# α-Cobratoxin: Proton NMR Assignments and Solution Structure

Remi Le Goas,[‡] Steven R. LaPlante,[§] Afaf Mikou,[‡] Marc-André Delsuc,[‡] Eric Guittet,[*,‡] Michel Robin,[‡] Isabelle Charpentier,[∥] and Jean-Yves Lallemand[‡]

*Institut de Chimie des Substances Naturelles, Centre National de la Recherche Scientifique, 91190 Gif-sur-Yvette, France, and Service de Biochimie, CEN, 91191 Saclay, Gif-sur-Yvette, France*

*Received December 2, 1991; Revised Manuscript Received March 10, 1992*

ABSTRACT: The solution structure of α-cobratoxin, a neurotoxin purified from the venom of the snake *Naja naja siamensis*, at pH 3.2 is reported. Sequence-specific assignments of the NMR resonances was attained by a combination of a generalized main-chain-directed strategy and of the sequential method. The NMR data show the presence of a triple-stranded β-sheet (residues 19–25, 36–41, and 52–57), a short helix, and turns. An extensive number of NOE cross peaks were identified in the NOESY NMR maps. These were applied as distance constraints in a molecular modeling protocol which includes distance geometry and dynamical simulated annealing calculations. A single family of structures is observed which fold in such a way that three major loops emerge from a globular head. The solution and crystal structures of α-cobratoxin are very similar. This is in clear contrast to results reported for α-bungarotoxin where significant differences exist.

α-Cobratoxin is one of a number of lethal neurotoxic proteins present in the *Naja naja siamensis* snake venom. Upon injection, this postsynaptic neurotoxin binds to the nicotinic acetylcholine receptors at the neuromuscular junction and thereby prevents binding of acetylcholine and coupled depolarization. The muscle then fails to respond, and the victim usually dies of respiratory arrest (Menez, 1987).

The postsynaptic neurotoxins are classified into two distinct families: the short neurotoxins having 60–62 residues and four disulfide bridges and the long neurotoxins having 71–74 residues and five disulfide bridges. α-Cobratoxin (71 residues) belongs to the latter. Both families show significant sequence homologies, and conformation-prediction studies suggest common structural features (Dufton & Hider, 1977; Menez et al., 1978). Furthermore, crystal structures have been solved for erabutoxin-b (Kimball et al., 1979; Low & Corfield, 1986) of the short family and for α-bungarotoxin (Love & Stroud, 1986) and α-cobratoxin (Walkinshaw et al., 1980) of the long family. Indeed, both families exhibit a common structural theme of a globular head with three protruding loops. Differences do exist, however, at the end of the middle loop and at the tail end. These could explain differences in binding rates to the acetylcholine receptors (Chicheportiche et al., 1975) and also the immunological crossreactivity within the families but not between them (Boquet et al., 1973).

The crystal structure of α-cobratoxin was obtained at 2.8-Å resolution (Walkinshaw et al., 1980) and later refined (Benzel, Lange, Pal, Maelicke, Wilson, and Saënger, unpublished results). Crystallization, however, is limited to harsh solvent conditions such as 75% methylpentanediol at pH 2.8. Hider et al. (1982) later investigated pH and solvent effects on the solution structure of α-cobratoxin by a partial structure analysis; two-dimensional NMR[1] methods had not been developed at that time. The preliminary 1D NMR and CD study concluded that subtle differences existed between aqueous and 75% methylpentanediol solvents. In contrast, significant differences were detected between the aqueous solution structures at low and physiological pH. This observation is important since it may provide an atomic level explanation of the structural differences of α-cobratoxin between its inactive state within the venom gland (at low pH) and its active state in the victim (at physiological pH). Kondakov et al. (1984) studied the pH-dependent structural differences by 2D NMR. Most but not all of the resonances were assigned, and crude structural comparisons were proposed. We have therefore chosen to further explore these pH-dependent conformations by a complete structural analysis at pH 3.2 and 7.0 using 2D NMR.

In this paper, the complete assignment of the $^1$H NMR spectrum of α-cobratoxin at pH 3.2 using two-dimensional NMR techniques is reported. Secondary structures have been observed, and a tertiary solution structure model has been calculated.

## Materials and Methods

*NMR.* α-Cobratoxin was isolated and purified from the venom of *Naja naja siamensis* as described elsewhere (Karlsson et al., 1971). The unbuffered solution contained 11 mg of protein in 0.4 mL of either 90% $H_2O$/10% $D_2O$ or 99.996% $D_2O$ at pH 3.2.

$^1$H NMR spectra were recorded on Bruker AM600 and AM400 spectrometers and were processed on an Aspect 3000 computer. Some spectra were processed on a μVAX II computer using the GIFA software, developed in our laboratory (Delsuc, Robin, and Stoven, unpublished). NOESY (Jeener et al., 1979; Macura et al., 1981) and DQF-COSY spectra (Shaka et al., 1983; Rance et al., 1983) were recorded in $D_2O$ at 20 °C with 100-, 150-, and 300-ms mixing times for the NOESY experiments. Spectra taken in 90% $H_2O$/10% $D_2O$ include DQF-COSY and TQF-COSY at 20 °C; NOESY (100-, 150-, and 300-ms mixing times) and HOHAHA spectra

* To whom correspondence should be addressed.
‡ Institut de Chimie des Substances Naturelles, CNRS.
§ Current address: Bio-Mēga Inc., 2100 rue Cunard, Laval, Quēbec H7S 2G5, Canada.
∥ Service de Biochimie, CEN.

[1] Abbreviations: NMR, nuclear magnetic resonance; 1D, one dimensional; 2D, two dimensional; 3D, three dimensional; CD, circular dichroism; $^1$H, proton; NOE, nuclear Overhauser effect; NOESY, 2D NOE spectroscopy; DQF-COSY, double-quantum filtered correlated spectroscopy; TQF-COSY, triple-quantum filtered COSY; HOHAHA, homonuclear Hartmann–Hahn spectroscopy; MCD, main chain directed; AMX, spin system of three different protons; RMSD, root mean square deviation.

0006-2960/92/0431-4867$03.00/0 

(30-, 50-, and 85-ms spin-lock times) (Braunschweiler & Ernst, 1983; Davis & Bax; 1985) were taken at 5, 20, and 45 °C. Water suppression was achieved by presaturation during the relaxation delay; for the NOESY experiments, presaturation was also applied during the mixing time. Spectra were acquired with 2048 complex data points in $t_2$ and 512 points in $t_1$ with spectral widths of 8024 Hz at 600 MHz and 5000 Hz at 400 MHz. All 2D spectra were recorded in the pure-phase absorption mode using the time-proportional incrementation method (Redfield & Kunz, 1975; Marion & Wüthrich, 1983). HOHAHA and NOESY data were apodized using a shifted cosine function in both dimensions; a pure sine function was applied to the DQF-COSY and TQF-COSY data. All data were zero-filled in $t_1$ before Fourier transformation. DQF-COSY data were additionally zero-filled in $t_2$.

*Distance Restraints.* NOEs were interpreted in terms of distance ranges of 1.8–2.6, 1.8–3.4, and 1.8–4.75 Å corresponding to strong, medium, and weak NOEs, respectively, derived from 150-ms NOESY spectra. Corrections for pseudoatom representation (Wüthrich et al., 1983) were implemented for nonstereospecifically assigned protons.

Backbone hydrogen bonds within the triple-stranded $\beta$-sheet were identified by amide proton exchange data, interstrand NOEs, and $^3J_{HN\alpha}$ coupling constants. Two constraints were used for each NH($i$)–CO($j$) backbone hydrogen bond with $r_{N-O}$ restricted to 2.4–3.3 Å and $r_{NH-O}$ to 1.7–2.3 Å.

Other distance constraints were obtained from the five known disulfide bridges; the $S^\gamma$–$S^\gamma$ distance was fixed to a range of 1.9–2.1 Å.

*Angle Restraints.* $^3J_{HN\alpha}$ coupling constants were measured in the fingerprint region of DQF-COSY spectra by measuring the peak-to-peak separation of the antiphase fine structure components in contour plots. For $^3J_{HN\alpha}$ smaller than 5.5 Hz, $\phi$ was restricted to the range (−90°, −40°). For $^3J_{HN\alpha}$ equal to 8 Hz, $\phi$ was restricted to the range (−170°, −70°); and for large couplings of more than 10 Hz, this range was narrowed to (−140°, −100°). Stereospecific assignment of the four Val $CH_3^{\gamma,\gamma'}$ groups and several $CH^{\beta,\beta'}$ groups was accomplished according to established procedures (Zuiderweg et al., 1985; Wagner et al., 1987). This allows the corresponding $\chi^1$ angle to be constrained to 60°, −60°, or 180°.

*Structure Computations.* Three-dimensional structures were calculated on the basis of the NMR data by using a combination of distance geometry (Crippen, 1977; Havel et al., 1983; Kuntz et al., 1979) and dynamical simulated annealing methods (Brünger et al., 1986). The metric matrix distance geometry part of the calculations was carried out with the VEMBED program, a vectorized version of the EMBED program (written by Crippen, Kuntz, and Havel). The dynamical simulated annealing calculations were performed with the X-PLOR program (Brünger, 1990), using part of the NMR refinement protocol (Nilges, 1990).

Fifty initial conformations were computed with the metric matrix distance geometry program VEMBED, with only 200 steps of optimization after the embedding. This procedure took 1 h per structure on an Alliant VFX40. The structures had an average of 10 distance violations greater than 0.5 Å, but some of them showed a nonuniform folding of the backbone in the "core" region of the protein, where four of the five disulfide bridges are located. We therefore decided to use the dynamical simulated annealing protocol of the X-PLOR program which circumvents these folding problems.

The total target function $F_{tot}$ is made up of the following terms:

$$F_{tot} = F_{covalent} + F_{repel} + F_{NOE} + F_{tor}$$

$F_{covalent}$ maintains correct bond lengths, angles, chirality, and planes, while the nonbonded interactions $F_{repel}$ are represented by a simple van der Waals repulsion term, with a variable force constant $k_{repel}$. The NOE distance and the torsion angle restraints, $F_{NOE}$ and $F_{tor}$, are represented by square-well potentials with variable force constants $k_{NOE}$ and $k_{tor}$, respectively.

The annealing protocol consists in setting the initial weight on the repulsive van der Waals term to a very low value to allow atoms to pass through each other easily. The structural folding is then determined mainly by the experimental restraints rather than by high van der Waals interactions. The structures thus obtained all exhibited a uniform folding. The dynamical simulated annealing protocol, which takes about 40 min on a IBM 3090-VF, proceeds in five phases: (i) 500 cycles of Powell minimization in order to retrieve bad nonbonded contacts in the VEMBED structures. This is followed by (ii) 15-ps dynamics with a time step of 2 fs at 1000 K and (iii) 10-ps dynamics at 1000 K. During this last period, the force constant $k_{repel}$ is increased from 0.002 up to 0.1 by multiplying it by a factor of 1.5 every picosecond. Finally, (iv) the temperature is cooled down to 300 K by 25 K steps of 0.1 ps while all the force constants are kept to their final values, and (v) the protocol is concluded by 400 cycles of Powell minimization.

Structures which presented too many violations of the experimental constraints, with a NOE term and/or a torsion term greater than, respectively, 300 and 70 kcal/mol, were eliminated at the end of the protocol. The analysis of the 18 structures satisfying these conditions showed that some errors remained in the constraint list. A second run of the X-PLOR protocol described above was then applied and delivered the final 18 structures.

The resulting computed structures were inspected on a Silicon Graphics Personal Iris 4D25 using the program Insight II from Biosym Technologies.

## RESULTS

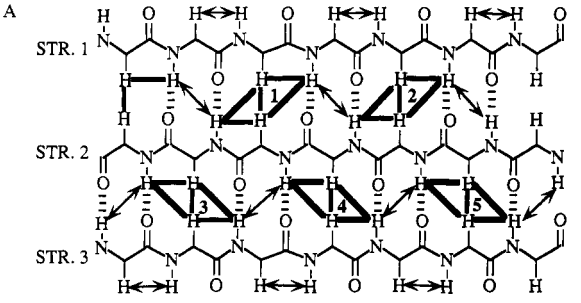
*Resonance Assignments and Secondary Structures.* Proton resonance assignments were made using a strategy which includes three major steps (see the Supplementary Material for more details concerning the assignments).

Step I consists in taking DQF-COSY,[2] NOESY,[2] and HOHAHA[2] spectra in $D_2O$ and $H_2O$.

Step II entails the identification and attribution of the secondary structures. The generalized MCD method (Saudek et al., 1989; LaPlante et al., 1990) begins by locating patterns (Englander & Wand, 1987; Di Stephano & Wand, 1987) which represent the characteristic NMR cross peak signature of antiparallel $\beta$-sheets and helices. One advantage of this approach is that pattern searches are directly focused on spin systems of residues involved in the secondary structures. This differs from the previous strategy (Englander & Wand, 1987) where all the NH–$\alpha$H–$\beta$H spin systems need to be identified before conducting searches for MCD patterns among all the main-chain spin systems.

*Antiparallel $\beta$-Sheet.* The generalized MCD approach for antiparallel $\beta$-sheets begins by locating the readily apparent $\alpha$H–$\alpha$H NOESY cross peaks. Each $\alpha$H–$\alpha$H cross peak and the associated motif defines an inner loop MCD pattern within

[2] DQF-COSY spectra provide connectivities between resonances of protons which are $J$-coupled (usually within two or three bonds). NOESY spectra provide connectivities between resonances of protons which are within 4.5-Å distance from each other. HOHAHA spectra can provide through-space connectivities between resonances of protons which are within the same residue (across covalent bonds).




B

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| STR.1 | Val 52 | Asp 53 | Ile 54 | Gln 55 | Cys 56 | Cys 57 | Ser 58 |
| STR.2 | Trp 25 | Thr 24 | Lys 23 | Thr 22 | Tyr 21 | Cys 20 | Val 19 |
| STR.3 | Arg 36 | Val 37 | Asp 38 | Leu 39 | Gly 40 | Cys 41 | Ala 42 |

FIGURE 1: Schematic representation of the triple-stranded antiparallel β-sheet of α-cobratoxin. The three strands (STR.) are indicated in panel A. The thick lines show the five inner loop MCD patterns. Other important NOESY connectivities are indicated by double-headed arrows. Hydrogen-bonding sites are indicated by slashes. Amino acid assignments of the sheet are given in panel B.

the antiparallel β-sheet shown in Figure 1. Extensive combinations of DQF-COSY and NOESY cross peaks were found in spectra taken in $H_2O$ and $D_2O$ which showed interconnecting full loop MCD patterns from the sheet (Englander & Wand, 1987; Di Stephano & Wand, 1987; LaPlante et al., 1990).

The second part of this strategy consists in assigning the residues within the β-sheet pattern shown in Figure 1. The task was much simpler than one would imagine. Rather than completely assigning all the individual spin systems of the residues involved in the β-sheet, only the readily determinable ones such as valine, isoleucine, threonine, alanine, and glycine were completely identified. The remaining residues were classified as AMX or long-chain systems (see the Supplementary Material for the detailed description of this classification).

In addition, a short sheet was found, where residues 2 and 3 lie antiparallel to residues 12 and 13 (not shown).

*Helix and Turns.* The generalized MCD approach for helices (Saudek et al., 1989; LaPlante et al., 1990) begins by locating the characteristic series of relatively strong, connected NH–NH NOESY cross peaks. This key feature shown in Figure 2A allows the targeting of the residue spin systems involved in the helix. For instance, the NH–NH "NOESY walk" in Figure 2A is a starting point for the analogous NOESY walk in the NH–βH region shown in Figure 2B. Taken together, both walks complete a series of helix MCD patterns which results from the short intraresidue and interresidue distances between NHs and βHs. Medium-range NOEs observed between the αH of Cys-30 and the NH of Gly-34, and between the αH of Cys-30 and the NH of Arg-33, further confirmed the presence of this small helix (from Phe-29 to Gly-34).

Turns also should exhibit only one or two relatively strong NH–NH cross peaks (Wagner et al., 1986). Thus, a preliminary analysis was conducted from the remaining unattributed strong NOESY cross peaks in Figure 2A, considered as resulting possibly from turns.

The third step (step III) involves applying the standard sequential method (Wüthrich, 1986). The assignment of secondary structures (step II) provided valuable starting points for this final step. The complete resonance assignment of α-cobratoxin is given in Table I. A summary of the short-range NOEs is given in Figure 3.

*Molecular Modeling.* There are two classes of modeling distance constraints available from NMR NOE cross peaks. The long-range constaints are crucial in modeling the global fold, since they force protons of amino acids separated by at least four sequential residues to be close to each other. The second class of constraints includes short- and medium-range contacts involving protons separated by less than five residues, which are important for relatively local conformations.

The structures displayed in Figure 4 are those which satisfy the 413 distance (including 170 long-range and 36 medium-range) and 76 dihedral constraints imposed during distance geometry, dynamical simulated annealing, and energy minimization calculations. Also included as constraints are the 13 hydrogen-bonding sites within the sheets (e.g., see slashes in Figure 1) and the five disulfide bridges. Among all the possible hydrogen-bonding sites which could be detected from

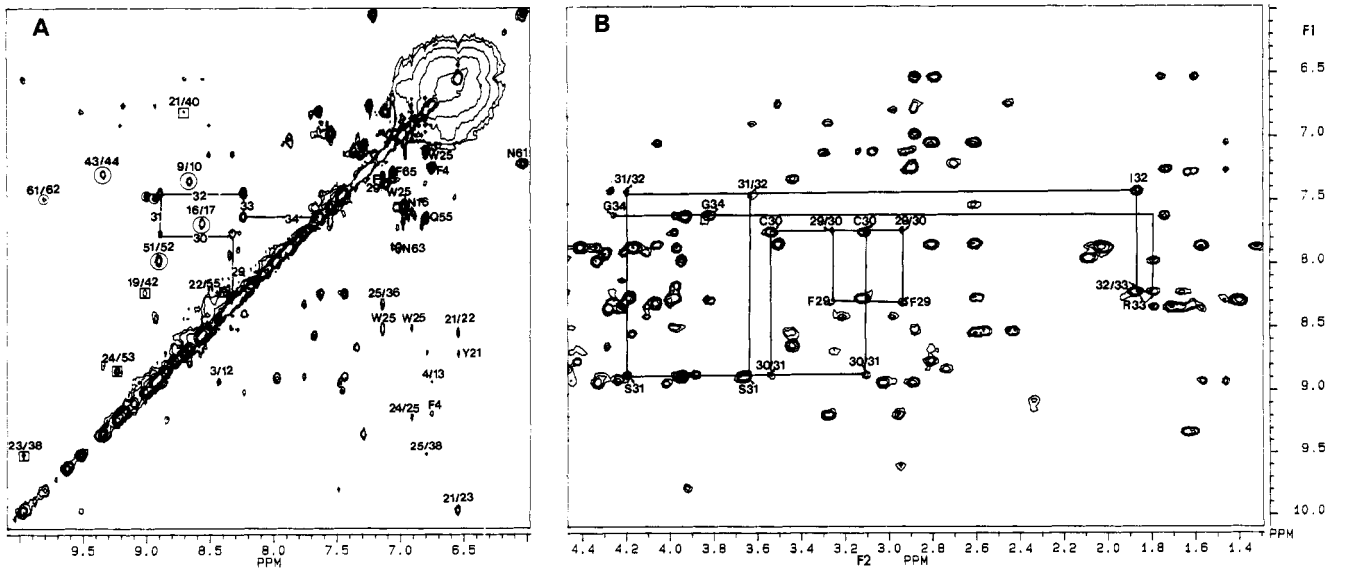



FIGURE 2: 600-MHz NOESY spectrum (150 ms) in $H_2O$ at 20 °C. The NH (aromatic) to NH (aromatic) region is shown in panel A and the NH (aromatic) to βH region in panel B. The $N^{\epsilon 1}H$ resonance of Trp at 10.43 ppm is not shown. The right side of the diagonal in panel A shows assignments of asparagine, glutamine, and aromatic protons (intrastrand side chains). NOEs involving the aromatic side chains are also shown; some long-range NOEs are indicated. On the left side of the diagonal, NOEs resulting from the helix are shown in the NOESY walk. The squares correspond to interstrand NOEs from the β-sheet. Finally, the circles show NOEs possibly resulting from turns. The sequential connectivities within the NH (aromatic) to βH region are shown in panel B.

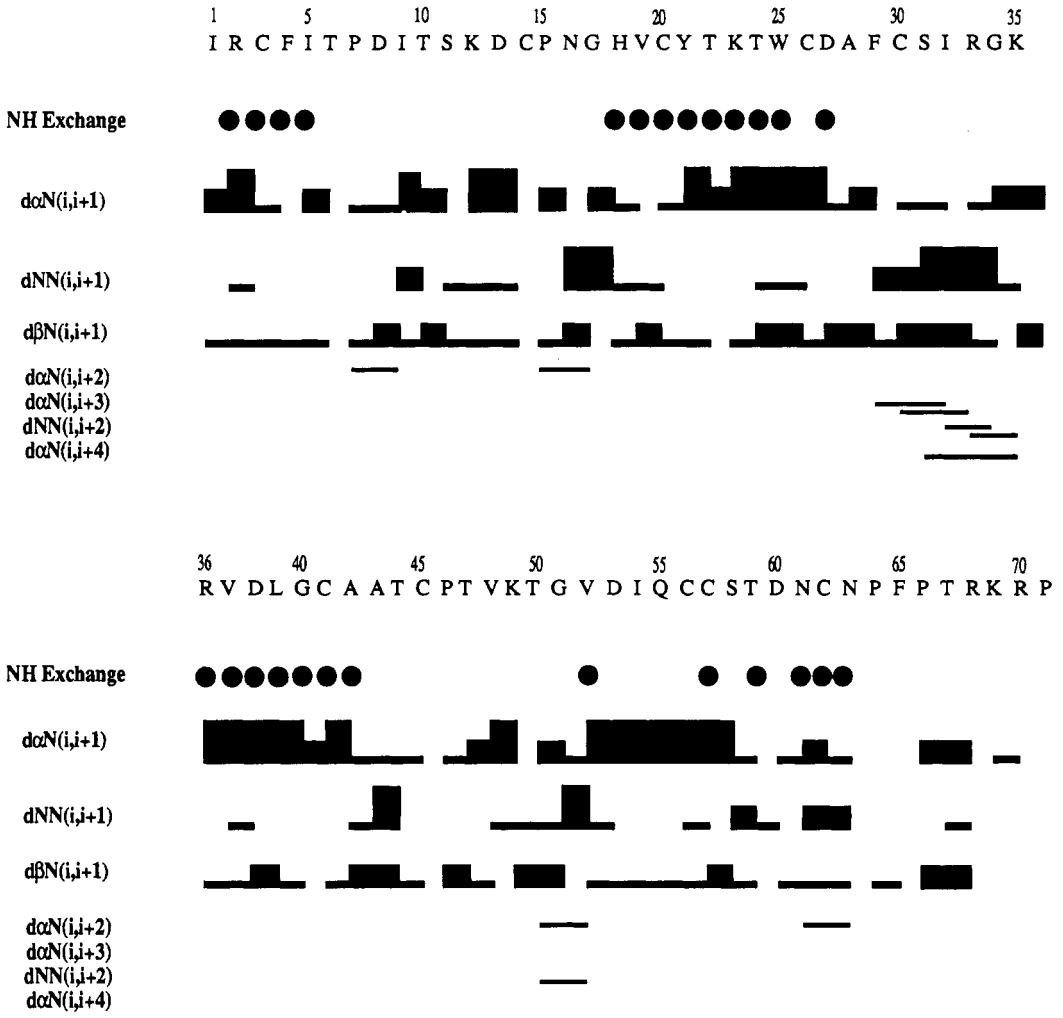



FIGURE 3: Primary sequence of $\alpha$-cobratoxin together with a summary of the short-range NOEs involving the NH, $\alpha$H, and $\beta$H protons, mostly from spectra taken at 20 °C. The NOEs are reported as strong, medium, or weak as shown by the thickness of the lines. Note that $d_{\alpha\delta}$, $d_{\beta\delta}$, and $d_{N\delta}$ connectivities for prolines are shown as $d_{\alpha N}$, $d_{\beta N}$, and $d_{NN}$, respectively. Amide protons that were still present in $D_2O$ solvent after two weeks are indicated by a filled circle.

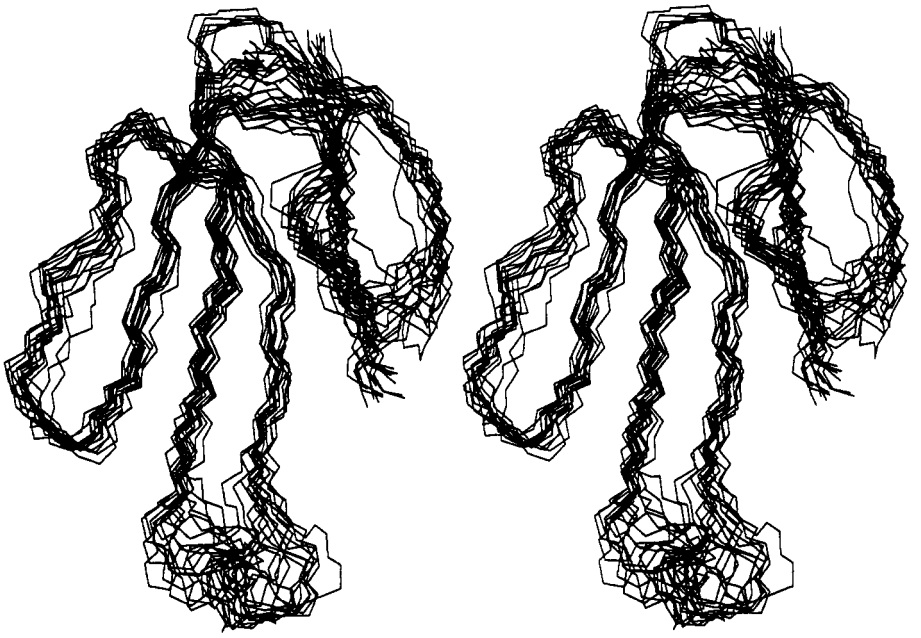


FIGURE 4: Stereopair of the best-fit superposition of the 18 solution structures of $\alpha$-cobratoxin as determined by NMR and dynamical simulated annealing calculations (backbone atoms only).

the early folding analysis, only slow exchangeable NH protons in the triple-stranded $\beta$-sheet were used as hydrogen-bond constraints during the protocol. The last five residues at the C terminal end are not shown in Figure 4. They were not taken into account in the calculations since only few short-range constraints were found in this region. This C-terminal

Table I: $^1$H NMR Chemical Shifts of α-Cobratoxin at 20 °C and pH 3.2[a]

| residue | chemical shifts NH | $C^{\alpha}H$ | $C^{\beta}H$ | $C^{\gamma}H$ | $C^{\delta}H$ | others |
|---|---|---|---|---|---|---|
| Ile-1 | 7.89 | 4.03 | | 1.20, 1.44 | 0.61 | |
| Arg-2 | 8.21 | 4.27 | 1.33,* 1.33* | 1.20,* 1.20* | 2.84, 2.84 | 7.05 ($N^{\epsilon}H$) |
| Cys-3 | 8.45 | 3.09 | 2.10, 3.04 | | | |
| Phe-4 | 9.22 | 4.72 | 2.52, 2.86 | | 6.80 | 7.27 ($C^{\epsilon}H$), 6.85 ($C^{\xi}H$)+ |
| Ile-5 | 8.11 | 4.76 | 1.90 | 1.40, 1.41 | 0.74 | 1.10 ($C^{\gamma}H_3$) |
| Thr-6 | 8.47 | 4.24 | | | | |
| Pro-7 | | 4.32 | 2.44, 2.25 | 2.04, 1.80 | 3.74, 3.60 | |
| Asp-8 | 8.78 | 4.44 | 2.79, 2.79 | | | |
| Ile-9 | 8.66 | 3.46 | 2.19 | 0.73, 1.40 | 0.45 | 0.86 ($C^{\gamma}H_3$) |
| Thr-10 | 7.40 | 4.70 | 4.14 | 1.20 | | |
| Ser-11 | 8.27 | 4.07+ | 3.05, 3.24 | | | |
| Lys-12 | 8.96 | 4.61 | 1.76, 1.76 | 1.65,* 1.65* | 1.40,* 1.40* | 2.84, 2.90 ($C^{\epsilon}H$), 7.52 ($N^{\xi}H$) |
| Asp-13 | 8.60 | 4.56 | 2.45, 2.63 | | | |
| Cys-14 | 8.97 | 5.12 | 2.86, 3.72 | | | |
| Pro-15 | | 4.21 | 2.20, 2.22 | 1.87, 1.97 | 3.57, 3.79 | |
| Asn-16 | 8.60 | 4.45+ | 2.67, 2.92 | | | 7.02, 7.60 ($N^{\delta}H$) |
| Gly-17 | 7.71 | 3.84, 3.07 | | | | |
| His-18 | 8.43 | 4.26 | 3.30, 3.45 | | | 8.64 ($C^{\epsilon 1}H$), 7.40 ($C^{\delta 2}H$) |
| Val-19 | 8.26 | 5.05 | 2.32 | 0.88, 0.65 | | |
| Cys-20 | 9.14 | 5.76 | 2.92, 3.30 | | | |
| Tyr-21 | 8.73 | 5.95 | 2.83, 2.91 | | 6.58 | 6.58 ($C^{\epsilon}H$) |
| Thr-22 | 8.57 | 5.02 | 4.01 | 1.03 | | |
| Lys-23 | 10.01 | 5.71 | 2.12, 1.82 | 1.24, 1.24 | 1.48, 1.41 | 2.40, 2.13 ($C^{\epsilon}H$), 7.54 ($N^{\xi}H$) |
| Thr-24 | 9.26 | 5.95 | 4.04 | 1.10 | | |
| Trp-25 | 8.57 | 5.17 | 3.64, 3.32 | | | 1.043 ($N^{\epsilon 1}H$), 7.43 ($C^{\xi 2}H$), 6.94 ($C^{\delta 1}H$), 7.19 ($C^{\eta 2}H$), 6.85 ($C^{\xi 3}H$), 7.20 ($C^{\zeta 3}H$) |
| Cys-26 | 9.24 | 5.06 | 3.35, 3.01 | | | |
| Asp-27 | 8.30 | 4.79 | 2.64, 3.12 | | | |
| Ala-28 | 8.34 | 4.03 | 1.04 | | | |
| Phe-29 | 8.39 | 4.40+ | 3.37, 2.97 | | 7.20 | 7.37 ($C^{\epsilon}H$) |
| Cys-30 | 7.76 | 4.03 | 3.57, 3.15 | | | |
| Ser-31 | 8.92 | 4.24 | 3.96, 3.99 | | | |
| Ile-32 | 7.48 | 4.13 | 1.92 | 1.28, 1.61 | 0.94 | 0.99 ($C^{\gamma}H_3$) |
| Arg-33 | 8.28 | 4.54 | 1.71, 1.93 | 1.71, 1.86 | 3.13, 3.13 | 7.19 ($N^{\epsilon}H$) |
| Gly-34 | 7.62 | 3.84, 4.30 | | | | |
| Lys-35 | 8.33 | 4.14 | 1.41, 1.41 | 1.16, 1.16 | 1.61,* 1.61* | 2.92, 2.92 ($C^{\epsilon}H$), 7.54 ($N^{\xi}H$) |
| Arg-36 | 8.38 | 4.00 | 0.61, −0.06 | 1.03, 1.22 | 2.97, 2.99 | 7.33 ($N^{\epsilon}H$) |
| Val-37 | 7.63 | 5.00 | 1.80 | 0.50, 0.60 | | |
| Asp-38 | 9.56 | 5.38 | 3.03, 2.98 | | | |
| Leu-39 | 8.99 | 5.27 | 1.60, 1.60 | 1.50 | 0.86, 0.86 | |
| Gly-40 | 6.85 | 4.41, 3.53 | | | | |
| Cys-41 | 8.59 | 5.77 | 3.32, 2.84 | | | |
| Ala-24 | 9.07 | 4.84 | 1.64 | | | |
| Ala-43 | 9.36 | 4.27 | 1.67 | | | |
| Thr-44 | 7.35 | 4.40 | 4.05 | 1.10 | | |
| Cys-45 | 8.98 | 4.97 | 2.97, 2.97 | | | |
| Pro-46 | | 4.16 | 1.34, 2.02 | 0.82, 0.82 | 3.10, 3.77 | |
| Thr-47 | 7.92 | 4.24 | 4.03 | 1.20 | | |
| Val-48 | 8.31 | 4.27 | 2.08 | 0.68, 0.78 | | |
| Lys-49 | 7.91 | 4.31 | 1.82, 1.82 | 1.40, 1.40 | 1.62, 1.62 | 2.91, 2.91 ($C^{\epsilon}H$), 7.45 ($N^{\xi}H$) |
| Thr-50 | 8.04 | 4.01 | 4.01 | 1.26 | | |
| Gly-51 | 8.91 | 4.40, 3.72 | | | | |
| Val-52 | 8.02 | 4.52 | 2.13 | 0.84, 0.64 | | |
| Asp-53 | 8.90 | 5.12 | 2.82, 2.86 | | | |
| Ile-54 | 8.74 | 5.44 | 1.47 | 1.33, 1.12 | 0.45 | 0.85 ($C^{\gamma}H_3$) |
| Gln-55 | 8.31 | 4.70+ | 1.91, 2.04 | 2.26, 2.26 | | 6.81, 7.64 ($N^{\epsilon}H$) |
| Cys-56 | 9.65 | 5.63 | 3.77, 3.00 | | | |
| Cys-57 | 9.20 | 5.14 | 3.60, 3.35 | | | |
| Ser-58 | 9.01 | 4.94 | 3.10, 4.15 | | | |
| Thr-59 | 7.51 | 4.74 | 4.15+ | 1.21 | | |
| Asp-60 | 8.37 | 4.70 | 2.49, 2.18 | | | |
| Asn-61 | 9.85 | 3.99 | 2.93, 2.74 | | | 7.25, 6.12 ($N^{\delta}H$) |
| Cys-62 | 7.54 | 4.61 | 3.54, 3.49 | | | |
| Asn-63 | 8.97 | 4.94 | 2.09, 2.84 | | | 7.03, 7.90 ($N^{\delta}H$) |
| Pro-64 | | 4.17 | 1.70,* 1.70* | 1.60,* 1.60* | 3.50, 3.29 | |
| Phe-65 | 7.89 | 4.13 | 2.64, 2.84 | | 7.10 | 7.33 ($C^{\epsilon}H$) |
| Pro-66 | | 4.27 | 2.09, 2.30 | 1.87, 1.87 | 3.32, 4.09 | |
| Thr-67 | 7.92 | 4.26 | 4.17 | 1.23 | | |
| Arg-68 | 8.40 | 4.61 | 1.78, 1.78 | 1.65, 1.59 | 3.21, 3.21 | 7.21 ($N^{\epsilon}H$) |
| Lys-69 | 8.38 | 4.34 | 1.79, 1.79 | 1.38, 1.38 | 1.60, 1.60 | 2.99, 2.99 ($C^{\epsilon}H$), 7.58 ($N^{\xi}H$) |
| Arg-70 | 8.18 | 4.52 | 1.67, 1.57 | 1.57, 1.45 | 3.17, 3.17 | 7.19 ($N^{\epsilon}H$) |
| Pro-71 | | 4.21 | 2.22, 2.20 | 1.87, 1.97 | 3.57, 3.69 | |

[a] Chemical shifts are relative to the $H_2O$ resonance, whichis 4.82 ppm at 20 °C relative to DSS. A plus (+) indicates that the chemical shift was taken from spectra at 45 °C. An asterisk (*) indicates that the assignment can be interchanged with another assignment within the same residue which also has an asterisk (*).

Table II: Structural Statistics

| | 18 structures | mean structure |
|---|---|---|
| $F_{NOE}$ (kcal/mol) | 128.2 ± 10.3 | 120.3 |
| $F_{tor}$ (kcal/mol) | 10.4 ± 5.4 | 4.0 |
| $F_{repel}$ (kcal/mol) | 76.7 ± 5.8 | 73.7 |
| $E_{L-J}$ (kcal/mol)[a] | −138.5 ± 9.7 | −136.4 |

[a] $E_{L-J}$ is the Lennard-Jones van der Waals energy calculated with the CHARMM empirical energy function (Brooks et al., 1983).

Table III: Average RMSDs between the 18 Refined Structures of α-Cobratoxin and, Respectively, the Mean of These 18 Structures and the Crystal Structure

| reference structure | atoms considered[a] | average RMSD (Å) |
|---|---|---|
| mean α-cobratoxin structure[b] | backbone 1–66 minus 25–36 | 1.07 ● 0.23 |
| | all heavy atoms minus 25–36 | 1.56 ± 0.28 |
| crystal structure of α-cobratoxin | backbone 1–66 minus 25–36 | 1.86 ± 0.21 |
| | all heavy atoms minus 25–36 | 2.66 ± 0.24 |

[a] Backbone atoms are N, $C^{\alpha}$, and C′. The numbers indicate the amino acid residues considered. [The conformations of the segment 25–36 and the terminal segment 67–71 are not well defined by NMR data (see the text)]. [b] The mean structure was obtained by taking the average of the atom coordinates in the 18 refined structures.

tail is most likely in high mobility, as emphasized by the crystallographic structure where the last four residues do not all lie in continuous electronic density.

None of the structures calculated exhibited distance violations greater than 0.3 Å or dihedral violations greater than 10°. Structural statistics for the converged structures and atomic RMSDs are given in Tables II and III, respectively. Figure 5 gives the variations of the RMSDs along the primary sequence of α-cobratoxin. One can easily see, in Figures 4 and 5, that the segment comprising residues 26–35 is rather disordered and seems to adopt several conformations. This will be discussed later.

## Discussion

*Description of the Structure.* The Results section describes how MCD patterns were used to identify a triple- and a double-stranded antiparallel β-sheet, a short helix, and turns. Additional data, given in Figure 3, support these results.[3] The characteristic strong $d_{\alpha N}(i,i+1)$ and weak $d_{NN}(i,i+1)$ NOE cross peaks are observed for the extended chain segments involved in the sheets. Large $^3J_{HN\alpha}$ coupling constants are also observed (data not shown). The residues involved in the short helix (29–34) show large $d_{NN}(i,i+1)$ and a couple of medium-range $d_{\alpha N}(i,i+3)$ and $d_{\alpha N}(i,i+4)$ cross peaks. A clear indicator of turns is the combination of one or two short-range $d_{NN}(i,i+1)$ NOEs with medium range $d_{\alpha N}(i,i+1)$ NOE cross peaks (Wagner et al., 1986). The data in Figure 3 are thus consistent with possible turns located at segments 7–10, 14–17, 49–52, and 60–63.

The nature of the distance constraints shows that only one family of structures exists rather than several greatly differing structures. The structures within this family are clearly very similar (Figure 4) and have a low RMSD value of 1.1 Å for the backbone atoms (Table III). The low-energy values also indicate that the structures converged well to acceptable stable conformations (Table II), confirmed by the plot of dihedral backbone angles in the $(\phi,\psi)$ conformation space (Ramachandran plot) of the mean NMR structure (Figure 6). A

[3] The nomenclature used is similar to that described by Wüthrich (1986).

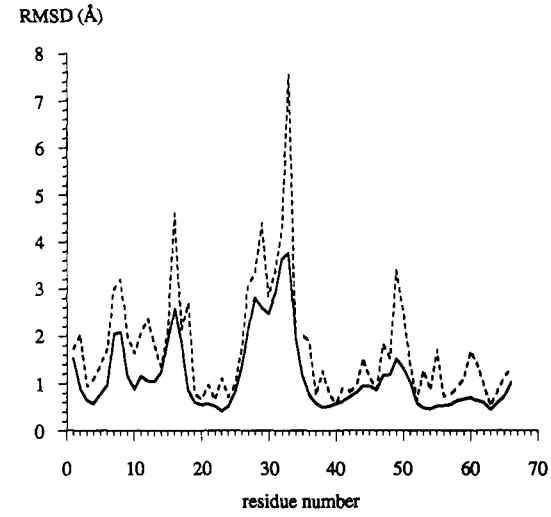



FIGURE 5: Distribution of the RMS difference of the 18 structures of α-cobratoxin shown in Figure 6 relative to the mean structure, plotted on a per residue basis (residues 1–66): backbone heavy atoms (solid line) and side-chain heavy atoms (dashed line).

single global fold family is outlined in Figure 4. Essentially, three loops emerge from a globular head region. Loop I (the right-most loop on the picture) involves the double-stranded sheet. Loop II (middle loop) involves STR.2 and STR.3 (see Figure 1) of the triple-stranded sheet. Loop III (the left-most third loop) involves STR.1 of the triple-stranded sheet.

The hydrophobic part of the molecule is well-defined due to the large number of constraints available, involving both backbone and side-chain atoms, and the numerous hydrogen bonds (unexchanged backbone amide protons from residues 2 to 5 and 61 to 63; see Figure 3). This structural region is composed of segments 1–5, 19–21, 39–43, and 57–65. The aromatic residues show many proximities with other residues in the area: Phe-4 with residues 10–12, 39, and 61–63; Tyr-21 with residues 40–42, 46, 54, 56, 57, and 63; Phe-65 with residues 9, 37, and 39.

The Tyr-21 is clustered in a hydrophobic environment provided by its neighboring residues: the first side-chain atoms of Lys-23, Leu-39, Ala-42 and Ala-43, Pro-46, and Ile-54. Only the hydroxyl group of Tyr-21 emerges from this cylindrical cavity. The plane of the Gly-40 amide grouping and the Tyr-21 ring are approximately parallel to one another, providing an explanation for the unusual chemical shift of the Gly-40 amide proton.

Loop I is partly hydrophobic (residues 1–5) and partly exposed to solvent (7–18), showing more flexibility in this latter part (Figure 5).

Loop II is the largest loop in the tertiary structure and is mainly composed of two strands from the β-sheet. The Trp-25 residue is sandwiched between Arg-36 and Val-52, this configuration being confirmed by the unusual chemical shifts of the β protons of Arg-36 located in front of the Trp-25 ring (Figure 2 of Supplementary Material).

At the end of the central loop lies the helical region identified by NMR. The structure seems conformationnally variable; this mobility is caused by the shortness of the helix (five residues), the high accessibility of the region to solvent (Ser-31, Arg-33, and Lys-35), and the flexibility of Gly-34. The region may also be less constrained by NMR data, as no long-range constraints are available in this segment.

Loop III is closed by the disulfide bond Cys-45–Cys-56. Nearly perpendicular to the β-sheet plane, this loop is exposed to solvent in the Cys-45–Thr-50 sequence, with mobile residues like Lys-49 which is involved in binding interactions with

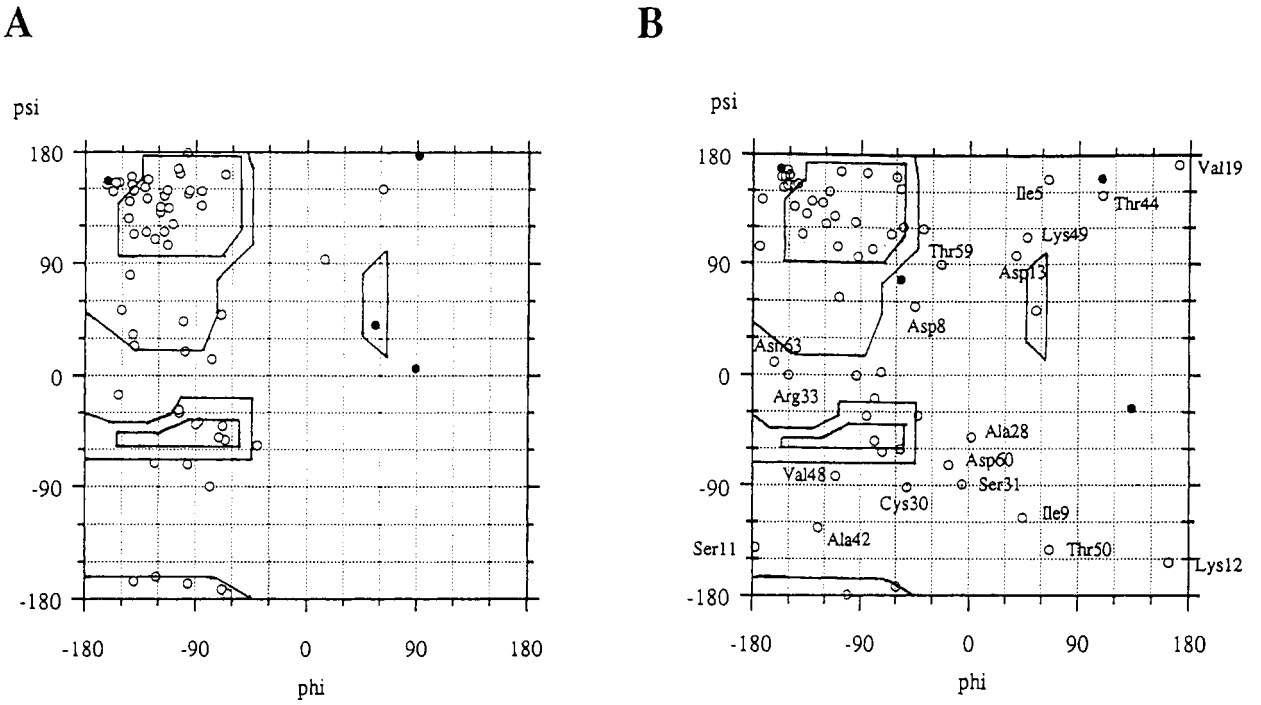



FIGURE 6: Ramachandran plots of the backbone conformational ($\phi$,$\psi$) angles of α-cobratoxin. (A) Plot for the mean NMR structure and (B) plot for the X-ray structure. Glycines are represented by filled circles.

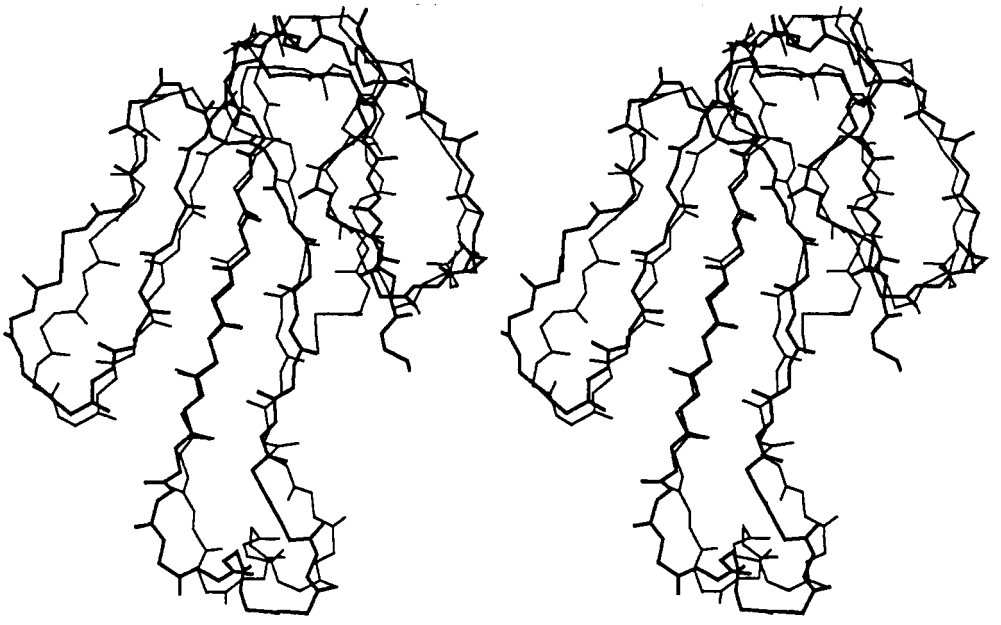


FIGURE 7: Least-squares fit superposition of α-cobratoxin backbone atoms (residues 1–66): mean NMR structure (thick line) and X-ray structure (thin line).

acetylcholine receptor. This loop has been identified as a structural crevice where a Trp from the receptor could fit in (Low & Corfield, 1986).

*Comparisons with X-ray Structures.* The X-ray structure of α-cobratoxin has been solved at 2.8-Å resolution (Walkinshaw et al., 1980). The global fold, sheets, and turns are very similar in solution and in the cristal as evidenced by the 1.8-Å RMSD value (Table III) for the backbone atoms and the superposition of the two structures (Figure 7). The C-terminal ends differ notably because the X-ray structure has this end tucked in behind the middle loop. If this would be the case for the solution structure, long-range NOEs should be observable. In fact, no long-range NOEs were observed, probably due to the mobility of this region, where the electronic density was also found to be rather stretched. The helix region differs notably between the solution and X-ray structures (Figure 8). Two hydrogen bonds between NH 30-O 27 and NH 33-O 30 are reported in the X-ray structure, where Walkinshaw et al. (1980) found an "intricate chain-folding pattern which involves the fifth disulfide bridge and two β-turns". In solution, all NHs from this region appeared to exchange rapidly with the solvent in $D_2O$ (Figure 3). Both structures may not necessarily be in disagreement; it may be more appropriate to conclude that it is most likely due to the flexibility of this region of the protein. This observation is based on the absence of a well-defined conformation deduced from NMR data available for the 26–35 segment (Figures 4 and 5). Furthermore, the crystallographic data report intermolecular interactions within the lattice at this region which could destabilize the helix: the Phe-29 experiences a hydrophobic environment provided by a neighboring molecule, consisting of Pro-64, Phe-65, and Phe-4. This region may be

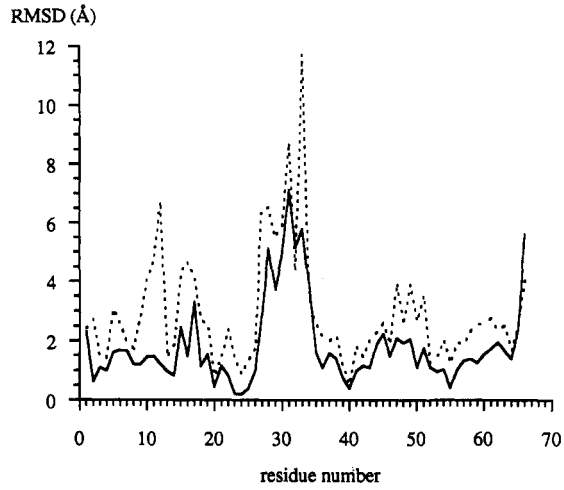



FIGURE 8: Distribution of the RMS difference per residue between the mean NMR and the X-ray structures on the primary sequence of α-cobratoxin (residues 1–66): backbone heavy atoms (solid line) and side-chain heavy atoms (dashed line).

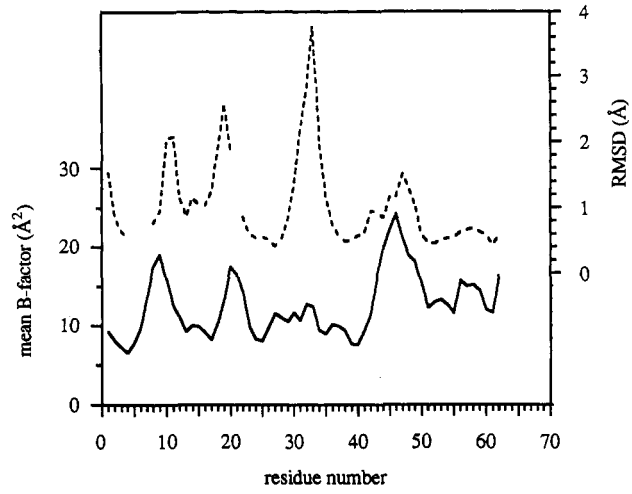



FIGURE 9: Comparison of mean *B* values for crystalline erabutoxin-b (solid line; Low & Corfield, 1986) and the distribution of the RMS difference from the mean NMR structure of α-cobratoxin (dashed line; same data as in Figure 5). The sequences have been aligned according to the erabutoxin-b primary sequence.

a site involved in the binding of α-cobratoxin to the acetylcholine receptor (Martin et al., 1983). It is interesting to note that, beside the extensive similarities between the long and short neurotoxins, one difference is that the former have additional residues within this region (Walkinshaw et al., 1980). This raises the question of whether this dynamic helical structure could account for the difference in binding rates to acetylcholine receptors observed between the toxin families (Chicheportiche et al., 1975).

Loop I also presents differences in the orientation of the side chains involved in the short β-sheet identified in the NMR structure. Lys-12 has thus an inverted position in the NMR structure.

Comparison of Figures 5 and 8 reveals some similarities between the two graphs; the RMSD value is particularly high for the 26–35 segment. This suggests that the NMR structure is disordered and that it is different from the crystal structure, probably distorted by intermolecular packing interactions. Both arguments imply that the 26–35 segment is intrinsically flexible. The other weakly determined segments in the NMR structure (loop I and segment 45–51) differ less from the X-ray structure. The Ramachandran plot of the crystalline structure (Figure 6) shows that these segments have unrealistic conformations which could not all be explained by intermolecular interactions in the crystal. Residues out of the favorable zones in the $(\phi,\psi)$ plot are from the helical region (Cys-30, Ser-31, and Arg-33), from turn regions (Lys-49, Thr-50 and Thr-59, Asp-60, and Asn-63), and from loop I (Ile-5, Ile-9, and Asp-13).

Crystallographic thermal *B* factors are not available for the X-ray structure of α-cobratoxin. We therefore decided to compare the solution structure with the crystallographic structure of erabutoxin-b, a homologous protein of the short-toxin family, which has been solved to a resolution of 1.4 Å (Low & Corfield, 1986). Figure 9 shows the comparison between the thermal *B* factors of the erabutoxin-b and the mean RMSD of the NMR structures of α-cobratoxin for backbone atoms. The sequences have been aligned according to previous homology studies, and the numbering refers to the erabutoxin-b sequence. It appears that the region of the helix which is present in α-cobratoxin but not in erabutoxin-b seems extremely mobile by comparison to the rest of the structure, as previously noticed. The other regions exhibit remarkably similar mobilities in the two structures, in spite of differences in the sequences of the two toxins. The antiparallel β-sheet, although shorter in α-cobratoxin, is the most stabilized part for both proteins and plays a structural function for these toxins. It is also interesting to note that Bourne et al. (1985) reported that the five-stranded β-sheet in erabutoxin-b is disrupted between a two-stranded loop and a three-stranded sheet, a structural feature present in solution but not in the X-ray structure of α-cobratoxin. Loop III, highly homologous in the two toxins, is in both cases a mobile part of the protein, an explanation for the high number of reactive-site residues (Pro-46, Val-48, Lys-49, Gly-51, and Val-52) present in this region.

It is clear that the solution and crystal structures of α-cobratoxin are very similar (Figure 7). This is in contrast to results reported for the family-related α-bungarotoxin (Basus et al., 1988; Love & Stroud, 1986). Basus et al. reported that the solution structure is more similar to α-cobratoxin than to the crystal structure of α-bungarotoxin, where Trp-28 lies on the opposite side of the β-sheet. In fact, the β-proton of Val-39 is found to be the most upfield resonance in the spectrum of α-bungarotoxin, which is consistent with a location in front of the Trp-28 ring. The same is observed in the case of α-cobratoxin where the β-protons of Arg-36 are located in front of the Trp-25 ring.

The differences between the solid-state and solution structures of α-bungarotoxin, along with our observation of conformational mobility in the helical region, suggest that this long-toxin family has a high degree of flexibility. This feature should be functionally important in order to accommodate specific interactions with receptors from several species.

## ACKNOWLEDGMENTS

We thank Dr. A Menez for providing the protein. We also thank Pr. W. van Gunsteren and T. Malliavin for valuable discussions. We are also very pleased to thank the Groupement Scientifique CNRS-IBM France, for a generous allottment of computer time on an IBM3090 at the CIRCE Orsay, on which the X-PLOR computations were performed. R.L.G. thanks Rhône-Poulenc Rorer for financial support.

## SUPPLEMENTARY MATERIAL AVAILABLE

Detailed procedure used to identify the proton resonance assignments of α-cobratoxin and the complete X-PLOR input files of constraints (24 pages). Ordering information is given on any current masthead page.

**Registry No.** α-Cobratoxin, 11119-60-1.

## REFERENCES


Basus, V. J., Billeter, M., Love, R. A., Stroud, R. M., & Kuntz, I. D. (1988) *Biochemistry 27*, 2763–2771.

Boquet, P., Poilleux, G., Dumarey, C., Izard, Y., & Ronsseray, A. M. (1973) *Toxicon 11*, 333–340.

Bourne, P. E., Sato, A., Corfield, P. W., Rosen, L. S., Birken, S., & Low, B. W. (1985) *Eur. J. Biochem. 153*, 521–527.

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.

Brünger, A. T. (1990) X-PLOR *Manual*, Yale University, New Haven, CT.

Brünger, A. T., Clore, G. M., Gronenborn, A. M., & Karplus, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3801–3805.

Chicheportiche, R., Vincent, J. P., Kopeyan, C., Schweitz, H., & Lazdunski, M. (1975) *Biochemistry 14*, 2081–2091.

Crippen, G. M. (1977) *J. Comput. Phys. 24*, 96.

Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820–2821.

Di Stephano, D. L., & Wand, A. J. (1987) *Biochemistry 26*, 7272–7281.

Dufton, M. T., & Hider, R. C. (1977) *J. Mol. Biol. 115*, 177–193.

Englander, S. W., & Wand, A. J. (1987) *Biochemistry 26*, 5953–5958.

Havel, T. F., Kuntz, I. D., & Crippen, G. M. (1983) *Bull. Math. Biol. 45*, 665–720.

Hider, R. C., Drake, A. F., Inagaki, F., Williams, R. J. P., Endo, T., & Miyazawa, T. (1982) *J. Mol. Biol. 158*, 275–291.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.

Karlsson, E., Arnberg, H., & Eaker, P. (1971) *Eur. J. Biochem. 21*, 1–16.

Kimball, M. R., Sato, A., Richardson, J. S., Rosen, L. S., & Low, B. W. (1979) *Biochem. Biophys. Res. Commun. 88*, 950–959.

Kondakov, V. I., Arseniev, A. S., Plushnikov, K. A., Tsetlin, V. I., Bystrov, V. F., & Ivanov, V. T. (1984) *Bioorg. Khim. 10*, 1606–1628.

Kördel, J, Forsén, S., & Chazin, W. J. (1989) *Biochemistry 28*, 7065–7074.

Kuntz, I. D., Crippen, G. M., & Kollman, P. A. (1979) *Biopolymers 18*, 939–957.

LaPlante, S. R., Mikou, A., Robin, M., Guittet, E., Delsuc, M. A., Charpentier, I., & Lallemand, J. Y. (1990) *Int. J. Pept. Protein Res. 36*, 227–230.

Love, R. A., & Stroud, R. M. (1986) *Protein Eng. 1*, 37–46.

Low, B. W., & Corfield, P. W. R. (1986) *Eur. J. Biochem. 161*, 579–587.

Macura, C., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259–282.

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

Martin, B. M., Chibber, A., & Maelicke, A. (1983) *J. Biol. Chem. 258*, 8714–8722.

Menez, A. (1987) *Recherche 18*, 888–893.

Menez, A., Langlet, G., Tamiya, N., & Fromageot, P. (1978) *Biochimie 60*, 505–516.

Nilges, M. (1990) X-PLOR *Manual*, Version 2.1 (A. T. Brünger), Yale University, New Haven, CT.

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

Redfield, A. G., & Kunz, S. D. (1975) *J. Magn. Reson. 19*, 250–254.

Saudek, V., Atkinson, R. A., Williams, R. J. P., & Ramponi, G. (1989) *J. Mol. Biol. 205*, 229–239.

Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson. 51*, 169–173.

Wagner, G., Neuhaus, D., Wörgötter, E., Vasaak, M., Kägi, J. H. R., & Wüthrich, K. (1986) *J. Mol. Biol. 187*, 131–135.

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

Walkinshaw, M. D., Saënger, W., & Maelicke, A. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 2400–2404.

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

Wüthrich, K., & Wagner, G. (1979) *J. Mol. Biol. 130*, 1–18.

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

Zuiderweg, E. R. P., Boelens, R., & Kaptein, R. (1985) *Biopolymers 24*, 601–611.