Puwainaphycin C, a Cardioactive Cyclic Peptide from the Blue-Green Alga *Anabaena* BQ-16-1. Use of Two-Dimensional $^{13}C-^{13}C$ and $^{13}C-^{15}N$ Correlation Spectroscopy in Sequencing the Amino Acid Units

Richard E. Moore,* Volker Bornemann, Walter P. Niemczura, John M. Gregson, Jian-Lu Chen, Ted R. Norton,[1] Gregory M. L. Patterson, and Gregory L. Helms

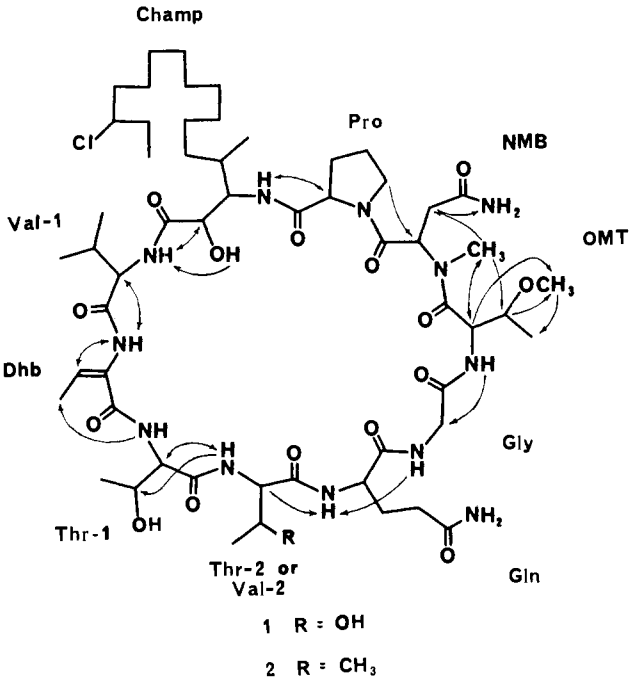
Contribution from the Department of Chemistry, University of Hawaii, Honolulu, Hawaii 96822.
Received December 8, 1988

Abstract: Puwainaphycin C, a cardioactive agent which elicits a strong, positive inotropic effect in isolated mouse atria, is a novel chlorine-containing cyclic decapeptide from a Hawaiian terrestrial *Anabaena* sp. (strain BQ-16-1). Detailed spectral analysis of the intact molecule and an acid hydrolysis study show that puwainaphycin C possesses an unusual β-amino acid unit, viz., a 3-amino-14-chloro-2-hydroxy-4-methylpalmityl residue, and nine known amino acid units. Using difference NOE spectroscopy and two-dimensional $^{13}C-^{13}C$ and carbon-detected $^{13}C-^{15}N$ correlation spectroscopy to sequence the 10 amino acid residues, the gross structure of puwainaphycin C has been shown to be cyclic(3-amino-14-chloro-2-hydroxy-4-methylpalmityl-valyl-2-aminobut-2(*E*)-enoyl-threonyl-threonyl-glutaminyl-glycyl-*O*-methylthreonyl-*N*-methylasparaginyl-prolyl). Puwainaphycin D, a closely related inactive compound, differs from puwainaphycin C in having a second valyl unit at the site occupied by Thr-2.

Introduction and Discussion

A high percentage of hydrophilic extracts of blue-green algae exhibit cardiotonic activity in isolated mouse atria. Tyramine is frequently the agent responsible for increased chronotropic activity.[2] Some extracts, e.g., those of *Tolypothrix byssoidea* (UH isolate H-6-2)[3] and the terrestrial *Anabaena* sp. (UH isolate BQ-16-1), elicit a strong, positive inotropic effect (PIE) in isolated mouse atria with virtually no concomitant chronotropic response. Tolypophycins A and B, two water-soluble substances of unknown structure, account for the inotropic activity of *T. byssoidea*.[3] The inotropic activity of *Anabaena* BQ-16-1, however, is due to a different, less polar substance, i.e., an unusual chlorine-containing cyclic decapeptide, puwainaphycin C (**1**).[4,5] Interestingly the algal extract contains a closely related compound, puwainaphycin D (**2**), which is essentially inactive. We describe here the studies leading to the gross structures of **1** and **2**.

Puwainaphycins C and D were isolated as white amorphous powders from the cultured alga[6] in respective yields of 0.15 and 0.1%. The molecular formulas $C_{56}H_{95}N_{12}O_{16}Cl$ and $C_{57}H_{97}N_{12}O_{15}Cl$ were deduced for **1** and **2** from detailed analyses of the ^{13}C and ^{1}H NMR spectra and the high-resolution positive FAB mass spectra [MH+ ions at m/z 1227.6763 (+0.7 mmu error) and 1225.6830 (−13.3 mmu error)]. Extensive proton–proton decoupling studies of puwainaphycin C, aided by homonuclear ($^{1}H-^{1}H$) phase-sensitive COSY, RCT[7] (Figure 1), and difference NOE experiments (example shown in Figure 2) and a heteronuclear ($^{1}H-^{13}C$) CSCM experiment, led to 10 partial structures (Chart I), eight of which were amino acid units inferred from acid hydrolysis (2 N HCl, 36-h reflux), viz., glycyl, two threonyl, valyl, *O*-methylthreonyl (OMT), glutaminyl, *N*-methylasparaginyl (NMB), and prolyl units. The remaining two partial structures, 2-aminobut-2*E*-enyl (Dhb) and 14-chloro-2-hydroxy-3-amino-4-methylpalmityl (Champ) residues, were consistent with 3-oxobutanoic acid and two C-14 isomers of 3-amino-2,14-dihydroxy-4-methylpalmitic acid, which were also produced on acid hydrolysis. Since ^{1}H NMR analysis indicated that the three hydroxyl groups were located in the two threonyl units and on C-2 of Champ, the chlorine had to be on C-14 of Champ by process of elimination.[8] The carbon chemical shifts for the C-12 to C-16 segment of Champ agreed with the placement of the chloro group at this position.[9]



(1) Department of Pharmacology and Pacific Biomedical Research Center, University of Hawaii.

(2) Norton, T. R.; Moore, R. E.; Patterson, G. M. L., unpublished results.

(3) Entzeroth, M.; Moore, R. E.; Niemczura, W. P.; Patterson, G. M. L.; Shoolery, J. N. *J. Org. Chem.* **1986**, *51*, 5307–5310.

(4) The ED_{50} of puwainaphycin C in mouse atria is 0.2 ppm; at 0.4 ppm (ED_{100}) the positive inotropic effect (PIE) is 325–550% with only a slight increase in chronotropic activity to 128%. The PIE is not inhibited by propranolol and phentolamine and only partially inhibited by verapamil.

(5) The puwainaphycins were named after the site of the Punchbowl National Cemetery (called Puowaina in Hawaiian) and the Hawaiian word for heart (puuwai).

(6) The alga was isolated and grown in mass culture by using the procedure described for *Hapalosiphon fontinalis* [Moore, R. E.; Cheuk, C.; Yang, X.-Q. G.; Patterson, G. M. L.; Bonjouklian, R.; Smitka, T. A.; Mynderse, J. S.; Foster, R. S.; Jones, N. D.; Swartzendruber, J. K.; Deeter, J. B. *J. Org. Chem.* **1987**, *52*, 1036–1043].

(7) (a) Wagner, G. *J. Magn. Reson.* **1983**, *55*, 151–156. (b) Bax, A.; Drobny, G. *J. Magn. Reson.* **1985**, *61*, 306–320.

(8) No explanation can be given at this time for the facile loss of the chloro group from C-14 during the acid hydrolysis.

0002-7863/89/1511-6128$01.50/0

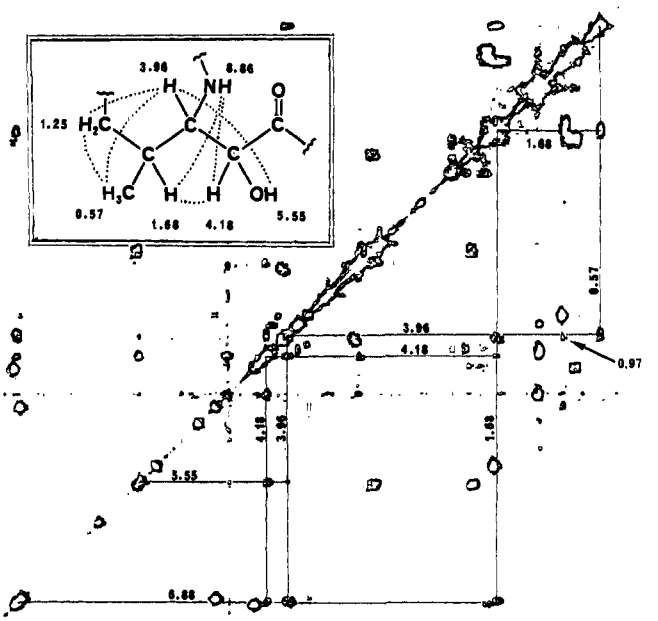


Figure 1. The 300-MHz ^{1}H RCT spectrum of puwainaphycin C in Me_2SO-d_6 from 0.5 to 7.0 ppm. Lines correlate chemical shifts via off-diagonal cross peaks for the C-2, OH on C-2, C-3, NH on C-3, C-4, and Me on C-4 protons in the Champ unit; cross peaks to the signal for the C-5 protons are not shown on this plot. Arrow indicates the relay cross peak connecting the signals for the protons on C-14 (4.00 ppm) and C-16 (0.97 ppm) of Champ.

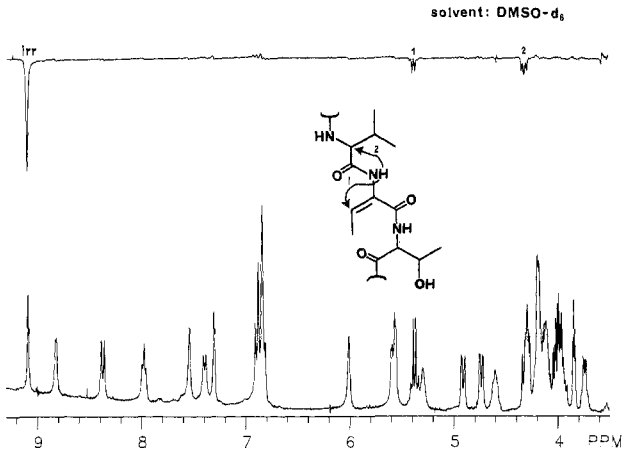


Figure 2. The 300-MHz ^{1}H NMR spectrum of puwainaphycin C in Me_2SO-d_6 from 3.5 to 9.3 ppm (lower trace). The difference NOE spectrum (upper trace) resulting from irradiation of the Dhb NH signal at 9.10 ppm shows strong, negative nuclear Overhauser effects in the proton signals at 5.37 (Dhb H-3) and 4.31 (Val H-2) ppm, indicating strongly that the Dhb NH group is cis to the Dhb C-3 proton and that it is attached to the Val carbonyl.

A similar analysis of **2** showed that puwainaphycin D differed in only one of the amino acid units. Puwainaphycin D possessed a second valyl unit (Val-2) at the Thr-2 site of puwainaphycin C.

Difference NOE spectra (example shown in Figure 2) suggested that the amino acid sequences in **1** and **2** were as depicted (NOEs indicated by arrows). Attempts to sequence the amino acids in **1** and **2** by mass spectral analysis of the intact peptides or suitable degradation products failed. Unlike scytonemin A,[10] **1** and **2** could

(9) Levy, G. C.; Nelson, G. L. *Carbon-13 Nuclear Magnetic Resonance for Organic Chemists*; Wiley-Interscience: New York, 1980. With parameters in Tables 3.2 and 3.7 (but a more appropriate chemical shift change of +32 for replacement of a methyl substituent on C-14 by a chloro substituent), the carbon-13 chemical shifts for C-12, C-13, C-14, C-15, and C-16 in Champ are calculated to be 26.3, 38.7, 66.6, 31.6, and 9.9 ppm, respectively. The observed chemical shifts in dimethyl sulfoxide-d_6 are 25.9, 37.4, 66.3, 30.9, and 10.7 ppm.

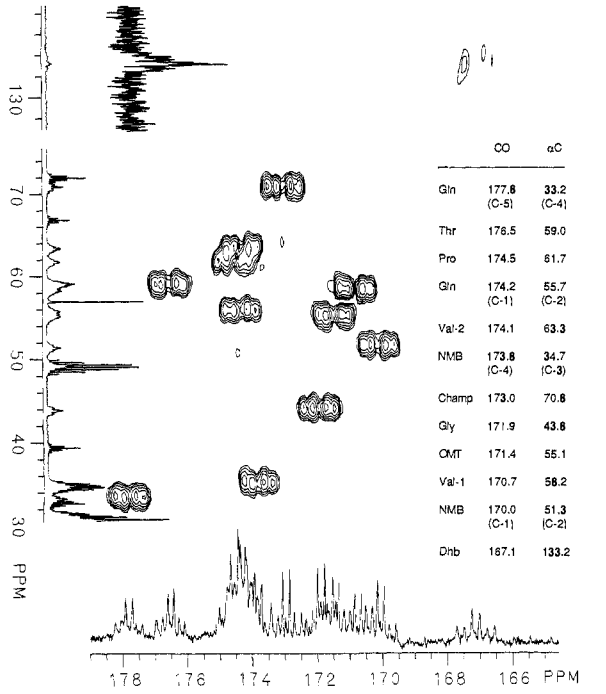


Figure 3. Expanded regions of the COSY-X spectrum of 50% ^{13}C- and 90% ^{15}N-labeled puwainaphycin D in MeOH-d_4 showing the cross peaks that correlate the carbonyl and α-carbon signals. With a standard COSY sequence with continuous Waltz-16 decoupling of protons, 256 4K data sets of 224 scans/block were recorded. The spectrum was apodized with double-exponential weighting functions in each dimension. The F1 dimension was zero-filled twice to yield a final data set of 1K × 2K real data points. The experiment was performed on a GE QE-300 operating at 75.48 MHz for ^{13}C.

Chart I. Gross Structures of Amino Acid Units in Puwainaphycin C Showing ^{1}H Chemical Shifts and Coupling Constants in Me_2SO-d_6

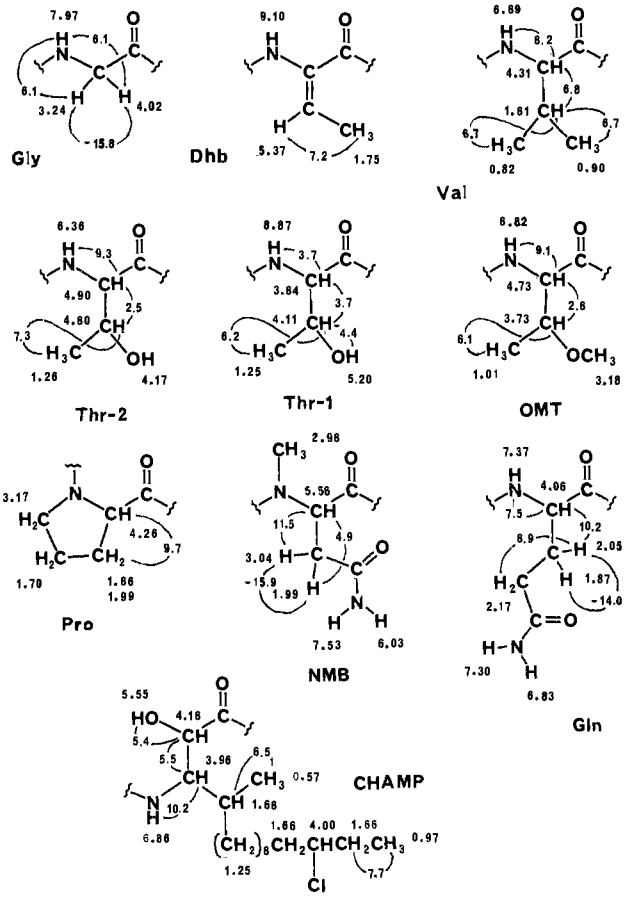


Table I. COSY-X and Carbon–Nitrogen Correlation Spectral Data in Methanol-d_4 Used in Sequencing Amino Acid Units of Puwainaphycin C

amino acid unit	carbon-13 chemical shifts			nitrogen-15 chemical shifts		
	C=O	αC	βC	N on C=O	N on αC	N on βC
Gln C5	177.8	33.2		264.9		
Thr-1	176.2	59.0		267.5	275.2	
Pro	174.4	61.7		277.2	290.2[a]	
Gln C1	174.2	55.0		263.2	273.2	
NMB C4	173.8	34.8		260.2		
Thr-2	173.6	63.6		273.2	267.5	
Champ	172.9	70.8	59.0	275.2		277.2
Gly	171.8	43.8		266.2	263.2	
OMT	171.4	55.0		268.4[b]	266.2	
Val	170.9	58.2		287.8	275.2	
NMB C1	170.2	51.3		290.2[a]	268.4[b]	
Dhb	166.2	133.2		275.2	287.8	

[a] This nitrogen also shows coupling to the *N*-methyl carbon (31.0 ppm). [b] This nitrogen also shows coupling to proline C-5 (49.2 ppm).

not be converted to acyclic decapeptides by mild acid hydrolysis.

To confirm the gross structures, puwainaphycins C and D were uniformly enriched to 90% ^{15}N and 50% ^{13}C and the amino acid sequences determined independently by ^{13}C–^{13}C and ^{13}C–^{15}N correlation spectroscopy.[11-13] First a COSY-X spectrum was obtained for each compound to correlate the carbonyl and adjacent α-carbon signals (correlations shown in Table I for **1** and Figure 3 for **2**). Next two carbon-detected ^{13}C–^{15}N refocused double-quantum coherence experiments were performed, the first on the carbonyl carbon region and the second on the α-methine carbon region of each cyclic peptide to correlate each carbonyl carbon with other carbons bound to the same nitrogen (correlations shown in Table I for **1** and Figure 4 for **2**). All of the spectra were determined in methanol-d_4 since the nitrogen signals were separated best in that solvent.

All of the carbonyl carbon signals could be unambiguously assigned from the COSY-X data. The carbonyl carbon signal at 170.0 ppm in the COSY-X spectrum of **2**, for example, belonged to C-1 NMB since it showed a cross peak to the α-carbon signal at 51.3 ppm (Figure 3). The latter signal had previously been assigned to C-2 of NMB from a ^1H–^{13}C CSCM experiment.

Cross peaks were observed between all of the carbonyl carbon signals and nitrogen signals in the carbon-detected ^{13}C–^{15}N correlation spectra of **1** and **2**.[11] Cross peaks could also be seen from most of the α-methine carbon signals (as well as from the N-methyl and Pro C-5 carbon signals) to the nitrogen signals. For example, in the ^{13}C–^{15}N spectrum of **2** (Figure 4) the NMB C-1 carbonyl signal at 170.0 ppm exhibited a cross peak to the nitrogen signal at 289.6 ppm. The Pro C-2 and C-5 signals at 61.7 and 49.2 ppm, respectively, as well as the NMB C-2 signal at 51.3 ppm, also showed cross peaks to the nitrogen signal at 289.6 ppm. The NMB C-1 carbonyl therefore had to be connected to the proline nitrogen. From the two-dimensional data, connectivities could also be readily made from Pro C-1 to Champ N, Thr C-1 to Val-2 N, Val-2 C-1 to Gln N on C-2, OMT C-1 to NMB N on C-2, Gln C-5 to Gln N on C-5, and NMB C-4 to NMB N on C-4.

Connectivities, however, could not be unambiguously made from Val-1 C-1 to Dhb N, Champ C-1 to Val-1 N, Dhb C-1 to Thr

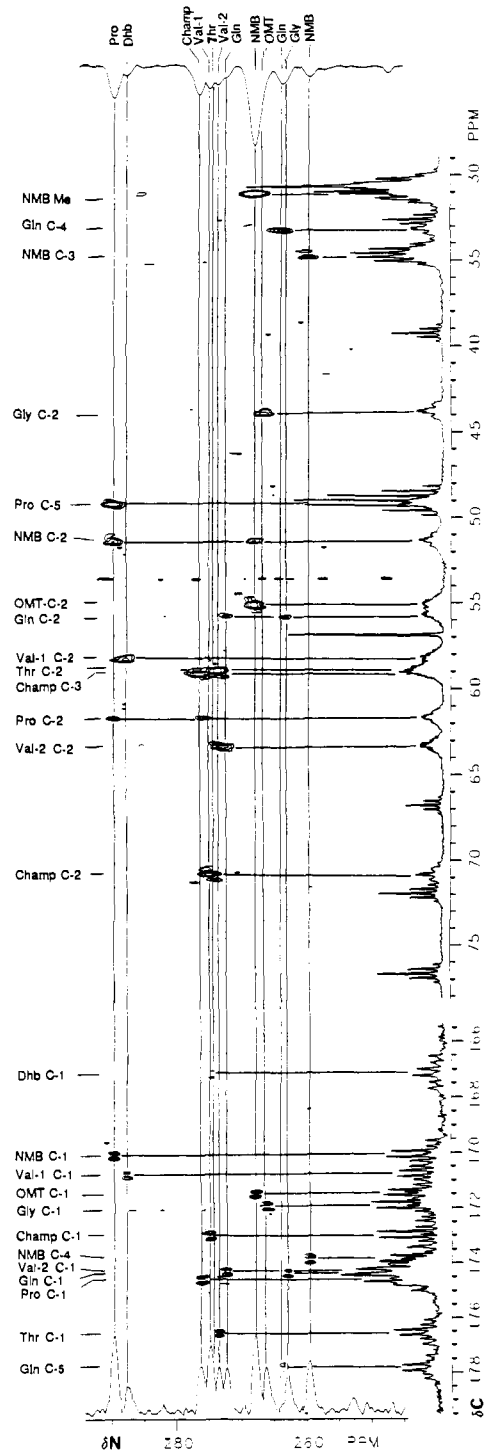


Figure 4. Carbon-detected carbon–nitrogen correlation experiment on 50% ^{13}C- and 90% ^{15}N-labeled puwainaphycin D in MeOH-d_4. The refocused double-quantum coherence experiment was performed separately for the carbonyl and aliphatic carbon regions of the spectrum because of the ^{13}C–^{15}N coupling constant dependence of the experiment. For the carbonyl region 128 1K data sets of 320 scans/block were recorded (total measurement time 18.5 h) by using a value of 33 ms for the $(2J_{C-N})^{-1}$ delay; the total data set was multiplied by a double-exponential weighting function in both time domains and zero-filled to yield a final matrix of 512 × 512 real points. For the aliphatic region 64 2K spectra of 1200 scans/block were recorded (total measurement time 36.8 h) by using a value of 67 ms for the $(2J_{C-N})^{-1}$ delay; the total data set was multiplied by a double-exponential weighting function in both time domains and zero-filled to yield a final matrix of 512 × 1K real points. Both experiments were performed on a modified Nicolet NT-300 narrow-bore spectrometer operating at 75.656 MHz for ^{13}C. Continuous low-power broad-band decoupling was used during the relaxation delay to maintain a NOE between ^1H and ^{13}C while high-power broad-band decoupling was used during the evolution and detection period of the experiment to decouple the protons.

(10) Helms, G. L.; Moore, R. E.; Niemczura, W. P.; Patterson, G. M. L.; Tomer, K. B.; Gross, M. L. *J. Org. Chem.* **1988**, *53*, 1298–1307.

(11) Niemczura, W. P.; Helms, G. L.; Chesnick, A. S.; Moore, R. E.; Bornemann, V. *J. Magn. Reson.* **1989**, *81*, 635–640.

(12) For protein ^{13}C- and ^{15}N-labeling with amino acids and use of ^{13}C–^{15}N correlation in sequencing, see: Kainosho, M.; Tsuji, T. *Biochemistry* **1982**, *21*, 6273–6279.

(13) Recently uniform ^{13}C and ^{15}N enrichment of blue-green algae has been used to study the structures of proteins such as flavodoxin from *Anabaena* 7120 [(a) Oh, B. H.; Westler, W. M.; Darba, P.; Markley, J. L. *Science (Washington, D.C.)* **1988**, *240*, 908–911. (b) Westler, W. M.; Kainosho, M.; Nagao, H.; Tomonaga, N.; Markley, J. L. *J. Am. Chem. Soc.* **1988**, *110*, 4093–4095. (c) Stockman, B. J.; Westler, W. M.; Darba, P.; Markley, J. L. *J. Am. Chem. Soc.* **1988**, *110*, 4095–4096. (d) Westler, W. M.; Stockman, B. J.; Markley, J. L.; Hosoya, Y.; Miyake, Y.; Kainosho, M. *J. Am. Chem. Soc.* **1988**, *110*, 6256–6258]. Uniform enrichment to 26% was achieved by aerating the blue-green algal culture with air containing [26% ^{13}C]CO_2 as the sole carbon source.

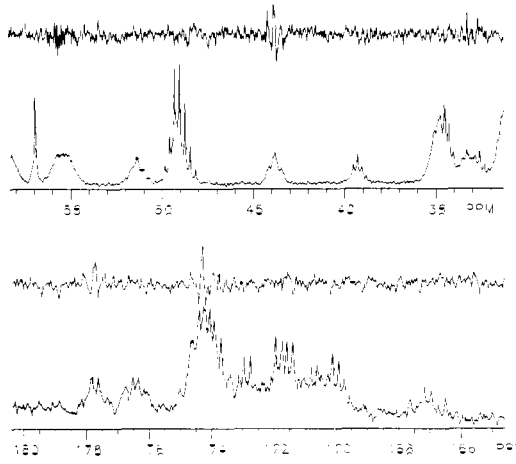


Figure 5. Single-frequency continuous-wave ^{15}N-decoupled ^{13}C spectra of 50% ^{13}C- and 90% ^{15}N-labeled puwainaphycin D in MeOH-d_4 resulting from irradiation at 262.8 ppm (Gly N). Four traces are shown for the carbonyl carbon region (bottom), the α-carbon region (third from bottom), and difference spectra showing the positions of decoupled carbon signals. Irradiation of the Gly nitrogen signal results in decoupling of the Gln C-1 (174.2 ppm), Gln C-2 (55.7 ppm), and Gly C-2 (43.8 ppm) signals [Gln C-5 (177.8 ppm) and Gln C-4 (33.2 ppm) signals are also affected due to overlap of the Gly and Gln nitrogen signals (signal for N on C-5 of Gln at 263.5 ppm)].

N, Gly C-1 to OMT N, and Gln C-1 to Gly N since cross peaks were lacking or difficult to see for six of the nitrogen to α-carbon correlations, viz., Dhb N to Dhb C-2, Val-1 N to Val-1 C-2, Thr N to Thr C-2 and Dhb C-2, OMT N to OMT C-2, and Gly N to Gly C-2. Single-frequency ^{15}N decoupling, however, readily established these correlations. For example, irradiation of the nitrogen signal at 262.8 ppm removed [in addition to the $^1J_{CO-N}$ coupling from the Gln C-1 signal at 174.2 ppm and the $^2J_{\alpha C-N}$ coupling from the Gln C-2 signal at 55.7 ppm (cross peaks are seen in Figure 4 for both of these correlations)] the $^1J_{N-\alpha C}$ coupling from the Gly C-2 signal at 43.8 ppm (Figure 5).[14] The Gln C-1 carbonyl therefore had to be attached to the Gly nitrogen.

Experimental Section

General Procedures. Spectra were determined at 300 MHz for proton NMR and 75 MHz for carbon-13 NMR. Proton chemical shifts are referenced in Me_2SO-d_6 to the residual Me_2SO-d_5 signal (2.52 ppm), in MeOH-d_4 to the residual [CH_3-d_2]MeOD signal (3.30 ppm), and in $CDCl_3$ to the residual $CHCl_3$ signal (7.24 ppm). ^{13}C chemical shifts are referenced in Me_2SO-d_6 and MeOH-d_4 to the solvent (39.5 and 49.0 ppm, respectively). ^{15}N chemical shifts are referenced in MeOH-d_4 externally to acetamide (269.5 ppm where δ = 0 for nitromethane). Qualitative homonuclear ^1H NOEs were obtained in Me_2SO-d_6 by selective irradiation for 2 s by using 30–32 dB of gated decoupler power (hetero mode), followed by data acquisition (decoupler off) with no recycle delay; subtraction of this on-resonance FID from an off-resonance FID resulted in a difference FID that after processing gave an NOE difference spectrum. Homonuclear ^1H and heteronuclear ^1H–^{13}C connectivities were determined by using phase-cycled 16-step COSY and CSCM experiments, respectively, as described by Bax.[15] Relayed coherence transfer (RCT) spectra were obtained by using the pulse sequence and phase cycling described by Wagner or Bax and Drobny;[7] refocusing delays of 20–25 ms were used for the single- and double-relay experiments.

Culture Conditions and Isolation of Puwainaphycins C and D. An unidentified species of *Anabaena* (Nostocaceae), designated strain number BQ-16-1, was isolated from a soil sample collected at the Punchbowl National Cemetery[5] in Honolulu, HI, in the fall of 1984 and grown in mass culture in the laboratory by using the conditions described for *Hapalosiphon fontinalis*.[6]

Freeze-dried alga (5 g) was extracted twice with 1-L portions of ethanol/water (7:3) for 12 h. The combined extracts were evaporated to dryness, and the residue (1.25 g) was chromatographed on a 5 × 40 cm column of Amberlite XAD-2 resin.[10] Fractions were eluted with water, ethanol/water (1:1), and ethanol. The ethanol fraction was evaporated, and the residue in methanol was subjected to flash, reverse-phase chromatography on a 4 × 20 cm column of C-18 silica. A mixture of puwainaphycins A–E was eluted with methanol/water (8:2) which was separated into the individual peptides by HPLC [C-18, 10 μm, 250 × 10 mm, acetonitrile/water (55:45, 1.5 mL/min), UV detection at 254 nm] to give 5.0 mg of A (t_R 14.2 min), 7.5 mg of B (t_R 17.0 min), 15.1 mg of C (t_R 20.0 min), 10.9 mg of D (t_R 23.0 min), and 0.5 mg of E (t_R 24.5 min).

Physical data for puwainaphycin C: $[\alpha]_D$ −39° (1:1 MeCN/H_2O, *c* 0.4); UV (MeCN) λ_{max} nm (ϵ) 207 (21 900), 243 (6200); IR ($CHCl_3$) ν_{max} 3408, 3358, 1659, 1535, 797 cm^{-1}. ^{13}C NMR (Me_2SO-d_6, 16 mM): *amino acid unit*, δ (carbon position); *Champ* 169.5 (1), 69.6 (2), 56.0 (3), 32.2 (4), 15.9 (Me on 4), 33.4 (5), 25.5 (6), 29.7 (7), 29.2 (8), 29.0 (9), 29.0 (10), 28.5 (11), 25.9 (12), 37.4 (13), 66.3 (14), 30.9 (15), 10.7 (16); *Val* 168.8 (1), 55.7 (2), 32.7 (3), 18.9 (4), 18.4 (4); *Dhb* 164.0 (1), 132.3 (2), 117.2 (3), 12.2 (4); *Thr*-1 170.4 (1), 61.9 (2), 65.3 (3), 20.6 (4); *Thr*-2 174.4 (1), 56.8 (2), 69.8 (3), 18.8 (4); *Gln* 171.8 (1), 53.1 (2), 26.4 (3), 31.7 (4), 173.9 (5); *Gly* 168.8 (1), 42.4 (2); *OMT* 169.7 (1), 53.1 (2), 74.8 (3), 15.0 (4), 55.7 (OMe on 3); *NMB* 167.5 (1), 49.6 (2), 33.9 (3), 171.5 (4), 30.3 (NMe on 2); *Pro* 171.0 (1), 59.8 (2), 30.0 (3), 23.3 (4), 46.8 (5).

Physical data for puwainaphycin D: $[\alpha]_D$ −37° (1:1 MeCN/H_2O, *c* 0.4), UV (MeCN) λ_{max} nm (ϵ) 207 (21 800), 243 (5900); IR ($CHCl_3$) ν_{max} 3410, 3320, 1659, 1541, 802 cm^{-1}. ^1H NMR (Me_2SO-d_6): *amino acid unit*, δ (carbon position, mult); *Champ* 4.18 (2, dd), 5.56 (OH on 2, d), 3.96 (3, ddd), 6.84 (NH on 3, d), 1.68 (4, m), 0.59 (Me on 4, d), 1.55 (5, m), 1.25 (6-12, m), 1.66 (13 m), 4.00 (14, p), 1.66 (15, dq), 0.98 (16, t); *Val*-1 4.32 (2, dd), 6.89 (NH on 2, d), 1.82 (3, m), 0.89 (4, d), 0.84 (4, d); *Dhb* 9.10 (NH on 2, s), 5.37 (3, q), 1.75 (4, d); *Thr* 4.86 (2, dd), 8.42 (NH on 2, d), 4.60 (3, dq), 4.08 (OH on 3, s), 1.26 (4, d); *Val*-2 3.84 (2, dd), 8.61 (NH on 2, d), 2.17 (3, m), 1.04 (4, d), 1.02 (4, d); *Gln* 4.10 (2, m), 7.52 (NH on 2, d), 1.20 (3, m), 2.20 (4, m), 6.83 and 7.37 (NH_2 on 5, two br s); *Gly* 3.99 and 3.23 (2, two dd), 7.97 (NH, dd); *OMT* 4.73 (2, dd), 6.82 (NH on 2, d), 3.73 (3, dq), 1.01 (4, d), 3.18 (OMe on 3, s); *NMB* 5.56 (2, dd), 2.96 (NMe on 2, s), 3.02 and 1.88 (3, two dd), 7.53 and 6.03 (NH_2 on 4, two br s); *Pro* 4.28 (2, dd), 1.99 and 1.88 (3, two m), 1.70 (4, m), 3.17 (5, m). ^{13}C NMR (Me_2SO-d_6, 16 mM): *amino acid unit*, δ (carbon position); *Champ* 169.6 (1), 69.3 (2), 56.7 (3), 32.3 (4), 16.0 (Me on 4), 33.4 (5), 25.5 (6), 29.7 (7), 29.2 (8), 29.0 (9), 28.8 (10), 28.6 (11), 26.0 (12), 37.5 (13), 66.7 (14), 31.0 (15), 10.8 (16); *Val*-1 168.9 (1), 57.1 (2), 32.8 (3), 19.1 (4), 18.3 (4); *Dhb* 164.6 (1), 131.9 (2), 118.2 (3), 12.4 (4); *Thr* 175.2 (1), 56.2 (2), 70.3 (3), 19.1 (4); *Val*-2 171.8 (1), 61.4 (2), 29.0 (3), 19.1 (4), 18.0 (4); *Gln* 170.4 (1), 53.6 (2), 26.0 (3), 31.8 (4), 175.1 (5); *Gly* 169.6 (1), 42.6 (2); *OMT* 172.4 (1), 53.2 (2), 74.9 (3), 15.1 (4), 56.0 (OMe on 3); *NMB* 167.9 (1), 49.5 (2), 33.8 (3), 172.1 (4), 30.4 (NMe on 2); *Pro* 171.9 (1), 59.9 (2), 30.2 (3), 23.4 (4), 47.1 (5).

Uniform ^{13}C and ^{15}N Enrichment of Puwainaphycins C and D. *Anabaena* sp. BQ-16-1 was grown in a 10-L glass vessel containing 8 L of an inorganic medium[6] from which buffer had been omitted, and 4.0 g $Na^{15}NO_3$ (99 at. %) added as the sole nitrogen source. The culture was stirred, incubated at 24 ± 2 °C, illuminated at an incident intensity of 150 μeinsteins m^{-2} s^{-1} with cool-white fluorescent lighting for a continuous period of 16 h per day, and aerated[16] at approximately 1 L/min with ordinary air (no extra CO_2 added). The culture vessel was equipped with acid (0.5 N HCl) and base ($NaH^{13}CO_3$ solution) addition ports and an autoclavable pH electrode. The pH was kept at 7.85 ± 0.05 by continuous monitoring with a pH controller and automatic addition of acid. A 1-L aqueous solution of 6.5 g $NaH^{13}CO_3$ (99 at. %)[17] was added continuously over 26 days.

After 28 days, the 8-L culture (medium and cells) was lyophilized and the solid residue extracted twice with 1 L of ethanol/water (7:3) for 12 h. Workup as described above resulted in the isolation of ^{13}C,^{15}N-labeled puwainaphycins A (3 mg), B (3 mg), C (8 mg), D (7 mg), and E (0.1 mg).

Inspection of the ^{13}C NMR spectra of the labeled puwainaphycins C and D indicated uniform ^{13}C enrichment to about 50% and ^{15}N enrichment to over 90%.[17]

Acid Hydrolysis. Acid hydrolysis of 50 mg of puwainaphycin C or D in 6 mL of 2 N HCl (24–36 h reflux) gave a hydrolyzate that was

(14) $^1J_{CO-N}$ = 13.8–17.7 Hz; $^1J_{N-\alpha C}$ and $^2J_{N-\alpha C}$ = 6.4–10.6 Hz where the α-carbon is sp^3; $^1J_{N-\alpha C}$ = 16.2 Hz where the α-carbon is sp^2 as in the Dhb unit.

(15) Bax, A. *Two-Dimensional Nuclear Magnetic Resonance in Liquids*; Delft University Press: Delft, Holland, 1982.

(16) *Anabaena* species have heterocysts, but atmospheric nitrogen does not appear to be incorporated appreciably, at least under these conditions.

(17) Generally use of 5 g of $NaH^{13}CO_3$ results in uniform ^{13}C enrichment to 20–25%, but only if the pH is held at 7.85 ± 0.05 and the aeration is maintained at 1 L/min throughout the experiment; use of 6.5–8.0 g of $NaH^{13}CO_3$ results in labeling to 40–50% ^{13}C, particularly if aeration is carried out at a lower flow rate (0.5 L/min or less).

separated into two fractions by reverse-phase chromatography on C-18 (Analytichem BondElut). The fraction eluted with water was subjected to HPLC on a Whatman PAC (amino-cyano) column with 0.1% trifluoroacetic acid (TFA) to give 2-oxobutanoic acid, glycine, threonine, glutamic acid, valine, *O*-methylthreonine, *N*-methylaspartic acid, and proline. Threonine and glutamic acid, which coeluted on the Whatman PAC column, were separated on a Whatman SCX (strong cation exchange) column with water; similarly, *N*-methylaspartic acid and proline coeluted on the PAC column but could be separated on the SCX column with 0.1% TFA in water.

The C-18 BondElut fraction eluted with MeOH was treated first with Ac_2O in pyridine followed by CH_2N_2 in CH_2Cl_2. Gradient HPLC of the resulting Champ-related *N*-acetyl methyl esters on silica with 10-50% EtOH in 1:1 hexane/CH_2Cl_2 gave a 1:1 mixture of methyl 3-acetamido-2,14-diacetoxy-4-methylpalmitates (**3**) which were isomeric at C-14.

The mixture of isomers of gross structure **3** had the following properties: $[\alpha]_D$ +17° (*c* 1.6, 1:1 MeCN/H_2O); EIMS *m/z* (relative intensity) 457 (0.1), 426 (5), 415 (3), 414 (9), 398 (6), 397 (17), 386 (15), 372 (10), 354 (14), 338 (15), 326 (96), 296 (11), 284 (15), 267 (49), 266 (100), 224 (43), 202 (25), 160 (49), 142 (50); high-resolution EIMS *m/z* 457.3014 (calcd for $C_{24}H_{43}NO_7$, -2.4 mmu error). ^1H NMR ($CDCl_3$) δ 5.50 (d, J = 10.1 Hz, NH), 5.00 (d, J = 6.3 Hz, H-2), 4.90 and 4.78 (two quintets, each of 0.5 H intensity, J = 6.3 Hz, H-14), 4.52 (ddd, J = 10.1, 6.3, 4.1 Hz, H-3), 3.72 (s, $COOCH_3$), 2.12 (s, CH_3CO), 2.04 and 2.02 (two singlets, each of 1.5 H intensity, CH_3COO on C-14), 2.00 (s, CH_3CO), 1.78 (m, H-4), 1.60 (m, 2 H on C-15), 1.25 (br m, 18 H on C-5 to C-13), 0.88 (d, J = 6.9 Hz, CH_3 on C-4), 0.86 (t, J = 7.5 Hz, H_3-16).

Acknowledgment. This research was supported by NSF Grant CHE-8800527.

Supplementary Material Available: ^{13}C and ^{15}N NMR data for 50% ^{13}C- and 90% ^{15}N-labeled **2** in methanol-d_4 and single-frequency continuous-wave ^{15}N-decoupled ^{13}C spectra of labeled **2** in methanol-d_4 (4 pages). Ordering information is given on any current masthead page.

Molecular Topology of Multiple-Disulfide Polypeptide Chains

Boryeu Mao

Contribution from Computational Chemistry, Upjohn Research Laboratories, Kalamazoo, Michigan 49001. *Received September 1, 1988*

Abstract: Molecular topology of the polypeptide chain in a stable folded protein is characterized from analyzing the graph representation of molecular covalent structure and the embedding of such a graph. A subset of topologies (i.e., embedded graphs) that do not contain knotted structures is enumerated and classified for some graphs. As a working hypothesis, it is proposed that topologies of the finite polypeptide chain in a stable, folded globular protein can be represented by this subset. Thus, for any polypeptide chain containing three or fewer disulfides, there is only *one topology* for the polypeptide. For some four-disulfide and some five-disulfide chains, their covalent structure graphs are intrinsically nonplanar (of genus 1), and in each case there are *two* enantiomorphic molecular topologies in the subset. Only one of the allowed topologies represents the stable folded tertiary structure of a protein; e.g., two mammalian-active neurotoxins from scorpions, variant 3 toxin from the North American species *Centruroides sculpturatus* Ewing and toxin II from North African scorpion *Androctonus australis* Hector, each a nonplanar four-disulfide chain, have the **D** topology. Topological stereoisomerism, i.e., the existence of multiple possibilities for the molecular topology of nonplanar polypeptide chains, indicates that the correct prediction of molecular topology must be a criterion for any scheme that predicts tertiary structure of these proteins.

Graph Representation of Polypeptide Chains

Proteins are polypeptide chains that are essentially linear structures except for disulfide bonds; each such disulfide bond links the side chains of two cysteine residues in the primary sequence. Disulfide bonds provide important structural stability in proteins, and the number of disulfide bonds found in single polypeptide chains of proteins varies from 0 to more than 12.[1] The covalent structure of a multiple-disulfide polypeptide chain is fully described by a graph in which each vertex represents the α-carbon atom of a disulfide-linked cysteine residue, and each edge represents a covalent linkage between two such cysteinyl C_α atoms.[1,2] An example of such a graph (henceforth *the covalent structure graph*) is shown in Figure 1a for a three-disulfide chain. In considering the topology of a covalent structure graph (i.e., an embedded covalent structure graph), labeling each vertex alphabetically and uniquely from the amino terminal to the carboxyl terminal essentially specifies that all edges are nonequivalent; i.e., they are edges of different colors. This is consistent with the fact that in general a partial amino acid sequence delimited by a pair of cysteine residues is different from one delimited by another pair; such labeling is also essential for the discussion of isomerism of molecular topology by topological graph theory (ref 3, and discussions in the Four-Disulfide Chains section).

Representation of a polypeptide covalent structure by a graph has been employed for studying knotting and loop penetration problems in protein structure. Crippen[2,4] estimated the probability of knotted structures in idealized polypeptide chains. Klapper and Klapper[1] described the knotting problem as a more general "loop penetration" phenomenon and also investigated the planarity/nonplanarity of different disulfide pairings in multiple-disulfide proteins. Connolly et al.[5] identified covalent and noncovalent loops in tertiary structure of proteins and studied linking and threading of such loops. Kikuchi et al.[6] devised schemes for identifying spatial arrangements of polypeptide fragments in tertiary structure of proteins. For any polypeptide chain containing three or fewer disulfide bonds (Figure 1a), the covalent structure

(1) Klapper, M. H.; Klapper, I. Z. *Biochim. Biophys. Acta* **1980**, *626*, 97-105.
(2) Crippen, G. M. *J. Theor. Biol.* **1974**, *45*, 327-338.
(3) (a) Walba, D. M. *Tetrahedron* **1985**, *41*, 3161-3212. (b) Walba, D. M. In *Graph Theory and Topology in Chemistry*; King, R. B., Rouvray, D. H., Eds.; Elsevier Science Publishers B. V.: Amsterdam, 1987; pp 23-42.
(4) Crippen, G. M. *J. Theor. Biol.* **1975**, *51*, 495-500.
(5) Connolly, M. L.; Kuntz, I. D.; Crippen, G. M. *Biopolymers* **1980**, *19*, 1167-1182.
(6) Kikuchi, T.; Némethy, G.; Scheraga, H. A. *J. Comput. Chem.* **1986**, *7*, 67-88.

0002-7863/89/1511-6132$01.50/0