a smaller temperature dependence of their amide proton chemical shifts. Seven additional H-bonds were introduced on the basis of nOe contacts, proximity to an acceptor oxygen in the prefinal structures, and the temperature dependence of the amide proton chemical shift. Four ideal β-turns were determined from characteristic coupling constants and nOe patterns: Type I at 7−10, 26−29, and 46−49, and Type I′ at 36−39. The turns at 7−10 and 26−29 were stablized by additional backbone H-bonds from 7 to 11 and 26 to 30, respectively. In addition, the 7−10 turn was also found to have H-bonds from the Ser-9 and Cys-11 amides to the Asp-7 side-chain oxygen atoms. The turn at 36−39 was determined to be Type I′ (positive φ angles for Ala-37 and Met-38) on the basis of coupling constants and nOe contacts. Two hydrophobic surfaces were observed: a large area consisting of residues Tyr-4, His-34, Tyr-36, Ala-37, Tyr-41, and Ala-55, and a smaller one comprised of Leu-13, Val-16, Ala-17, and Tyr-20. Their presence was evident from the abundance of NOESY contacts involving aromatic to aromatic and aromatic to other nonpolar residues. Tyr-4 and Tyr-41 in the patch align in a "herringbone" configuration (*30*, *31*), as confirmed by the unusual chemical shifts for the ring protons of Tyr-41, i.e., the $C^{\delta}H$ protons are shifted upfield of the $C^{\delta}H$ protons. Whereas most of the polar side chains were found near the surface either completely exposed or involved in a hydrogen bond, the side chains of Lys-1 and Lys-12 were found to be buried in the hydrophobic core. The H^{ξ} of Lys-1 was within H-bonding distance of the carbonyl oxygens of Glu-47 and Ala-49, and was presumed to be H-bonded to these. Lys-12 H^{ξ} had no obvious acceptor atoms. The side chain of Lys-28 was quite well-defined, with long-range NOE contacts to Pro-5 and Leu-13. This well-defined orientation may be a result of a hydrogen bond to the carbonyl of the Gln-24 side chain, which is also well defined.

Modeling Results. Two-hundred distance geometry structures were generated, which were then used as input for the dynamical simulated annealing calculations. Of these, 120 structures were found to be acceptable with constraint violations less than or equal to 0.2 Å for NOE, 5° for torsion angles, 0.002 Å for bond lengths and 0.52° for bond angles. From this output, a family of 20 structures with bond-length and bond−angle violations less than 0.0015 Å and 0.5°, respectively, was selected as the final structures constituting the $\langle SA \rangle_k$ family. These structures were run through PROCHECK and found to be well behaved. In particular, all nonglycine residues were found within allowed regions of ϕ-ψ space (Figure 5). The average structure was computed from these, and energy minimized to obtain the $\langle \overline{SA} \rangle_{kr}$ structure. Tables 1 and 2 gives the structural statistics for the $\langle SA \rangle_k$ and $\langle \overline{SA} \rangle_{kr}$ structures. The $\langle \overline{SA} \rangle_{kr}$ structure exhibited an rms deviation of 0.66 Å for the backbone, and 0.97 Å for all heavy atoms, with respect to the family. When the C-terminal five residues (residues 56−60) were excluded, the rmsd values decreased to 0.45 and 0.76 Å for the backbone and all heavy atoms, respectively. In general, as expected the rmsd values showed inverse correlation with the number of constraints per residue (Figure 4). The helix and sheet regions had the largest number of constraints per residue, and consequently the lowest rmds. These trends are evident from the family of 20 backbone $\langle SA \rangle_R$ structures (Figure 6). This view shows the front face containing the

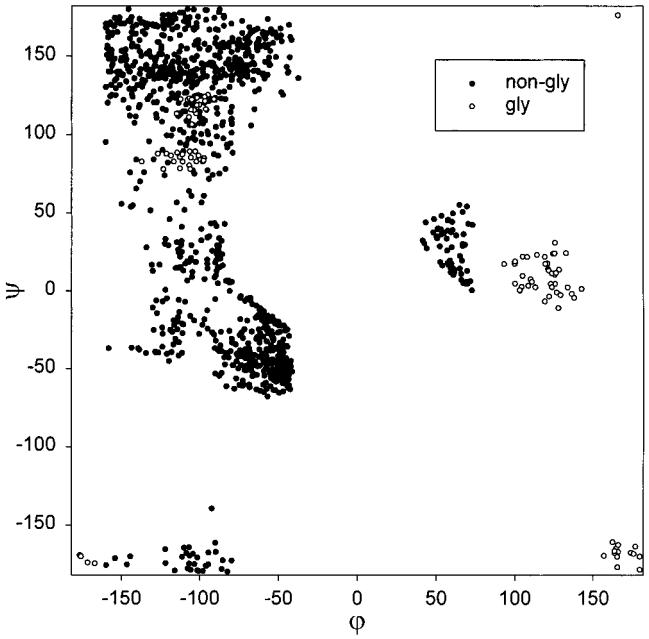

FIGURE 5: Ramachandran Plot for CsE-v5. Data are shown for the 20 structures used in the statistical analysis.

Table 1: Structural Statistics for CsE-v5

	$\langle SA\rangle_k$	$\langle\overline{SA}\rangle_{kr}$
rms deviations from experimental distance constraints (Å)		
all (919)	0.009 83	0.005 00
sequential ($\|i-j\|=1$) (216)	0.0129	0.007 38
medium-range ($1<\|i-j\|<5$) (66)	0.007 69	0.001 67
long-range ($\|i-j\|>5$) (99)	0.007 54	0.000 86
intraresidue (339)	0.008 39	0.004 94
H-bonds (70)	0.006 80	<0.001
rms deviations from experimental dihedral angle constraints (deg.) (91)	0.2	0.127
energies (kcal mol^{-1})		
E(nOe)	1.0	0.877
E(tor)	0.1	0.024
E(repel)	0.009	0.007
E(L-J)	−260	−271
deviations from idealized covalent geometry		
bonds (Å)	0.002	0.001 04
angles (deg)	0.6	0.475
impropers (deg)	0.4	0.360

Table 2: Atomic RMS Deviations (Å) for the CsE-I and CsE-v3

	backbone	all atoms
$\langle SA\rangle_k$ vs $\langle\overline{SA}\rangle_k$	1.10	1.55
$\langle SA\rangle_k$ vs $\langle\overline{SA}\rangle_{kr}$	1.09	1.71
$\langle\overline{SA}\rangle_k$ vs $\langle\overline{SA}\rangle_{kr}$	0.40	0.75
$\langle\overline{SA}\rangle_{kr}$ vs $\langle SA\rangle_{dr}$[a]	0.22	0.28
$\langle\overline{SA}\rangle_{kr}$ vs CsE-v3 (1−64)	1.48	
$\langle\overline{SA}\rangle_{kr}$ vs CsE-v3 (helix and sheet)	0.86	

[a] $\langle\overline{SA}\rangle_{dr}$ is the average of the $\langle SA\rangle_k$ structures energy minimized without restraints.

hydrophobic patch of residues with the α-helix at the top. A comparison of the backbone $\langle\overline{SA}\rangle_{kr}$ structure with a few representative toxins (obtained from PDB) is shown in Figure 7.

Electrostatic Potential Surfaces. Electrostatic potential surfaces were calculated for the energy-minimized average

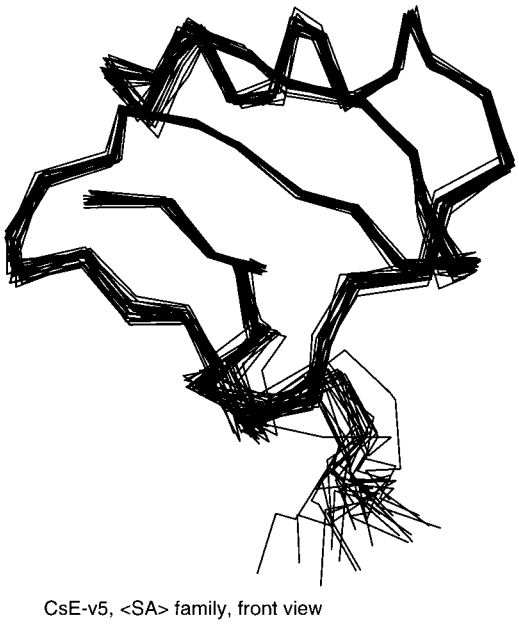

FIGURE 6: Twenty $\langle SA\rangle_k$ structures of CsE-v5 (Cα trace).

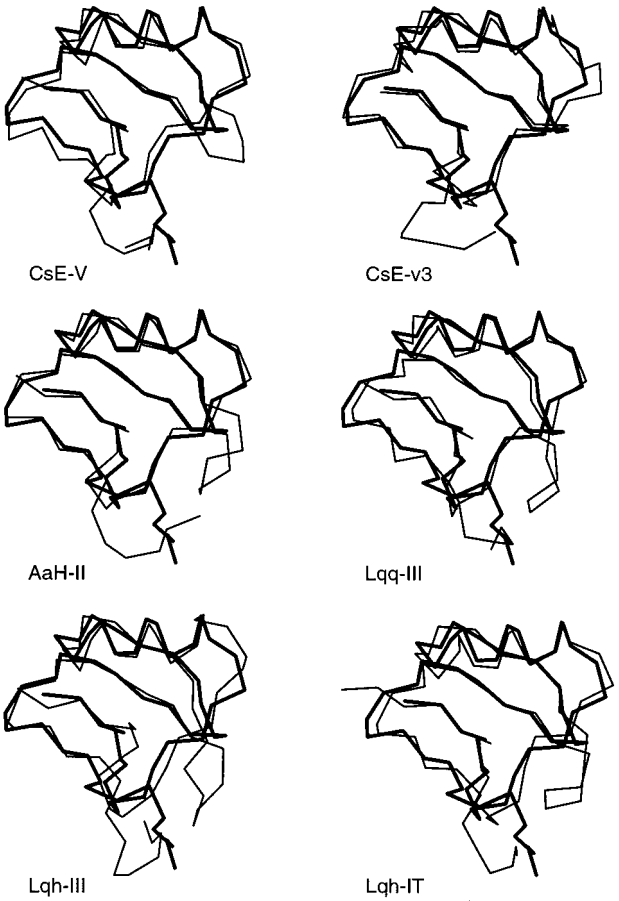

FIGURE 7: Best fit superposition of backbone structures of CsE-v5 (thick line) and CsE-V, CsE-v3, AaH-II, Lqq-III, Lqh-III, and Lqh-αIT. The coordinates were obtained from the protein data bank.

structures ($\langle\overline{SA}\rangle_{kr}$) of the two Old World-like toxins, CsE-v5 and CsE-V (Figure 8). CsE-v5 shows areas of both positive (blue) and negative (red) charge associated with the exposed charged side chains, and large neutral areas (white or light gray color) from the nonpolar regions. A large area of negative potential makes up the center of the face, mainly associated with residues Asp-2 and Glu-43. The C-terminus

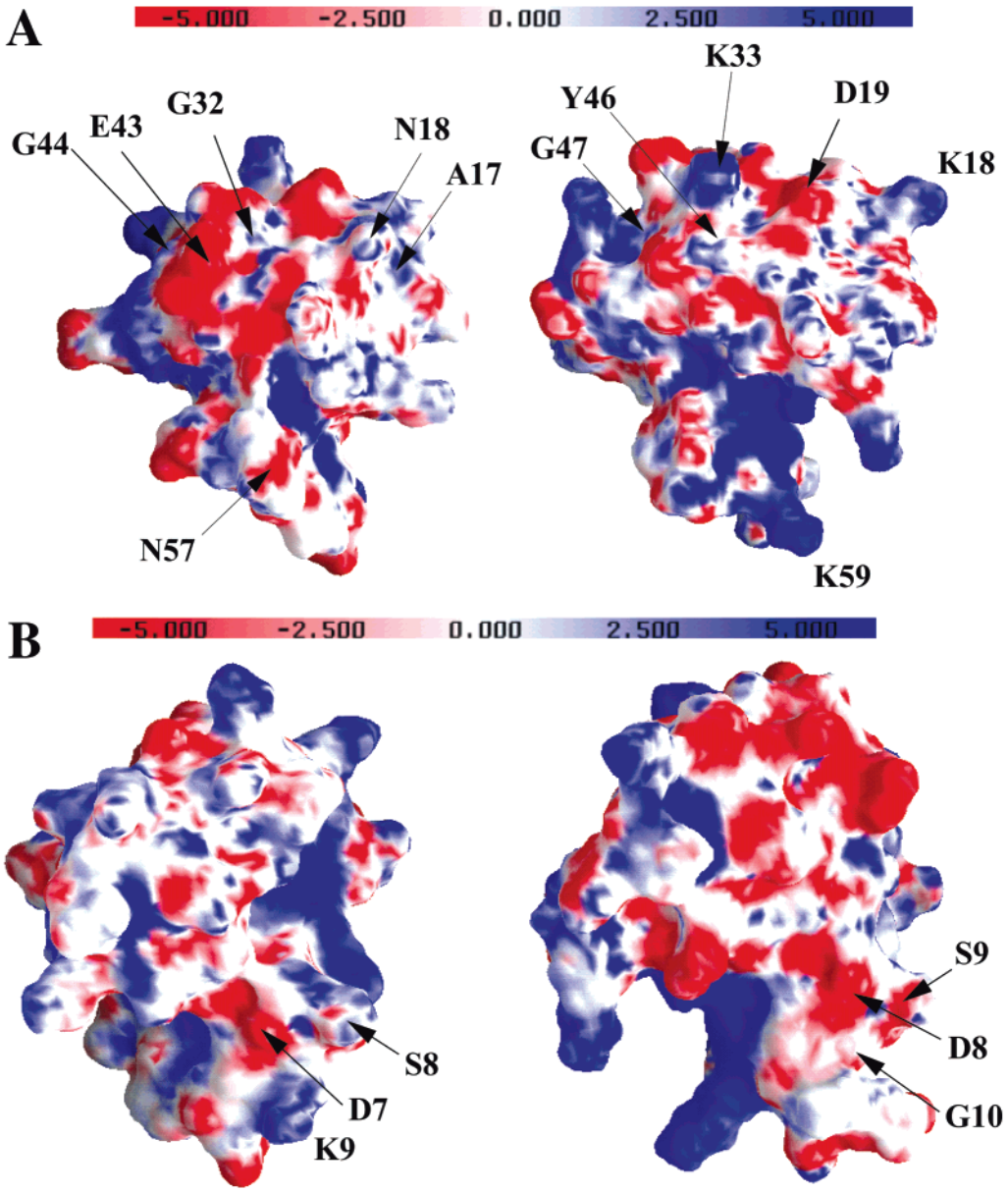

FIGURE 8: Comparison of electrostatic potential surfaces for CsE-v5 (on the left) and CsE-V (on the right) toxins. The front surface is shown in Figure 8A and the side view (looking edge-wise from the right of panel A) is shown in panel 8B. Some areas of significant difference are indicated in both the toxins. The potentials are shown at ±5 kT, with red representing the negative and blue the positive surfaces.

of CsE-v5 (bottom) is predominantly nonpolar, whereas it is positively charged in CsE-V.

Coordinates for the 20 ⟨SA⟩ structures and the $\langle\overline{SA}\rangle_{kr}$ structures were deposited in the Protein Data Bank (accession codes 1I6F and 1I6G). The chemical shifts were deposited in the BMRB (accession code BMRB-5029).

DISCUSSION

CsE-v5 Solution Structure. In solution, CsE-v5 exists as a compact protein consisting of an α-helix and an antiparallel *β*-sheet as the core secondary structures with several loops extending out from this core. The structure is highly ordered for all but the last five residues at the C-terminus. The rmsd for the family of structures was 0.66 and 0.97 Å for the backbone and heavy atoms, respectively. This reduced to 0.45 and 0.76 Å, respectively, when the last five residues were excluded. In addition to the presence of stable secondary structures such as the helix, *β*-sheet, and *β*-turns, additional factors that contribute to the stability include seven long-range hydrogen bonds not associated with the secondary structures, and packing of several hydrophobic residues extending from the front face to the backside. The stability of the protein (less residues 56−60) is evident from (1) the long exchange lifetimes exhibited by nearly 50% of the amide protons, (2) presence of all expected nOes observed in secondary structures, (3) no evidence of isomerization of prolines or cysteines, (4) a large number of hydrophobic side chain−side chain nOe contacts, and (5) some exceptionally well-defined side chains (rmsd < 0.5 Å): Tyr-4, Tyr-41, Lys-1, and Lys-12. The large number of very slowly exchanging amides is remarkable for a protein of this small size. In particular, as expected the chemical shifts of these slowly exchanging amide protons exhibited smaller temperature dependence compared to the more rapidly exchanging

amides. Exceptions were Asn-19, Gly-33, and Gly-44 which exchanged rapidly in the H/D exchange experiment. These amides were within acceptable distances of acceptor atoms in the initial round of calculations and were constrained as hydrogen bonds. All amides, except Tyr-4, which are either slowly exchanging or have other evidence for being hydrogen bonded such as small temperature dependence of the amide chemical shift or nOe contacts, were found to have acceptor oxygens within allowed distance and angle to form stable hydrogen bonds.

The protein core is made of primarily hydrophobic residues with some buried polar side chains, and consists of a 2.5 turn α-helix which is attached to the central strand of a three strand β-sheet by two disulfide bridges in an αβDB motif (*11*). Two β-turns which give rise to unusual φ angles were found in the core. The hairpin turn linking the second and third strands at 36/39 was determined to be Type I′. This type of turn, though energetically less favorable than a type-I turn, is common when connecting two β-strands (*29*) where the additional hydrogen bonding of the β-sheet "ladder" offsets the high-energy turn. The 26/29 turn from the α-helix into the β-sheet has Lys-29 with a positive φ at the $i+3$ position. In general, it is rare for a lysine to occupy this position in a Type I turn, as it does not help stablize the turn. This turn is also stablized by an additional hydrogen bond between Cys-25 and Ala-30. All the cysteines in the disulfide configurations were well-defined except for Cys-59, for which the degeneracy of its β-proton resonances precluded a determination of its side-chain orientation. Despite this, its partner, Cys-11, was well-defined as t^2g^3. The hydrophobic residues form a continuous patch extending from the front to the backside of the protein. This area is well-defined with the side-chain heavy atoms having less than 0.5 Å rmsd. The rings of Tyr-4 and -41 align perpendicularly in "herringbone" fashion. This orientation minimizes the exposed surface area of the rings and results in unusual chemical shifts for Tyr-41. Several of the lysine side chains, Lys-1, Lys-12, and Lys-28, are also well-defined, possibly due to the involvement of the amine group in hydrogen bonding. Only the Lys-1 amino group was found near an acceptor (O of Glu-47 and Ala-49) in the initial structures. Lys-28 had a possible acceptor at the Glu-24 side chain, but conclusive nOe contacts were not observed. No possible acceptor was found for Lys-12, even though this side chain was buried in the core.

Comparison of CsE-v5 to Other Scorpion Toxins. To better understand the nature of the toxin-sodium channel interactions, we compared sequence homology and the structure of CsE-v5 with some selected scorpion toxins, including CsE-V, the first example of an Old World-like toxin from CsE venom whose structure was determined (*2*). The sequence homology may be defined in terms of number of residues without regard to type, to exact matches, and to conservative substitutions, e.g., Lys for Arg. CsE-v5 has the features of both Old World α- and α-like toxins (a shortened turn out of the first β-strand, and shortened J and M loops) and New World toxins (shortened B loop), which is the same length as the other weak α-toxins (CsE-v1 to -v4, and -v6) and the β-toxins in CsE, and a longer F loop). These shortened loops make for a very compact protein whereas in other toxins these loops project out from the core. Differences in specific residues have been implicated in specific activity functions such as α- vs β-, insect vs mammal, and arthropod vs crustacean toxicity. Of particular interest to us is the comparison of the two Old World-like toxins from *C. scupturatus*, CsE-v5 and CsE-V that show significant differences in their activity (vide infra). This comparison is dicussed later on the basis of their electrostatic potential surfaces.

Table 3: Side-Chain Configurations of Aromatics and Cysteines in Selected Scorpion Neurotoxins

aromatics on the hydrophobic face				
CsE-v5	CsE-V	CsE-I	CsE-v3	AaH-II
Y4 tg	Y5 tg	Y4 tg	Y4 tg	Y5 tg
		Y37 gt	Y39 gt	
H34 gg	Y35 gg	Y39 gg	Y40 gg	Y35 gg
Y36 gt	Y37 gt	Y41 gt	Y42 gt	
				Y42 tg
Y41 gt	Y44 gt	Y46 gt	W47 gt	Y47 gt
	Y46 tg			Y49 tg
		W57 gt	Y58 gt	

cysteines shown as disulfide pairs in register with CsE-v5					
	CsE-v5	CsE-V	CsE-I	CsE-v3	AaH-II
C11	tg	tg	tg	tg	tg
C60	NA	gt	gt	gt	gt
C15	gg	gg	gg	gg	gg
C35	gt	gt	gt	gt	gt
C21	tg	tg	tg	gt	tg
C40	gt	tg	tg	tg	tg
C25	tg	tg	gt	tg	tg
C42	tg	tg	gt	tg	tg

Figure 7 shows a best-fit superposition of the backbone conformation of CsE-v5 with a few typical long-chain neurotoxins from the New World and the Old World scorpion venoms. The rmsd calculation was restricted to the backbone atoms of amino acids in register with residues 1−52 of CsE-v5 only since the last segment of residues 53−60 in CsE-v5 shows a different conformation with respect to other toxins and is also somewhat poorly defined. Further, loops of dissimilar lengths were excluded. The toxins and their rmsd deviations (in brackets) are CsE-V (1.39 Å), CsE-v3(1.37 Å), AaH-II (1.24 Å), Lqq-III (1.51 Å), Lqh-III (1.83 Å), and Lqh-IT (1.31 Å). It is of interest that the backbone rmsd for the helix and sheet residues are similar, even though the C40−C25 disulfide bond conformation in the αβDB motif for CsE-v5 is different compared to that in other toxins. CsE-v5 has a similar overall fold and shows similar secondary structural elements of other long-chain scorpion toxins, as well as similar side-chain orientations for the surface aromatic residues and most of the disulfide-bonded cysteines (Table 3). However, it also displays some significant differences in some of the loops, the conformation of the C-terminal domain, and the side chain orientation of C-40 which is bonded to C-25 (Table 3) with respect to these other proteins.

As described above, the secondary and tertiary structural features of CsE-v5 are very stable as shown by the slow amide exchange rates for nearly 50% of the backbone amide protons, a property which varies significantly among the CsE toxins. For example, no slowly exchanging amide peaks were observed for the β-toxin CsE-I (*7*), whereas CsE-v3 and -V showed the slowly exchanging amides essentially only in the secondary structural regions, and in long-range hydrogen

bonds (*2*, *11*). The exceptional stability of amide hydrogens in CsE-v5 may be a reflection of the tight packing of the protein core from the shortened loops.

All residues with slowly exchanging amides had acceptors except Tyr-4, which is in the first β-strand, but has its amide facing away from the sheet. This amide may form a hydrogen-bonded network with one or more buried waters and other nearby residues. In the CsE-v3 crystal structure (*32*), a bound water is found near the amide of Tyr-4 and the carbonyl of Tyr-58 (Ser-52 in CsE-v5). A nearby long-range hydrogen bond Ser-52 HN (Tyr-58)/Tyr-4 CO is found in both proteins. All expected slowly exchanging amides were seen except Asn-21, Gly-33, Cys-35, Tyr-36, and Gly-44. These amides are within bonding distance of acceptors in the initial structures, and had small changes in the amide chemical shifts with temperature. The amine of Lys-12 was found to be buried but had no obvious acceptor for a hydrogen bond. A homologous lysine is found in CsE-v3, and forms a network with a bound water and the carbonyl of Ala-43 (Ala-37 in CsE-v5). It is likely that the rigidity of Lys-12 might be due to the formation of a similar network in CsE-v5.

Electrostatic Potential Surface and Implications on Biological Activities. Figure 8 shows a comparison of the electrostatic potential surfaces for the two Old World-like neurotoxins from the New World scorpion *C. sculpturatus*, viz., CsE-v5 which is highly specific to insects, and CsE-V which exhibits both high anti-insect and high anti-mammal activities. Significant variations on the electrostatic potential surface of CsE-v5 (CsE-V) between the two toxins is observed around residues K9(G10), A17 (K18), N18(D19), G32(K33), E43(Y46), N57(K59). Considering the variations in target specificity of toxins within the same species (*1*) and from different species including those between the Old World and the New World (*17*, *18*) and the recognition by some toxins of distinct or overlapping receptor sites on the sodium channel (*17*), a direct translation of results of mutational studies from one toxin to another can be non-trivial. Nevertheless, the results of such mutational studies may be useful in gaining some general understanding of the approximate location of functional residues. On the basis of site-directed mutagenesis studies, residues Y-49, A-50, and N-54 in LqhαIT toxin have been implicated in anti-mammal activity (*14*). In CsE-V, which exhibits anti-mammal activity, the corresponding residues are Y-46, G-47, and N-51, that have a somewhat similar charge properties as in LqhαIT. However, in CsE-v5, the corresponding residues are E-43, G-44, and N-48. The replacement of a Tyr by a Glu at position 43 results in a significant alteration of the electrostatic potential, being considerably more negatively charged at and in the vicinity of residue 43 in CsE-v5, and it is likely that this highly negatively charged surface might be responsible for the loss of anti-mammal activity in CsE-v5. Residues 8−10 and 64 in LqhαIT have been implicated as essential for anti-insect activity (*14*). Among the New World α-toxins in Table 2, the anti-insect activity (vide supra) decreases from CsE-V to CsE-v3 in the order CsE-V > CsE-v5 > CsE-v1 > CsE-v3 (*1*). A comparison of the corresponding three residues in these toxins which are in register with residues 7−9 in CsE-v5 shows that this variation in insect toxicity correlates approximately with the average charge among these residues, with CsE-V being the most negative and CsE-v1 and -v3 being the most positive. Figure 8B shows a comparison of electrostatic potential surface centered around residues 7−9 in CsE-v5 and 8−10 in CsE-V, and the differences in electrostatic potential surface is clear. Additionally, a positively charged surface near the C-terminus was also suggested as essential for receptor interaction on insect neuronal receptors (*9*). Once again, a comparison of the potential surfaces near the C-terminus shows that CsE-v5 is less positively charged than CsE-V due to the N57(K59) substitution and might explain in part the slightly less insect activity of this protein compared to CsE-V. However, the C-terminal amidation of CsE-v5 which makes it less negative at the C-terminus may also contribute to the higher activity of CsE-v5 compared to CsE-v1 and -v3. Some of these conclusions might be considered as tentative in view of the complexity of the interaction of toxins with sodium channels (vide supra). Like other Old World toxins, it is interesting that the two Old World-like toxins, CsE-v5 and CsE-V, also are more toxic to insects than the remaining CsE variant toxins.

ACKNOWLEDGMENT

The authors also wish to express their gratitude to Dr. Dean D. Watt, Professor Emeritus of Biomedical Science at Creighton University (Omaha, NE) for his collaboration and consultation over the many years on our laboratory's effort on the structural biology of neurotoxins from the scorpion *C. sculpturatus* Ewing.

REFERENCES

1. Simard, J. M., Meves, H., and Watt, D. D. (1992) Neurotoxins in venom from the north American scorpion *Centruroides sculpturatus* Ewing. In *Natural Toxins: Toxicology, Chemistry, and Safety* (Keeler, R. F., Mandava, N. B., and Tu, A. T., Eds.) pp 236−263, Alaken Inc., Fort Collins, CO.
2. Jablonsky, M. J., Watt, D. D., and Krishna, N. R. (1995) Solution structure of an Old-World-like neurotoxin from the venom of the New World scorpion *Centruroides sculpturatus* Ewing. *J. Mol. Biol. 248*, 449−458.
3. Landon, C., Cornet, B., Bonmatin. J.-M., Kopeyan, C., Rochat, H., Vovelle, F., and Ptak, M. (1996) 1H NMR derived secondary structure and the overall fold of the potent anti-mammal and anti-insect toxin III from the scorpion *Leiurus quinquestriatis quinquestriatus*. *Eur. J. Biochem. 236*, 395−404.
4. Jover, E., Couraud, F., and Rochat, H. (1980) Two types of scorpion neurotoxins characterized by their binding to two separate receptor sites on rat brain synaptosmoes. *Biochem. Biophys. Res. Commun. 95*, 1607−1614.
5. Meves, H., Rubly, N., Watt, D. D. (1982) Effects of toxins isolated from the venom of the scorpion *Centruroides sculpturatus* on the Na current of the node of ranviers. *Pflügers Arch 393*, 56−62.
6. Gordon, D., Savarin, P., Gurevitz, M., and Zinn-Justin, S. (1998) Functional anatomy of scorpion toxins affecting sodium channels. *J. Toxicol., Toxin Rev. 17*, 131−159.
7. Jablonsky, M. J., Jackson, P. L., Watt, D. D., and Krishna, N. R., (1999) Solution structure of a β-neurotoxin from the New World scorpion *Centruroides sculpturatus* Ewing. *Biochem. Biophys. Res. Commun. 254*, 406−412.
8. Possani, L. D., Becerril, B., Delepierre, M., and Tytgat, J. (1999) Scorpion toxins specific for Na^+-channels. *Eur. J. Biochem. 264*, 287−300.
9. Krimm, I., Gilles N., Sautierre, P., Stankiewicz. M., Pelhate. M., Gordon, D., and Lancelin. J. M. (1999) NMR structures and activity of a novel α-like toxin from the scorpion *Leiurus quinquestriatus hebraeus. J. Mol. Biol. 285*, 1749−1763.

10. Pintar, A., Possani, L. D., and Delepierre, M. (1999) Solution structure of toxin-2 from *Centruroides noxius Hoffmann*, a beta-scorpion neurotoxin acting on sodium channels. *J. Mol. Biol. 287*, 359−367.
11. Lee, W., Moore. C. H., Watt, D. D., and Krishna, N. R. (1994) Solution structure of the variant-3 neurotoxin from *Centruroides sculpturatus* Ewing. *Eur. J. Biochem. 218*, 89−95.
12. Darbon, H. (1990) Scorpion neurotoxin: a molecular model of interaction with the voltage-dependent sodium current. In *Protein structure−function* (Zaidi, Abbasi, and Smith, Eds.) pp 169−181, TWEL publishers.
13. Loret, E. P., Martin-Eauclaire, M.-F., Mansuelle, P., Sampieri, F., Granier, C., and Rochat, H. (1991) An anti-insect toxin purified from the scorpion *Androctonus australis* Hector also acts on the α- and β-sites of the mammalian sodium channel: sequence and circular dichroism study. *Biochemistry 30*, 633−640.
14. Zilberberg, N., Froy, O., Loret, E., Cestele, S., Arad, D., Gordon, D., and Gurevitz, M. (1997) Identification of structural elements of a scorpion α-neurotoxin important for receptor site recognition. *J. Biol. Chem. 272*, 14810−14816.
15. Tugarinov, V., Kustanovich, I., Zilberberg, N., Gureritz, M., and Anglister, J. (1997) Solution structure of a highly insecticidal recombinant scorpion α-toxin and a mutant with increased activity. *Biochemistry 36*, 2414−2424.
16. Kharrat, R., Darbon, H., Rochat, H., and Granier, C. (1989) Structure/activity relationships of scorpion α-toxins: Multiple residues contribute to the interaction with receptors. *Eur. J. Biochem. 181*, 381−390.
17. Gordon, D., Martin-Eauclaire, M.-F., Cestele, S., Kopeyan, C., Carlier, E., Khalifa, R. B., Pelhate, M., and Rochat, H. (1996) Scorpion toxins affecting the sodium current inactivation bind to distinct homologous receptor sites on rat brain and insect sodium channels. *J. Biol. Chem. 271*, 8034−8045.
18. Wheeler, K. P., Watt, D. D., and Lazdunski, M. (1983) Classification of Na Channel receptors specific for various scorpion toxins. *Pflügers Arch. 397*, 164−165.
19. Gilson, M. K., Sharp, K. A., and Honig, B. H. (1988) Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comput. Chem. 9*, 327−335.
20. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct., Funct., Genet. 11*, 281−296.
21. Babin, D. R., Watt, D. D., Goos S. M., and Mlejnek R. V. (1974) Amino acid sequences of neurotoxic protein variants from the venom of *Centruroides sculpturatus* Ewing. *Arch. Biochem. Biophys. 164*, 694−706.
22. Montelione, G. T., Wüthrich, K, Burgess, A. W., Nice, E. C., Wagner, G., Gibson, K. D., and Scheraga, H. A. (1992) Solution structure of murine epidermal growth factor determined by NMR spectroscopy and refined by energy minimization with restraints *Biochemistry 31*, 236−249.
23. Wang, Y., Nip, A. M., and Wishart, D. S. (1997) A simple method to quantitatively measure polypeptide $J_{HN-H\alpha}$ coupling constants from TOCSY or NOESY spectra *J. Biomol. NMR 10*, 373−382.
24. Nilges, M., Clore, G. M., and Gronenborn, A. M., (1990) 1H NMR stereospecific assignments by conformational database searches. *Biopolymers 29*, 813−822.
25. Constantine, K. L., Friedrichs, M. S., and Mueller, L. (1994) Simple approaches for estimating vicinal ^{1}H-^{1}H coupling constants and for obtaining stereospecific resonance assignments in leucine side chains. *J. Magn. Reson. B104*, 62−68.
26. Xu, Y., Jablonsky, M. J., Jackson, P. L, Braun, W., and Krishna, N. R. (2001) Automatic Assignment of NOESY Cross Peaks and Determination of the Protein Structure of a New World Scorpion Neurotoxin Uaing NOAH/DIAMOD. *J. Magn. Res. 148*, 35−46.
27. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
28. Driscoll, P. C., Clore, G., M., Beress, L., and Gronenborn, A. M. (1989) Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from the sea anemone *Anemonia sulcata*, A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* 28, 2188−2198.
29. Wilmot, C. M., and Thornton, J. M. (1988) Analysis of the different types of β-turns in proteins. *J. Mol. Biol. 203*, 221−232.
30. Burley, S. K., and Petsko, G. A. (1985) Aromatic−aromatic interaction: a mechanism of protein structure stabilization. *Science 229*, 23−28.
31. Krishna, N. R, Nettesheim, D. G., Klevit, R. E., Drobny, G., Watt, D. D., and Bugg, C. E. (1989) Proton Nuclear Magnetic Resonance characterization of the aromatic residues in the variant-3 neurotoxin from *Centruroides sculpturatus* Ewing. *Biochemistry*, 28, 1556−1562.
32. Zhao, B., Carson, M., Ealick, S. E., and Bugg, C. E. (1992) Structure of scorpion toxin variant-3 at 1.2 Å resolution. *J. Mol. Biol. 227*, 239−252.
33. Rochat, H., Rochat, C., Sampieri, F., Miranda, F., and Lissitzky, S. (1972) The amino acid sequence of neurotoxin II of *Androctonus australis* Hector. *Toxicon 22*, 695−703.
34. Kopeyan, C., Mansuelle, P., Sampieri, F., Brando, T., Bahraoui, E. M., Rochat, H., and Granier, C. (1990) Primary structure of scorpion anti-insect toxins isolated from the venom of *Leiurus quinquestriatus. FEBS Lett. 261*, 423−426.
35. Kopeyan, C., Mansuelle, P., Martin-Eauclaire, M. F., Rochat, H., and Miranda, F. (1993) Toxin III of the scorpion *Leiurus quinquestriatus quinquestriatus*, an α-toxin highly toxic both on mammals and insects: sequence and pharmacology. *Nat. Toxins. 1*, 308−312.
36. Sautiere, P., Cestele, S., Kopeyan, C., Martinage, A., Drobecq, H., Doljansky, Y., and Gordon, D. (1998) New toxins acting on sodium channels from the scorpion *Leiurusquinquestriatus hebraeus* suggest a clue to mammalian vs insect selectivity. *Toxicon 36*, 1141−1154.
37. Eitan, M., Fowler, E., Herrmann, R., Duval, A., Pelhate, M., and Zlotkin, E. (1990) A scorpion venom neurotoxin paralytic to insects that affects sodium current inactivation: purification, primary structure, and mode of action. *Biochemistry 29*, 5941−5947.

BI010223H