Phytochemistry, Vol. 23, No. 5, pp. 1204–1205, 1984.
Printed in Great Britain.

0031–9422/84 $3.00 + 0.00

COUMESTRIN, A COUMESTAN DERIVATIVE FROM SOYBEAN ROOTS

NGO LE-VAN

Monsanto Agricultural Products Company, 800 North Lindbergh Boulevard, St. Louis, Missouri 63167, U.S.A.

(*Revised received* 29 *September* 1983)

Key Word Index—*Glycine max*; Leguminosae; soybean; coumestrin; coumestrol 3-β-D-glucopyranoside.

Abstract—A new coumestrol glycoside, coumestrin, has been isolated from soybean roots together with its aglycone, coumestrol, and the known isoflavones genistin, genistein, daidzin and daidzein. Their structures were determined by spectroscopic technique (^{1}H NMR, UV, IR, EIMS, CIMS and FDMS) and by some chemical transformations.

INTRODUCTION

Soybean (*Glycine max* L.) has been shown to contain two isoflavone glycosides, genistin and daidzin and their respective aglycones [1]; a 7,4-dihydroxy-6-methoxy isoflavone (glycitein) and its glucoside [2]. Recently, 6″-*O*-acetyl daidzin has been isolated and reported [3].

The present study of soybean roots led to the isolation and characterization of a new coumestrol glucoside, for which the name coumestrin is proposed, as well as other constituents.

RESULTS AND DISCUSSIONS

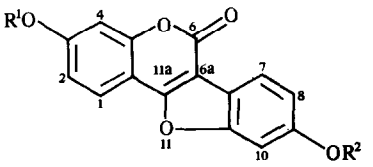
The ether–petrol and ethyl acetate extracts of soybean roots yielded the known coumestrol **1**, genistein and daidzein. The methanolic extract contained genistin, daidzin and a new coumestrol glucoside (**3**), which was characterized by UV, IR, ^{1}H NMR and mass (EI, CI and FDMS) spectral means. The structure was confirmed by enzymatic hydrolysis of **3** with β-glucosidase, by conversion to its respective tetra- and pentaacetate derivatives **4** and **5** and by correlation of their spectroscopic data with those of an authentic sample of coumestrol diacetate (**2**).

Coumestrin (**3**) showed characteristic UV maxima of a coumestrol derivative at 244, 303 and 341 nm. The ^{1}H NMR spectrum of **3** (Table 1) exhibited characteristic patterns of coumestrol. In addition, the multiplets between 3 and 4 ppm in the ^{1}H NMR spectrum of **3** suggested the presence of a sugar moiety. This was consistent with the greater polarity of **3** on reversed-phase HPLC, as compared to coumestrol (**1**). The EIMS of **3** showed a base peak at *m/z* 268.036 but the nature of the glycosidic unit could not be deduced. In contrast, the FDMS exhibited a strong molecular ion at *m/z* 430 consistent with the structure **3**, and a fragment at *m/z* 267 due to the loss of a hexose sugar. Acetylation of **3** yielded the chloroform-soluble tetraacetate **4** and pentaacetate **5**. The ^{1}H NMR spectrum of **5** revealed in addition to the methyl singlets of four sugar acetyl signals an aromatic acetyl signal at 2.37 ppm (Table 1). The sugar protons of the pentaacetate **5** could only be satisfactorily resolved in acetone-d_6. The large coupling constant of the doublet at 5.64 (J = 7.5 Hz) which was assigned to the anomeric proton of the sugar, indicated a β-*O*-glycosidic linkage of the hexose to ring A or D. Although a molecular ion was not observed in the EIMS of **5**, the mass fragment *m/z* 310.047 could be rationalized by the loss of a hexose unit from the $[M]^+$ 640; this was confirmed by CIMS. The base peak at 331 and the fragments *m/z* 271, 211 and 169 were characteristic of the acetylated hexose unit. The relatively intense mass fragment *m/z* 268 was due to the $[aglyc]^+$.

Enzymatic hydrolysis of **3** with β-glucosidase followed by acetylation with acetic anhydride and 4-pyrrolidinopyridine [4] yielded the aglycone diacetate **2** and D-glucose pentaacetate, which were identical with authentic samples (^{1}H NMR, MS, TLC). The location of the glucose

Table 1. ^{1}H NMR data of coumestrin (**3**), its tetra- and pentaacetates **4** and **5**, and the related coumestrol diacetate **2** (100 MHz, TMS as int. standard)

H	**2** ($CDCl_3$)	**3** (CD_3OD)	**4** ($CDCl_3$)	**5** ($CDCl_3$)
1	8.03 *d* (8.3)	7.64 *d* (J = 8.5 Hz)	7.93 *d* (8.3)	7.94 *d* (8.5)
2	7.2 *m*	6.84 *dd* (J = 8.5 + 2 Hz)	7.0 *m*	7.04 *dd* (8.5 + 2)
4	7.31 *d* (2)	6.98 *d* (J = 2 Hz)	7.12 *m*	7.14 *d* (2)
7	8.12 *d* (8.3)	7.81 *d* (J = 9.5 Hz)	7.91 *d* (8.5)	8.08 *d* (8.5)
8	7.2 *m*	7.10 *dd* (J = 9.5 + 2.2 Hz)	6.96 *d* (*br*) (8.5)	~7.2 m
10	7.50 *d* (1.7)	7.13 *d* (J = 2.2 Hz)	7.12 *m*	7.47 *d* (1.7)
1′		~5 *m*	5.3 *m*	5.3 *m*
2′–6′		3–4 *m*	4.26 *m*	4.26 *m*
OAc	2.37 *s* (20 Ac)		2.14 *s*	2.37 *s*
			2.09 *s*	2.14 *s*
			2.08 *s*	2.09 *s*
				2.07 *s*
			2.05 *s*	2.05 *s*



1 $R^1 = R^2 = H$
2 $R^1 = R^2 = Ac$
3 $R^1 = \beta-$glucosyl; $R^2 = H$
4 $R^1 = \beta-$glucose tetraacetate; $R^2 = H$
5 $R^1 = \beta-$glucose tetraacetate; $R^2 = Ac$

moiety in **3** was deduced by comparison of the ^1H NMR data of its tetra- and pentaacetates (**4**, **5**) with those of the coumestrol diacetate **2** (Table 1). The relatively high-field doublet of H-4 of the pentaacetate **5** (7.14, *d*, $J = 2$ Hz) suggested that the glucose moiety was located at C-3.

Therefore, the structure of coumestrin (**3**) is coumestrol 3-β-D-glucopyranoside.

EXPERIMENTAL

UV: Et_2O or MeOH; IR: $CHCl_3$; ^1H NMR: 100 MHz; EIMS: direct inlet probe at 70 eV; CIMS: isobutane.

Freshly harvested soybean roots (*G. max* var. Wayne, 1200 young plants) were lyophilized and extracted successively with Et_2O–petrol (1:1), then EtOAc and finally with MeOH. The Et_2O–petrol and EtOAc extracts were combined and separated first by CC (silica gel) and then by repeated TLC, to yield coumestrol (12 mg), genistein (11 mg) and daidzein (17 mg). The crude methanolic extract was diluted with H_2O and extracted ($\times 4$) with Et_2O to remove the non-polar materials. HPLC of the MeOH–H_2O soluble phase was carried out on two semi-prep reversed-phase C_{18} columns (Waters associates, μ-Bondapak C_{18}, 10 μm particle, 7.8 mm × 30 cm, UV 254 nm detector) using a gradient system of H_2O–MeCN, pH 3, as the eluent to give in order of elution daidzin (33 mg), genistin (24 mg) and coumestrin (**3**, 10 mg). *Coumestrin* (**3**): colourless oil, UV λ_{max}^{MeOH} nm (rel. absorbance): 244 (1.33), 265 (sh), 303 (0.46), 341 (1.36); ^1H NMR: Table 1; EIMS *m/z* (rel. int.): 268.036 [aglycone]$^+$ (100) (calc. for $C_{15}H_8O_5$, 268.037), 254 (7.26), 253 (5.43), 240 [aglycone $-$ CO]$^+$ (19.2), 212 (7.57), 211 (8.35), 184 (10.1).

Acetylation of **3**. 8 mg in 0.5 ml pyridine, 0.5 ml Ac_2O and 8 mg 4-pyrrolidinopyridine were stirred for 18 h at room temp. After usual work-up and TLC (EtOAC–toluene 1:1, run × 3) 3 mg **4** and 4 mg **5** were obtained. *Coumestrin tetraacetate* (**4**): UV $\lambda_{max}^{Et_2O}$ nm (rel. absorbance): 243 (0.78), 304 (0.34), 340 (0.79), 358 sh; CIMS *m/z*: 599; EIMS *m/z* (rel. int.): 598 [M]$^+$ (0.2), 268.035 [aglycone]$^+$ (19.43) (calc. for $C_{15}H_8O_5$, 268.037), 240 [aglycone $-$ CO]$^+$ (2.53), 331 [glucose tetraacetate]$^+$ (4.64), 271 [331 $-$ AcOH]$^+$ (3.23), 211 [331 $-$ 2AcOH]$^+$ (4.07), 169 [211 $-$ ketene]$^+$ (33.56), 127 [169 $-$ ketene]$^+$ (15.13), 109 [169 $-$ AcOH]$^+$ (31.20). *Coumestrin pentaacetate* (**5**): UV $\lambda_{max}^{Et_2O}$ nm (rel. absorbance): 237 (1.09), 260 (0.43), 285 (0.40), 297 (0.64), 320 sh, 330 (1.33), 339 (1.01) and 347 (1.02); CIMS *m/z*: 641; EIMS *m/z* (rel. int.): 310.047 [M $-$ glu(OAc)$_4$ + H]$^+$ (3.44) (calc. for $C_{17}H_{10}O_6$, 310.048), 268.037 [aglycone]$^+$ (40.44) (calc. for $C_{15}H_8O_5$, 268.037), 240 [aglycone $-$ CO]$^+$ (6.07), 331 [glucose tetraacetate]$^+$ (12.73), 271 [331 $-$ AcOH]$^+$ (9.62), 211 [271 $-$ AcOH]$^+$ (10.63), 169 [271 $-$ ketene]$^+$ (70.31), 127 [169 $-$ ketene]$^+$ (34.17), 109 [169 $-$ AcOH]$^+$ (66.12).

Acknowledgements—I thank Drs. Steve J. Wratten and Terry L. Graham for helpful discussions.

REFERENCES

1. Waltz, E. (1931) *Ann.* **489**, 118.
2. Naim, M., Gestetner, B., Kirson, I., Birk, Y. and Bondi, A. (1973) *Phytochemistry* **12**, 169.
3. Ohta, N., Kuwata, G., Akahori, H. and Watanabe, T. (1980) *Agric. Biol. Chem.* **44**, 469.
4. Steglich, W. and Hofle, G. (1969) *Angew. Chem.* **81**, 1001.