

## Solution structure of candoxin, a novel three-finger toxin from the venom of *Bungarus candidus*

Rani Parvathy V.<sup>a</sup>, Kandala V. R. Chary<sup>a,\*</sup>, R. Manjunatha Kini,<sup>b,c</sup> and Girjesh Govil<sup>a</sup>

<sup>a</sup>Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India

<sup>b</sup>Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore 119260, Singapore

<sup>c</sup>Department of Biochemistry and Molecular Biophysics, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298-0614, USA

E-mail: [chary@tifr.res](mailto:chary@tifr.res)

Dedicated to Professor P.T. Narasimhan on his 75<sup>th</sup> birthday

---

### Abstract

Candoxin, a novel toxin purified from the venom of the Malayan krait (*Bungarus candidus*), is a 66-residue polypeptide containing five disulfide bridges, and is a reversible antagonist of postjunctional nicotinic acetylcholine receptors of the neuromuscular junction. A family of structures were calculated using a combination of distance geometry and simulated annealing with nOe, hydrogen bond and dihedral angle constraints. After refinement 19 structures, which satisfy the experimental constraints, were obtained. A comparison of each of the final structures with the average structure, gives an RMSD of 1.20 Å for backbone atoms. Candoxin has an overall conformation similar to other three-finger toxins with a two-stranded and a three-stranded  $\beta$ -sheets. In spite of the low sequence homology, the tertiary fold of candoxin closely resembles those of erabutoxin b and cobrotoxin, with the exception of the disulfide bridge in the Loop I. Though candoxin lags the most conserved Tyr at the origin of the central loop it posses the most popular 3-dimensional array of residues (W31, R35 and K49), which is reported to be critical for curaremimetic neurotoxicity. The presence of the kink at the tip of the first loop (Loop I) caused by the C6-C11 disulphide bridge, shortens the Loop and hence may isolate this portion of the molecule and prevent its interference with the rest of the molecule, and facilitate specific interaction with receptor/acceptor protein. These observations suggest that candoxin may bind to acetylcholine receptor with lower affinity compared to other neurotoxins and hence could account for its relatively lower toxicity.

**Keywords:** Toxin, 2D NMR spectroscopy, structure

---

## Introduction

Toxicity of venoms from snakes of elapidae family (cobras, kraits and mambas) arises from a complex mixture of ingredients, which include neurotoxins and cardiotoxins<sup>1-3</sup>. In general, toxins target specific receptors, ion channels and enzymes and interfere in their normal physiological processes. For example, some neurotoxins specifically interfere at neuromuscular junctions, acting at both pre- and post-synaptic sites. Presynaptic neurotoxins act, by inhibiting the release of acetylcholine and consequently impairing neuromuscular transmission. There are two major targets in the postsynaptic site namely acetylcholine receptor and acetylcholinesterase. Monomeric  $\alpha$ -neurotoxins bind with high affinity to the nicotinic site of acetylcholine receptors of the neuromuscular junction. In most cases, the binding is stronger to the acetylcholine receptor<sup>4</sup> ( $K_d \sim 10^{-9}$  M- $10^{-11}$  M) than their physiological ligand acetylcholine ( $K_d \sim 10^{-6}$  M). The  $\alpha$ -neurotoxins can be classified into two groups based on the chain length and number of disulfide bridges. Short chain neurotoxins have 58-62 residues and 4 disulfide bridges in the core region, while long chain neurotoxins have 65-74 residues and 5 disulfide bridges.

Structures of several neurotoxins have been solved by X-ray crystallography and 2D NMR spectroscopy<sup>5-9</sup>. They have long anti-parallel  $\beta$ -sheets forming three loops protruding from a core containing the four conserved disulfide bridges. Similar 3D structural topology is found in other classes of toxins such as  $\kappa$ -toxins<sup>10, 11</sup>, cardiotoxins (or cytotoxins)<sup>12</sup>, calciseptine and related toxins<sup>13</sup>, fasciculins<sup>14</sup> (inhibitors of acetylcholinesterase), muscarinic toxins<sup>15</sup> and mambin<sup>16</sup>. The structural fold of these toxins is classified under a super family of three-finger toxins because of their uncanny resemblance to three fingers (loops) stretched out of the palm (core). Despite topological similarities, these toxins differ significantly in their specific targets, receptor/acceptor interactions and consequently in pharmacological effects<sup>17</sup>. Each member of the family is distinct due to differences in sizes of loops interloop interactions, stability and flexibility<sup>18</sup>.

A small number of three-finger toxins possess a fifth disulfide bridge in the first loop (loop I)<sup>19</sup>. The Cys positions in the primary sequence of such proteins indicate that they have disulfide bridging patterns different from that normally seen in  $\alpha$ -neurotoxins. This group of polypeptides tends to have lower toxicity<sup>20</sup>. To date, data on the pharmacological properties or the tertiary structure of these toxins are scarce.

We have recently purified a novel toxin called candoxin, from the venom of the Malayan krait, *Bungarus candidus*. It has 66 amino acid residues<sup>21-23</sup>, intermediate between the long and short chain neurotoxins. Its sequence alignments with the long and short  $\alpha$ -neurotoxins show deletions of several amino acid residues at the C-terminal end and in the middle of the primary sequence; several insertions are seen at the N-terminal [Figure 1]. However, some sequence homology is observed between candoxin and short or long chain  $\alpha$ -neurotoxins. Candoxin produces reversible, postsynaptic neuromuscular blockade of nicotinic acetylcholine receptors at the avian and mammalian neuromuscular junction<sup>23</sup>. In comparison with erabutoxin b from *Laticauda semifasciata* or  $\alpha$ -bungarotoxin from *Bungarus multicinctus*, which are poorly reversible, candoxin is found to be completely reversible, although 8-12 times less potent. Interestingly, it is poorly

reversible antagonist of neuronal  $\alpha 7$  nicotinic acetylcholine receptors. As a step towards understanding the structure-function relationship of this toxin, we report its 3D solution structure, determined by 2D-NMR spectroscopy and molecular dynamics simulation.

candoxin	MKCKICNFDT	CRAGELKVCA	SGEKYCFKES	...WREARG
long neurotoxin homologue	MKCKICHFDT	CRAGELKVCA	SGEKYCFKES	...WREARG
CM-9a	.RCLNCPMEF	CGK..FQICR	NGEKICFKKL	H...QRRPLS
CM-10	LRCLNCPVEF	CRN..FHTCR	NGEKICFKRF	D...QRKLLG
erabutoxin	RICFNHQSSQ	PQT..TKTCS	PGESSCYHKQ	...WSDFRG
cobratoxin	LECHNQSSQ	TPT..TTGCS	GGETNCYKKR	...WRDHRG
bucandin	MECYRCGVSG	CHLK..ITCS	AEETF CYKWL	.....NKIS
CD59	LQCYNCPNPT	ADCKTAVNCS	SDFDACLITK	AGLQVYNKCW
candoxin	TRIERGCAAT	CPKGSVYGLY	VLCCTTDDCN	.....
long neurotoxin homologue	TRIERGCAAT	CPKGSVYGLY	VLCCTTDDCN	..... (98.48%)
CM-9a	R.YIRGCADT	CPVGYPK.EM	IECCSTDKCN	R..... (33.33%)
CM-10	KRYTRGCAVT	CPVAKPR.EI	VECCSTDGCM	R..... (29.23%)
erabutoxin	TIIERGCG..	CPTVKPG.IK	LSCCESEVCN	N..... (29.03%)
cobratoxin	YRTERGCG..	CPSVKNK.IE	INCCTTDRCN	N..... (33.87%)
bucandin	NERWLGCAKT	CTEIDTWNVY	NKCCTTNLCN	T..... (25.40%)
CD59	KFEHCNFNDV	TTRLRENELT	YYCCKKDLN	FNEQLN (10.61%)

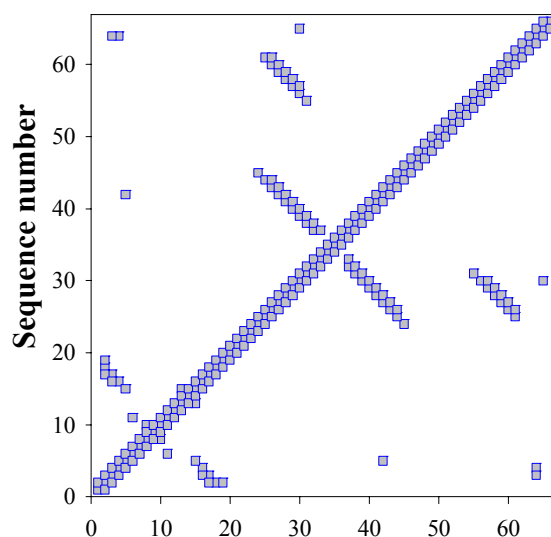
**Figure 1.** Sequence alignment of candoxin with other neurotoxins.

## Results and Discussion

### Secondary structure of candoxin

Sequence specific  $^1\text{H}$  and  $^{13}\text{C}$  resonance assignments were obtained by standard procedures as described elsewhere<sup>21-23</sup>. The  $^1\text{H}$  resonance assignments, nOe patterns, chemical shift differences for  $^1\text{H}^\alpha$ ,  $^{13}\text{C}^\alpha$  and  $^{13}\text{C}^\beta$  from the corresponding random coil values (CSI) and deuterium exchange rates for amide protons ( $^1\text{H}^\text{N}$ ) have been used to determine the secondary structural elements in candoxin. The nOe contact map for the backbone atoms for each residue in candoxin is shown in Figure 2. The complete absence of  $d_{\alpha\text{N}}(i, i+3)$ ,  $d_{\alpha\text{N}}(i, i+4)$  and the presence of very few  $d_{\text{NN}}$  and  $d_{\alpha\text{N}}(i, i+2)$  nOes indicate the absence of  $\alpha$ -helical segments in candoxin. The fact that most of the self  $d_{\alpha\text{N}}$  nOes are found to be weak in intensity and the sequential  $d_{\alpha\text{N}}$  connectivities are strong, support the presence of extended ( $\beta$ -strand) conformations along the polypeptide chain. In all, five  $\beta$ -strands ( $\beta 1$ – $\beta 5$ ) could be identified.

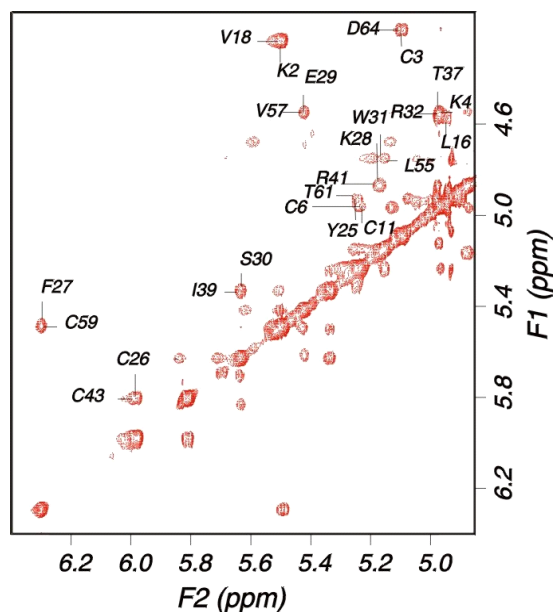
The  $\beta$ -sheet patterns are characterized by the long range ( $d_{\alpha\alpha}$ ,  $d_{\alpha N}$  and  $d_{NN}$ ) nOes. In antiparallel  $\beta$ -sheet structures long range  $d_{\alpha\alpha}$  nOes form the best signatures. In candoxin, 12  $d_{\alpha\alpha}$  nOes are observed in the NOESY spectrum (Figure 3). These have been characterized as arising from the residues K2-V18, K4-L16, C3-D64, T37-R32, S30-I39, K28-R41, C26-C43, K24-A45, W31-L55, E29-V57, F27-C59 and Y25-T61. Such long range nOes provide direct evidence for the hydrogen bonded amino acid partners between various  $\beta$ -strand structures as shown in Figure 4. Assignment of several long-range  $d_{NN}$  and  $d_{\alpha N}$  nOes (Figure 4), establishes that the five  $\beta$ -strands form two independent  $\beta$ -sheet structures. The hydrogen bonds shown by dotted lines in Figure 4 are confirmed from deuterium exchange studies. Most of the hydrogen bonded  $H^N$  protons in the  $\beta$ -strands exchange very slowly. The nOe connectivities (Figure 4), establish that strands  $\beta 1$  and  $\beta 2$  form a double stranded anti-parallel  $\beta$ -sheet structure (I). Strands  $\beta 3$ ,  $\beta 4$  and  $\beta 5$  form a triple stranded structure with  $\beta 3$  being antiparallel to  $\beta 4$  and  $\beta 5$  (II).



**Figure 2.** Diagonal plot representing the nOes between different backbone protons in candoxin, used for the delineation of secondary structure. Both axes are calibrated with the amino-acid sequence number. Self and sequential nOes form the diagonal. The long range nOes, which form perpendicular lines to the diagonal, arise from  $\beta$ -sheets.

An important secondary structure element in proteins is  $\beta$ -turn. It is usually characterized by a unique nOe pattern and supported by specific hydrogen exchange rates for the individual backbone  $H^N$  protons present in such turn. In candoxin, 6 such tight turns ( $T1$ - $T6$ ) have been identified from the nOe data. These are N7-F8-D9-T10 ( $T-1$ ), R12-A13-G14-E15 ( $T-2$ ), G22-E23-K24-Y25 ( $T-3$ ), E33-A34-R35-G36 ( $T-4$ ), A44-A45-T46-C47 ( $T-5$ ) and D63-D64-C65-N66 ( $T-6$ ). In the segment N7-F8-D9-T10 ( $T-1$ ), we observe strong  $d_{NN}$  (9,10) connectivity (i.e. between the 3<sup>rd</sup> and 4<sup>th</sup> residues of the turn) and medium  $d_{\alpha N}$ (8,10) and  $d_{\beta N}$ (8,10) connectivities which establish that this stretch adopts a characteristic type II turn. This is confirmed by the

absence of  $d_{\text{NN}}$  (8,9) connectivity. Similarly, the stretches *T*-2 and *T*-5 have been characterized as type II  $\beta$ -turns from the observation of strong  $d_{\text{NN}}$  and  $d_{\alpha\text{N}}$  connectivities between the 3<sup>rd</sup> and 4<sup>th</sup> residues and 2<sup>nd</sup> and 3<sup>rd</sup> residues, respectively. On the other hand, *T*-4 and *T*-6 stretches are found to adopt type I  $\beta$ -turn conformation. This is established for *T*-4 by the strong sequential  $d_{\text{NN}}$  connectivities between A34 and R35 and R35 and G36 and the absence of  $d_{\alpha\text{N}}$  (*i*, *i*+2) connectivities and further confirmed from the medium  $d_{\alpha\text{N}}$  (*i*, *i*+1) connectivities along this stretch, a characteristic of type I turn. Similarly, *T*-6 has been characterized as a type I  $\beta$ -turn. Finally, *T*-3 is found to adopt a tight turn from the observed strong  $d_{\text{NN}}$  connectivities between G22 and E23 and K24 and Y25. The strong  $d_{\text{NN}}$  (24-25) connectivity is characteristic of both type I and type II  $\beta$ -turns in this segment. In *T*-3, distinction between these two turns could not be made because of the near degeneracy of the  $^1\text{H}^{\text{N}}$  chemical shifts for E23 and K24 even though the nOe corresponding to  $d_{\alpha\text{N}}$  is not observed. Weak  $d_{\text{NN}}$  connectivities between S51-V52, V52-Y53 and G54-L55 are observed indicating another turn, but this stretch could not be fully characterized.



**Figure 3.** Contour plot of a selected region of the NOESY spectrum ( $\tau_{\text{m}} = 200\text{ms}$ ) of candoxin dissolved in 99%  $\text{D}_2\text{O}$  and at 298 K showing the observed long range  $\text{H}^{\alpha}$ - $\text{H}^{\alpha}$  nOes. These cross peaks were used to establish the  $\beta$ -sheet structures found in candoxin.

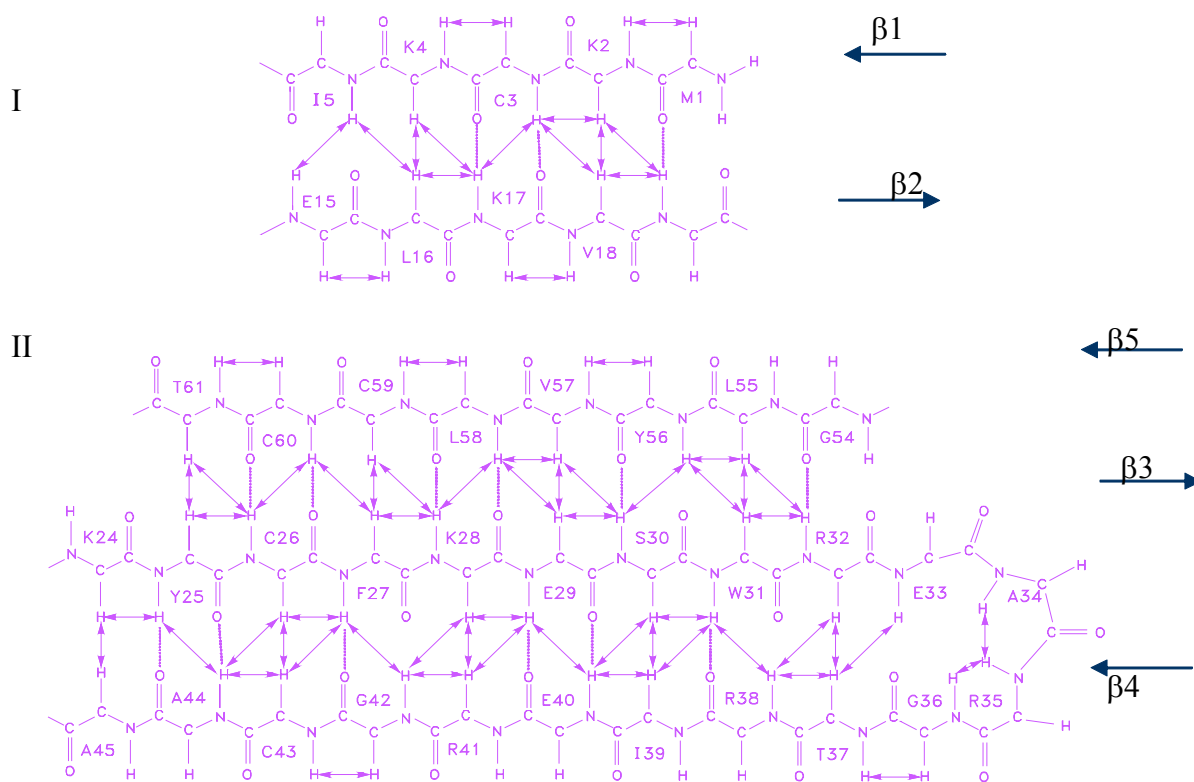
### Identification of disulfide bridges

For the 10 Cys residues, the downfield shift of individual  $^{13}\text{C}^{\beta}$  resonances, in the range 38-48 p.p.m., indicates the oxidized state of all the residues<sup>24</sup>. In their reduced state, respective  $^{13}\text{C}^{\beta}$  spins of the Cys residue resonate in the range 24-36 p.p.m.<sup>24</sup>. All the 5 disulfide bridges present in the candoxin have been identified with the aid of inter-cysteine  $\text{H}^{\beta}$ - $\text{H}^{\beta}$  nOes in the NOESY spectrum. The  $\text{H}^{\beta}$  spins of the disulfide linked cysteine residues (which are typically  $\sim 4\text{-}5 \text{ \AA}$

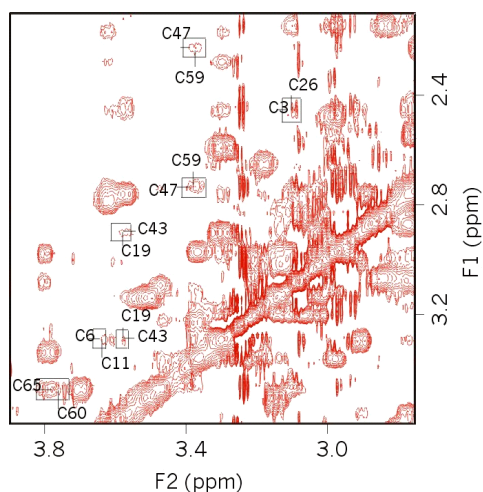
apart) showed nOes at long mixing times (200 ms) and thus helped in establishing all the 5 disulfide bridges. Figure 5 shows the Cys  $H^{\beta}$ - $H^{\beta}$  region of the NOESY spectrum in candoxin. For example, both C47 ( $H^{\beta}$ ) protons show nOes to one of the C59 ( $H^{\beta}$ ) protons indicating a disulfide bridge between C47 and C59 (Figure 5). Likewise, one of the C60 ( $H^{\beta}$ ) protons shows nOe to one of the C65 ( $H^{\beta}$ ) protons, confirming a disulfide bridge between C60 and C65. Similarly, the other 3 disulfide bridges between C3 and C26, C6 and C11, and C19 and C43 were characterized.

### 3D Structure simulation

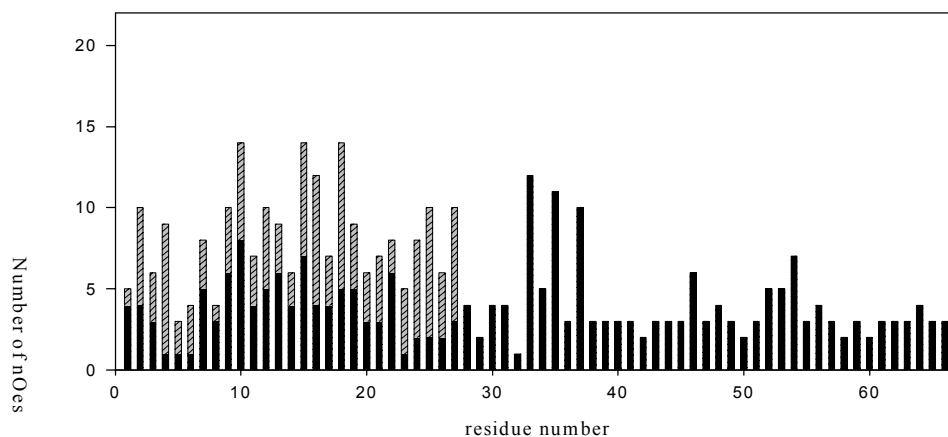
The 3D structure of candoxin was calculated using distance geometry and simulated annealing as described in Experimental Section. Initially, structures were calculated with 560 distance restraints. Standard pseudo-atom distance corrections were incorporated to account for centre averaging<sup>25</sup> for methyl protons and nonstereospecifically assigned methylene protons. From the folding pattern, ambiguities in the constraints due to resonance overlap were eliminated. The nOes, which could not be assigned unambiguously, were included in the final set of calculations.



**Figure 4.** Schematic representation of the two  $\beta$ -sheets in candoxin with various long range and sequential nOe connectivities observed in the NOESY spectrum: (a) Sheet I is two stranded and (b) Sheet II is three stranded. The dotted lines indicate hydrogen bonds. Arrows show the strand directions.



**Figure 5.** Contour plot of a selected region of Cys ( $H^{\beta}$ - $H^{\beta}$ ) cross peaks in the NOESY spectrum ( $\tau_m = 200\text{ms}$ ) of codoxin recorded in 99%  $D_2O$  and at 298 K. The cross peaks between various  $H^{\beta}$  protons belonging to different Cys residues were used to establish disulfide bridges.



**Figure 6.** Plot of nOe number of restraints per residue used in structure calculation. The self, sequential and long range nOes are marked separately.

At each stage, the distance constraints, which were consistently violated in all the structures were checked and corrected. Once the restraints were finalized, the final set of structures was calculated as described previously<sup>26</sup> with 597 nOe restraints (295 intra, 255 sequential and 177 long-range), 32 hydrogen bond restraints, 40 backbone dihedral restraints and 16 disulfide constraints. Figure 6 shows a plot of number of nOe restraints used for each residue (self, sequential and long range) in the structure calculation. 65% of the amino acid residues show long-range nOes in the toxin. On an average, 9 nOe restraints have been used per residue, which are fairly well distributed (Figure 6). Only three residues, F8, P48 and S51 show 4 nOes each. The fact that long-range nOes were observed for most of the amino acid residues,

provide confidence in the final structures reached from molecular dynamics simulations. After refinement, 19 structures with lowest energy were chosen for analysis. All atom pair-wise RMSDs were computed using MOLMOL<sup>27</sup> (Table 1). The quality of these structures was analyzed using PROCHECK<sup>27, 28</sup>. The corresponding PDB files for an ensemble of 19 structures (PDB code: 1JFJ) have been deposited. An average structure using the ensemble of 19 structures was also calculated using MOLMOL.

**Table 1.** Structural parameter statistics for the 19 energy-minimized conformers of candoxin calculated using XPLOR 3.1

<b>(1) XPLOR parameters:</b>	
Root mean square deviation from ideal geometry	
Bond length <sup>1</sup>	1.11*-3 Å
Bond angle <sup>1</sup>	0.47°
Dihedral angle <sup>1</sup>	0.36°
<u>Energy terms (kcal/mol)</u>	
E <sub>L-J</sub>	72
E <sub>NOE</sub>	7.87*-2
E <sub>dihed</sub>	0.15
<b>(2) RMSD values (Å)<sup>2</sup></b>	
Two stranded β-sheet (2-5, 15-18)	0.23±0.09 Å
Three stranded β-sheet (24-32, 37-44, 55-61)	0.49±0.11 Å
Entire back bone (Residues 1-66)	1.20±0.15 Å
Tip of Loop I (6-14)	1.04±0.39 Å
C-terminal residues(61-66)	0.91±0.69 Å
<b>(3) Ramachandran plot statistics<sup>3</sup></b>	
Residues in most favorable regions	57.2 %
Residues in allowed regions	33.3 %
Residues in generously allowed regions	5.7 %
Residues in disallowed regions	3.8 %

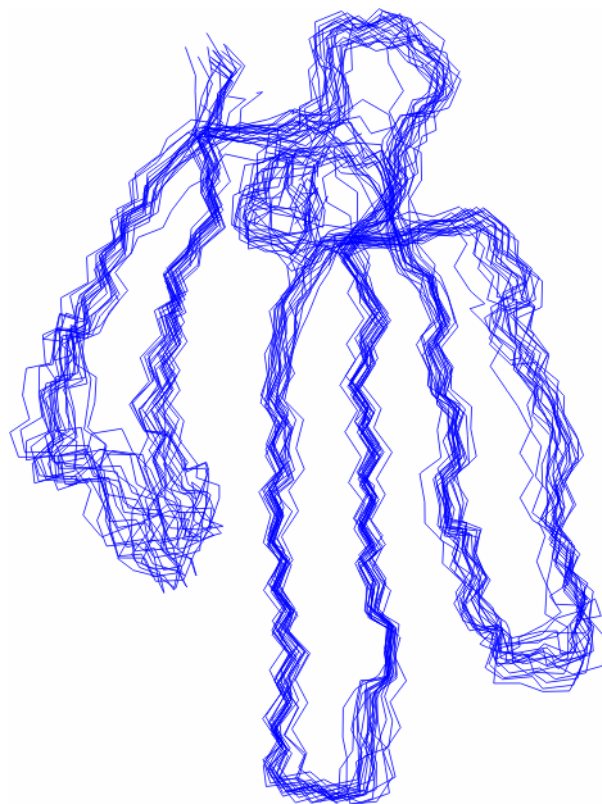
<sup>1</sup>Bond length, bond angle and dihedral angle deviations are represented for all the atoms including hydrogen atoms. <sup>2</sup>RMSDs from the average NMR structure were evaluated using MOLMOL for the backbone atoms. <sup>3</sup>Statistics for the Ramachandran plot were obtained using PROCHECK.

### 3D structure of candoxin

Figure 7A shows 19 structures with the lowest energy superimposed on one another. These structures do not have any nOe distance restraint violation greater than 0.5Å and dihedral angle violation greater than 5.0 degrees. The structures have an empirical energy within 10 kcal of each



other. A summary of the structural characteristics is given in Table 2. PROCHECK analysis<sup>27,28</sup> of these structures shows that for 90.5% of the residues, backbone dihedral angles are in the allowed regions of Ramachandran plot<sup>29</sup>. The residues with unfavorable  $\phi$ - $\psi$  angles are mainly at the tip of Loop I, in the unstructured polypeptide stretch 50-55 of Loop III and in the C-terminal end. These regions in the molecule are found to be less defined as reflected by relatively higher RMSD values from the mean structure, which are 1.04, 0.72 and 0.91 Å, respectively (Table 1).



**Figure 7A.** Superimposition of 19 energy minimized NMR structures of candoxin.

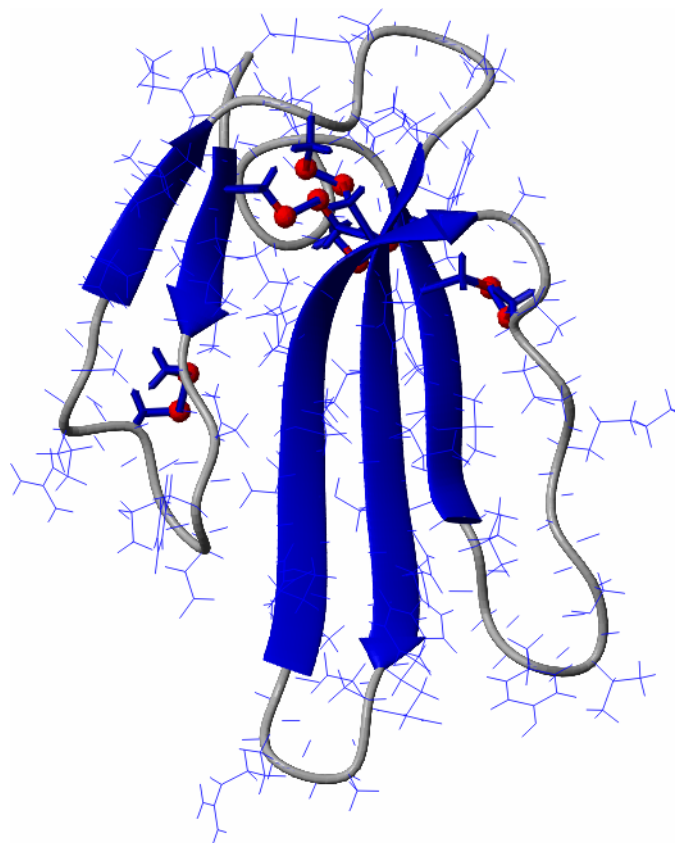
A ribbon diagram of the final average structure of candoxin is shown in Figure 7B. The structure has 2 anti-parallel  $\beta$ -sheets; a two stranded  $\beta$ -sheet (composed of  $\beta$ 1 and  $\beta$ 2 strands) and a three stranded  $\beta$ -sheet (composed of  $\beta$ 3,  $\beta$ 4 and  $\beta$ 5 strands). These  $\beta$ -strands are interleaved by 6 tight  $\beta$ -turns ( $T1$  to  $T6$ ). The polypeptide stretch at the N-terminal end forms the first loop. This loop consists of a double stranded  $\beta$ -sheet, with polypeptide stretches of K2-I5 ( $\beta$ 1) and E15-V18 ( $\beta$ 2) forming the two strands. This highly ordered  $\beta$ -sheet has a very low RMSD value from the mean structure (0.23 Å) and is stabilized by the following hydrogen bonds: E15 (CO)-I5 ( $H^N$ ), K17 ( $H^N$ /CO)-C3 (CO/ $H^N$ ) and C19 ( $H^N$ )-M1 (CO) (Figure 4). Further, the linker between the first two  $\beta$ -strands is made up of 9 residues C6-G14, where in C6 is disulfide bridged with C11, leaving two short stretches of amino acid residues, namely N7-T10 and R12-G15, which form two type II  $\beta$ -turns. These two turns appear as a double toe of the first

loop. The loop is connected to the central anti-parallel  $\beta$ -sheet by a 5 residues long peptide segment, C19-E23. The segment G22-Y25 adopts a type II  $\beta$ -turn. The second  $\beta$ -sheet is triple stranded, which is also highly ordered with a low RMSD value from the mean structure (0.49 Å) with the individual strands consisting of stretches, K24-R32, T37-A45 and L55-T61. The  $\beta$ -strands K24-R32 and T37-A45 together form the second finger of the toxin, and the residues E33-G36 connecting these two strands form a type I turn. Further, three residues of the T37-A45  $\beta$ -strand and T46 together adopt type II turn. The stretch C47-Y53 adopts an extended conformation facilitating G54-T61 stretch of amino acid residues to form part of the third loop, which runs anti-parallel to the K24-R32 stretch and is stabilized by several inter-strand hydrogen bonds (Figure 4). Both the C- and N- terminals of candoxin are close in space, as supported by the observed nOes between C3 ( $H^\alpha$ ) and D64 ( $H^\alpha/H^N$ ).

**Table 2.** List of 2D NMR experiments carried out on candoxin. The terms, np1 and np2 refer to number of points along  $\omega_1$  and  $\omega_2$  axes, respectively, while nt refers to number of scans

Experiment	$\tau_m$ (ms)	Temp (in °C)	np <sub>1</sub>	np <sub>2</sub>	nt
<b>(1) In H<sub>2</sub>O</b>					
TOCSY	50, 80	25, 10, 30	600	4096	32
Watergate NOESY	50, 75, 100, 200	25, 10, 30	700	4096	96
2QF-COSY		25, 30	600	4096	48
Relayed COSY		25	600	4096	96
<b>(2) In D<sub>2</sub>O</b>					
TOCSY	80, 50	25, 10, 30	600	4096	32
NOESY	200, 75, 50	25, 10, 30	700	4096	48
2QF-COSY		25	600	4096	48
3QF-COSY		25, 30	600	4096	96
ECOSY		25	640	4096	194
Exchange study with a series of TOCSY	80ms	25, 10	400	2048	8

Hydrogen bonding is observed between fingers II and III which hold them together. Fingers I and II do not show any inter-finger hydrogen bonding. It has previously been reported that, a longer length of loop II confers greater affinity to the acetylcholine receptor. In candoxin the length of the Loop II is ~28 Å.



**Figure 7B.** Ribbon diagram of candoxin. The disulfide bridges are marked in blue and red.

The disulfide bridges between C3 and C26, C19 and C43, C47 and C59, and C60 and C65 are homologous to the four conserved disulfide bridges found in other members of the three-finger toxin family. Expectedly all these disulfide bridges are in the core region of the candoxin. The two extra Cys residues at 6<sup>th</sup> and 11<sup>th</sup> positions form the fifth disulfide bridge in candoxin and are located at the tip of the loop I. This disulfide bridge forms a kink in the loop and stabilizes two type II  $\beta$ -turns (N7-T10 and R12-E15) at the tip of the Loop I, as described earlier.

### **Sequence homology of candoxin with other neurotoxins**

Snake toxins are short proteins consisting of 60-70 amino acid residues. To date, primary sequences of over 200 toxins from different species have been determined. They all show a similar tertiary fold, namely a predominantly anti-parallel  $\beta$ -sheet structure with three loops and a highly conserved core, which is rich in disulfide bridges. Yet, they have widely different biological functions. For example, long and short  $\alpha$ -neurotoxins bind with high affinity to the nicotinic acetylcholine receptor, thus blocking cholinergic transmission<sup>30</sup>. Cytotoxins have the ability to lyse a number of different cells<sup>31</sup>, to inhibit protein aggregation<sup>32</sup>, and, fasciculins possess a strong anticholinesterasic action<sup>33</sup>. Such distinct biological properties of toxins belonging to different groups are reflected in subtle differences in the composition of the amino

acid residues, size of loops, inter-loop interactions and the flexibility of individual loops. Thus, understanding differences in the primary sequences and the 3D structure of specific toxins should provide an insight into the structure-function relationships of three-finger toxins.

Unlike long chain- and  $\kappa$ -neurotoxins, which have an additional (fifth) disulfide bridge in the second loop (Loop II) (e.g.  $\alpha$ -bungarotoxin and  $\kappa$ -toxin)<sup>34</sup> (1KBA), candoxin has it in the first loop. To date, only two proteins with this disulfide bridge scaffold have been structurally characterized. They are bucanadin (pdb: 1F94 (X-ray) and 1IJC (NMR)), another toxin from *Bungarus candidus* venom<sup>35-36</sup> and Cd59 (pdb: 1CDQ), which is a cell surface protein<sup>37</sup>.

As evident from Figure 1, Candoxin is highly homologous to a long-chain neurotoxin homologue from *Bungarus multicinctus* venom, with 98% sequence identity. They differ only at the seventh location in the primary sequence, with a His in the long-chain neurotoxin homologue, instead of Asn present in candoxin. However, to date, little is known about this long-chain neurotoxin homologue. Other toxins listed in Figure 1 show a similarity of only 41-47% and an identity of 38-40% with candoxin. This degree of sequence homology primarily arises from the conserved cysteines, which constitutes 10-15% of the sequence identity, forming disulfide bridges and some additional residues forming the central core in which the functional residues are located.

As the N-terminal amino-acid residues in candoxin are the most non-conserved residues, they are unique to candoxin (Figure 1). In the entire N-terminal polypeptide stretch from M1 to V18, C3 is the only conserved residue. The non-conserved residues C6 and C11 show sequence homology at their respective locations, only to bucanadin, is another toxin from *Bugarus Candidus*. In spite of such non-conserved composition, this N-terminal stretch adopts a standard two stranded  $\beta$ -sheet conformation with individual  $\beta$ -strands K2-I5 ( $\beta_1$ ) and E15-V18 ( $\beta_2$ ). The linker region between the two strands adopts two type II  $\beta$ -turns (N7-T10 and R12-G14), and has a disulfide bridge between C6 and C11. Such a disulfide bridge shortens Loop I. The individual conformations and lengths of Loop II and III in candoxin are remarkably similar to those in erabutoxin b (pdb: 1FRA) and cobrotoxin (pdb: 1COD). Approximately half of the residues in these two loops (C26, W31, R32, E33, R35, G36, T37, I39, E40, R41, G42, C43, T46, C47, P48, G54, L55, V57, C59 and C60) are conserved. However, a highly conserved Tyr (Y25 in erabutoxin b, cobrotoxin and bucanadin) at the origin of the central loop, which participates in an extensive hydrophobic interaction with the adjacent residues is replaced by a Phe residue (F27) in candoxin. Though candoxin lags the conserved Tyr at the origin of the central loop, it still posses the most popular 3-dimensional array of residues (W31, R35 and K49), which is reported to be critical for curaremimetic neurotoxicity. Whereas erabutoxin b possesses fourteen residues identified as being critical for curaremimetic neurotoxicity, only six of them are present in candoxin (W31, R35, G36, G40, G42 and P48). Out of the remaining eight locations, three are conservatively substituted i.e. V57 for a Leu, L55 for an Ile and E33 for an Asp. The presence of the kink at the tip caused by the C6-C11 disulphide bridge in candoxin, shortens the Loop and may isolate this portion of the molecule and prevent its interference with the rest of the molecule, thereby facilitating specific interaction with receptor/acceptor protein. Taken together, these

observations suggest that the candoxin may bind to acetylcholine receptor with lower affinity compared to other neurotoxins and hence could account for its relatively lower toxicity.

## Experimental Section

**Materials.** Lyophilized *Bungarus candidus* venom was obtained from Venom Supplies (Tanunda, SA, Australia). Prepacked columns, Superdex 30 and Nucleosil C18 were purchased from Pharmacia Biotech (Uppsala, Sweden) and Phenomenex (Torrance, CA, USA). Reagents for N-terminal sequencing were from Applied Biosystem (Foster City, CA, USA), chromatographic reagents acetonitrile and TFA were from Fisher Scientific (Fair Lawn, NJ, USA) and Fluka (Buchs, Switzerland), respectively. Other chemicals used were of analytical status.

Candoxin was isolated and purified from *Bungarus candidus* venom by a combination of gel filtration and reverse-phase HPLC. The purified protein (MW 7334.69) was found to be homogenous by electrospray ionization, MALDI-TOF mass spectroscopy and capillary electrophoresis. About 10 mg of candoxin was dissolved in 0.6 ml to get approximately 4.5 mM solution in 99.9%  $^2\text{H}_2\text{O}$  (pH=3.0). For experiments in  $^2\text{H}_2\text{O}$ , the sample was lyophilized thrice from  $^2\text{H}_2\text{O}$  to deprotonate all the exchangeable protons, prior to its dissolution in 0.6 ml of 99.9%  $^2\text{H}_2\text{O}$ .

### NMR

NMR experiments were carried out on a Varian Unity<sup>+</sup> 600 MHz NMR spectrometer equipped with pulsed field gradient unit and triple resonance probe with actively shielded Z-gradients. 2D experiments recorded in  $^2\text{H}_2\text{O}$  include two and three quantum-filtered correlation spectroscopy (2QF-COSY, 3QF-COSY)<sup>39,40</sup>, clean total correlation spectroscopy (clean-TOCSY)<sup>41</sup> with a mixing time ( $\tau_m$ ) of 50 and 80 ms and nuclear Overhauser enhancement spectroscopy (NOESY)<sup>42</sup> with a  $\tau_m$  of 50, 75, 100 and 200 ms. Spectra in 90%  $\text{H}_2\text{O}$  + 10%  $^2\text{H}_2\text{O}$  include Watergate-NOESY<sup>43</sup> with a  $\tau_m$  of 50, 75 and 200 ms and clean-TOCSY with a  $\tau_m$  of 50 and 80 ms. Data were acquired with spectral widths of 7200 Hz for  $^2\text{H}_2\text{O}$  samples and 8000 Hz for  $\text{H}_2\text{O}$  samples (Table 2). Deuterium exchange studies were carried out by recording a series of 1D  $^1\text{H}$  NMR spectra followed by a series of 2D TOCSY spectra immediately after the lyophilized, fully protonated toxin was dissolved in  $^2\text{H}_2\text{O}$ . Data transformation and processing were done on a Silicon Graphics workstation (R10000 based Indigo II Solid Impact Graphics) using Felix 97 (MSI, San Diego, USA) and on a SUN workstation using VNMR. During the course of NMR experiments, the sample was found stable and did not change or degrade with time. This was tested by recording 1D  $^1\text{H}$  NMR spectra before and after recording 2D NMR spectra.  $^1\text{H}$  chemical shift calibrations were carried out with respect to the methyl signal (at 0.0 p.p.m.) of 3-(trimethylsilyl)[3,3,2,2- $^2\text{H}$ ] propionate- $d_4$  (TSP), used as an external reference. Carbon chemical shifts were calibrated indirectly relative to DSS.

### NMR based constraints for structure calculations

**(a) nOe derived distance constraints.** Upper bounds for nOe distance restraints were derived from the  $^1\text{H}$  watergate NOESY spectra recorded with 100 ms mixing time for exchangeable protons and 75 ms mixing time for non-exchangeable protons. Interproton distances were classified into three categories based on the nOe cross peak integral volumes<sup>44,45</sup>). For weak cross peaks, the interproton distance was set between 1.8-5.0 Å. For medium and strong cross peaks interproton distances were restrained to 1.8-3.6 Å and 1.8-2.8 Å, respectively. In all cases, the lower distance limit (1.8 Å) is chosen close to the 'hard contact' distance.

**(b) Hydrogen bond constraints.** Hydrogen bond constraints were added only for residues which were involved either in  $\beta$ -sheet or  $\beta$ -turn, as revealed by  $^3\text{J}(\text{H}^{\text{N}}-\text{H}^{\alpha})$  values, slow  $\text{H}^{\text{N}}$  exchange rates and  $^1\text{H}^{\alpha}$ ,  $^{13}\text{C}^{\alpha}$  and  $^{13}\text{C}^{\beta}$  chemical shift indices (CSI)<sup>21-23, 46-48</sup>. For each hydrogen bond, N-O and  $\text{H}^{\text{N}}-\text{O}$  distances were constrained to 2.8-3.1 Å and 1.8-2.1 Å, respectively.

**(c) Disulfide bridge constraints.** The 5 Cys-Cys disulfide bridges present in candoxin, as established by the observation of long-range  $\text{H}^{\beta}-\text{H}^{\beta}$  nOes in the NOESY spectrum, were included as covalent bond restraints.

**(d) Dihedral angle constraints.** Dihedral angle constraints were used only in regions where the secondary structure could be established from well-defined nOes. The  $\phi$  angle in the case of  $\beta$ -strands was restricted to a range of -80 to -160. For type I  $\beta$ -turns,  $\phi_2$ ,  $\phi_3$ ,  $\psi_2$  and  $\psi_3$  were restricted to 0 to -120, -30 to -150, 0 to -90 and -150 to 30, respectively. For type II  $\beta$ -turns  $\phi_2$ ,  $\phi_3$ ,  $\psi_2$  and  $\psi_3$  were restricted to 0 to -120, 30 to 150, 60 to 180 and -150 to 30, respectively.

### 3D Structural calculations

3D structure calculations were achieved with an XPLOR 3.1 package<sup>26,49</sup> using a combination of distance geometry and simulated annealing protocols, on a Dec Alpha system. The final energy was calculated as a function of bond lengths, bond angles and torsion angles. Square well potentials were used for nOe and torsion angle restraints. Initially 200 structures were generated by random simulated annealing protocol. Simulated annealing at 1000K generated random co-ordinates, and regularized co-ordinates were obtained by distance geometry calculation. These structures were refined by 9 ps dynamics at 2000K. System was cooled to 100K over a time period of 18 ps. The dihedral force constant was taken as 5 kcal/M.deg<sup>2</sup> for the first 9 ps, which was increased to 200 kcal/M.deg<sup>2</sup> there after. After cooling, 200 steps of Powell minimization were performed. The second step in refinement was performed by the 'refine' protocol in XPLOR (35). This included a 6 ps high temperature dynamics and 24 ps of cooling with dihedral angle force constants held at 200 kcal/M.deg<sup>2</sup>. The final energy minimization included 200 steps of Powell minimization. The structures were energy minimized using CHARMM<sup>50</sup> force field and individually analyzed with Insight II (MSI, San Diego, USA) and MOLMOL<sup>25</sup> on an R 12000 based Octane SGI workstation. Structures with lowest energy and good non-bonded contacts were selected. An average structure of the best 19 structures was obtained and the RMSD deviations from this structure were calculated.

## Acknowledgements

The facilities provided by the National Facility for High Field NMR, supported by Department of Science and Technology (DST), Department of Biotechnology (DBT), Council of Scientific and Industrial Research (CSIR), and Department of Atomic Energy (DAE), are gratefully acknowledged. It is a great pleasure for two of us (K.V.R. Chary and G. Govil) to contribute this article for the special volume in honor of Prof. P.T. Narasimhan, whose pioneering work in the areas of NMR and theoretical chemistry has been a source of inspiration to us.

## References

1. Rang, H. P. *Q. Rev. Biophys.* **1975**, *7*, 283.
2. Bougis, P. E.; Marchot, P.; Rochat, H. *Biochemistry* **1986**, *25*, 7235.
3. Harvey, A. L.; Hider, R. C.; Khader, F. *Biochem. Biophys. Acta* **1983**, *728*, 215.
4. Rees, B.; Bilwes, A. *Chem. Res. Toxicol.* **1993**, *6*, 385.
5. Love, R. A.; Stroud R. M. *Prot. Eng.* **1986**, *1*, 37.
6. Brown, L. R.; Wuthrich, K. *J. Mol. Biol.* **1992**, *227*, 1118.
7. Peng, S. S.; Kumar, T. K. S.; Jayaraman, G.; Chang, C. C.; Yu, C. *J. Biol. Chem.* **1997**, *272*, 7817.
8. Zinn-Justin, S. Z.; Roumestand, C.; Gilquin, B.; Bontems, F.; Menez, A.; Toma, F. *Biochemistry* **1992**, *31*, 11335.
9. Yu, C.; Bhaskaran, R.; Yang, C. C. *J. Toxicol.-Toxin Reviews* **1994**, *13*, 291.
10. Grant, G. A.; Chiappinelli, V. A. *Biochemistry* **1985**, *24*, 1532.
11. Connolly P. J.; Stern A. S.; Hoch J. C. *Biochemistry* **1996**, *35*, 418.
12. Kumar, T. K. S.; Jayaraman, G.; Lee, C. S.; Arunkumar, A. I.; Sivaraman, T.; Samuel, D.; Yu, C. *J. Biomol. Struct. & Dyna.* **1997**, *15*, 431.
13. De Weille, J. R.; Schweitz, H.; Maes, P.; Tartar, A.; Lazdunski, M. *Proc. Natl Acad. Sci. USA* **1991**, *88*, 2437.
14. Du, M. H. L.; Marchot, P.; Bougis, P. E.; Fontecilla-Camps, J. C.; *J. Biol. Chem.* **1992**, *267*, 22122.
15. Cervenansky, C.; Dajas, F.; Harvey, A. L.; Karlsson, E. In *Snake Toxins*, Harvey, A. L., Ed., Pergamon Press: New York., 1991; pp 303-332.
16. Sutcliffe, M. J.; Jaseja, M.; Hyde, E. I.; Lu, X.; Williams, J. A. *Nat. Struct. Biol.* **1994**, *1*, 802.
17. Tsetlin, V. *Eur. J. Biochem.* **1999**, *264*, 281.
18. Ohno, M.; Menez, R.; Ogawa, T.; Danse, J. M.; Shimohigashi, Y.; Fromen, C.; Ducancel, F.; Zinn-Justin, S.; Le Du M. H.; Boulain, J.-C.; Tamiya, T.; Menez, A. *Progress Nucl. Acid Res. Mol. Biol.* **1998**, *59*, 307.

19. Chang, L. S.; Lin, S. R.; Wang, J. J.; Hu, W. P.; Wu, B.; Huang, H. *Biochim. Biophys. Acta* **2000**, *1480*, 293.
20. Joubert, F. J.; Taljaard, N. *Hoppe-Seyler's Z. Physiol. Chem.* **1980**, *361*, 425.
21. Rani, P. V.; Chary, K. V. R.; Kini, M.; Govil, G. *Current Science*. **2000**, *79*, 219.
22. Rani, P. V.; Chary, K. V. R.; Kini, M.; Govil, G. *Mag. Res. in Chem.* **2001**, *39*, 577.
23. Nirathanan, S.; Charpantier, E.; Gopalakrishnakone, P.; Gwee, M. C. E.; Khoo, H. E.; Cheah, L. S.; Bertrand, D.; Kini, R. M. *J. Biol. Chem.* **2002**, *277*, 17811.
24. Atreya, H. S.; Sahu, S. C.; Chary, K. V. R.; Govil, G. *J. Biomol. NMR* **2000**, *17*, 125.
25. Wüthrich, K.; Billeter, M.; Braun W. *J. Mol. Biol.* **1983**, *169*, 949.
26. Brünger, A. T. *X-PLOR Software Manual, Version 3.1*, Yale University, New Haven, CT, 1992.
27. Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graph.* **1996**, *14*, 51.
28. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Cryst.* **1993**, *26*, 283.
29. Ramachandran G. N.; Sasisekharan, V. *Adv. Prot Chem.* **1968**, *23*, 283.
30. Harvey, A. L.; Hider, R. C.; Hodges S. J.; Joubert F. *J. Pharm. J.* **1984**, *82*, 709.
31. Changeux, J. -P. *Trends Pharmacol. Sci.* **1990**, *11*, 485.
32. Buisson, B.; Bertrand, D. *J. Physiol. Paris.* **1998**, *92*, 89.
33. Bilwes, A.; Rees, B.; Moras, D.; Menez, R.; Menez, A. *J. Mol. Biol.* **1994**, *239*, 122.
34. Cervenansky, C.; Dajas, F.; Harvey, A. L.; Karlsson, E. In *Snake Toxins*, 1991.
35. Utkin, Y. N.; Kukhtina, V. V.; Kryukova, E. V.; Chiodini, F.; Bertrand, D.; Methfessel, C.; Tsetlin, V. I. *J. Biol. Chem.* **2001**, *276*, 15810.
36. Kuhn, P.; Deacon, A. M. Comoso, S., Rajaseger, G., Kini, R. M., Uson, I., and Kolatkar, P. R., *Acta Crystallogr., Sect. D* **2000**, *56*, 1401.
37. Torres, A. M.; Kini, R. M.; Selvanayagam, N.; Kuchel, P. W. *Biochem. J.* **2001**, *360*, 539.
38. Fletcher, C. M.; Harrison, R. A.; Lachmann, P. J.; Neuhaus, D. *Structure* **1994**, *2*, 185.
39. Piantini, U.; Sörenson, O. W.; Ernst, R. *J. Am. Chem. Soc.* **1982**, *104*, 6800.
40. Muller, N.; Ernst, R. R.; Wüthrich, K. *J. Am. Chem. Soc.* **1996**, *108*, 6482.
41. Griesinger, C.; Otting, G.; Wüthrich, K.; Ernst, R. R. *J. Am. Chem. Soc.* **1988**, *110*, 7870.
42. Kumar, A.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *96*, 1156.
43. Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661.
44. Clore, G. M.; Gronenborn, A. M.; Brünger, A. T.; Karplus, M. *J. Mol. Biol.* **1985**, *186*, 435.
45. Williamson, M. P.; Havel, T. F.; Wüthrich, K. *J. Mol. Biol.* **1985**, *182*, 295.
46. Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490.
47. Wishart, D. S.; Sykes, B. D. *Methods in Enzymology* **1994**, *239*, 363.
48. Wishart, D. S.; Richards, F. M.; Sykes, B. D. *J. Mol. Biol.* **1991**, *222*, 311.
49. Nilges, M.; Kuszewski, J.; Brünger, A. T. *Computational Aspects of the Study of Biological Macromolecules by NMR*, Hoch, J. C. Ed., New York: Plenum Press, 1991.
50. Charm Englander, S. W.; Wand, A. J. *Biochemistry* **1987**, *26*, 5953.