Pergamon

0031-9422(95)00851-9

Phytochemistry, Vol. 42, No. 2, pp. 443–445, 1996

0031-9422/96 $15.00 + 0.00

NONPROTEIN AMINO ACIDS FROM SEEDS OF *CYCAS CIRCINALIS* AND *PHASEOLUS VULGARIS*

CONG-JUN LI, DELIA M. BROWNSON, TOM J. MABRY,* CHRIS PERERA† and E. ARTHUR BELL†

Department of Botany, University of Texas at Austin, Austin, TX 78713-7640, U.S.A.; †Pharmacology Group, King's College London, Manresa Road, London SW3 6LX, U.K.

(*Received* 19 *September* 1995)

Key Word Index—*Cycas circinalis*; Cycadaceae; *Phaseolus vulgaris*; Leguminosae; seeds; *N*-(3′-one-5′-methyl)-hexylalanine; leucine betaine; *N*-methylisoleucine.

Abstract—Our chemical studies on *Cycas circinalis* seeds from Guam has provided two new nonprotein amino acids, *N*-(3′-one-5′-methyl)-hexylalanine and leucine betaine. *N*-methylisoleucine, previously reported as a component of naturally occurring peptides, has been isolated as a free amino acid from the seeds of *Phaseolus vulgaris* (pinto bean), together with *S*-methylcysteine, pipecolic acid and a dipeptide, γ-glutamyl-leucine.

INTRODUCTION

The medicinal and dietary use of the cycad, *Cycas circinalis*, has been implicated in the aetiology of the neurodegenerative disease, amyotrophic lateral sclerosis–parkinsonism–dementia complex (ALS-PDC), in native populations on western Pacific islands, notably Guam [1–4]. Over the last 40 years, ALS–PDC among the Chamorro people of Guam has declined in a manner consistent with the decreased traditional utilization of the cycad plant [1, 5]. ALS–PDC may be an example of a chronic neurodegenerative disease induced by exposure to excitotoxins. Previous chemical studies only reported one nonprotein amino acid, α-amino-β-methylaminopropionic acid (L-BMAA), from cycads [6]. Our chemical research on endosperm from Guam cycad seeds has led to the isolation of two previously unreported nonprotein amino acids, *N*-(3′-one-5′-methyl)-hexylalanine (**1**) and leucine betaine (**2**).

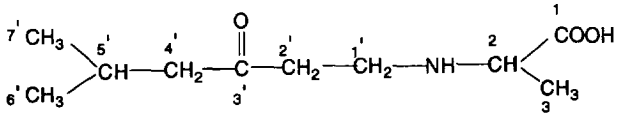
The seeds of various varieties of *Phaseolus vulgaris* are known [7–10] to contain the nonprotein amino acids: *S*-methylcysteine, *S*-methylcysteine sulphoxide, pipecolic acid, γ-glutamyl-*S*-methylcysteine, γ-aminobutyric acid and γ-glutamyl-leucine. We now report the isolation of an additional nonprotein amino acid, *N*-methylisoleucine (**3**), a compound previously reported only as a component of naturally occurring peptides [11–13].

RESULTS AND DISCUSSION

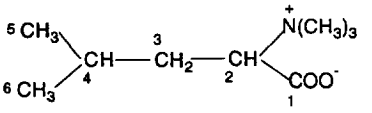
Compound **1** from seeds of *C. circinalis*, a white powder, reacted positively with ninhydrin and its infrared absorption at 1730 cm^{-1} indicated the presence of a ketone function. A chemical ionization mass spectrum gave a molecular ion peak at 202 $[M + 1]^+$. These findings, together with ^1H and ^{13}C NMR data, confirmed its formula as $C_{10}H_{19}NO_3$. The ^1H–^1H COSY clearly demonstrated three independent proton-coupling systems in **1**. First, a six-proton doublet signal (δ 0.88) coupled with a one-proton multiplet signal (δ 2.06). This multiplet is further coupled with a two-proton doublet peak at δ 2.44. Therefore, the partial structure $(CH_3)_2CHCH_2$– was deduced. Second, the coupling correlation between a two-proton triplet signal at δ 3.0 and a two-proton multiplet signal at δ 3.25 gave the partial structure $-CH_2CH_2-$. Third, a three-proton doublet δ 1.47 was coupled to a downfield one-proton quartet at δ 3.66, which was considered to be a methine group directly connected with both an amine group and a carboxyl moiety, as found in α-amino acids. Thus, the partial structure =N–$CH(CH_3)COOH$ was deduced. The relative downfield chemical shifts of the signals at δ 3.0 and δ 2.44 were considered to be in accordance with methylene groups adjacent to a ketone function, while the proton giving the signal at δ 3.25 should be near the nitrogen atom. Therefore, structure **1** was assigned, which is consistent with the mass spectral fragment ion peak at *m/z* 117 $[(M + 1) - (CH_3)_2CHCH_2CO]^+$ and also the ^{13}C NMR data.

Compound **2**, also from *C. circinalis*, was obtained as an amorphous powder. Mass spectral analysis indicated its formula to be $C_9H_{19}NO_2$. One- and two-dimensional NMR evidence suggested the partial structural components $(CH_3)_2CH$–, $-CH_2-$, >CH– and $-N(CH_3)_3$, together with a carboxyl group. Therefore, structure **2** was assigned. Not only were the two geminal protons nonequivalent (δ 1.41 and 1.65) but

*Author to whom correspondence should be addressed.



1



2

the two geminal methyl groups also had different chemical shifts (δ 0.73 and 0.75). Leucine betaine (see Experimental) was synthesized to confirm the proposed structure **2**. The ^1H and ^{13}C NMR data and TLC properties confirmed that the isolated compound was identical to the synthetic sample. Both **1** and **2** are nonprotein amino acids not previously reported.

Compound **3** from seeds of *Phaseolus vulgaris* was identified as *N*-methylisoleucine using ^1H and ^{13}C NMR spectroscopy, and CI-mass spectrometry. *S*-methylcysteine, pipecolic acid and γ-glutamyl-leucine from this same plant were identified by NMR data and direct TLC and electrophoresis comparison.

EXPERIMENTAL

1D and 2D-NMR chemical shifts are reported in ppm using dioxane as ext. ref. and D_2O as solvent. Drs Ulla-Katrina Craig and Thomas Marler, University of Guam, identified the *C. circinalis* L. plants and provided the seeds from near Mangilao, Guam, collected on January 1, 1994. Seeds of *P. vulgaris* (pinto beans) were obtained on the open market in Austin, Texas, U.S.A., October 1994. Seed vouchers, Mabry No. 10 and No. 11, respectively, are deposited in the Department of Botany, University of Texas at Austin.

Extraction and isolation. After the outer coats were removed from the *C. circinalis* seeds, the combined endosperm material was ground to a powder and the powder extracted with 75% EtOH. The concentrate was redissolved in H_2O and the aq. soln extracted sequentially with $CHCl_3$ and EtOAc. Excess EtOAc in the soln was removed using a rotary evaporator and the remaining aq. soln (1 l) was then subjected to an AG50W-X8 cation-exchange resin column. The column was washed with 2 l of H_2O and then eluted with 0.5 N NH_4OH soln to obtain a mixt. of amino acids. This mixt. was then chromatographed on a microcrystalline cellulose column using isoPrOH with increasing amounts of H_2O. The eluate from frs 3–6 (isoPrOH-H_2O, 100:1) was evapd and the residue washed with EtOH to remove pigments; yield of **1** was 3.2 mg. The concentrate from frs 8–9 (isoPrOH-H_2O, 50:1) was subjected to further prep. cellulose TLC, developed with isoPrOH-H_2O (6:1); yield of **2**: 4.6 mg.

Finely ground pinto bean seeds (7.6 kg) were extracted with 80% EtOH (10 l × 8). The comb. extracts were concd and filtered. The filtrate was extracted sequentially with $CHCl_3$ and EtOAc to remove lipids and the partially purified soln was applied to a cation-exchange resin column (AG 50W-X8, H^+ form). The column was washed with H_2O and the amino acids then displaced with 1 N NH_4OH. The resulting amino acid mixt. was applied to a series of microcrystalline cellulose columns for separation using solvent systems of gradually increasing polarity, including isoPrOH-H_2O, *n*-BuOH-EtOH-2N NH_4OH and MeCOMe-H_2O. Further purification was carried out on a column of Sephadex LH-20 using 80% EtOH; yield of **3** (crystals): 1.8 mg.

N-(3′-one-5′-methyl)-hexyl-alanine (**1**). Amorphous solid. IR υ(KBr): 3600~3200 (*br*), 3100~3020 (*br*), 2960, 2940, 1730, 1630, 1580, 1430, 1630 cm^{-1}. $[\alpha]_D$ +99.9° (*c* 1.1, H_2O). CI-MS *m/z*: $[M+1]^+$ 202, $[(M+1)-(CH_3)_2CHCH_2CO]^+$ 117. ^1H NMR (D_2O, 360 MHz): δ 3.66 (1H, *q*, J = 7.2 Hz, H-2), 3.25 (2H, *m*, H-1′), 3.0 (2H, *t*, J = 6.4 Hz, H-2′), 2.44 (2H, *d*, J = 7.0 Hz, H-4′), 2.06 (1H, *m*, H-5′), 1.47 (3H, *d*, J = 7.2 Hz, H-3), 0.88 (6H, *d*, J = 6.7 Hz, H-6′ and CH_3-5′). ^{13}C NMR (D_2O, 90 MHz): δ 174.6 (C-1), 58.1 (C-2), 14.9 (C-3), 51.1 (C-1′), 38.2 (C-2′), 213.0 (C-3′), 40.8 (C-4′), 24.2 (C-5′), 21.6 (C-6′ and CH_3-5′).

Leucine betaine (**2**). Amorphous powder. IR υ(KBr): 3040, 2950, 2870, 1650(sh), 1580, 1475, 1400, 1135, 895 cm^{-1}. $[\alpha]_D$ +125° (*c* 1.5, H_2O). HR-CIMS *m/z*: 174.1492 ($C_9H_{19}NO_2$ + H) (calcd for 174.1494). ^1H NMR (D_2O, 360 MHz): δ 3.45 (1H, *dd*, J_1 = 12.2 Hz, J_2 = 2.5 Hz, H-2), 2.93 (9H, *s*, $N(CH_3)_3$), 1.65 (1H, *td*, J_1 = 12.2 Hz, J_2 = 3.2 Hz, H-3), 1.41 (1H, *td*, J_1 = 11.3 Hz, J_2 = 2.8 Hz, H-3), 1.35 (1H, *m*, H-4), 0.75 (3H, *d*, J = 6.5 Hz, H-5), 0.73 (3H, *d*, J = 6.4 Hz, CH_3-4). ^{13}C NMR (D_2O, 90 MHz): δ 172.2 (C-1), 77.9 (C-2), 51.8 ($N(CH_3)_3$), 35.0 (C_3), 25.0 (C-4), 22.9 and 20.2 (C-5 and CH_3-4). Assignments based on ^1H–^1H COSY and ^{13}C–^1H COSY spectra.

Synthesis of **2**. L-leucine (100 mg) was dissolved in 20 ml 90% EtOH and 20 ml of MeI added to the soln. This mixt. was maintained at room temp. with stirring for 48 hr, then 5 ml of 30% NH_4OH added to stop the reaction. The reaction mixt. was then concd and subjected to microcrystalline cellulose CC, eluting with

n-BuOH - EtOH - 1N NH_4OH (4:1:0.1); frs 3–5 gave 15 mg of an amorphous powder identical with the naturally occurring **2** based on ^1H. ^{13}C NMR data and co-TLC comparison. ^1H NMR (D_2O, 360 MHz): δ 3.44 (1H, *dd*, $J_1 = 12.3$ Hz, $J_2 = 3.0$ Hz, H-2), 2.93 (9H, *s*, $N(CH_3)_3$), 1.65 (1H, *td*, $J_1 = 12.2$ Hz, $J_2 = 3.2$ Hz, H-3), 1.40 (1H, *td*, $J_1 = 12.5$ Hz, $J_2 = 2.9$ Hz, H-3), 1.34 (1H, *m*, H-4), 0.75, 0.72 (3H each, *d*, $J = 6.5$ Hz, H-5 and CH_3-4). ^{13}C NMR (D_2O, 90 MHz): δ 172.2 (C-1), 77.9 (C-2), 51.6 ($N(CH_3)_3$), 34.9 (C-3), 25.0 (C-4), 22.9 and 20.2 (C-5 and CH_3-4).

N-methylisoleucine (**3**). Mp 250–252°. $[\alpha]_D$ +250° (*c* 0.6, H_2O). ^1H NMR (D_2O, 360 MHz): δ 3.25 (1H, *d*, $J = 4.1$ Hz, H-2), 2.48 (3H, *s*, NCH_3), 1.70 (1H, *m*, H-3), 1.30, 1.05 (1H each, *d*, $J = 4.1$ Hz, H-2), 0.75 (3H, *d*, $J = 6.93$ Hz, CH_3-3), 0.72 (3H, *d*, $J = 7.37$ Hz, H-5). ^{13}C NMR (D_2O, 92.5 MHz): δ 172.6 (C-1), 68.4 (C-2), 32.6 (C-3), 25.4 (C-4), 11.0 (C-5), 14.2 (CH_3-3). IR (KBr): 3050, 2950, 2920, 2880, 1580, 1470, 1405, 1330, 1250, 1180, 1140, 1120, 850, 765, 695 cm^{-1}. CIMS *m/z*: 146 [M + 1], 100.

Acknowledgements—This research was supported by the National Institute on Aging (Grant AG-10637) and the Robert A. Welch Foundation (Grant F-130). C-JL also thanks Dr Ben A. Shoulders of the Department of Chemistry, University of Texas at Austin, for helpful suggestions.

REFERENCES

1. Kurland, L. T. (1972) *Fed. Proc.* **31**, 1540.
2. Spencer, P. S., Nunn, P. B., Hugon, J., Ludolph, A. C., Ross, S. M., Roy, D. N. and Robertson, R. C. (1987) *Science* **237**, 517.
3. Spencer, P. S., Palmer, V. S., Herman, A. and Asmedi, A. (1987) *Lancet*, **ii**, 1273.
4. Whiting, M. G. (1963) *Econ. Botany* **17**, 270.
5. Stone, R. (1993) *Science* **261**, 424.
6. Vega, A. and Bell, E. A. (1967) *Phytochemistry* **6**, 759.
7. Fowden, L. (1964) *Ann. Rev. Biochem.* **33**, 173.
8. Casimir, J. and Le Marchant, G. (1966) *Bull. Jardin. Bot. Etat. Brux.* **36**, 53.
9. Bell, E. A., Qureshi, Y., Charlwood, B. V., Pilbeam, D. J. and Evans, C. S. (1977) in *Genetic Diversity in Plants* (Muhammed, A., Aksel, R. and von Borstel, R. C., eds), p. 397. Plenum Press, New York.
10. Perera, K. P. W. C., Bell, E. A. and Nunn, P. B. in preparation.
11. Godfredsen, W. O., Vangel, S. and Thomas, D. W. (1970) *Tetrahedron* **26**, 4931.
12. Audhyya, T. K. and Russell, D. W. (1974) *J. Chem. Soc. Perkin Trans.* **1**, 743.
13. Kitagawa, I., Lee, N. K., Kobayashi, M. and Shibuya, H. (1991) *Tetrahedron* **47**, 2169.