Phytochemistry, 1977, Vol. 16, pp. 1953–1956. Pergamon Press. Printed in England.

WATER-SOLUBLE GLYCOPROTEINS FROM *CANNABIS SATIVA* (THAILAND)*

AGNES HILLESTAD, JENS K. WOLD and THOR ENGEN

Department of Pharmacognosy, Institute of Pharmacy, University of Oslo, Oslo 3, Norway

(*Revised received* 13 *June* 1977)

Key Word Index—*Cannabis sativa*; Cannabinaceae; marijuana; glycoprotein; cell wall protein; serine-*O*-galactoside.

Abstract—Glycoproteins were extracted with water from leaves of *Cannabis sativa* grown from seeds of Thailand origin. By ion exchange chromatography the material was separated into a neutral and an acidic fraction. Both glycoprotein fractions contained arabinose, galactose, glucose, mannose and xylose, and in addition rhamnose and galacturonic acid were present in the acidic fraction. The carbohydrate moieties were investigated by methylation analysis and Smith-degradation, whereas the glycopeptide linkage was studied by alkaline hydrolysis in the presence of $NaBH_4$ and Na_2SO_3, respectively. This linkage was shown to be of the serine-*O*-galactoside type. The carbohydrate structure is highly branched, the majority of branches terminating in arabinofuranose end groups. Arabinose is also present in the chain, predominantly (1→4)- and/or (1→5)-linked. Galactose makes up most of the main chain as (1→3)-linked residues but also constitutes end groups and branch points, as do mannose and/or glucose. Xylose and rhamnose are present as (1→4)- and (1→2)-linked units, respectively. Galacturonic acid is assumed to be (1→4)-linked with some branching at 3 position. The amino acid hydroxyproline, present in the glycoprotein of South African *Cannabis* leaves, was absent in the corresponding Thailand material.

INTRODUCTION

It has been a matter of debate whether *Cannabis sativa* L. is one or more species [1]. *Cannabis* plants originating from different parts of the world have a somewhat different appearance and also contain different proportions of secondary metabolites like the cannabinoids. From a chemotaxonomic viewpoint it would therefore be of interest to examine and compare compounds of primary importance to the plant, since basic chemical differences in such constituents would strongly indicate species differences.

In the preceding papers [2, 3] the leaf glycoprotein of South African type *Cannabis* was isolated and characterized, and the present work is a corresponding study on the leaf glycoprotein from plants of Thailand origin.

RESULTS AND DISCUSSION

In conformity with previous work on glycoprotein from *Cannabis* South Africa [2, 3] the present material was also isolated by aqueous extraction, without any pretreatment with enzymes or alkali. The extract was separated into two fractions by chromatography on DEAE-cellulose. Fraction A $[\alpha]_D^{20}$ − 12°, passed unretarded through the DEAE-cellulose column in 0.05 M Tris–HCl buffer and was obtained in 17.4% yield, whereas fraction B $[\alpha]_D^{20}$ + 14° was eluted with buffer containing 0.25 M NaCl and obtained in 40.9% yield. A and B contained the same sugars as the corresponding glycoprotein fractions of *Cannabis* South Africa: arabinose, galactose, glucose, mannose and xylose; B also contained rhamnose and galacturonic acid.

The IR spectrum of A showed no absorption bands at 1735 and 1250 cm^{-1}, indicating the absence of *O*-acetyl groups, whereas that of B showed very small bands at the above frequencies.

Estimation of MWs was carried out by molecular sieve chromatography on a calibrated column of Sepharose 4B. Calculations based on the elution volume of the peaks of the curve resulted in a MW for the material in the two peaks of fraction A of *ca* 70000 and 16500. Fraction B was also divided partly into two subfractions on the Sepharose column, corresponding to a MW of *ca* 110000 and 28000 (Figs 1a and b). Since the partial fractionation of A and B achieved on the Sepharose column was due to a difference in MW rather than chemical composition, the original fractions A and B as obtained from the DEAE-cellulose column were used for the further studies.

The residual glycoprotein of B after the first Smith-degradation (B_1) was subjected to incubation with pronase. The digest was chromatographed on a Sephadex G-100 column and the effluent analysed for carbohydrate [4] and protein [5]. From Figs 2a and b it appears that the material was degraded significantly by pronase, the elution pattern (Fig. 2b) being characteristic of a proteolysed glycoprotein.

Table 1 shows the chemical composition of A and B and also of the residual glycoproteins after one (A_1 and B_1) and two (A_2 and B_2) Smith-degradations. The con-

* This investigation was carried out in collaboration with the United Nations Narcotics Laboratory in Geneva, as part of the U.N. Cannabis research programme established by resolution 8 (XIV) of the Commission on Narcotic Drugs.

For Part 2 in this series see ref. [3].

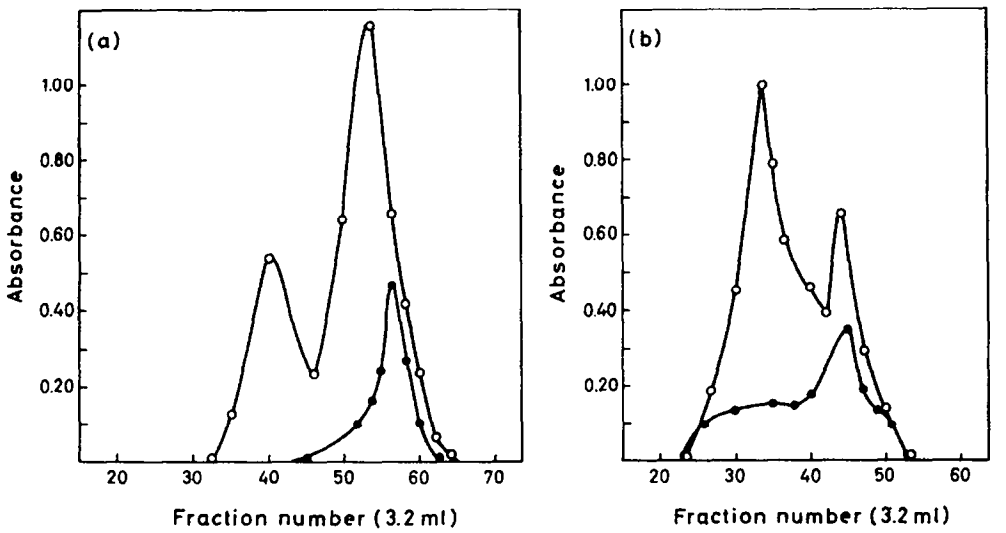

Fig. 1. Gel filtration of (a) fraction A and (b) fraction B on a Sepharose 4B column (37 × 2.5 cm) with 0.025 M Tris–HCl buffer pH 7.2 as the eluting agent. —○—○— carbohydrate —●—●— protein.

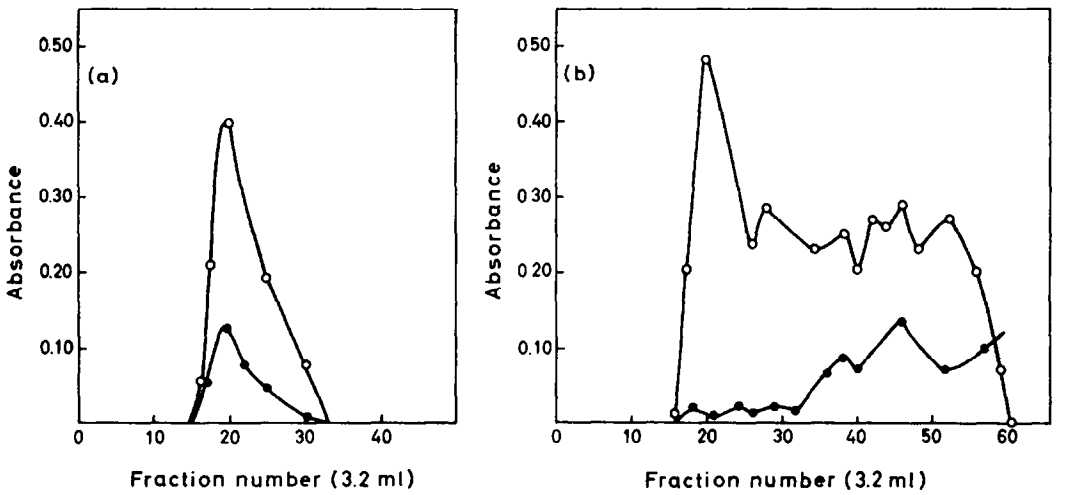

Fig. 2. Gel filtration of (a) fraction B_1 (residual glycoptotein after the first Smith-degradation of B) and (b) pronase-treated fraction B_1 on a Sephadex G-100 column (37 × 2.5 cm) with 0.05 M Tris–HCl buffer pH 7.2 as the eluting agent, —○—○— carbohydrate —●—●— protein.

Table 1. Optical rotation and % composition of fractions A and B

	A	A_1	A_2	B	B_1	B_2
$[\alpha]_D^{20}$	−12°			+14°		
Total carbohydrate	75.0	73.8	71.2	59.3	51.3	55.4
Protein	19.9	16 5	10.4	25.0	32.1	22.6
Uronic acid	—	—	—	18.6	12.5	7 0
Arabinose	23 6	10.9	2.8	12.7	7 5	2.8
Rhamnose	—	—	—	7.3	3.0	1.4
Xylose	1 6	tr	tr	2.5	3.6	0.9
Mannose	11.2	1.7	tr	2.4	1.2	tr
Galactose	25.4	49.7	65.6	20.3	37.9	52.6
Glucose	6.6	2.4	2.5	3.5	1.3	tr
Galacturonic acid	—	—	—	17.3	7.5	3.4

A_1 and B_1: Residual glycoprotein after the first Smith-degradation of A and B.

A_2 and B_2: Residual glycoprotein after the second Smith-degradation of A and B.

sumption of periodate and the yields obtained after the two degradations are summarized in Table 2. The results are similar to those obtained on *Cannabis* glycoproteins grown from South African seeds [2, 3]. Arabinose and galactose are the predominant sugars although galacturonic acid is also a major constituent in B. The sugars are destroyed to a large extent by two consecutive Smith-degradations except for galactose most of which proved resistant to periodate. This sugar is largely located to the inner core region of the polymer, either (1 → 3)-linked, or as branch points.

The low MW alcohols obtained from the Smith-degradations were analysed as their acetates by GLC. Glycerol was present in each fraction whereas erythritol and threitol were both absent. Propylene glycol was presumably present in the dialysable fraction after the second Smith-degradation of B. Propylene glycol would arise from terminal or (1 → 2)-linked rhamnose, while glycerol would originate mainly from arabinofuranose

Table 2. Consumption of periodate and yields (mg) obtained after the first and the second Smith-degradation of fractions A and B

	A	B
Consumption of IO_4^- in first Smith-degradation	0.60	0.60*
Consumption of IO_4^- in second Smith-degradation	0.35	0.28
Material before first Smith-degradation	160	160
Residual glycoprotein after first Smith-degradation	27	52
Material before second Smith-degradation	20	45
Residual glycoprotein after second Smith-degradation	6	16
Dialysable fraction after first Smith-degradation	57	38
Dialysable fraction after second Smith-degradation	10	17

* Mole IO_4^-/anhydrohexose unit.

and hexopyranose end groups. As mentioned previously [3] methylation of uronic acid-containing polysaccharides may lead to β-elimination in a strongly alkaline medium and subsequent degradation by acid hydrolysis. The galacturonic acid-containing fraction B was divided into two after methylation. One half was carboxyl-reduced before hydrolysis, reduction and acetylation, whereas the other half was not carboxyl-reduced. The results obtained by GC–MS analysis (Table 3) show that arabinofuranose end groups make up a high proportion (*ca* 21 %) of fraction A and a somewhat smaller part (*ca* 8 %) of B. 2,3-Di-*O*-methylarabinitol, the major methylated product from A (*ca* 26 %), originates from (1 → 4)-linked arabinopyranose and/or (1 → 5)-linked arabinofuranose. Xylose occurs as (1 → 4)-linked xylopyranose while rhamnose, only present in B, is (1 → 2)-linked and also represents branch points. This corresponds well with

Table 3. Methylation analysis of fraction A, B, and B after carboxyl-reduction of the methylated polymer (B')

RTMG	Alditol acetate	Primary fragments MS *m/e*				Area in % of total peak area A	B	B'
0.51	2,3,5-Tri-*O*-methylarabinitol	45	118	161		20.7	8.4	
0.63	2,3,4-Tri-*O*-methylarabinitol	117	118	161	162	2.0	0.9	
0.84	3,5-Di-*O*-methylarabinitol	45	161	190		2.0	1.9	
0.90	2,5-Di-*O*-methylarabinitol	45	118	233		3.6	5.4	
0.90	3,4-Di-*O*-methylrhamnitol	131	190			—		
1.00	2,3,4,6-Tetra-*O*-methylglucitol/mannitol	45	118	161	162 205	3.1	1.6	
1.07	2,3-Di-*O*-methylarabinitol	118	189			25.8	10.2	
1.13	2,3,4,6-Tetra-*O*-methylgalactitol	45	118	161	162 205	4.6	4.9	
1.18	2,3-Di-*O*-methylxylitol	118	189			3.1	3.0	
1.38	2-Mono-*O*-methylrhamnitol	118				—	1.1	
1.55	3-Mono-*O*-methylrhamnitol	190	203			—	1.2	
1.63	2,4,6-Tri-*O*-methylglucitol	45	118	161	234	4.1	2.6	
1.74	2,4,6-Tri-*O*-methylmannitol	45	118	161	234	12.1	9.2	
1.78	2,4,6-Tri-*O*-methylgalactitol	45	118	161	234		7.7	
1.89	2,3,6-Tri-*O*-methylgalactitol	45	118	162	233	—	2.9	
1.96	2,3,6-Tri-*O*-methylglucitol	45	118	162	233	5.4	1.0	
2.30	Degr. product of galacturonic acid					—	26.0	12.9
2.36	2,3,4-Tri-*O*-methylgalactitol	118	162	189	233	2.7	4.4	6.3
3.63	2,3-Di-*O*-methylgalactitol	118	261			—	0.4	1.7
3.87	2,4-Di-*O*-methylgalactitol	118	189			6.8	4.7	4.8
4.72	2-Mono-*O*-methylgalactitol	118				3.8	2.6	3.1

RTMG: R_t relative to 1,5-Di-*O*-acetyl-2,3,4,6-tetra-*O*-methylglucitol.

the formation of propylene glycol by the second Smith-degradation of B.

The small amount formed of 2,3,6-tri-*O*-methylhexitols indicates a very low content of (1 → 4)-linked hexose units. This is in agreement with the absence of erythritol and threitol in the low MW alcohol fractions obtained by Smith-degradation. The high proportion of (1 → 3)-linked galactose residues in both fraction A and B is in accord with the resistance to periodate demonstrated for galactose by Smith-degradation.

By carboxyl-reduction of galacturonic acid in fraction B, the amount of galactose would increase. Table 3 shows a significant increase in the proportion of 2,3,4-tri-*O*-methyl-, 2,3-di-*O*-methyl, and 2-mono-*O*-methylgalactitol in the methylated carboxyl-reduced fraction B (B') compared to the methylated non-reduced fraction B. This indicates that the galacturonic acid occurs as end groups and as (1 → 4)-linked units with some branching at C_3. Apart from constituting end groups glucose and mannose are present as (1 → 3)-linked and glucose also as (1 → 4)-linked residues.

The glycoprotein fraction B of *Cannabis* Thailand is devoid of hydroxyproline (Table 4), which is a common component of plant glycoproteins. Fraction A and B were analysed for hydroxyproline by the spectrophotometric method of Blumenkrantz and Asboe-Hansen [6] and both glycoprotein fractions were found to be entirely free of this amino acid. For this reason the degradation procedure with $Ba(OH)_2$, as proposed by Lamport [7] for detection of glycopeptide linkages between hydroxyproline and arabinose, was omitted.

Two methods were employed to establish if the glycoprotein contained the serine-*O*-galactoside type linkage. In alkaline solution containing $NaBH_4$ such a bond will be cleaved with the simultaneous conversion of the sugar to the corresponding alditol. As with the results obtained previously [3] galactitol was recovered as the only alditol by GLC.

Alkaline hydrolysis in the presence of Na_2SO_3 converts the glycosidically linked serine and/or threonine to cysteic acid and 2-amino-3-sulphonyl butyric acid, respectively. TMSi-cysteic acid was the only sulphonic acid identified by GLC by the method used. Thus it can be concluded that a serine-*O*-galactosidic bond represents the mode of linkage between protein and carbohydrate in the present glycoprotein.

The experiments clearly demonstrate the complexity of the carbohydrate structure in the present glycoprotein. A definite structure has not been proposed although it is obvious that arabinofuranose is responsible for the main

Table 4. Amino acids in fraction B (μg/mg)

Asp	14.5
Glu	17.4
Thr	9.5
Ser	10.6
Pro	7.7
Gly	9.4
Ala	11.4
Val	7.3
Met	2.6
Ile	4.4
Leu	8.3
Tyr	3.4
Phe	2.9

part of the end groups as well as it is constituting (1→4)-linked arabinopyranose and/or (1→5)-linked arabinofuranose units in the chain. Galactose is located to the inner part of the molecule, predominantly (1→3)-linked with branching to some extent at the other positions.

A distinct difference between the South African and the Thailand type *Cannabis* glycoprotein occurs with the absence of hydroxyproline in the latter material. Also there are minor differences in the quantitative composition of the various sugars. However, the qualitative and quantitative similarities between the two *Cannabis* glycoproteins are far more striking than are the differences. Thus the results obtained are regarded as being consistent with the general concept that the species *Cannabis sativa* comprises a number of subspecies exhibiting minor variations, also in chemical composition.

EXPERIMENTAL

Materials. Cannabis sativa L. (Thailand) was grown and identified as described previously [3]. The dried leaves were milled and extracted with $CHCl_3$–MeOH–H_2O (10:10:1) and (30:20:1.5) consecutively at 50° for 1 hr and finally with $CHCl_3$–MeOH (4:1) at 50° for 1 hr. Other materials were obtained as in [2] and [3].

Aqueous extraction and purification. The pretreated plant material (90 g) was suspended in H_2O (2 l.), stirred for 3 hr at 60° and filtered by suction. The procedure was repeated and the two filtrates combined, concd and dialysed. The crude polymer was recovered by freeze-drying. Yield: 0.9%. The lyophilized material (0.4 g) was separated into two fractions (A and B) on a column (45 × 3 cm) of DEAE-cellulose DE 52 (Cl form) as previously described [2]. Yield A: 17.4%, B: 40.9%.

General methods. The analytical methods used are given in [2] and [3]. Uronic acid was estimated according to ref. [8] and hydroxyproline by the method of ref. [6] with 4-hydroxyproline as standard.

Gel filtration. The sample (40 mg) in 0.025 M Tris–HCl buffer pH 7.2 was applied on a column (37 × 2.5 cm) of Sepharose 4B. The column was eluted with the same buffer; 3.2 ml fractions being collected and analysed for carbohydrate and protein. For estimation of MW the column was calibrated with dextrans of known MW.

Incubation with pronase. The experiment was carried out as in [3] but the digest was chromatographed on a Sephadex G-100 column (37 × 2.5 cm).

Smith-degradation. The sample (160 mg) was oxidised with 25 mM $NaIO_4$ (100 ml) at 5° in the dark for 24 hr. The consumption of periodate was followed spectrophotometrically at 223 nm [9]. The reaction was stopped by addition of ethylene glycol and the mixture dialysed for 18 hr. $NaBH_4$ (50 mg) was added and the mixture left at room temp. for 6 hr. The reaction mixture was acidified with HOAc and the polyol of A and B, respectively, dialysed for 16 hr and freeze-dried. Yield A: 100 mg, B: 106 mg. The polyols were hydrolysed with 0.05 N H_2SO_4 (5 ml) for 70 min at 80°. After neutralisation with $Ba(OH)_2$ (40 mg), centrifugation and dialysis of the supernatant for 5 hr, the retentate and dialysate fractions were concd. The dialysable fractions were deionized on a Dowex 50 × 8 (20–50 mesh) H^+ column and the eluate evapd to dryness. The retentate fractions were freeze-dried. For the second Smith-degradation the 3 steps (oxidation–reduction–mild hydrolysis) were repeated. The residual glycoproteins obtained (retentate fractions) were analysed by GLC under the conditions given in [3] following methanolysis and trimethylsilylation of the resulting Me glycosides [2]. The low MW alcohol fractions (dialysable fractions) were likewise analysed as acetylated derivatives on the 3% OV-225 column (400 × 0.2 cm) at 130° followed by an increase of 2°/min to 210°.

Methylation analysis and alkaline hydrolysis in the presence of $NaBH_4$ and Na_2SO_3, respectively. These methods were carried out as described in [3].

Acknowledgement—The authors are indebted to Institute of Clinical Biochemistry, Rikshospitalet, Oslo, for carrying out the GC–MS.

REFERENCES

1. Small, E. and Cronquist, A. (1976) *Taxon.* **25**, 405.
2. Hillestad, A., Wold, J. K. and Smestad Paulsen, B. (1977) *Carbohydr. Res.* **57**, 135.
3. Hillestad, A. and Wold, J. K. (1977) *Phytochemistry* **16**, 1947.
4. Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. and Smith, F. (1956) *Anal. Chem.* **28**, 350.
5. Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265.
6. Blumenkrantz, N. and Asboe-Hansen, G. (1973) *Anal. Biochem.* **55**, 288.
7. Lamport, D. T. A. (1967) *Nature* **216**, 1322.
8. Bitter, T. and Muir, H. M. (1962) *Anal. Biochem.* **4**, 330.
9. Aspinall, G. O. and Ferrier, R. J. (1957) *Chem. Ind. (London)* 1216.