Proton Nuclear Magnetic Resonance and Distance Geometry/Simulated Annealing Studies on the Variant-1 Neurotoxin from the New World Scorpion *Centruroides sculpturatus Ewing*[†,‡]

Weontae Lee,[§,‖] Michael J. Jablonsky,[§] Dean D. Watt,[⊥] and N. Rama Krishna*[,§]

Departments of Biochemistry and Physics, and Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, Alabama 35294, and Department of Biochemistry, Creighton University, Omaha, Nebraska 68178

Received July 15, 1993; Revised Manuscript Received December 27, 1993[⊗]

ABSTRACT: The sequence-specific proton resonance assignments for the variant-1 (CsE-v1) neurotoxin from the venom of the New World scorpion *Centruroides sculpturatus Ewing* (range Southwestern United States) have been performed by 2D ^{1}H NMR spectroscopy at 600 MHz. The stereospecific assignments for the β-methylene protons of 19 non-proline residues have been determined. A number of short-, medium-, and long-range NOESY contacts as well as the backbone and the side-chain vicinal coupling constants for several residues have been determined. Slowly exchanging amide hydrogens from a number of residues have been identified. On the basis of the NMR data, the solution structure of this protein has been determined by a hybrid procedure consisting of distance geometry and dynamical simulated annealing refinement calculations. Distance constraints from the NOESY data and torsion angle constraints from proton vicinal coupling constant data were used in the simulated annealing calculations. The three-dimensional structure of CsE-v1 is characterized by a three-stranded antiparallel β-sheet, a short α-helix, a *cis*-proline, and intervening loops. A comparison with the solution NMR data of a homologous protein (CsE-v3) from the *Centruroides* venom, shows that the structures are essentially similar, except for some minor differences. Some of the NMR spectral perturbations are felt in regions far removed from sites of amino acid substitutions. The hydrophobic surface in CsE-v1 is slightly more extended than in CsE-v3.

The venom from the scorpion *Centruroides sculpturatus* Ewing (CsE)[1] (range Southwestern United States) contains many small basic proteins, each of which contains four disulfide bridges. Each toxin is about 7000 Da and contains 65–70 residues. There is a degree of species specificity exhibited by these toxins in that the various proteins have dramatic differences in toxicities to mammals and insects (Watt et al., 1978; Watt & Simard, 1984; Zlotkin et al., 1972a,b). Biochemical and electrophysiological characterization of the scorpion neurotoxins indicate that two distinct classes of sodium channel binding toxins are present in the venom (Meves et al., 1982; Simard et al., 1986). Variants 1–6 (CsE-v1 to -v6) and CsE-V, named α-toxins, slow the inactivation of the sodium channels, but with no effect on the Na activation. The CsE toxins I, III, IV, and VII, so called β-toxins, affect Na activation. The α- and β-toxins were also shown to bind to different sites on the sodium channel (Jover et al., 1980; Meves et al., 1982). Interestingly, α-toxins from the New World species of scorpions do not compete with the Old World α-toxin AaH II from *Androctonus australis Hector* in binding to rat brain synaptosomes (Wheeler et al., 1983). [For a summary of binding to synaptosomes by the scorpion toxins, see Meves et al. (1986)]. This was considered to be a result of species specificity of these polypeptides rather than the presence of a separate binding site for the New World α-toxins (Wheeler et al., 1983).

Our laboratory has previously characterized the secondary structure of CsE-v3 toxin by 2D-NMR spectroscopy (Nettesheim et al., 1989; Krishna et al., 1989) and reported the solution structures (Krishna et al., 1990; Lee et al., 1993). All other scorpion toxins for which solution structures were reported are from the Old World species and include the neurotoxin M_9 from *Buthus eupeus* (Pashkov et al., 1988), AaH IT (Darbon et al., 1991), and Toxin III (Mikou et al., 1992) from *A. australis Hector*, the K-channel binding charybdotoxin (Bontems et al., 1991a,b; Lambert et al., 1990) and leiurotoxin I (Martins et al., 1990) from *Leiurus quinquestriatus hebraeus*, and the short-chain toxin Be I_5A (Arseniev et al., 1984). Crystal structures were reported for the variant-3 from the New World species (Fontecilla-Camps et al., 1980, 1982; Almassy et al., 1983; Zhao et al., 1992), for AaH II from the Old World species (Fontecilla-Camps et al., 1988), and for the insect-directed toxin from *A. australis* venom (Fontecilla-Campus, 1989).

Even though the CsE-v1 and -v3 toxins have nearly 90% sequence homology, the former is 20 times more toxic to insects (Watt & Simard, 1984). They differ at positions 15, 17, 18, 21, 24, and 26 in the amino acid sequences (see Figure 1). The region 12–26 is referred to as the J-loop. In this paper, we report the sequence-specific proton NMR assignments of this protein and details of its secondary structure. We also report

[†] This work was supported by the grant MCB-9118503 from the National Science Foundation and by the grant CA-13148 from the National Institutes of Health.

[‡] The atomic coordinates for 26 simulated annealing structures and the energy minimized average structure ($\langle\overline{SA}\rangle_{kr}$), have been deposited (file names 1VNA and 1VNV, respectively) with the Protein Data Bank, Brookhaven National Laboratories, Upton, Long Island, New York 11973.

* Address all correspondence to this author at the Department of Biochemistry.

[§] University of Alabama at Birmingham.

[‖] Present address: Division of Molecular and Structural Biology, Department of Medical Biophysics, Ontario Cancer Institute, University of Toronto, Toronto M4X 1K9, Ontario, Canada.

[⊥] Creighton University.

[⊗] Abstract published in *Advance ACS Abstracts,* February 1, 1994.

[1] Abbreviations: CsE, *Centruroides sculpturatus Ewing*; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; SA, simulated annealing; DG, distance geometry.

0006-2960/94/0433-2468$04.50/0

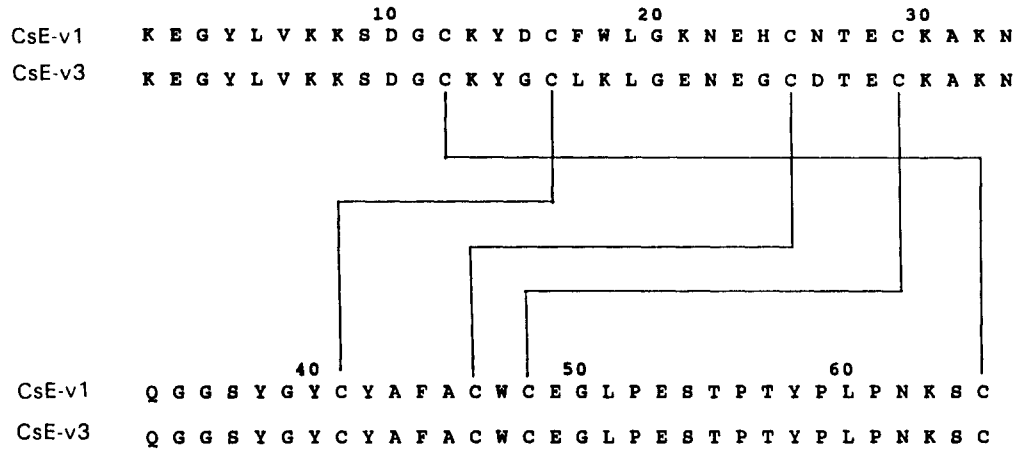

FIGURE 1: Amino acid sequences of CsE-v1 and CsE-v3 toxins from *Centruroides sculpturatus Ewing*. The solid lines indicate the disulfide bridges.

the detailed three-dimensional structure of this protein in solution determined by distance geometry and dynamical simulated annealing refinement calculations. A comparison is also made of the chemical shifts of the CsE-v1 and CsE-v3 toxins.

MATERIALS AND METHODS

(*i*) *Sample Preparation.* The isolation and purification of the variant-1 toxin has been described earlier (Babin et al., 1974). The NMR experiments were performed at sample concentrations of 0.5–1.0 mM. The solution pH was maintained at 3.0. Experiments at higher pH values (>4.5) were attempted but abandoned since many of the backbone amide hydrogens exhibited significant exchange broadening that resulted in lower sensitivity for the amide protons (due to solvent saturation transfer) in the 2D-NMR spectra obtained with solvent presaturation. Further, the exchange broadening obscured coupling constant information in phase-sensitive COSY spectra. The solution structure obtained at pH 3 was considered to be relevant since we did not see any evidence for pH-dependent structural changes when compared with NOESY spectra obtained at pH 6.5. Furthermore, the lethality of the crude venom incubated at different pH values (2.0–11.0) was not affected when compared to control samples at pH 7.4 (McIntosh & Watt, 1967).

(*ii*) *NMR Spectroscopy.* The NMR experiments were performed on a Bruker AM 600 spectrometer equipped with an Aspect 3000 computer and operating in quadrature detection mode. Most of the experiments were performed at 303 and 313 K, whereas the amide hydrogen exchange experiment was performed at 293 K. Mixing times of 100, 150, and 200 ms were employed in collecting NOESY spectra (Macura & Ernst, 1980). TOCSY spectra (Bax, 1989) in H_2O and in D_2O solutions were recorded with mixing times of 35 and 78 ms using an MLEV17 pulse sequence for the spin lock. DQF-COSY (Rance et al., 1983) experiment was performed at 313 K in D_2O solution, and phase-sensitive COSY (Marion & Wuthrich, 1983) experiment was performed in H_2O solution. All phase-sensitive mode measurements were performed using time proportional phase incrementation (TPPI) (Marion & Wuthrich, 1983) with 2K data points in the t_2 dimension and 512 t_1 increments. For the phase-sensitive COSY and DQF-COSY experiments, the time domain data size was 4K (t_2) × 1K (t_1). The exchange experiment to identify slowly exchanging amide hydrogens was performed as a 2D-COSY in the absolute value mode on a freshly prepared D_2O solution of the protein after lyophilization of a H_2O sample.

All NMR data were transferred to a microVAX II computer and processed using the FTNMR program (Hare Research Inc., Woodinville, WA). Prior to Fourier transformation in the t_1 dimension, the first row was half-weighted (Otting et al., 1986) to suppress t_1 ridges. The phase-sensitive COSY and DQF-COSY data sets were zero filled to 4K × 4K real data matrices to obtain a final digital resolution of 1.62 Hz/pt in the F_2 dimension (1.47 Hz/pt for D_2O solutions). The proton chemical shifts were referenced to sodium 4,4-dimethyl-4-silapentane 1-sulfonate (DSS).

(*iii*) *Simulated Annealing Calculations.* The three-dimensional structures were generated using a hybrid procedure that employed distance geometry (Kuntz et al., 1989) and dynamical simulated annealing refinement. The calculations were done using the X-Plor/QUANTA/CHARm modeling package (version 3.0, Molecular Simulations, Inc.) on a Silicon Graphics Iris workstation. The methodology employed by us for protein structure refinement was based on the original work of Clore and Gronenborn and their co-workers (Nilges et al., 1988a-c; Driscoll et al., 1989) but with minor modifications that were described elsewhere (Lee et al., 1993).

(*iv*) *Experimental Constraints.* On the basis of cross-peak intensities in the NOESY spectra obtained with mixing times of 100, 150, and 200 ms, the NOE constraints were classified into three distance ranges, strong (1.8–2.7 Å), medium (1.8–3.3 Å), and weak (1.8–5.0 Å). Corrections for pseudoatom representations were used for nonstereospecifically assigned methylene protons, methyl groups, and the ring protons of Tyr and Phe (Wuthrich et al., 1983). In addition, 0.5 Å was added, as an intensity correction, to the upper limit of distance constraints involving methyl group resonances (Clore et al., 1987; Wagner et al., 1987). The NOESY spectra of the protein obtained in H_2O and D_2O solutions yielded a total of 492 NOE constraints. Of these, 212 were intraresidue, 130 were sequential ($|i-j| = 1$), 29 were short-range ($1 < |i-j| < 5$), and 121 were long-range ($|i-j| > 5$) NOESY contacts. Figure 2 shows a summary of the interresidue NOESY contacts established in CsE-v1.

Each of the four disulfide bonds was represented by three distance constraints: $r_{S(i)-S(j)} = 2.0 \pm 0.1$ Å and $r_{C\beta(i)-S(j)} - 3.0 \pm 0.1$ Å. On the basis of the expected NOESY contacts [$\alpha(i) - \beta(j)$] between two cysteine residues forming the disulfide bridge, the published original amino acid sequence (Babin et al., 1974) was revised by interchanging Cys64 and the C-terminal Ser65 to be Ser64 and Cys65. Such a revision is consistent with the correction made by Fontecilla-Camps et al. (1980) for the CsE-v3 toxin on the basis of diffraction data. Hydrogen bond distance constraints were introduced

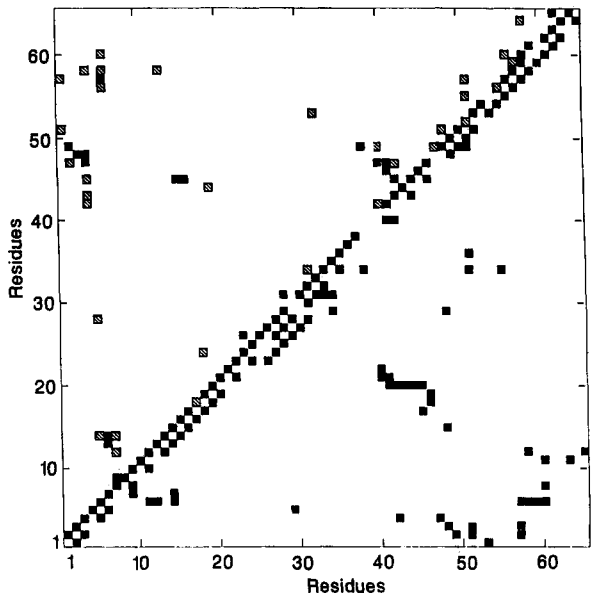

FIGURE 2: Summary of interresidue NOESY contacts established in CsE-v1. The backbone–side-chain contacts (solid boxes) are indicated in the lower right half, while the upper left half shows backbone–backbone (solid boxes) and side-chain–side-chain (striped boxes) contacts.

by defining two constraints for each bond, $r_{N(i)-O(j)} < 3.3$ Å and $r_{NH(i)-O(j)} < 2.3$ Å. To be on the conservative side, in the initial stages hydrogen bond constraints were introduced only for 17 of the amide hydrogens for which the acceptor oxygen atoms could be readily inferred (e.g., in α-helix, β-sheet, and β-turn). An additional hydrogen bond between the indole NH of Trp-47 and the side-chain carboxylate group of Glu2 was also included. The presence of this hydrogen bond was inferred on the basis of a strong NOE contact between the indole NH and the $C^{\beta}H_2$ protons of Glu2, as well as the anomalous pH-dependent chemical shift exhibited by the indole NH of a highly homologous toxin, CsE-v3, studied earlier by this laboratory (Krishna et al., 1989). From an inspection of the average structures in some preliminary set of calculations, suitable acceptor oxygen atoms capable of forming hydrogen bonds were identified unambiguously for an additional seven amide hydrogens, and the average structure was further energy minimized by including the additional hydrogen bonds as constraints. The additional hydrogen bonds introduced were K7NH-C12CO, Y58NH-Y4CO, C16NH-F44CO, T55NH-Y52CO, K63NH-L60CO, G35NH-A31CO, and C41NH-G20CO. No constraints were introduced for four slowly exchanging amide hydrogens (L60, Y14, F17, and T57); in the final structure each was in close proximity to two oxygen atoms, and hence were capable of forming stable hydrogen bonds.

Constraints for the ϕ torsion angles of 31 residues were deduced on the basis of the $^3J_{N\alpha}$ coupling constants obtained from phase-sensitive COSY spectra in H_2O solution. For $^3J_{N\alpha} > 8$ Hz, ϕ was constrained to $-120° \pm 50°$; for $^3J_{N\alpha} < 6.5$ Hz, ϕ was constrained to $-55° \pm 45°$. The χ torsion angle constraints were deduced for 19 of the residues on the basis of stereospecific assignments of the β-methylene protons using standard procedures (Wagner et al., 1987; Basus, 1989). For Pro59, which was identified to be in a cis conformation (vide infra), the ω torsion angle for Tyr58–Pro59 peptide bond was constrained to be $0° \pm 5°$. The remaining prolines were constrained to trans peptide bond conformations ($180° \pm 5°$).

RESULTS AND DISCUSSION

(*i*) *Sequence-Specific Assignments.* The sequence-specific assignments of proton resonances from each residue in CsE-v1 were made using standard procedures (Wuthrich, 1986) from 2D-NMR spectra recorded at 303 and 313 K. Figure 3 shows the fingerprint region in the PH-COSY spectrum of CsE-v1. The sequential connectivities and the short- and medium-range contacts established are essentially similar to those observed for the CsE-v3. The detailed chemical shift assignments are included in Table 1. We have also been able to establish the stereospecific assignments for the β-methylene protons of 19 residues using standard procedures that involve the use of NH–$C^{\beta}H$ and $C^{\alpha}H$–$C^{\beta}H$ NOE contacts and the αH–βH coupling constants. On the basis of this data, the side-chain torsion angles, χ, were also deduced for these residues (vide supra).

(*ii*) *Secondary Structural Elements.* The observation of sequential d_{NN} contacts together with the $\alpha N(i, i+3)$ and

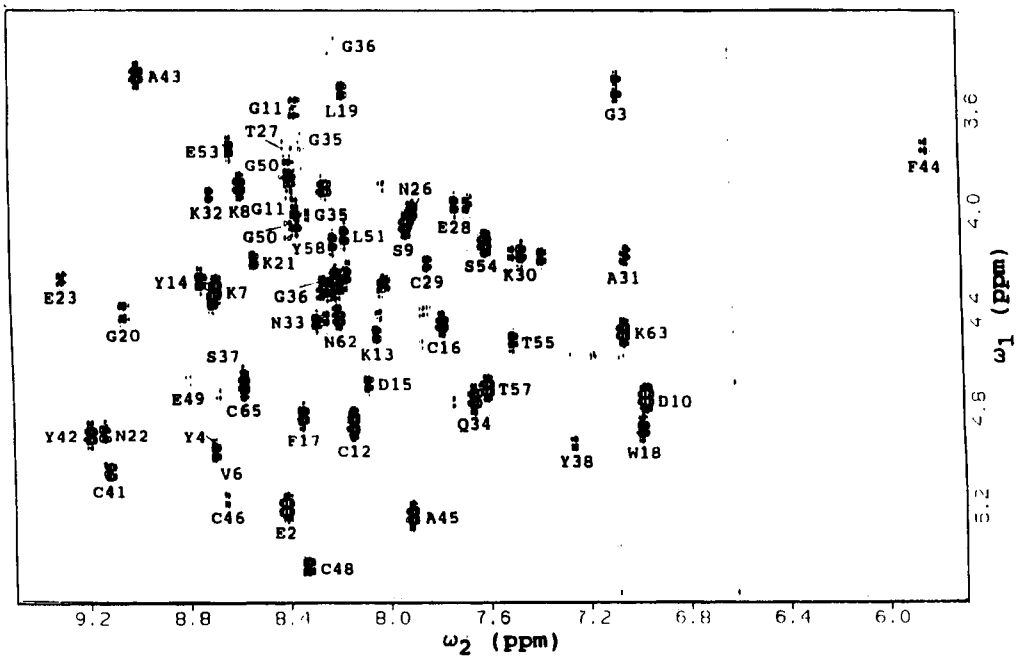

FIGURE 3: Fingerprint region in the PH-COSY spectrum of CsE-v1/H_2O.

Table 1: ^{1}H Resonance Assignments of the Individual Residues in the CsE-v1 Neurotoxin (pH 3.0, 313 K)[a]

residue	NH	$C^{\alpha}H$	$C^{\beta}H$	other
Lys1		4.17	1.88, 1.78	$C^{\gamma}H_2$ 1.05, 1.01, $C^{\beta}H_2$ 1.60, 1.37; $C^{\epsilon}H_2$ 2.64, 3.03; $N^{\zeta}H$ 7.24
Glu2	8.41	5.20	2.23, 1.98	$C^{\gamma}H_2$ 2.49, 2.39
Cly3	7.06	3.50		
		1.65		
Tyr4	8.70	4.97	3.12,* 3.89	(2, 6) 7.43, (3, 5) 7.33
Leu5	7.15	4.41	1.57, 1.28*	$C^{\gamma}H$ 1.45; $C^{\delta}H_3$ 0.61, 0.37
Val6	8.69	4.98	1.74	$C^{\gamma}H_3$ 0.92, 0.65
Lys7	8.68	4.30	1.98, 1.74	$C^{\gamma}H_2$ 0.96, 0.85; $C^{\delta}H_2$ 1.33; $C^{\epsilon}H_2$ 2.42, 2.29
Lys8	8.58	3.89	1.82	$C^{\gamma}H_2$ 1.59, 1.40; $C^{\delta}H_2$ 1.71; $C^{\epsilon}H_2$ 2.99
Ser9	7.92	4.06	3.75, 3.65	
Asp10	6.97	4.75	3.13, 2.58*	
Gly11	8.35	4.03		
		3.58		
Cys12	8.15	4.85	3.33,* 2.89	
Lys13	8.04	4.50	1.72, 1.29	$C^{\gamma}H_2$ 0.57; $C^{\delta}H_2$ 1.54; $C^{\epsilon}H_2$ 2.67
Tyr14	8.74	4.26	2.81, 2.72*	(2, 6) 7.05, (3, 5) 6.49
Asp15	8.07	4.69	2.62, 2.55*	
Cys16	7.77	4.45	3.18, 2.95*	
Phe17	8.34	4.81	3.21, 2.66*	(2, 6) 7.05, (3, 5) 7.18, (4) 7.11
Trp18	6.98	4.88	3.38, 3.18	(NH) 10.14, (2) 7.40, (4) 7.77 (5) 7.27, (6) 7.29, (7) 7.57
Leu19	8.16	3.51	1.59	$C^{\gamma}H$ 1.29; $C^{\delta}H_3$ 0.62, –0.12
Gly20	9.05	4.40		
		3.61		
Lys21	8.53	4.20	1.83, 1.44	$C^{\gamma}H_2$, 1.05, 1.02; $C^{\delta}H_2$ 1.75; $C^{\gamma}H_2$ 3.04, 2.95; $N^{\zeta}H$ 7.24
Asn22	9.14	4.89	2.94, 2.42	
Glu23	9.29	4.27	2.16, 2.12*	$C^{\gamma}H_2$ 2.91, 2.50
His24	8.30	3.93	3.44, 3.54	(2) 8.72, (4) 7.15
Cys25	8.09			
Asn26	7.89	4.00	2.88	
Thr27	8.35	3.73	4.26	$C^{\gamma}H_3$ 1.24
Glu28	7.72	3.98	2.23, 2.07	
Cys29	7.83	4.20	4.04, 2.85	
Lys30	7.46	4.17	1.88, 1.45	$C^{\gamma}H_2$ 1.21; $C^{\epsilon}H_2$ 2.99
Ala31	7.04	4.17	1.52	
Lys32	8.70	3.92	1.86, 1.80	$C^{\gamma}H_2$ 1.49; $C^{\delta}H_2$ 1.68; $C^{\epsilon}H_2$ 2.98; $N^{\zeta}H$ 7.56
Asn33	8.28	4.43	2.88, 2.73*	
Gln34	7.66	4.75	2.09, 1.97*	$C^{\gamma}H_2$ 2.49, 2.28
Gly35	8.32	3.86		
		3.72		
Gly36	8.20	4.28		
		3.33		
Ser37	8.58	4.64	3.86, 3.78*	
Tyr38	7.25	4.94	3.38, 2.55	(2, 6) 7.12, (3, 5) 6.69
Gly39				
Tyr40		4.64	2.99, 2.82*	(2, 6) 6.70, (3, 5) 6.59
Cys41	9.11	5.04	3.88,* 2.78	
Tyr42	9.19	4.90	2.85, 2.52*	
Ala43	8.98	3.45	1.02	
Phe44	5.83	3.73	3.54, 3.16	(2, 6) 7.17; (3, 5) 7.26, (4) 7.03
Ala45	7.91	5.23	1.25	
Cys46	8.65	5.16	2.85, 2.58	
Trp47	9.43	4.59	2.75, 2.34	(NH) 9.81, (2) 6.45, (4) 5.76, (5) 6.63 (6) 7.05, (7) 7.12
Cys48	8.34	5.43	2.64, 2.44	
Glu49	8.80	4.67	2.24, 2.09	$C^{\gamma}H_2$ 2.64
Gly50	8.38	4.05		
		3.84		
Leu51	8.16	4.09	1.30, 1.15	$C^{\gamma}H$ 1.25; $C^{\delta}H_3$ 0.68, 0.28
Pro52		4.48	2.20, 1.97	
Glu53	8.61	3.75	2.04, 1.90	$C^{\gamma}H_2$ 2.45
Ser54	7.60	4.12	4.03, 3.76	
Thr55	7.50	4.52	4.01	$C^{\gamma}H_3$ 1.48
Pro56		4.43	2.22, 1.71	$C^{\gamma}H_2$ 2.20, 1.93; $C^{\delta}H_2$ 4.09, 3.89
Thr57	7.60	4.71	4.33	$C^{\gamma}H_3$ 0.52
Tyr58	8.21	4.12	2.91, 2.75	(2, 6) 7.09, (3, 5) 7.09
Pro59		3.67	1.66, 1.07	$C^{\gamma}H_2$ 2.15, 1.69; $C^{\delta}H_2$ 3.55, 3.37
Leu60	8.99	4.58	1.99, 1.79	$C^{\gamma}H$ 1.68; $C^{\delta}H_3$ 0.93, 0.89
Pro61		4.20	1.84	$C^{\gamma}H_2$ 2.02; $C^{\delta}H_2$ 3.93, 3.67
Asn62	8.19	4.43	3.82	
Lys63	7.05	4.49	1.54, 1.19	$C^{\gamma}H_2$ 0.78; $C^{\delta}H_2$ 1.54, 1.45; $C^{\epsilon}H_2$ 2.88; $N^{\zeta}H$ 7.38
Ser64	8.40	4.49	3.75	
Cys65	8.57	4.69	3.24,* 3.04	

[a] Chemical shifts were referenced from internal DSS. An asterisk (*) indicates that the chemical shift corresponds to the $C^{\beta 3}H$ and the other value to the $C^{\beta 2}H$ proton from stereospecific assignment results (IUPAC notation).

the $\alpha\beta(i, i + 3)$ contacts suggest an α-helical conformation for the segment of residues 23 to 31. This observation is further confirmed by the small NH–$C^{\alpha}H$ couplings for residues in this segment.

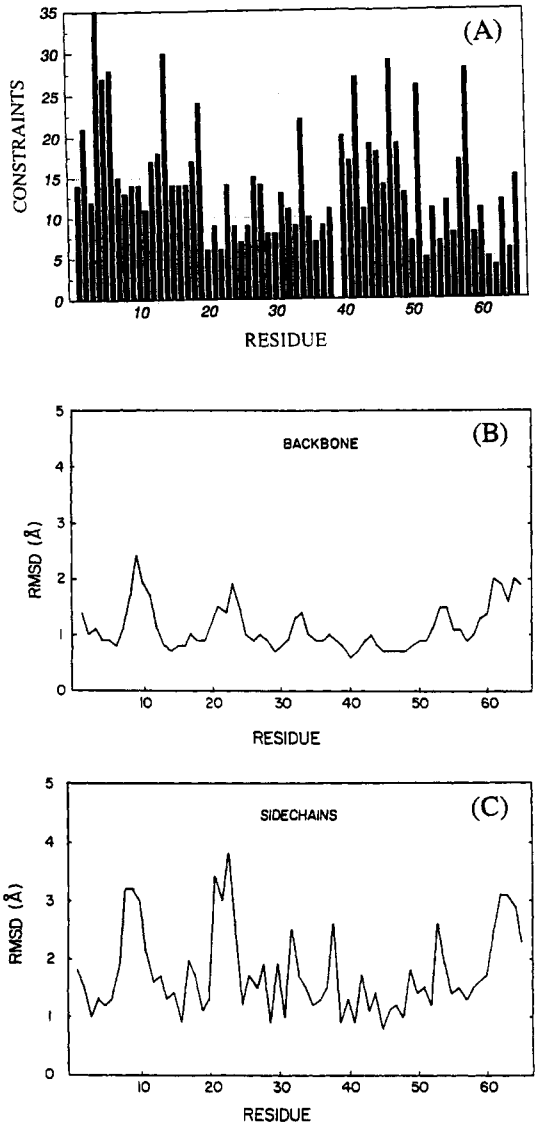

FIGURE 4: (A) Distribution of experimental constraints among the residues in CsE-v1, (B) distribution of the average atomic RMS deviations (w.r.t. $\langle\overline{SA}\rangle_k$) of the backbone atoms in the final simulated annealing structures ($\langle SA\rangle_k$) for the CsE-v1, and (C) distribution for the side-chain atoms.

An antiparallel β-sheet has also been identified in CsE-v1 on the basis of long-range interstrand NOESY contacts between residues in the middle strand 46–50 and the two terminal strands 1–5 and 36–41. Slowly exchanging amide from these strands together with large vicinal NH–C$^\alpha$H coupling constants ($\geq$8 Hz) further confirm the β-sheet structure. The strands 36–41 and 45–50 are connected by a β-turn composed of residues 42–45. A kink in the β-strand 36–41 is indicated due to the d_{NN} contact between Ser37 and Tyr38. The α-helix and β-sheet are joined by a turn involving residues 32–36. An additional turn involves residues 8–12.

A number of NOESY contacts between the different aromatic rings (residues 4, 38, 40, 42, 47, and 58) have been observed as in the case of CsE-v3 (Krishna et al., 1989). This observation suggests that these aromatic rings are clustered together on one side of the protein.

(iii) Computation of Chemical Shift Index (CSI) Values. The CSI values (Wishart et al., 1991, 1992) were deduced as digitized indicators from the differences between the observed and "coil" chemical shift values for the C$^\alpha$H protons. This procedure has been successfully applied by our laboratory recently in the study of a predominantly α-helical polypeptide (Sakai et al., 1993) and a left-handed polyproline II (PPII) helical peptide (Curto et al., 1993) characterized by positive CSI values.

Table 2: Structural Statistics for the Final Simulated Annealing Structures of CxE-v1 Neurotoxin

	$\langle SA\rangle_k$	$\langle\overline{SA}\rangle_{kr}$
(A) RMS deviations from experimental distance restraints (Å)		
all (527)	0.0262	0.0178
sequential ($\|i-j\|=1$) (130)	0.0332	0.0255
short range ($1<\|i-j\|\leq 5$) (29)	0.0195	0.0037
long range ($\|i-j\|>5$) (121)	0.0147	0.0035
intraresidue (212)	0.0258	0.0189
H-bond (35)	0.0201	0.0109
(B) RMS deviations from experimental dihedral restraints (deg) 53	0.290	0.162
(C) energies		
E_{NOE} (all) (kcal mol^{-1})	19.31	8.33
E_{tor} (kcal mol^{-1})	0.410	0.087
E_{repel} (kcal mol^{-1})	11.86	7.32
$E_{L\text{-}J}$ (kcal mol^{-1})[a]	−208.3	−212.3
(D) deviations from idealized covalent geometry		
bonds (Å)	0.00439	0.00377
angles (deg)	0.5745	0.5190
impropers (deg)	0.2701	0.1973

[a] $E_{L\text{-}J}$ is the Lennard-Jones/van der Waals potential calculated using CHARMm empirical energy function.

(iv) Modeling Calculations. A total of 570 distance constraints and 50 torsion angle constraints (31 for φ and 19 for χ) were used for structural calculations. Figure 4A shows the distribution of experimental constraints among the 65 residues of CsE-v1. No constraints were used for Gly39 since the signals from this residue could not be located under any of the experimental conditions.

A total of 60 DG substructures were generated, and each was subjected to one cycle of the dynamical simulated annealing protocol described earlier in the Materials and Methods section [also see Lee et al. (1993)]. Of these, 23 converged to a family of structures with no constraint violations (i.e., distance violations less than 0.5 Å and torsion angle violations less than 5°), and five structures violated one constraint. These five structures were subjected to a second cycle of the refinement protocol. This generated three additional structures with no violations, yielding a total of 26 final structures with no constraint violations. By using the notation of Driscoll et al. (1989), the family of structures sampled at the end of stage 3 are designated as $\langle SA\rangle_k$ and the average of these structures as $\langle\overline{SA}\rangle_k$. Since this structure is based on averaging the coordinates from 26 structures, it can have bond length and bond angle distortions. These were removed by an additional step of restrained energy minimization, and the resulting structure is designated as $\langle\overline{SA}\rangle_{kr}$. The family of structures at the end of stage 4 (i.e. after energy minimization without NOE and torsion angle constraints) are designated as $\langle SA\rangle_d$, indicating that the disulfide bridge constraints are still retained. The average of this family is indicated as $\langle\overline{SA}\rangle_d$, and the energy minimized average structure as $\langle\overline{SA}\rangle_{dr}$. The energies and structural statistics (RMS deviations with respect to experimental constraints) for $\langle SA\rangle_k$ and $\langle\overline{SA}\rangle_{kr}$ are shown in Table 2. It is noted that there is good convergence of these structures, resulting in acceptable energies for the final structures. The experimental constraints are satisfied within acceptable limits (<0.5 Å for distances and <5° for angles). The deviations from idealized

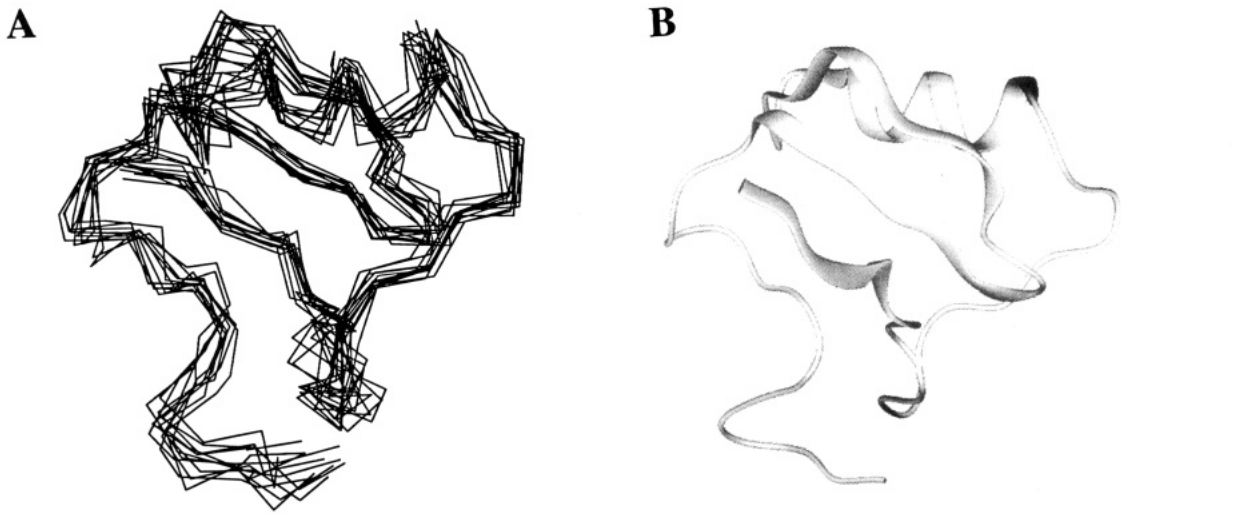

FIGURE 5: (A) Best-fit superposition of some typical backbone conformations from the final simulated anneling structures ($\langle SA \rangle_k$) and (B) cartcon representation of the energy-minimized average structure ($\langle \overline{SA} \rangle_{kr}$) of the CsE-v1 toxin.

Table 3: Atomic RMS Deviations for the Final Simulated Annealing Structures of CsE-v1 (All Residues)

	backbone atoms (Å)	all atoms (Å)
$\langle SA \rangle_k$ vs $\langle \overline{SA} \rangle_k$	1.28	1.82
$\langle SA \rangle_{kr}$ vs $\langle \overline{SA} \rangle_k$	0.46	0.84
$\langle SA \rangle_k$ vs $\langle \overline{SA} \rangle_{kr}$	1.36	2.00
$\langle SA \rangle_{kr}$ vs $\langle \overline{SA} \rangle_{dr}$	0.31	0.48

covalent geometry are very small (average RMSD for bonds is 0.00439 Å, for angles 0.5745°, and 0.2701° for improper torsion angles), suggesting that the structures have very good stereochemistry as well. Table 3 gives a comparison of the atomic RMS deviations for the different types of final simulated annealing structures. In Figure 4, parts B and C, the variations in average RMS deviations among 26 structures in the $\langle SA \rangle_k$ with respect to the family average $\langle \overline{SA} \rangle_k$ are plotted for each residue in CsE-v1. Some representative simulated annealing structures together with a protein cartoon of the average energy-minimized structure ($\langle \overline{SA} \rangle_{kr}$) are shown in Figure 5. Some portions of the sequence are more well defined (e.g., α-helix and β-sheet) than other regions, and the degree of convergence at any specific segment of the protein sequence correlates with the number of constraints (Figure 4A) available in that segment. We also note that the RMSD values are generally smaller in the vicinity of the cysteine residues (12, 16, 25, 29, 41, 46, and 48), suggesting the stabilizing influence of the disulfide bridges. A comparison of the average structures $\langle \overline{SA} \rangle_{kr}$ and $\langle \overline{SA} \rangle_{dr}$ was also made by best-fit superposition (not shown), and these two were found to be practically identical, with an RMS deviation of 0.35 Å for the backbone atoms. A stereoview of the backbone structure is shown in Figures 6, parts A and B.

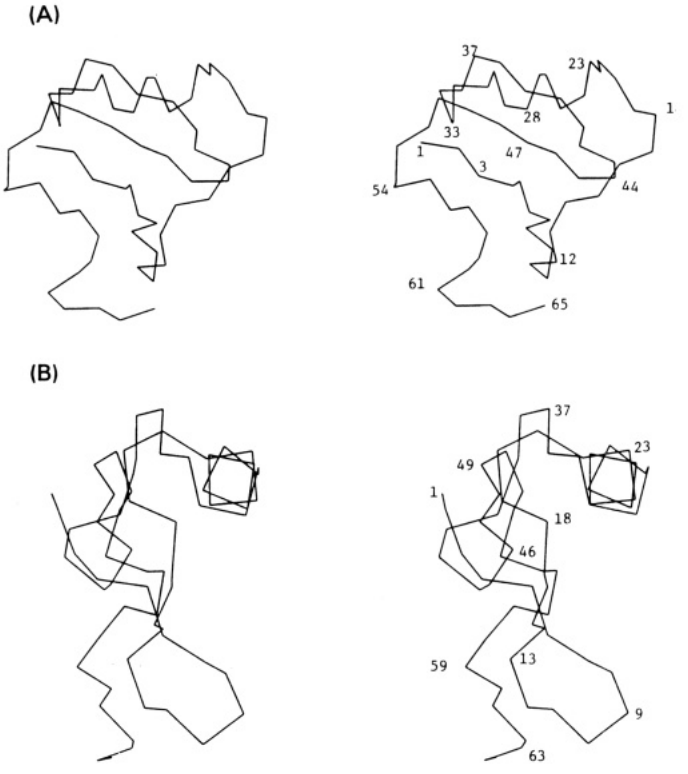

FIGURE 6: Stereoviews of the α-carbon backbone $\langle \overline{SA} \rangle_{kr}$ structure of CsE-v1: (A) front view and (B) side view. The locations of some representative residues are indicated.

FINAL REMARKS

In this work, we have determined the detailed sequence-specific resonance assignments for the CsE-v1 protein from the venom of the New World scorpion *C. sculpturatus Ewing*. This protein has nearly 90% sequence homology with the protein CsE-v3, for which sequence specific assignments were reported by us earlier (Nettesheim et al., 1989; Krishna et al., 1989), and its solution structure was determined (Krishna et al., 1990; Lee et al., 1993). The main secondary structural features of CsE-v1 are an α-helix formed by residues 23–31 and an antiparallel three-stranded β-sheet structure formed by the residues 1–5, 46–50, and 36–41, with 46–50 being the middle strand. A β-turn connects the last two strands. The remaining secondary structural elements consists of loops and turns. The d_{NN} contacts expected from the α-helical region are relatively week, possibly because of the conformational flexibility of this region under the experimental conditions (pH 3.0, 40 °C) employed in the present study. The effect of charged group residues (acidic residues near the N-terminus and basic residues near the C-terminus of an α-helix) on the stability of an α-helix conformation has been well studied (Shoemaker et al., 1985; Presta & Rose, 1988). The helix dipole is expected to be much more stable when the N-terminal acidic residues are deprotonated. Under the conditions used in the present study, the protonation of Glu23 could contribute

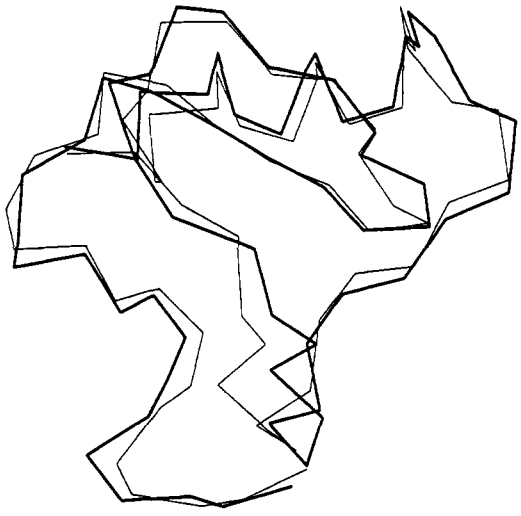

FIGURE 7: Comparison of the α-carbon backbone conformations for the CsE-v1 NMR structure (thick line) and the CsE-v3 crystal structure (thin line).

to the destabilization of the α-helix. This destabilization could be further accentuated by the basic residue, His24, which is positively charged. 2D-NMR measurements (TOCSY) at higher pH values (pH 5) where the α-helix is likely to be more stable, were attempted but were found to be very low in information content presumably due to line broadening of several of the amide hydrogens (e.g., from the various loop structures). The picture that emerges is that of a protein that has a well-defined folded structure stabilized by four disulfide bridges and by two secondary structural elements (an α-helix and a β-sheet), but nevertheless exhibits conformational flexibility.

The overall folding pattern as well as the specific secondary structural features of the CsE-v1 toxin (the α-helix, the three-stranded antiparallel β-sheet, *cis*-Pro59, and the turns) is similar to those found in the CsE-v3 toxin by NMR spectroscopy (Nettesheim et al., 1989; Krishna et al., 1990; Lee et al., 1993) and by crystallography (Fontecilla-Camps et al., 1980; Zhao et al., 1992). The backbone solution structure ($\langle\overline{SA}\rangle_{kr}$) of CsE-v1 showed an average RMS difference of 1.43 Å with respect to crystal structure and 1.20 Å with respect to the solution was structure of CsE-v3. Figure 7 shows a superposition of the v1 solution structure and the v3 crystal structure. Some differences between the two structures are apparent in the segment of residues 6–12. This segment is relatively poorly defined in both the crystal (high-temperature factors) and the solution structures (high average RMS deviations). We have also previously shown that minor differences existed in the arrangement of the side chains of aromatic residues between the solution and crystal structures of the CsE-v3 toxin (Krishna et al., 1989). In particular, Tyr42 and Trp47 move slightly closer in solution, whereas in the crystal structure, they rearrange slightly to accommodate a molecule of methylpentanediol (MPD). The CsE-v1 toxin also shows similar minor differences with respect to the crystal structure of CsE-v3. A second difference involves regions of apparent flexibility. The crystal structure shows largest temperature factors for segments centered around residue 33 followed by a segment centered around residue 8 (Zhao et al., 1992). The larger variation around residue 33 was ascribed to the presence of two enantiomers of a bound MPD molecule in the crystal structure. In the solution structure, the large RMSD values are located in segments of protein centered around residues 9, 23, and 63. The region centered around residues 32 is relatively well defined compared to these.

The sequences of CsE-v1 and the homologous CsE-v3 differ only in six locations, and these are located in the region of residues 15–26 (the so-called J-loop). Residues L-17 and K-18 of v3 are replaced by F-17 and W-18 in CsE-v1. An examination of the space-filling structures (not shown) of v1 and v3 shows that these two residues in v1 extend the hydrophobic patch formed by the aromatic residues and other hydrophobic residues. Nine out of 11 aromatic residues, viz., Y-4, F-17, W-18, Y-38, Y-40, Y-42, F-44, W-47, and Y-58 are part of this hydrophobic patch (the remaining two aromatic residues, Y-14 and H-24, are located on the backside of the protein, rich in hydrophilic residues). This extended hydrophobic surface in CsE-v1 may play a role in the 20-fold increase of its toxicity toward insects over the v3 toxin. Of the scorpion toxins sequenced so far, the substitution of K-18 with a tryptophan is unique. Lysine at position 18 is quite common in the New World scorpion toxins whereas, in the Old World toxins there is no lysine at this position, and likewise no leucine at position 17. The J-loop is extensively substituted in these two positions in the Old World toxins.

A qualitative comparison of the amide hydrogen exchange life times (from the intensities of the cross peaks in exchange spectra recorded under similar conditions) suggests that even though both the proteins have *similar* three-dimensional conformations in solution, they exhibit subtle differences in the stability of their three-dimensional structures in solution state. For example, the amide hydrogens of residues 12, 18, 44, 62, and 63 in CsE-v3 have longer exchange life times than the corresponding residues in CsE-v1, while the reverse is true for residues 17 and 57. This observation is particularly interesting since the amino acid differences are located primarily in the segment 15–26, but the stability of the hydrogen bonds at locations far away from this segment are also affected. The cysteine disulfide bridges (12–65, 16–41, 25–46, and 29–48) may be responsible, at least in part, in transmitting minor structural perturbations to parts of the protein far from the site of amino acid substitutions. Similarly, a comparison of the chemical shifts of the NH and $C^{\alpha}H$ protons in CsE-v1 and v3 shows, in addition to the expected differences near the substituted residues, chemical shift perturbations at distal residues 44, 47, 61, 62, and 64. Some of these may have, as their origin, small perturbations in the ring-current fields from aromatic residues, while the others arise presumably from structural perturbations transmitted by the disulfide bridges, as suggested earlier. It is currently difficult to distinguish between these two possibilities because of the limited precision in the current solution structures as well as the inherent flexibility of the v1 and v3 proteins.

The solution structure determined in this study for the CsE-v1 toxin is also consistent with the chemical modification studies reported in literature for the toxins Css II from *Centruroides suffusus suffusus*, Lqq V from *Leiurus quinquestriatus quinquestraitus*, and AaH II and III from *Androctonus australis Hector* (Darbon, 1990; El Ayeb et al., 1986; Sampieri & Haberstezer-Rochat, 1978; Darkon & Angelides, 1984). In general, amino acids in these proteins, which result in a loss of physiological activity on chemical modifications, have a spatial distribution similr to the one expected from the CsE-v1 toxin and correspond to residues located on the side of the protein that is rich in hydrophobic residues. For example, modification of R2 in AaH III or K2 in Lqq V results in a significant decrease in pharmacological activity. These residues are in register with K1 in CsE-v1 which is located on the front side of the protein. On the other hand, K13 in CsE-v1 is adjacent to Y58 which itself is in

register with K58 in Lqq V. Modification of this latter residue results in a decrease of pharmacological activity.

These studies on the solution structures of scorpion toxins serve as a prelude for our future investigations on the interaction of these protein probes with voltage-sensitive sodium channels. The sequences in the extracellular regions on the sodium channel α-subunit that comprise the receptor for α-scorpion toxins have been determined by antibody mapping (Thomsen and Catterall, 1989). Future investigations will be directed at a study of complexes of neurotoxins with protein fragments representing the receptor domains in the sodium channel.

REFERENCES

Almassy, R. J., Fontecilla-Camps, J. C., Suddath, F. L., & Bugg, C. E. (1983) *J. Mol. Biol. 170*, 497–527.

Arseniev, A. S., Kondakov, V. I., Maiorov, V. N., & Bystrov, V. F. (1984) *FEBS Lett. 165*, 57–62.

Babin, D. R., Watt, D. D., Goos, S. M., & Mlejnek, R. V. (1974) *Arch. Biochem. Biophys., 164*, 694–706.

Basus, V. J. (1989) *Methods Enzymol. 177*, 132–149.

Bax, A. (1989) *Methods Enzymol. 176*, 151.

Bontems, F., Roumestand, C., Boyot, P., Gilquin, B., Doljansky, Y., Menez, A., & Toma, F. (1991a) *Eur. J. Biochem. 196*, 19–28.

Bontems, F., Roumestand, C., Gilquin, B., Menez, A., & Toma, F. (1991b) *Science 254*, 1521–1523.

Clore, G. M., Gronenborn, A. M., Nilges, M., and Ryan, C. A. (1987) *Biochemistry 26*, 8012–8023.

Curto, E. V., Jarpe, M. J., Blalock, J. E., Borovsky, D., & Krishna, N. R. (1993) *Biochem. Biophys. Res. Commun. 193*, 688–693.

Darbon, H. (1990) in *Protein Structure-Function. Proc. Int. Symp.*, (Zaidi, Z. H., Abbasi, A., & Smith, D. L., Eds.) pp 169–181, TWEL Publishers, Karachi, Pakistan.

Darbon, H., & Angelides, K. J. (1984) *J. Biol. Chem. 259*, 6074–6084.

Darbon, H., Weber, C., & Braun, W. (1991) *Biochemistry 30*, 1836–1845.

Driscoll, P. C., Gronenborn, A. M., Beress, L., & Clore, G. M. (1989) *Biochemistry 28*, 2188–2198.

El Ayeb, M., Darbon, H., Bahraouni, H., Vargas, O., & Rochat, H. (1986) *Eur. J. Biochem. 155*, 289–294.

Fontecilla-Camps, J. C. (1989) *J. Mol. Evol. 29*, 63–69.

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*, 6496–6500.

Fontecilla-Camps, J. C., Almassy, R. j., Ealick, S. E., Suddath, F. L., Watt, D. D., Feldman, R., & Bugg, C. E. (1981) *Trends. Biochem. Sci. 6*, 291–296.

Fontecilla-Camps, J. C., Habersetzer-Rochat, C., & Rochat, H. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7443–7447.

Jover, E., Couraud, F., & Rochat, H. (1980) *Biochem. Biophys. Res. Commun. 95*, 1607–1614.

Krishna, N. R., Nettesheim, D. G., Klevit, R. E., Drobny, G., Watt, D. D., & Bugg, C. E. (1989) *Biochemistry 28*, 1556–1562.

Krishna, N. R., Moore, C. H., Narasimhan, S., & Watt, D. D. (1990) in *Peptides: Chemistry, Structure and Biology* (Rivier, J. E., & Marshall, G. R., Eds.) pp 625–627, ESCOM, Leiden.

Kuntz, I. D., Thomason, J. F., & Oshiro, C. M. (1989) *Methods Enzymol. 177*, 159–203.

Lambert, P., Kuroda, H., Chino, N., Watanabe, T. X., Kimura, T., & Sakakibara, S. (1990) *Biochem. Biophys. Res. Commun. 170*, 684–690.

Lee, W., Moore, C. H., Watt, D. D., & Krishna, N. R. (1993) *Eur. J. Biochem.*, In Press.

Macura, S., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95.

Marion, D., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Martins, J. C., Zhang, W., Tartar, A., Lazdunski, M., & Borreman, F. A. M. (1990) *FEBS Lett. 260*, 249–253.

McIntosh, M. E., & Watt, D. D. (1967) *Animal Toxins* pp 47–58, Pergamon Press, New York.

Meves, H., Rubly, N., & Watt, D. D. (1982) *Pfluegers Arch. Eur. J. Physiol. 393*, 56–62.

Meves, H., Simard, H., & Watt, D. D. (1986) *Ann. N.Y. Acad. Sci. 479*, 113–132.

Mikou, A., LaPlante, S. R., Guittet, E., Lallemand, J., Claire, M., & Rochat, H. (1992) *J. Biomol. NMR 2*, 57–70.

Nettesheim, D. G., Klevit, R. E., Drobny, G., Watt, D. D., & Krishna, N. R. (1989) *Biochemistry 28*, 1548–1555.

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988a) *FEBS Lett. 229*, 317–324.

Nilges, M., Gronenborn, A. M., Brunger, A. T., & Clore, G. M. (1988b) *Protein Eng. 2*, 27–38.

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988c) *FEBS Lett. 239*, 129–136.

Otting, G., Widmer, H., Wagner, G., & Wuthrich, K. (1986) *J. Magn. Res. 66*, 187–193.

Pashkov, V. S., Maiorov, V. N., Bystrov, V. F., Hoang, A. N., Volkova, T. M., & Grishin, E. V. (1988) *Biophys. Chem. 31*, 121–131.

Presta, L. G., & Rose, G. D. (1988) *Science 240*, 1632–1641.

Rance, M., Sorenson, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.

Sakai, T. T., Jablonsky, M. J., DeMuth, P. A., Krishna, N. R., Jarpe, M. A. & Johnson, H. M. (1993) *Biochemistry 32*, 5650–5655.

Sampieri, F., & Habersetzer-Rochat, C. (1978) *Biochim. Biophys. Acta 535*, 100–109.

Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2349–2353.

Simard, J. M., Meaves, H., & Watt, D. D. (1986) *Pflugers Arch. 406*, 620.

Thomsen, W. J., & Catterall, W. A. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 10161–10165.

Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., & Wuthrich, K. (1987) *J. Mol. Biol. 196*, 611–639.

Watt, D. D., & Simard, J. M. (1984) *J. Toxicol., Toxin Rev. 3*, 181–221.

Watt, D. D., Simard, J. M., Babin, D R., Mlejnek, R. V. (1978) in *Toxins: Animal, Plant and Bacterial. Proceedings of the 5th International Symposium* (Rosenberg, P., Ed.) pp 647–660, Pergamon Press, New York.

Wheeler, K. P., Watt, D. D., & Lazdunski, M. (1983) *Pflugers Arch. 397*, 164–165.

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311–333.

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992) *Biochemistry 31*, 1647–1651.

Wuthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Wuthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949–961.

Zhao, B., Carson, M., Ealick, S. E., & Bugg, C. E. (1992) *J. Mol. Biol. 227*, 239–252.

Zlotkin, E., Miranda, F., & Lissitsky, S. (1972a) *Toxicon 10*, 207–209.

Zlotkin, E., Miranda, F., & Lissitsky, S. (1972b) *Toxicon 10*, 211–213.