Volume 269, number 2, 413-420 FEBS 08810 September 1990

^{1}H nuclear magnetic resonance study of the solution conformation of an antibacterial protein, sapecin

Hiroyuki Hanzawa[1,*], Ichio Shimada[1], Takashi Kuzuhara[1], Hiroto Komano[1], Daisuke Kohda[2], Fuyuhiko Inagaki[2], Shunji Natori[1] and Yoji Arata[1]

[1]*Faculty of Pharmaceutical Sciences, University of Tokyo, Bunkyo-ku, Hongo, Tokyo 113, Japan* and [2]*Department of Molecular Physiology, The Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Komagome, Tokyo 113, Japan*

Received 25 June 1990

The solution conformation of an antibacterial protein sapecin has been determined by ^{1}H nuclear magnetic resonance (NMR) and dynamical simulated annealing calculations. It has been shown that the polypeptide fold consists of one flexible loop (residues 4–12), one helix (residues 15–23), and two extended strands (residues 24–31 and 34–40). It was found that the tertiary structure of sapecin is completely different from that of rabbit neutrophil defensin NP-5, which is homologous to sapecin in the amino acid sequences and also has the antibacterial activity. The three-dimensional structure determination has revealed that a basic-residue rich region and the hydrophobic surface face each other on the surface of sapecin.

Antibacterial activity; Cardiolipin; Dynamical simulated annealing; Sapecin; Tertiary structure; 2D NMR

1. INTRODUCTION

Sapecin, which is an antibacterial protein isolated from the culture medium of an embryonic cell line of *Sarcophaga peregrina* (flesh fly), is toxic to various Gram-positive and Gram-negative bacteria [1,2]. Sapecin consists of 40 amino acid residues including three disulfide bonds [3,4]. It has been shown that the disulfide bonds are essential for its antibacterial activity [4].

Antibacterial proteins, which are homologous in the amino acid sequences to sapecin, have been reported. Phormicin was purified from the hemolymph of immunized larvae of dipteran insect *Phormia terranova* [5]. Defensins are a group of peptides of mammalian macrophages and neutrophils [6–8].

The solution conformation of rabbit neutrophil defensin NP-5 has been determined on the basis of ^{1}H NMR spectral data and distance geometry calculations [9,10]. NP-5 is different from sapecin in the position of the cysteine residues and in the combination of disulfide bonds [4,9,10]. In view of the common biological function shared by these two proteins, it is of interest to compare the tertiary structure of sapecin with that of NP-5. In the present paper, we report a two-dimensional ^{1}H NMR study of sapecin. The solution conformation of sapecin will be presented and compared with that reported for NP-5.

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

**Present address:* Analytical and Metabolic Research Laboratories, Sankyo Co., Ltd., Hiromachi, Shinagawa-ku, Tokyo 140, Japan

Abbreviations: DANTE, delays alternating with nutation for tailored excitation; DQF-COSY, double quantum filtered correlated spectroscopy; HOHAHA, homonuclear Hartmann-Hahn spectroscopy; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, 2-dimensional NOE spectroscopy

2. MATERIALS AND METHODS

Sapecin was isolated and purified as described previously [1]. NMR spectra were recorded on a JEOL JNM-GX500 spectrometer operating at 500 MHz. For the NMR measurements, sapecin was dissolved in CD_3OH at an approximate concentration of 3 mM. All two-dimensional spectra (DQF-COSY [11], HOHAHA [12] and NOESY [13]) were recorded in the pure-phase mode [14]. 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 [15]. A complete set of the two-dimensional spectra was recorded at 30°C.

HOHAHA spectra were recorded with mixing times of 45 msec and 60 ms. The rf pulses used in the HOHAHA experiments were generated by the observed channel (25 μs for the 90 pulse), and anisotropic mixing was carried out with the MLEV-17 pulse sequence sandwiched between 2.5-ms trim pulses. NOESY spectra were recorded with mixing times of 100 ms, 400 ms. The solvent resonance of CD_3OH was suppressed by selective irradiation with DANTE pulse [16] during the relaxation delay. 512 increments of 2K 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. The relaxation delay used was 1.2 s. All dynamical simulated annealing calculations were carried out on an IRIS-4D computer using the program X-PLOR. Calculation strategy employed was divided into the following two steps. First, a set of 'variable target function protocol' substructures was generated [17]. Second, by using this set as the initial structures, the dynamical simulated annealing

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

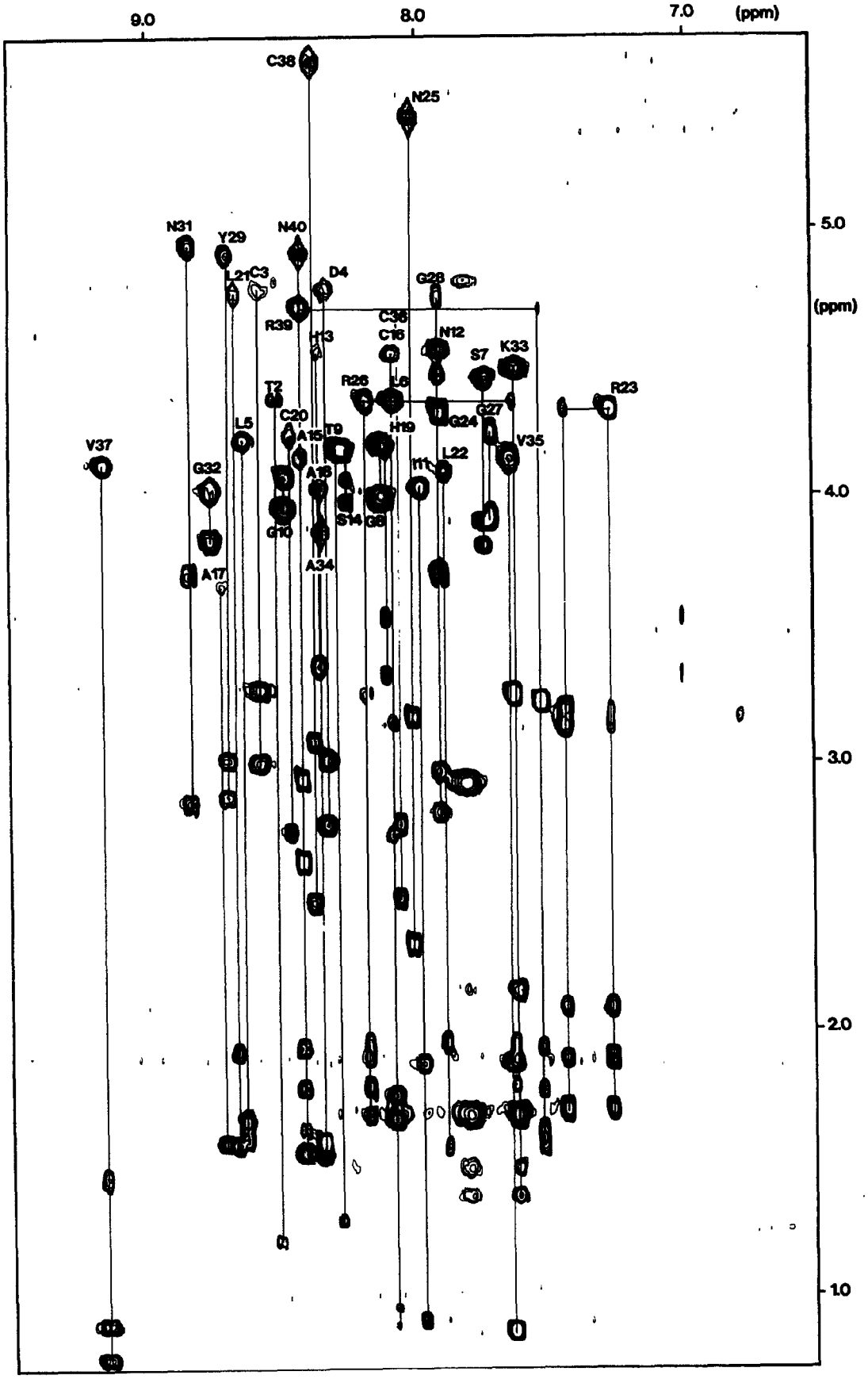

Fig. 1. Cross peaks observed at 30°C between the amide protons and the aliphatic protons of a 60-ms HOHAHA spectrum of sapecin. The cross peaks arise from direct and relayed connectivities and are indicated by continuous lines with the labels.

calculation was performed [18]. Details of the calculation will be described elsewhere [19].

3. RESULTS AND DISCUSSION

Two-dimensional spectra of sapecin were recorded in CD_3OH, which provides the protein with a less polar environment mimicking the hydrophobic region of receptor proteins and the phospholipid membrane. Sequence specific resonance assignments were made according to the method established by Wüthrich and coworkers [20]. Briefly, resonances were assigned to the spin systems of specific amino acid residues using DQF-COSY and 2D HOHAHA spectra. The spin systems were then aligned along the primary structure on the basis of the distance information obtained from the NOESY spectra.

The 2D-HOHAHA spectrum of sapecin in the region between the amide protons and the side chain protons is shown in Fig. 1. The spin systems of sapecin were identified prior to the sequential assignments. The magnetization from NεH of Arg was effectively transferred to CαH. The developed patterns from the back bone NH could be matched with those from NεH. Thus, the spin systems of Arg were identified. The spin systems of 6 Gly residues in sapecin were identified by the characteristic cross peak patterns between CαH-CαH′ due to a large spin-spin coupling. This was con-

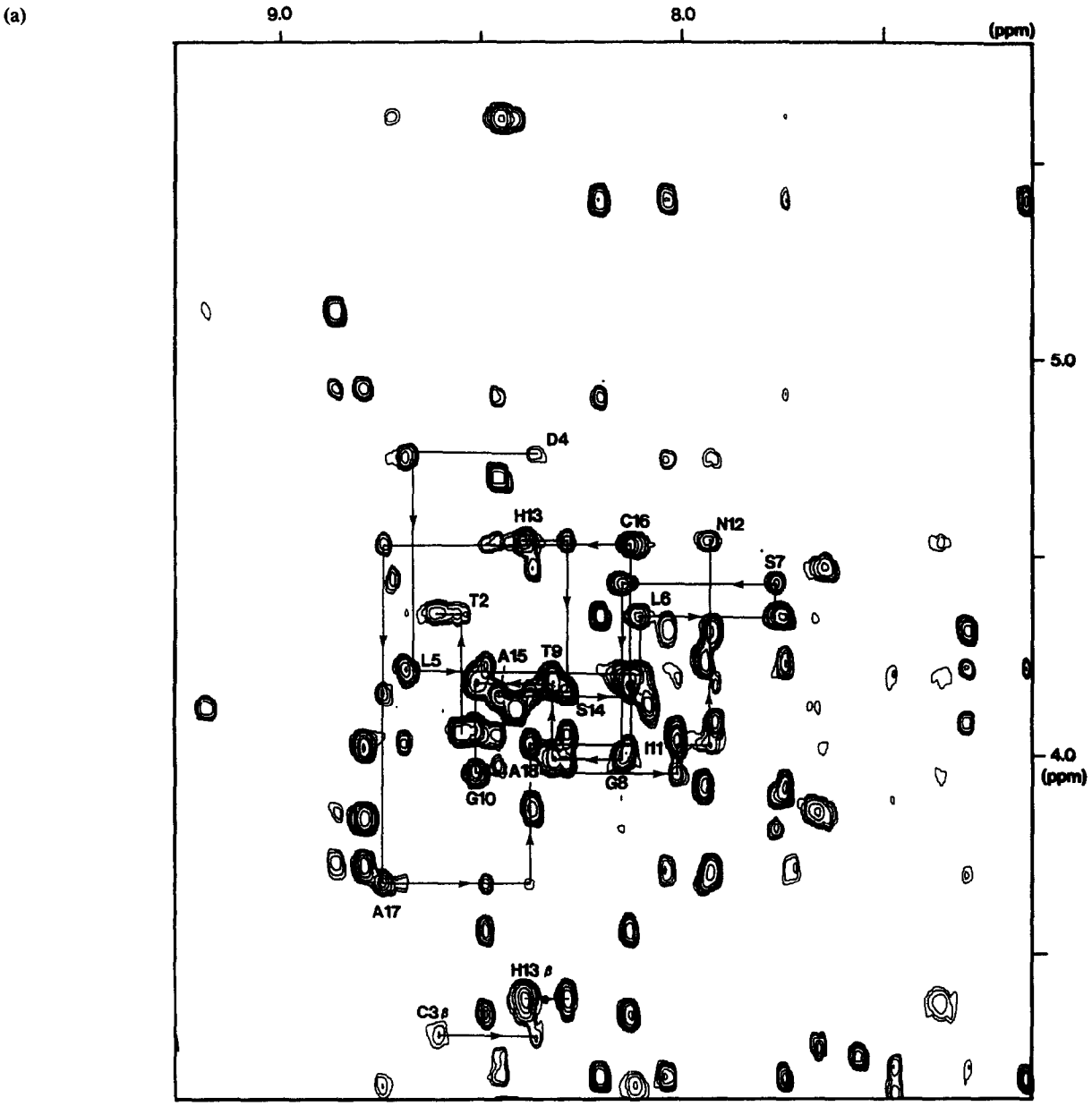

Fig. 2. dαN connectivities in the NOESY spectrum of sapecin observed at 30°C. Recorded with a mixing time of 400 ms. (a) A1 to A18; (b) A18 to N40. Instead of dαN, dβN was used for the connectivities of C3-D4.

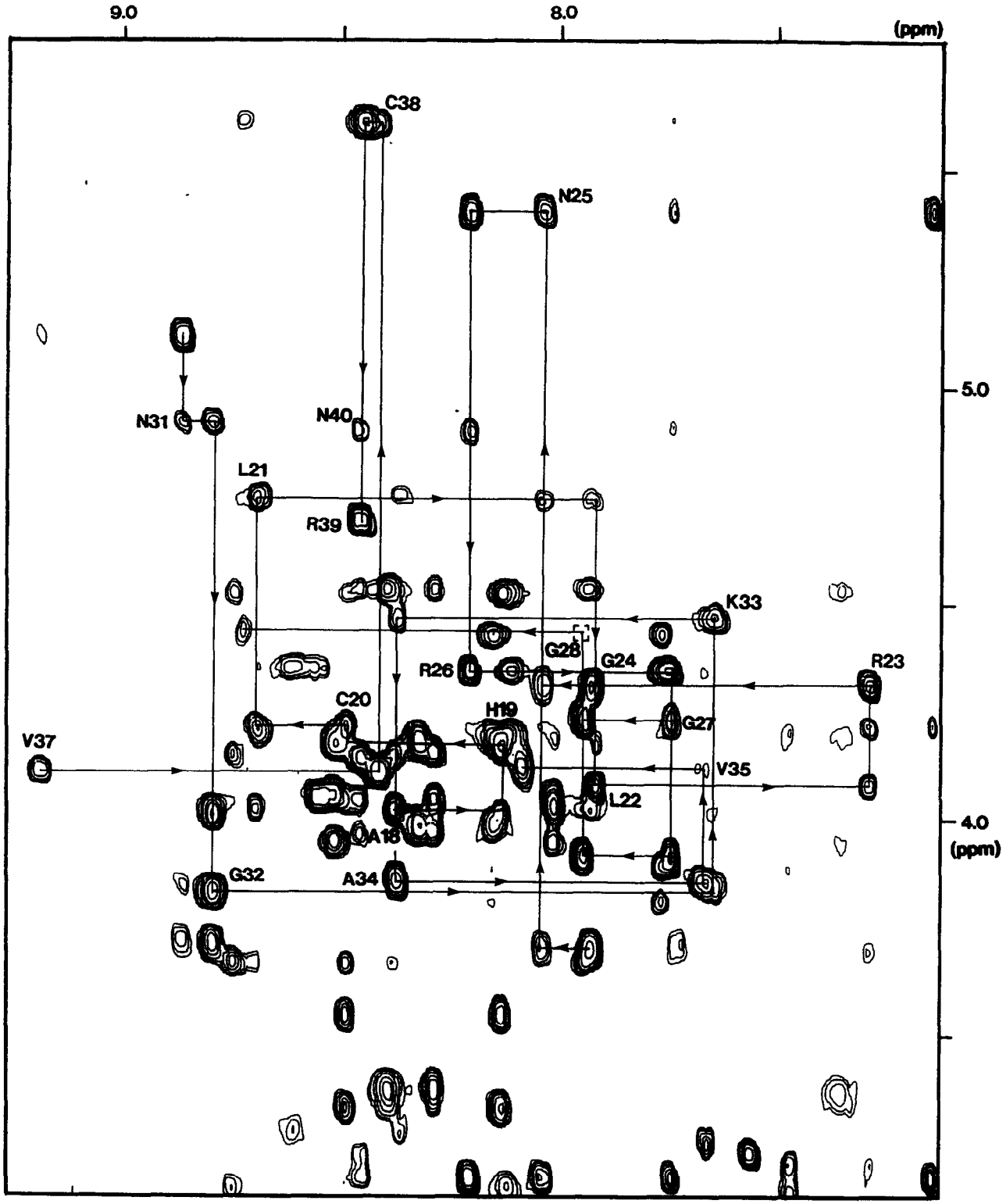

Fig. 2. (b)

firmed by two cross peaks between NH and CαH, CαH′ in the HOHAHA spectrum. In the region between CαH and the methyl protons of the DQF-COSY, the cross peaks CαH-CβH3 for Ala and CβH-CγH3 for Thr have strong intensities. The spin systems of Ala/Thr were readily identified. However, it was not possible to discriminate Thr from Ala due to the overlapping of the chemical shifts of CαH and CβH of Thr in this particular case. For Ile, the magnetization transfer from CαH through CβH to CγH3 and CδH3

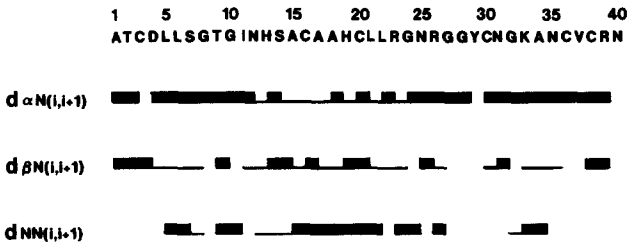

Fig. 3. The amino acid sequence of sapecin and the summary of all the short-range NOEs. The NOEs are classified as strong, medium, and weak by the height of the lines.

was observed in the region between CαH and the side chain protons in the HOHAHA spectrum. The Ile methyl proton resonances can be discriminated from those of Val and Leu by a characteristic pattern of the cross peaks in the methyl region of DQF-COSY spectrum. The spin system of Val and Leu were analyzed by using the pair of cross peaks between the δ methyl groups and the γ methene proton for Leu and between the γ methyl groups and the β methene proton for Val in the methyl region of the DQF-COSY spectrum. The chemical shifts of CβH and CγH of Leu were slightly different in this case. Thus, 4 spin systems of Leu and one of Val were identified. It was not possible to confirm the remaining two spin systems of Leu/Val due to the overlapping of the chemical shifts.

Sapecin has 16 AMX spin systems, i.e., one Tyr, two His, two Ser, 4 Asn, one Asp and 6 Cys. Identification of the Asn spin systems was established by observing NOE cross peaks between the β protons and the Nδ protons in the NOESY spectrum in H_2O. Identification of the Tyr and His spin systems was achieved by observing NOE cross peaks between the β protons and the aromatic δ protons. For the assignments of the Ser spin systems, we had to rely on the chemical shifts of β proton resonances and finally on the sequential assignments.

At this point in the assignment procedure, all of the 24 spin systems of Arg (3), Lys (1), Gly (6), Ile (1), Leu (3), Val (1), Asn (4), His (2) and Tyr (1) in sapecin have been identified and the two spin systems of Ser (2) were tentatively identified. The remaining 16 spin systems were grouped into [Cys(6),Asp(1)], [Ala(5),Thr(2)] and [Val(1),Leu(1)]. Since sapecin contains one of each Ile, Tyr and Lys, sequence specific assignments of Ile-11, Tyr-29 and Lys-33 have been automatically established.

Fig. 2a shows dαN and dβN connectivities from Ala-1 to Ala-18. The connectivity between Cys-3 and Asp-4 relied on dβN, instead of dαN due to the overlapping of CαH of C3 with the solvent signal. The connectivity between Asn-12 and His-13 relied on dNN, instead of dαN due to the degeneracy of the α proton of Asn-12 and His-13. Fig. 2b shows dαN connectivities from Ala-18 to Asn-40. It was observed that the connectivity between Arg-39 and Asn-40 existed at a low temperature of 20°C. Connectivity between Tyr-29 and Cys-30 cannot be identified because of the failure of observing the amide proton of Cys-30. This is the only part of the sequence where there is a break in the sequence connectivity. The NOE cross peaks due to dNN connectivities were sequentially connected (data not shown). The intensities of the cross peaks due to dNN from His-19 to Leu-21 and from Arg-23 to Asn-25 are significantly strong. Thus, by combining all the 2D NMR data, the assignments for sapecin were accomplished, and the results are shown in Table I. Fig. 3 summarizes the distance connectivities for sapecin.

The dynamical simulated annealing calculation was performed with distance constraints obtained from the NOESY spectra with mixing times of 100 ms and 400 ms. Quantitative determination of the cross-peak intensities was based on the counting of the exponentially spaced counter levels. For prochiral methylene protons, or the prochiral methyl groups of Val or Leu residues, we used pseudo atoms [21]. Observed NOE data were classified into three distance ranges, 1.8–2.5 Å, 1.8–3.5 Å and 1.5–5.0 Å, corresponding to strong, medium and weak NOEs, respectively. A total of 117 restraints were obtained. There were 27 long-range distances and 90 short-range distances (i.e. ; $|i-j| < 5$), in which intraresidue distances were not included. Additional nine restraints were included for the three disulfide bonds that are present in sapecin. For each disulfide bond,

Fig. 4. Stereo view of a superposition of 4 converged structures of sapecin. This is the result of best fit of C, N, O and Cα atoms for all residues excluding Asp-4-Asn-12 segment. Only the backbone is shown.

there are three distance restraints, Si-Sj, Si-Cjβ and Sj-Ciβ, whose target values were set to 2.02 ± 0.02, 2.99 ± 0.5 and 2.99 ± 0.5 Å, respectively [18].

We carried out the dynamical simulated annealing by 100 initial structures for sapecin. As a result of this calculation, 34 solutions, which had almost a similar global polypeptide fold, were obtained. However, some of them showed the large violation. Therefore, we selected 9 structures, which had the small violation up to the ninth lowest one. The average backbone atomic root mean square (r.m.s.) difference among the nine structures was 3.0 Å. There were no medium/long-range NOEs observed in the region from Asp-4 to Asn-12. When this segment was excluded in the calculation, an average backbone atomic r.m.s. difference of 1.5 Å was obtained. As for the first 3 residues, 4 long-range NOEs were observed; A1(β)-H19(ring), A1(β)-C36(β), T2(γ)-C36(NH), C3(NH)-A34(β). Due to these constraints, the first 3 residues were tied down. Fig. 4 shows a superposition of the backbone structures of the nine calculations. This superposition is a best fit of C, N, O and Cα atoms for all residues excluding the Asp-4-Asn-12 segment. The polypeptide fold can be described as one loop (residues 4–12), one helix (residues 15–23) and two extended strands (residues 24–31 and 34–40). As Figure 4 shows, the loop comprising residues Asp-4-Asn-12 is less well defined than other parts of the molecule. This is due to the lack of NOE data for these residues. We suggest that this loop is flexible in solution.

Table I

Chemical shifts of the proton resonances of sapecin observed at 30° C

Residue	NH	CαH	CβH	Others	
Ala-1	N.D.	4.02	1.24		
Thr-2	8.51	4.32	4.32	CγH 1.20	
Cys-3	8.56	4.74	3.25, 2.97		
Asp-4	8.33	4.73	3.00, 2.74		
Leu-5	8.65	4.18	1.63, 1.57	CγH 1.72	CδH 0.91, 0.85
Leu-6	8.08	4.33	1.70, 1.70	CγH 1.64	CδH 0.93, 0.86
Ser-7	7.73	4.41	3.86, 3.78		
Gly-8	8.12	3.97, 4.17			
Thr-9	8.29	4.15	4.15	CγH 1.27	
Gly-10	8.49	4.03, 3.93			
Ile-11	7.98	3.99	1.85	CγH 1.77	CδH 0.89, 0.94
Asn-12	7.90	4.51	2.95, 2.79	NδH 7.45, 7.32	
His-13	8.36	4.51	3.33, 3.33	C2 8.53 C4 7.33	
Ser-14	8.26	4.14	4.03, 3.94		
Ala-15	8.43	4.12	1.42		
Cys-16	8.09	4.50	3.13, 2.70		
Ala-17	8.71	3.65	1.56		
Ala-18	8.35	4.01	1.51		
His-19	8.10	4.16	3.51, 3.30	C2 8.42 C4 6.99	
Cys-20	8.46	4.21	2.72, 2.72		
Leu-21	8.66	4.71	1.90, 1.54	CγH 1.71	CδH 0.95, 0.90
Leu-22	7.88	4.06	1.94, 1.94	CγH 1.51	CδH 0.87, 0.87
Arg-23	7.26	4.29	2.06, 1.87	CγH 1.68, 1.68 NHε 7.43	CδH 3.13, 3.13
Gly-24	7.90	4.28, 3.68			
Asn-25	8.01	5.39	3.15, 2.29	NδH 7.11, 6.78	
Arg-26	8.17	4.33	1.93, 1.76	CγH 1.87, 1.65 NεH 7.62	CδH 3.63, 3.63
Gly-27	7.71	4.21, 3.90			
Gly-28	7.91	4.70, 4.41			
Tyr-29	8.70	4.88	2.98, 2.86	C26 6.76 C35 6.54	
Cys-30	N.D.	5.09	3.02, 2.77		
Asn-31	8.83	4.91	3.68, 2.82	NδH 7.69, 7.02	
Gly-32	8.75	4.00, 3.82			
Lys-33	7.62	4.45	2.11, 1.64	CγH 1.45, 1.35 CεH 2.91, 2.89	CδH 1.67, 1.65 NζH 7.81
Ala-34	8.34	3.84	1.57		
Val-35	7.64	4.11	1.85	CγH 0.83, 0.83	
Cys-36	8.05	4.80	2.75, 2.46		
Val-37	9.16	4.09	1.44	CγH 0.87, 0.74	
Cys-38	8.38	5.60	3.06, 2.45		
Arg-39	8.42	4.67	1.91, 1.75	CγH 1.55, 1.55 NεH 7.53	CδH 3.20, 3.20
Asn-40	8.43	4.90	2.91, 2.61	NδH 7.56, 7.11	

Chemical shifts expressed in ppm relative to DSS. N.D., not detectable.

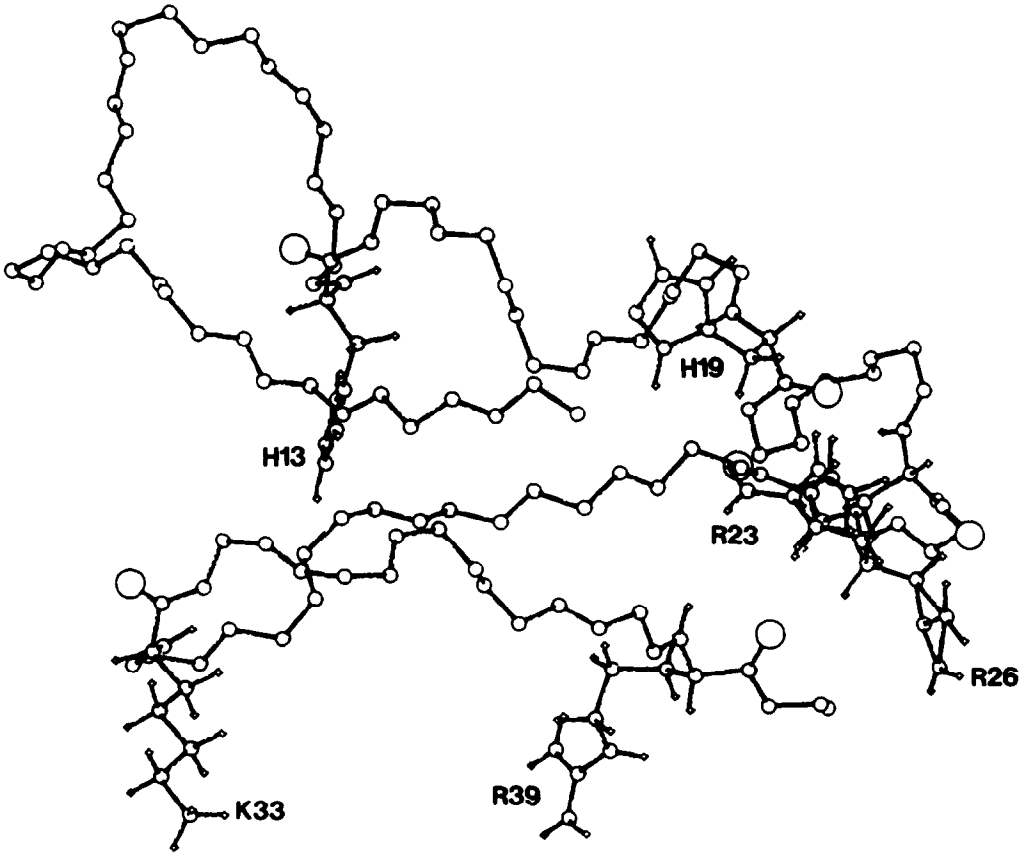

Fig. 5. Average structure of the peptide backbone of nine converged structures of sapecin. Side-chain atoms are shown for the basic residues.

The three disulfide bonds are known to play a crucial role in the expression of the antibacterial activity [4]. The present result shows that (1) the disulfide bond between Cys-3 and Cys 30 connects the terminal of the loop with the 24–31 strand and (2) the disulfide bonds between Cys-16 and Cys-36 and between Cys-20 and Cys-38 connect the 15–23 helix with the other 34–40 strand. The three disulfide bonds presumably stabilize the tertiary structure of sapecin. Cleavage of the disulfide bonds would result in disruption of the tertiary structure, leading to the reduction of the antibacterial activity [4]. The results of a ^{1}H NMR study have indicated that NP-5, which is another antibacterial protein, possesses a different structure, in which an antiparallel β sheet exists and the N- and C-terminal segments of NP-5 are in close spatial contact to each other [9,10]. It appears that the difference in structure between sapecin and NP-5 is primarily due to the mode of connection of the disulfide bonds.

It has been suggested that sapecin has a specific affinity to cardiolipin [2]. Sapecin contains three Arg residues at positions 23, 26 and 39 and one Lys at position 33. As Fig. 5 shows, these basic residues are located on one side of the molecule, forming a basic-residue rich region. Several NOEs were observed between the side chain protons of Arg-23, -39 and Lys-33 and the protons of other residues. No NOEs were observed between the side chain protons of Arg 26 and the protons of other residues. On the basis of these distance constraints, orientation of the side chains of Arg-23, -39 and Lys-33 was roughly defined (Fig. 5). The distances between Lys-33 and Arg-39 and between Arg-23 and Arg-39 were approximately 8 Å, which is quite similar to that between the two phosphate moieties of cardiolipin. In the sapecin molecule, Ala-15, Ala-17, Ala-18, Leu-21 and Leu-22 form a hydrophobic surface, which is opposite to the basic residue rich region. Existence of an interesting combination of the hydrophobic surface and the basic residue rich region on the tertiary structure of sapecin might have some biological relevance to the function of this antibacterial protein.

REFERENCES

[1] Matsuyama, K. and Natori, S. (1988) J. Biol. Chem. 263, 17112–17116.
[2] Matsuyama, K. and Natori, S. (1990) J. Biochem. (in press).
[3] Matsuyama, K. and Natori, S. (1988) J. Biol. Chem. 263, 17117–17121.
[4] Kuzuhara, T., Nakajima, Y., Matsuyama, K. and Natori, S. (1990) J. Biochem. 107, 514–518.

[5] Lambert, J., Keppi, E., Dimarcq, J.-L., Wicker, C., Dorsselaer, A.V., Hoffmann, J., Fothergill, J. and Hoffmann, D. (1989) Proc. Natl. Acad. Sci. USA 86, 262–266.
[6] Selsted, M.E., Brown, D.M., DeLange, R.J. and Lehrer, R.I. (1983) J. Biol. Chem. 258, 14485–14489.
[7] Selsted, M.E., Brown, D.M., DeLange, R.J., Harwig, S.S.L. and Lehrer, R.I. (1983) J. Biol. Chem. 260, 4579–4584.
[8] Selsted, M.E., Harwig, S.S.L., Ganz, T., Schgilling, J.W. and Lehrer, R.I. (1983) J. Clin. Invest. 76, 1436–1439.
[9] Pardi, A., Hare, D.R., Selsted, M.E., Marrison, R.D., Bassolino, D.A. and Back, II.A.C. (1988) J. Mol. Biol. 201, 625–636.
[10] Levy, R.M., Bassolino, D.A., Kitchen, D.B. and Pardi, A. (1989) Biochemistry 28, 9361–9372.
[11] Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Erunst, R.R. and Wüthrich, K. (1983) Biochem. Biophys. Res. Commun. 117, 479–485.
[12] Bax, A. and Davis, D.G. (1985) J. Magn. Reson. 65, 355–360.
[13] Jeener, J., Meier, B.N., Bachmann, P. and Ernst, R.R. (1979) J. Chem. Phy. 71, 4546–4553.
[14] State, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) J. Magn. Reson. 48, 286–292.
[15] Davis, D.G. (1989) J. Magn. Reson. 81, 603–607.
[16] Morris, G.A. and Freeman, R. (1978) J. Magn. Reson. 29, 433–462.
[17] Nilges, M., Gronenborn, A.M., Brünger, A.T. and Clore, G.M. (1988) Protein Eng. 1, 27–38.
[18] Nilges, M., Clore, G.M. and Gronenborn, A.M. (1989) FEBS Lett. 229, 317–324.
[19] Kohda, D. and Inagaki, F. to be published.
[20] Wüthrich, K., Wider, G., Wagner, G. and Brown, W. (1982) J. Mol. Biol. 155, 311–319.
[21] Wüthrich, K., Billeter, M. and Brawn, W. (1983) J. Mol. Biol. 169, 949–961.