Eur. J. Biochem. 217, 639–644 (1993)

Solution conformation of an antibacterial peptide, sarcotoxin IA, as determined by ^{1}H-NMR

Hideo IWAI, Yuki NAKAJIMA, Shunji NATORI, Yoji ARATA and Ichio SHIMADA
Faculty of Pharmaceutical Sciences, University of Tokyo, Japan

(Received April 19/August 9, 1993) – EJB 93 0573/2

The solution conformation of sarcotoxin IA, which is an antibacterial peptide isolated from *Sarcophaga peregrina* with a molecular mass of 4 kDa, was determined by NMR spectroscopy and hybrid distance geometry/dynamical simulated annealing calculations. On the basis of 227 experimental constraints, including 185 distance constraints obtained from NOE and 42 constraints associated with 21 hydrogen bonds, a total of 18 converged structures of sarcotoxin IA were obtained. The final 18 converged structures exhibit backbone-atomic root-mean-square differences about the averaged coordinate positions of 0.070 ± 0.027 nm for residues 3–23 and 0.040 ± 0.017 nm for residues 28–38. It has been indicated that sarcotoxin IA consists of two amphiphilic α-helical regions, i.e. helix I (Leu3–Gln23) and helix II (Ala28–Ala38), with a hinge region (Gly24–Ile27), which connects helix I and helix II. We conclude that these two amphiphilic helical segments of sarcotoxin IA are of importance for the expression of the antibacterial activity.

Humoral antibacterial proteins are known to be induced in some insects when they are immunized with dead or live bacteria [1–3]. In *Sarcophaga peregrina* (flesh fly) it has been indicated that an antibacterial peptide, sarcotoxin IA, is induced in the hemolymph of the larvae when the body wall is injured with a hypodermic needle [4, 5]. Sarcotoxin IA, which consists of 39 amino acid residues with an amidated C-terminus, is toxic to various Gram-positive and Gram-negative bacteria and has a molecular mass of 4 kDa [6]. The primary target of sarcotoxin IA has been found to be the cytoplasmic membrane of bacteria [7].

It has been found that the amino acid sequence of sarcotoxin IA is similar to that of cecropin B, which is the antibacterial peptide isolated from *Hyalophora cecropia* [8]. Sarcotoxin IA and cecropin B have been compared in terms of the antibacterial spectrum [9]. On the basis of the antibacterial activity of analogues of sarcotoxin IA and cecropin B, it has been suggested that the interaction of sarcotoxin IA with the bacterial membrane is different from that of cecropin B [9].

In the present paper, we report the solution conformation of sarcotoxin IA as determined by ^{1}H NMR and hybrid distance geometry/dynamical simulated annealing calculation. We will briefly discuss the relationship between the structure and the antibacterial activity of sarcotoxin IA.

MATERIALS AND METHODS

Sarcotoxin IA was synthesized chemically in a fully automated peptide synthesizer (Applied Biosystems 430A). NMR spectra were recorded on a Jeol JNM-GSX500 spectrometer operating at 500 MHz. For the NMR measurements, sarcotoxin IA was dissolved in CD_3OH at an approximate concentration of 1 mM.

CD spectra were recorded on a Jasco J-600 spectropolarimeter at 25°C. All two-dimensional spectra, i.e. double-quantum-filtered correlated spectroscopy (DQF-COSY) [10], two-dimensional homonuclear Hartmann-Hahn spectroscopy (HOHAHA) [11], and two-dimensional NOE spectroscopy (NOESY) [12], were recorded in the pure-phase mode [13]. In order to eliminate the base line distortion due to the electronic filter effect, all pulse sequences used in this study were modified according to the method of Davis [14]. A complete set of two-dimensional spectra was recorded at 30°C. HOHAHA spectra were recorded with mixing times of 45 ms and 60 ms. The radiofrequency pulses used in the HOHAHA experiments were generated by the observe channel (25 μs for 90° pulse) and anisotropic mixing was carried out with an MLEV-17 pulse sequence sandwiched between trim pulses of 2.5 ms. NOESY spectra were recorded with mixing times of 150 ms and 400 ms. 512 increments of 2 K data points were recorded with 64–128 transients, each of which yields after zero-filling, a spectrum with a digital resolution of 6.0 Hz. Suppression of undesirable t_2 ridges arising from the strong solvent resonance was achieved by linear base-line correction of the F_2 cross section prior to Fourier transformation in t_1. The Gauss function was used for the apodization function. The relaxation delay used was 1.2 s.

All calculations were carried out on a Silicon Graphics Iris 4D/25G workstation with the programs EMBOSS [15] and X-PLOR [16]. The structures were calculated by the hybrid distance geometry/dynamical simulated annealing method [17] with a few modifications [18]. The calculation strategy employed was divided into the following two steps. First, using the program EMBOSS, a set of geometrically correct substructures was generated by the metric matrix al-

Correspondence to I. Shimada, Faculty of Pharmaceutical Sciences, University of Tokyo, Bunkyo-ku, Hongo, Tokyo, Japan 113

Abbreviations. DQF-COSY, double-quantum-filtered correlated spectroscopy; HOHAHA, two-dimensional homonuclear Hartmann-Hahn spectroscopy; NOESY, two-dimensional NOE spectroscopy; RMSD, root-mean-square difference.

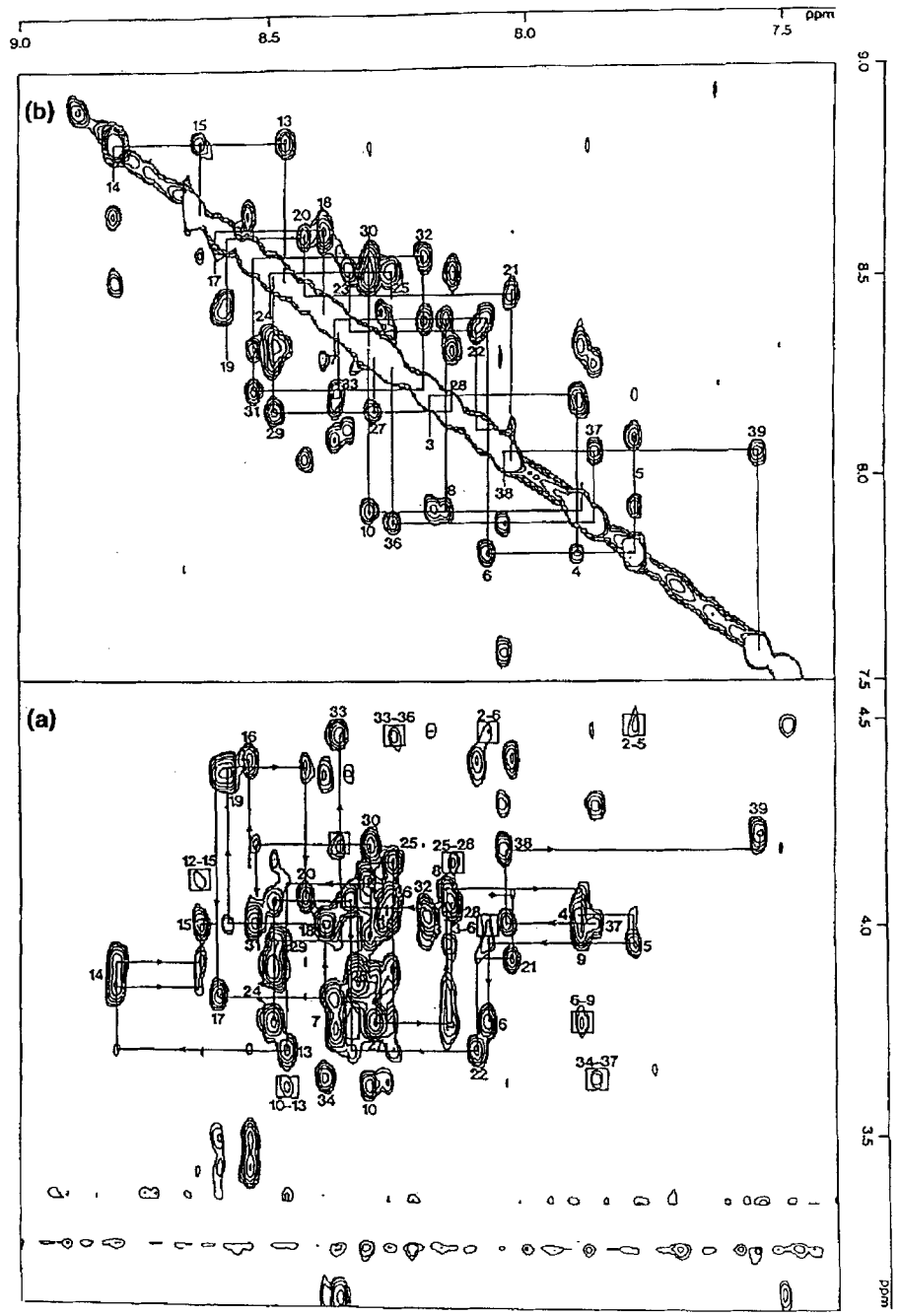

Fig. 1. NOESY spectra of sarcotoxin IA observed in CD_3OH at 30°C with sequential NOE connectivities. (a) CαH-NH region; (b) NH-NH region. The spectra were recorded with a mixing time of 400 ms. The cross peaks due to non-sequential $d\alpha N(i,i+3)$ connectivities are boxed.

gorithm. Second, starting from this set, a dynamical simulated annealing calculation was performed using the program X-PLOR.

RESULTS AND DISCUSSION

CD spectra of sarcotoxin IA were measured under various solution conditions. They indicated that sarcotoxin IA adopted an α-helical structure in methanol, 80% methanol/20% H_2O and in an aqueous lyso-phosphatidyl choline solution that mimics a hydrophobic membrane-like environment; in H_2O, it adopted random-coiled conformation (data not shown). Therefore, in order to investigate the solution conformation of sarcotoxin IA, two-dimensional spectra were recorded in CD_3OH, where it adopted an ordered structure. Sequence-specific resonance assignments were made according to the standard method for small proteins [19, 20].

Fig. 1 shows the CαH-NH fingerprint region (Fig. 1a) and the NH-NH region (Fig. 1b) of the NOESY spectrum with sequential $d\alpha N(i,i+1)$, $dNN(i,i+1)$ connectivities. It was possible to distinguish interresidue from intraresidue NOE cross peaks in the NOESY spectrum, by comparing the same region of the DQF-COSY and HOHAHA spectra. On the basis of the $d\alpha N(i,i+1)$, $dNN(i,i+1)$ NOE connectivit-

Table 1. ^{1}H chemical shifts of sarcotoxin IA at 30°C. Chemical shifts of ^{1}H are expressed relative to sodium 2,2-dimethyl-2-silapentane-5-sulfonate; n.d. = not detectable.

Residue	Chemical shift of			
	NH	CαH	CβH	others
	ppm			
Gly1	n.d.	n.d.		
Trp2	8.80	4.47	3.29, 3.29	2H 7.22, 4H 7.51, 5H 7.00, 6H 7.11, 7H 7.36
Leu3	8.20	4.02	1.54, 1.54	CγH 1.54, CδH$_3$ 1.54, 1.54
Lys4	7.91	4.05	1.83, 1.85	CγH 1.44, 1.51, CδH 1.83, 1.85, CεH 2.90, 2.90
Lys5	7.80	3.95	1.90, 1.95	CγH 1.40, 1.40, CδH 1.63, 1.63, CεH 2.86, 2.86
Ile6	8.08	3.78	1.95	CγH 1.65, 1.21, CγH$_3$ 0.92, CδH$_3$ 0.86
Gly7	8.38	3.76, 3.83		
Lys8	8.16	4.10	1.88, 1.99	CγH 1.63, 1.63, CδH 1.69, 1.69, CεH 2.89, 2.89
Lys9	7.90	4.00	2.03, 2.08	CγH 1.59, 1.59, CδH 1.68, 1.68, CεH 2.88, 2.88
Ile10	8.32	3.61	2.07	CγH 1.15, 1.83, CγH$_3$ 0.94, CδH$_3$ 0.86
Glu11	8.28	4.06	2.19, 2.30	CγH 2.40, 2.61
Arg12	8.32	4.10	2.00, 2.10	CγH 1.75, 1.87, CδH 3.26, 3.26 NεH 7.72
Val13	8.48	3.70	2.27	CγH$_3$ 1.01, 1.12
Gly14	8.82	3.87, 3.92		
Gln15	8.65	4.01	2.12, 2.16	CγH 2.40, 2.51, CONH$_2$ 6.77, 7.40
His16	8.55	4.40	3.43, 3.49	C2H 7.56, C4H 6.87
Thr17	8.61	3.84	4.36	CγH$_3$ 1.24
Arg18	8.40	4.00	1.84, 1.99	CγH 1.60, 1.60, CδH 3.20, 3.20, NεH 7.35
Asp19	8.59	4.36	2.77, 2.99	
Ala20	8.44	4.07	1.40	
Thr21	8.03	3.92	4.40	CγH$_3$ 1.22
Ile22	8.10	3.71	2.04	CγH 1.24, 1.79, CγH$_3$ 0.96, CδH$_3$ 0.88
Gln23	8.35	4.06	2.08, 2.20	CγH 2.33, 2.51, CONH$_2$ n.d., n.d.
Gly24	8.50	3.77, 3.90		
Leu25	8.27	4.17	1.59, 1.92	CγH 1.53, 1.53, CδH$_3$ 0.81, 0.89
Gly26	8.34	3.87, 3.87		
Ile27	8.30	3.77	2.04	CγH 1.19, 1.18, CγH$_3$ 0.96, CδH$_3$ 0.89
Ala28	8.15	4.07	1.55	
Gln29	8.50	3.87	2.16	CγH 2.30, 2.50, CONH$_2$ 6.57, 7.40
Gln30	8.31	4.21	2.17, 2.29	CγH 2.39, 2.52, CONH$_2$ 6.89, 7.52
Ala31	8.54	4.02	1.52	
Ala32	8.21	4.05	1.54	
Asn33	8.37	4.46	2.74, 3.14	CONH$_2$ 6.89, 7.52
Val34	8.41	3.64	2.24	CγH$_3$ 0.95, 1.12
Ala35	8.29	4.01	1.51	
Ala36	8.27	4.08	1.55	
Thr37	7.88	4.01	4.29	CγH$_3$ 1.29
Ala38	8.05	4.20	1.47	
Arg39	7.56	4.24	1.88, 1.98	CγH 1.68, 1.81, CδH 3.20, 3.20, NεH 7.37
CONH$_2$	7.11, 7.18			

ies, all of the proton resonances of sarcotoxin IA have been assigned in a sequential manner except for the resonances originating from Gly1 and the side chain amide group of Gln23. The resonance assignments of sarcotoxin IA in CD_3OH are summarized in Table 1.

A summary of the short-range ($|i-j| < 4$) NOEs involving the NH, CαH, and CβH protons is shown in Fig. 2 and provides the basis for a qualitative interpretation of the secondary structure of sarcotoxin IA. The backbone and side-chain NOE patterns indicate that two α-helices exist in sarcotoxin IA.

Interproton distance constraints were obtained from the NOESY spectra observed with mixing times of either 150 ms or 400 ms. Quantitative determination of the intensities of NOE cross peaks was based on the counting of the exponentially spaced contour levels. The distance constraints obtained by the NOESY experiments were calibrated by using the intensity of the NOE cross peak between two CαH protons originating from Gly24. Observed NOE data were classified into three distance ranges 0.18−0.27 nm, 0.18−0.33 nm (0.18−0.35 nm for distances involving NH protons), and 0.18−0.50 nm, corresponding to strong, medium and weak NOEs, respectively. Pseudo-atoms were used for the methyl proton or prochiral methylene protons [21]. Correcting factors for the use of pseudo-atoms were added to the distance constraints. In addition, 0.05 nm was added to the distance constraints involving methyl protons [22].

In order to determine the slowly exchangeable amide protons, HOHAHA spectrum of sarcotoxin IA observed in CD_3OD was measured: 21 amide proton resonances were still detectable after 24 h in CD_3OD. Fig. 2 summarizes the slowly exchangeable amide groups for sarcotoxin IA. We initially calculated the polypeptide fold by using the distance constraints obtained as described above. The positions of the hydrogen bonds have been determined on the basis of the preliminary polypeptide fold, along with the results of the

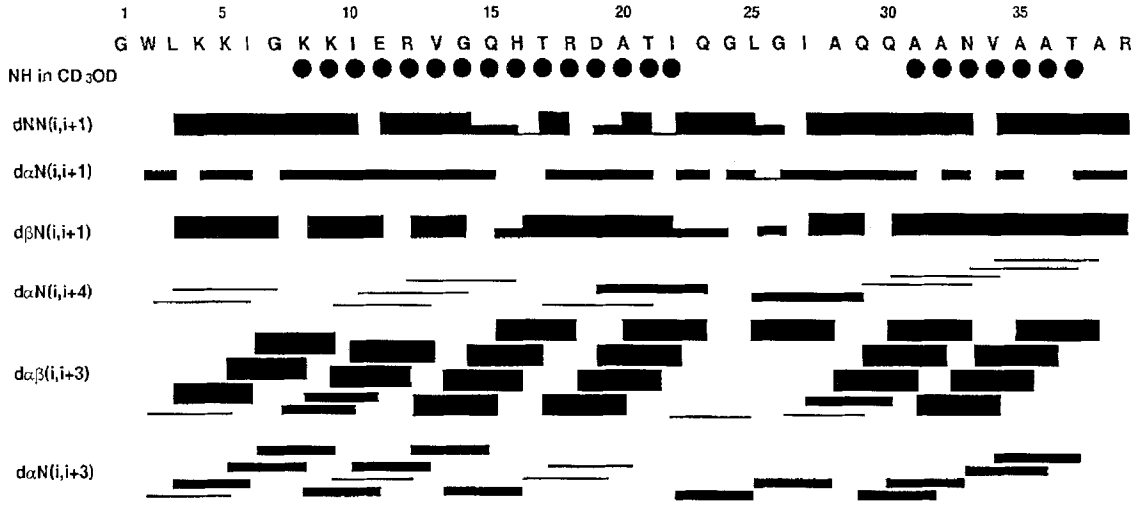

Fig. 2. The summary of all the short-range NOEs and NH exchange data. The NOEs are classified as strong, medium, and weak as indicated by the thickness of the lines. A closed circle indicates the residue of which the amide proton resonance is detectable after 12 h in CD_3OD.

Table 2. Structural statistics and atomic RMS differences of 18 converged structures. F_{NOE}, F_{tor}, and F_{repul} are the energies related to the NOE violations, the torsion angle violations, and the van der Waals repulsion term, respectively. The force constants of these terms used were the standard values [18]. E_{L-J} is the Lennard-Jones van der Waals energy calculated with the CHARMm empirical energy function [30]. E_{L-J} was not used in the dynamical simulated annealing calculations. The mean structures of residues 3–23 and 28–38 were obtained by best fitting the coordinates of the 18 converged structures to the residues.

Parameter	Value for 18 converged structures	
RMS deviations from experimental distance constraints	0.0086 ± 0.0005 nm	
F_{NOE}	356.2 ± 46.1 kJ mol^{-1}	
F_{tor}	23.80 ± 1.47 kJ mol^{-1}	
F_{repel}	277.4 ± 15.5 kJ mol^{-1}	
E_{L-J}	−477.2 ± 42.3 kJ mol^{-1}	
RMS deviations from idealized covalent geometry:		
bonds	0.0010 ± 0.00003 nm	
angles	2.589 ± 0.519 °	
impropers	1.287 ± 0.040 °	
	Value for	
	residues 3–23	residues 28–38
	nm	
RMSD for backbone (N, Cα, C)	0.070 ± 0.027	0.040 ± 0.017
RMSD for all heavy atoms	0.138 ± 0.025	0.087 ± 0.012

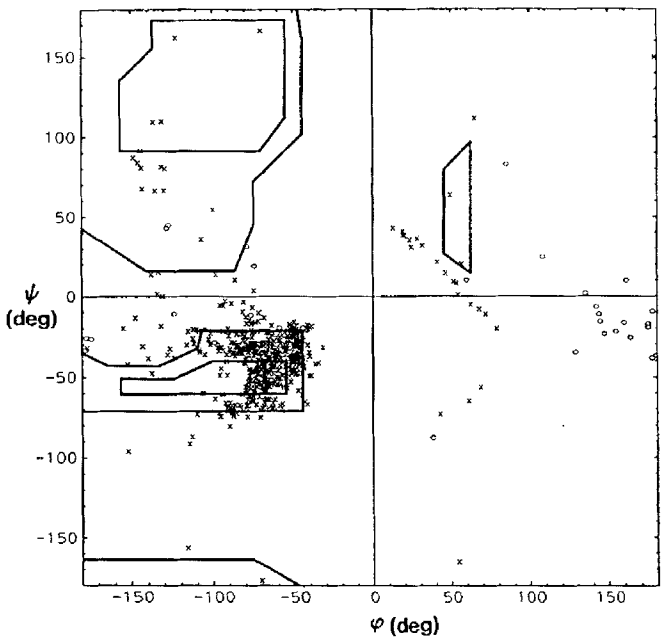

Fig. 3. Ramachandran-type plot for the 18 converged structures of sarcotoxin IA. (×) All residues except for glycine; (○), glycine residues.

amide proton exchange experiments in CD_3OD. As a result of this analysis, 21 NH($i+4$)–O(i) hydrogen bonds were identified. For each hydrogen bond, the NH($i+4$)–O(i) and N($i+4$)–O(i) distances were restrained to distance ranges of 0.17–0.23 nm and 0.24–0.33 nm, respectively. The final structures were calculated on the basis of 227 experimental constraints, which comprised 185 interproton distance constraints obtained by the NOESY experiments and 42 distance constraints related to hydrogen bonds in the α-helices. The interproton distance constraints consisted of 100 sequential ($|i-j| = 1$), 84 short-range ($1 < |i-j| < 5$), and one long-range ($|i-j| > 5$) interresidue NOEs. No torsion angle constraints were used for the structure calculations.

The dynamical simulated annealing calculations were performed by using 57 geometrically correct structures generated by the metric matrix algorithm. Of the 57 solutions obtained by dynamical simulated annealing calculations, we selected 18 solutions that had distance violations smaller than 0.07 nm.

Structural statistics for the 18 converged structures and atomic root-mean-square differences (RMSDs) are given in Table 2. The deviations from idealized covalent geometry were small and the Lennard-Jones van der Waals energy were large and negative, indicating that no distortions and bad

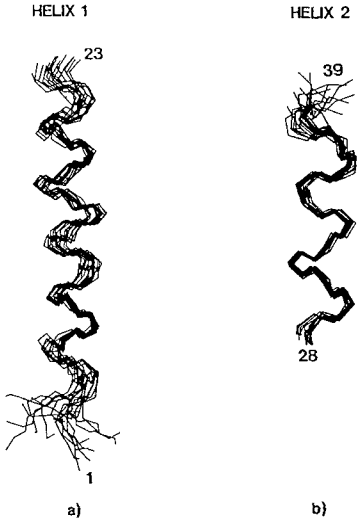

Fig. 4. Superpositions of the 18 converged structures of sarcotoxin IA for two selected segments. These are the results of the best fit of C, N, and Cα atoms for residues 3–23 (a) and for residues 28–38 (b). Only the backbones are shown.

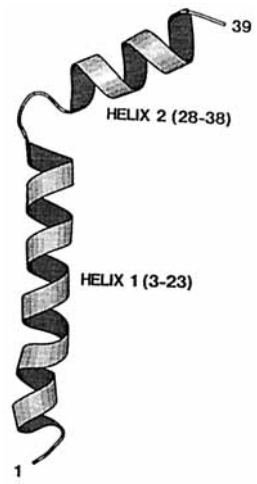

Fig. 5. Schematic ribbon drawing of sarcotoxin IA. This figure is drawn by the program MOLSCRIPT [29]. It should be noticed that the orientation of helices I and II has not been defined (see text).

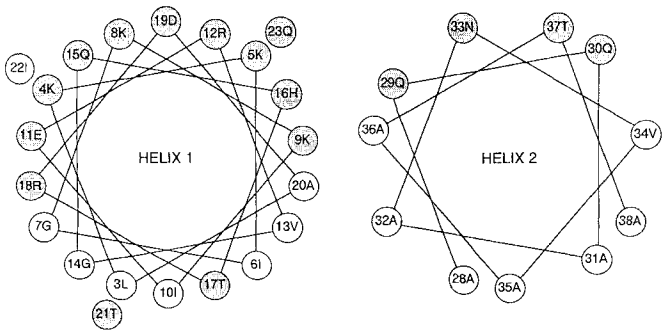

Fig. 6. Wheel projection of helix I (residues 3–23) and helix II (residues 28–38) of sarcotoxin IA. Open circles indicate hydrophobic amino acids and hatched circles indicate hydrophilic amino acids.

contacts exist in the converged structures. In the Ramachandran-type plot (Fig. 3), the dihedral backbone angles for the final 18 converged structures fall either in the helical region or in the generally allowed regions.

The structures of the backbone of residues 3–23 and of residues 28–38 are well defined. The region from Gly24 to Ile27 is less well defined than other parts of sarcotoxin IA. This is due to the lack of NOE data for these residues.

The final 18 converged structures exhibit backbone atomic RMSDs about the averaged coordinate positions of 0.070 ± 0.027 nm for residues 3–23 and 0.040 ± 0.017 nm for residues 28–38. Fig. 4a and 4b shows that the best-fit superposition of the backbone atoms of residues 3–23 and residue 28–38 of the 18 converged structures of sarcotoxin IA, respectively. Analysis of the 18 converged structures indicates that sarcotoxin IA consists of two α-helical regions, i.e. helix I (Leu3–Gln23) and helix II (Ala28–Ala38), which are connected by a hinge region (Gly24–Ile27). Schematic ribbon drawing of sarcotoxin IA is shown in Fig. 5.

The importance of amphiphilic helices has been discussed in a number of antibacterial peptides [23–27]. On the basis of the NMR data and the structure calculation as described above, it has been found that sarcotoxin IA also consists of two amphiphilic helices with the hydrophilic and hydrophobic faces (Fig. 6). The lengths of helix I and helix II are approximately 300 nm and 150 nm, respectively. The length of helix I is quite similar to the thickness of the hydrophobic core of phospholipid membrane, suggesting that helix I might penetrate into the bacterial membrane.

Cecropins A and B are potent antibacterial peptides isolated from the giant silk moth *Hyalophora cecropia* [28]. Fig. 7 compares the amino acid sequences of sarcotoxin IA, cecropin A, and cecropin B. The antibacterial activities of sarcotoxin IA, cecropin B, and the analogues of sarcotoxin IA and cecropin B have been studied by Boman and coworkers [9]. Boman's group has synthesized chemically an analogue of sarcotoxin IA with a deletion of Leu3 and Lys4. The amino acid sequence of the N-terminal region of the deletion analogue is similar to that of cecropin B (Fig. 7). However, it has been demonstrated that the deletion analogue reduces the antibacterial activity significantly. In contrast, it has also been demonstrated that an analogue of cecropin B with an insertion of the Leu-Lys dipeptide between Trp2 and Lys3 retains full antibacterial activity [9].

Although the solution conformation of cecropin B has not been determined, that of cecropin A has been determined by ^{1}H NMR [27]. It was concluded that two helical regions from Phe5 to Lys21 and from Pro24 to Lys37 exist in cecropin A. From the comparison between the primary structures of cecropin A and cecropin B, it is suggested that the conformation of the N-terminal region of cecropin B is similar to that of cecropin A.

The difference in antibacterial behavior between the two analogues may be explained as follows. As for the insertion analog of cecropin B, the Leu-Lys dipeptide is inserted into the disordered region of cecropin B. It should be noticed that the region from Gly1 to Val4 takes a disordered structure in cecropin A. It is suggested that the insertion of the Leu-Lys dipeptide between Trp2 and Lys3 dose not change the stability of the N-terminal helix of cecropin B (Fig. 7). Therefore, we conclude that the insertion analogue of cecropin B shows the same activity as the parent compound. In the case of the deletion analogue of sarcotoxin IA, it is possible that the deletion of Leu2 and Lys3 of sarcotoxin IA, which are

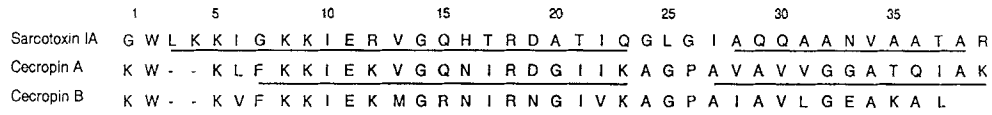

Fig. 7. Amino acid sequences for sarcotoxin IA, cecropin A and cecropin B. Location of the helical regions of sarcotoxin IA and cecropin A [26] are indicated by underlining.

located in helix I, reduces the stability of helix I, leading to the loss of the antibacterial activity (Fig. 7).

REFERENCES

1. Whitcomb, R. F., Shapiro, M. & Granados, R. P. (1974) in: *The physiology of Insectia*, 2nd edn, vol. 5, pp. 447–536, Academic Press, New York.
2. Boman, H. G. (1981) in *Microbial control of pests and plant disease* (Bruges, H. G., ed.) pp. 769–784, Academic Press, New York.
3. Chadwick, J. S., Deverno, P. J., Chung, K. L. & Aston, W. P. (1982) *Dev. Comp. Immunol. 6*, 433–440.
4. Natori, S. (1977) *J. Insect Physiol. 23*, 1169–1173.
5. Okada, M. & Natori, S. (1983) *Biochem. J. 211*, 727–734.
6. Okada, M. & Natori, S. (1985) *J. Biol. Chem. 260*, 7174–7177.
7. Okada, M. & Natori, S. (1984) *Biochem. J. 222*, 119–224.
8. Boman, H. G. (1991) *Cell 65*, 205–207.
9. Li, Z.-Q., Merrifiled, R. B., Boman, A. & Boman, H. G. (1988) *FEBS Lett. 231*, 299–302.
10. Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R. & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
11. Bax, A. & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
12. Jeener, J., Meier, B. N., Bachmann, P. & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
13. State, D. J., Haberkorn, R. A. & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
14. Davis, D. G. (1989) *J. Magn. Reson. 81*, 603–607.
15. Nakai, T., Kidera, A. & Nakamura, H. (1993) *J. Biomol. NMR 3*, 19–40.
16. Brünger, A. T., Clore, G. M. & Karplus, M. (1987) *Protein Eng. 1*, 399–406.
17. Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317–324.
18. Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M. & Gronenborn, A. M. (1986) *EMBO J. 5*, 2729–2735.
19. Wüthrich, K., Wider, G., Wagner, G. & Brown, W. (1982) *J. Mol. Biol. 155*, 311–319.
20. Gibbons, W. A., Crepaux, D., Delayre, J., Dunand, J. J., Hajdukovic, G. & Wyssbrod, H. R. (1975) in *Peptides: chemistry, structure, biology* (Walter, R. & Meienhofer, J., eds) pp. 127–137, Ann Arbor Science, Ann Arbor MI.
21. Wüthrich, K. (1986) in *NMR of proteins and nucleic acids*, pp. 130–161, John Wiley & Sons, New York.
22. Clore, G. M., Gronenborn, A. M., Nigles, M. & Ryan, C. A. (1987) *Biochemistry 26*, 8012–8023.
23. Fukushima, D., Kupferberg, J. P., Yokoyama, S., Kroon, D. J., Kaiser, E. T. & Kezdy, F. J. (1979) *J. Am. Chem. Soc. 101*, 3703–3704.
24. DeGrado, W. F., Kezdy, F. J. & Kaiser, E. T. (1981) *J. Am. Chem. Soc. 103*, 679–681.
25. Lee, K. H., Filton, J. E. & Wüthrich, K. (1987) *Biochim. Biophys. Acta. 911*, 144–153.
26. Marion, D., Zasloff, M. & Bax, A. (1988) *FEBS Lett. 227*, 21–26.
27. Holak, T. A., Engström, Å., Kraulis, P. J., Lindeberg, G., Bennich, H., Jones, T. A., Gronenborn, A. M. & Clore, G. M. (1988) *Biochemistry 27*, 7620–7629.
28. Steiner, H., Hultmark, D., Engström, Å., Bennich, H. & Boman, H. G. (1981) *Nature 292*, 246–248.
29. Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946–950.
30. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., State, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem. 7*, 165–175.