Chemistry of Natural Compounds, Vol. 33, No. 6, 1997

COMPARATIVE INVESTIGATION OF THE MOLECULAR STRUCTURES OF THE SPECIES-SPECIFIC PROTEINS OF THE SEEDS OF *Gossypium hirsutum* AND *G. barbadense*

A. M. Érmatov, D. A. Khashimov, and P. Kh. Yuldashev

UDC 547.96:582.796

The N-terminal amino acids and the amino acid compositions of four species-specific proteins of two species of the cotton plant — Gossypium hirsutum *and* G. barbadense — *have been determined. Peptide maps of tryptic hydrolysates of these proteins have been obtained and compared. Individual tryptic peptides have been isolated and purified. The N-ends, amino acid compositions, and amino acid sequences of individual peptides have been determined. The N-terminal sections of the proteins of* G. hirsutum *and* G. barbadense *as far as the 16th amino acid residue have been sequenced. A comparative analysis has been made of the peptides and proteins of these cottonplant species.*

We have previously carried out a comparative study of the physicochemical [1] and immunochemical [2] properties of species-specific proteins of the seeds of two species of the cotton plant, *G. hirsutum* (variety Tashkent-1) and *G. barbadense* (variety S-6030).

In the present paper we give the results of a comparative investigation of the amino acid compositions of the proteins and peptides and the primary structures of individual tryptic peptides of the H.0.43 and H.0.51 proteins of *G. hirsutum* and B.0.37 and B.0.48 of *G. barbadense*.

In the final stage of purification of the proteins from these species we used electrophoresis in 12% PAAG in the presence of sodium dodecyl sulfate (DDS-Na) according to Laemmli [3], the transfer of the proteins to a poly(vinylidene fluoride) (PVDF) membrane being conducted by Matsudaiza's procedure [4].

The degrees of homogeneity of the *G. hirsutum* proteins H.0.43 and H.0.51 and the *G. barbadense* proteins B.0.37 and B.0.48 were established by determining their N-terminal amino acid residues in the form of their 1-dimethylamino-naphthalene-5-sulfonyl (DNS) derivatives [5]. The analytical results showed that each of the proteins under investigation had only one free α-amino group and that the N-terminal amino acid residue of each of the four proteins was isoleucine (Ile). The C-terminal amino acids of the species-specific proteins of *G. hirsutum* and *G. barbadense* were determined with the aid of carboxypeptidases A, B, and Y [6] and also by hydrazinolysis [7] and each was identified for the H.0.43 and H.0.51 proteins of *G. hirsutum* and the B.0.37 and B.0.48 proteins of *G. barbadense* as aspartic acid.

By automatic gas-phase Edman degradation we established partial N-terminal amino acid sequences of the *G. hirsutum* H.0.51 and *G. barbadense* B.0.48 proteins:

H.0.51 10 15
Ile-Ala-Gly-Gly-Tyr-Ile-Asp-Ser-Gln-Gln-Gln-Gln-Leu-Glu-Lys-Met-
B.0.48
Ile-Ala-Gly-Gly-Tyr-Ile-Asp-Ser-Gln-Gln-Gln-Gln-Leu-Glu-Lys-Met-

The results obtained showed the complete identity of the N-terminal amino acid sequences of the two proteins as far as the 16th amino acid residue.

A comparison of the amino acid sequences found and those of a computer data bank of protein amino acid sequences revealed no clear homology with other known protein amino acid sequences. It followed from this that the proteins under study were new, not previously investigated.

Institute of the Chemistry of Plant Substances, Academy of Sciences of the Republic of Uzbekistan, Tashkent, fax (3712) 40 64 75. Translated from Khimiya Prirodnykh Soedinenii, No. 6, pp. 858-864, November-December, 1997. Original article submitted April 28, 1997.

 0009-3130/97/3306-0670$18.00

TABLE 1. Amino Acid Compositions of the *G. hirsutum* Proteins H.0.43 and H.0.51 and the *G. barbadense* Proteins B.0.37 and B.0.48

Amino acid	H. 0.43	H 0.51	B.0.37	B.0.48
Asp	5.2	5.8	5.2	5.2
Thr	3.0	3.5	3.7	3.6
Ser	5.3	5.7	5.9	5.6
Glu	28.0	28.9	30.1	29.7
Pro	1.8	1.3	1.4	1.9
Gly	13.0	13.2	12.2	12.4
Ala	2.2	2.6	2.7	2.7
Val	0.3	0.3	0.4	0.7
Met	9.0	9.2	9.1	9.2
Ile	3.0	3.3	3.5	3.4
Leu	2.3	2.5	3.6	2.9
Tyr	4.5	4.0	3.8	4.0
His	2.2	2.5	2.7	2.5
Lys	5.9	6.0	7.0	6.8
Arg	1.1	1.1	3.8	4.1

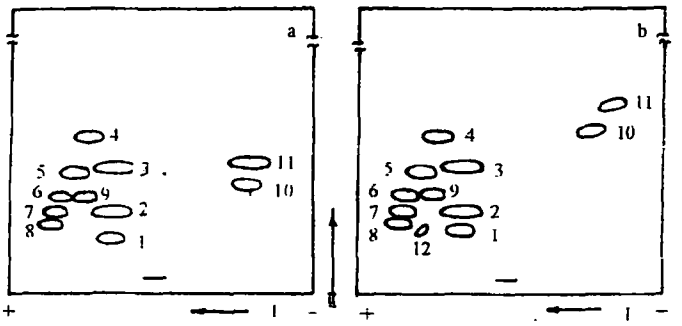

Fig. 1. Peptide maps of tryptic hydrolysates of the *G. hirsutum* (a) and *G. barbadense* (b) proteins: I) Electrophoresis in pyridine–acetate buffer, pH 5.6 at 800 V for 1 h; II) chromatography in the pyridine–butane-1-ol–acetic acid–water (10:15:3:12) system.

The results of a comparative analysis of the amino acid compositions of the *G. hirsutum* proteins H.0.43 and H.0.51 and the *G. barbadense* B.0.37 and B.0.48 proteins (Table 1) showed that they had a high degree of homology. It must be mentioned that in their amino acid compositions the *G. hirsutum* proteins H.0.43 and H.0.51 and the *G. barbadense* B.0.37 and B.0.48 proteins scarcely differed from one another. We may also note the high levels of glutamic acid, glycine, and methionine and the complete absence of such amino acids as phenylalanine and cysteine.

In order to elucidate the differences found between the relative electrophoretic mobilities and isoelectric points of these proteins, we needed to make more far-reaching structural investigations.

On the basis of the results of analysis of amino acid compositions, for the fragmentation of the proteins we chose enzymatic hydrolysis with trypsin. Hydrolysis was conducted in 0.1 N $NaHCO_3$ buffer, pH 8.0, at 37°C for 4 h. After hydrolysis, the peptides were separated by the production of peptide maps on cellulose [8] (Fig. 1). In the first direction we performed horizontal electrophoresis in pyridine buffer, pH 5.6. After the cellulose plate had been dried, chromatography was conducted in the perpendicular direction in the pyridine–butane-1-ol–acetic acid–water system. The dried chromatogram was then sprayed with a 0.2% solution of ninhydrin in acetone.

A comparison of the peptide maps of the dansylated tryptic hydrolysates revealed their clear homology. The tryptic hydrolysates of the *G. hirsutum* proteins H.0.43 and H.0.51 each gave 11 peptides, and the arrangements of the peptides on the peptide maps were absolutely identical when they were superposed on one another. An analogous pattern was observed in a comparison of the maps of the 12 tryptic peptides in each case from the B.0.37 and B.0.48 proteins of *G. barbadense*. From these results and those on the amino acid compositions of the proteins it could be concluded that the *G. hirsutum* proteins, and, respectively, the *G. barbadense* proteins were identical and that these peptides could be analyzed without their preliminary separation.

TABLE 2. Amino Acid Compositions of Individual Tryptic Peptides of Proteins of the Cottonplant Species *G. hirsutum*

Amino acid	Tryptic peptides						
	H-t-3	H-t-4	H-t-5	H-t-6	H-t-9	H-t-10	H-t-11
Asp		0.9(1)				1.1(1)	2.1(2)
Thr		0.7(1)		0.8(1)		0.9(1)	
Ser		0.8(1)				0.9(1)	0.8(1)
Glu	2.2(2)	2.1(2)	0.9(1)		1.1(1)	3.2(3)	2.9(3)
Pro							
Gly		1.1(1)				1.9(2)	3.1(3)
Ala		1.0(1)			0.9(1)	1.1(1)	
Val				0.9(1)			
Met	0.8(1)	0.7(1)	0.9(1)			0.8(1)	0.9(1)
Ile	0.9(1)					1.1(1)	
Leu		0.9(1)			1.1(1)		0.8(1)
Tyr						0.7(1)	2.8(3)
His		1.2(1)			1.1(1)		
Lys	1.1(1)			0.9(1)	0.9(1)	0.8(1)	1.1(1)
Arg		0.9(1)	1.1(1)				

TABLE 3. Amino Acid Compositions of Individual Tryptic Peptides of Proteins of the Cottonplant Species *G. barbadense*

Amino acid	Tryptic peptides						
	B-t-3	B-t-4	B-t-5	B-t-6	B-t-9	B-t-10	B-t-11
Asp		0.9(1)				1.1(1)	1.9(2)
Thr		0.7(1)		0.8(1)			
Ser		0.8(1)				0.9(1)	0.9(1)
Glu	2.1(2)	2.2(2)	1.2(1)		1.1(1)	3.8(4)	4.1(4)
Pro							
Gly		0.9(1)				2.9(3)	2.1(2)
Ala		0.8(1)			0.9(1)		
Val				0.8(1)			
Met	0.9(1)	0.7(1)	0.8(1)			0.9(1)	0.8(1)
Ile	1.1(1)						
Leu		0.9(1)			1.1(1)	1.0(1)	0.9(1)
Tyr						0.9(1)	2.9(3)
His		1.1(1)			0.9(1)		1.1(1)
Lys	0.9(1)			1.1(1)	1.0(1)	0.8(1)	0.9(1)
Arg		0.9(1)	1.1(1)				

On the peptide maps of *G. hirsutum* and *G. barbadense* it was observed that the positions of 9 out of the 11 peptides of the *G. hirsutum* proteins coincided completely with the positions of 9 out of the 12 peptides of the *G. barbadense* proteins. The positions of peptides H-t-10 and H-t-11 from *G. hirsutum* differed from those of the *G. barbadense* peptides B-t-10 and B-t-11, while peptide B-t-12 was absent from the peptide map of the proteins of *G. hirsutum*. It may be concluded that the amino acid compositions and, consequently, the amino acid sequences of peptides H-t-10 and H-t-11 differed from those of peptides B-t-10 and B-t-11.

To determine arginine-containing peptides, the peptide maps were sprayed with a solution of α-naphthol in ethanol according to Sakaguchi [8]. There were five spots stained pink on each of the peptide maps of the *G. hirsutum* and *G. barbadense* proteins.

The peptide spots were scraped off and extracted with 50% pyridine. The peptides were dried in a rotary evaporator until the last traces of pyridine had been eliminated. They were then purified by chromatography on FN paper in the pyridine–butan-1-ol–acetic acid–water system. Some of the peptides were additionally purified by paper electrophoresis in pyridine acetate buffer, pH 5.6. The N-terminal amino acid residues of the peptides were determined by the dansyl method. Amino acid compositions were analyzed on an amino acid analyzer after the acid hydrolysis of the peptides (Tables 2 and 3). The amino acid sequences of the peptides were determined by Edman's method on a gas-phase sequenator.

Amino acid sequences of peptides H-t-5 and B-t-5:

H-t-5 Gln-Met-Arg
B-t-5 Gln-Met-Arg

Amino acid sequences of peptides H-t-6 and B-t-6:

H-t-6 Val-Thr-Lys
B-t-6 Val-Thr-Lys

Amino acid sequences of peptides H-t-9 and B-t-9:

H-t-9 Gln-Ala-His-Leu-Lys
B-t-9 Gln-Ala-His-Leu-Lys

Amino acid sequences of peptides H-t-10 and B-t-10:

H-t-10 Tyr-Met-Glu-Glu-Ala-Ser-Gly-Asn-Glu-Gly-Thr-Ile-Lys
B-t-10 Tyr-Met-Glu-Glu-Glu-Leu-Gly-Gly-Glu-Gly-Ser-Asp-Lys

Amino acid sequence of peptide H-t-11:

H-t-11 Tyr-Gly-Met-Glu-Glu-Leu-Gly-Asp-Asn-Tyr-Glu-Gly-Tyr-Ser-Lys

A comparison of the results obtained from peptide maps of tryptic hydrolysates of *G. hirsutum* and *G. barbadense* proteins and of the amino acid compositions and amino acid sequences of the peptides investigated showed that peptides arranged similarly on the peptide maps, such as H-t-5 and B-t-5, H-t-6 and B-t-6, and H-t-9 and B-t-9, had the same amino acid compositions and sequences, while peptides H-t-10 and B-t-10, and H-t-11 and B-t-11, with dissimilar positions on peptide maps, differed from one another in their amino acid compositions and sequences, even though only slightly. These amino acid substitutions in peptides H-t-10 and H-t-11 of the *G. hirsutum* proteins as compared with peptides B-t-10 and B-t-11 of the *G. barbadense* proteins have probably led to an alteration in their helical sections and in the packing of their tertiary structures that have resulted in changes in the surface charges of the protein molecules and, consequently, to differences in their relative electrophoretic mobilities and their isoelectric points [1].

Thus, the species-specific proteins of *G. hirsutum* and *G. barbadense* that have been investigated are similar in many of their characteristics and are apparently very close in their phylogenetic relationship, have common ancestors, and at some stage of evolution, having suffered genetic changes, have produced forms from which the cottonplant species under study have developed.

EXPERIMENTAL

Purification of the *G. hirsutum* and *G. barbadense* Proteins. The electrophoresis of the proteins was performed under Laemmli's denaturing conditions [3] on a Midget model 2050 apparatus (LKB, Sweden) using vertical plates (100 × 82 × 1.5 mm). After electrophoresis, the gel was soaked in the buffer for transferring the proteins from the gel to a PVDF membrane (10 mM 3-cyclohexylaminopropane-1-sulfonic acid (CAPS), 10% methanol, pH 11.0) for 5 min to decrease the amounts of Tris and glycine. The PVDF membrane was washed with 100% methanol and was immersed in the buffer for transfer. The gel was placed between a sheet of PVDF membrane and several sheets of blotting paper and was installed in an electroblotting instrument (Midget, LKB, Sweden) and electroeluted at 0.5 A in the transfer buffer for 20 min. Then the PVDF membrane was washed with deionized water for 5 min, stained with 0.1% Coomassie Brilliant Blue R-250 (Bio-Rad, USA) in 50% methanol for 5 min, and washed with 50% methanol in 10% CH_3COOH for 10 min at room temperature. Finally, the membrane was rinsed with deionized water for 5-10 min, dried in the air, and stored at −20°C. Transfer buffer: CAPS, 2.21 g; methanol, 100 ml; deionized water to 1 liter.

Determination of N-Terminal Amino Acid Residues. The N-terminal amino acid residues of the proteins and peptides were determined in the form of DNS derivatives, followed by two-dimensional chromatography in a thin layer of silica gel [5] or on Micropolyamid F 1700 polyamide plates (Schleicher und Schull, FRG) [9].

To samples* of the proteins or peptides were added 3 μl of 0.1 N $NaHCO_3$ and 3 μl of DNS-Cl (Serva, FRG) in acetone (6 mg/ml), pH 8.0. The reaction was conducted at 37°C for 1 h, the residue after evaporation was treated with 30 μl of 5.7 N HCl (twice redistilled over $SnCl_2$), the mixture was sealed into a tube under vacuum, and hydrolysis was performed at 110°C for 16 h. The hydrolysate was evaporated in a rotary evaporator until the last traces of HCl had been eliminated.

For two-dimensional chromatography we used 6 × 6 cm plates coated with type KSK silica gel from the Salavat factory, with the following solvent systems: 1) acetone–isopropanol–25% ammonia (9:7:0.5); 2) acetone–isopropanol–25% ammonia (9:7:0.7); and 3) chloroform–benzyl alcohol–ethyl alcohol–CH_3COOH (6:4:5:0.2).

For the polyamide plates we used the following systems: 1) 1.5% HCOOH; 2) benzene–CH_3COOH (9:1); and 3) ethyl acetate–methanol–CH_3COOH (20:1:1).

Determination of C-Terminal Amino Acids. Carboxypeptidase Method. To 500 pmole of a protein in 0.1 M N-ethylmorpholine acetate buffer, pH 8.5, was added 0.05 μg of carboxypeptidase B, previously dissolved in the same buffer, in a ratio of enzyme to substrate of 1:500. The reaction mixture was kept at 37°C for 30 min and was then evaporated. The amino acids split out were analyzed on a Waters amino acid analyzer.

Hydrazinolysis. A mixture of 2 nmole of a protein and 50 μl of anhydrous hydrazine was sealed into a tube under vacuum. Hydrazinolyis was conducted at 105°C for 8 h, after which the contents of the tube were evaporated. The amino acid hydrazides were extracted from aqueous solution (20 μl in each case) with freshly distilled benzaldehyde. The aqueous layer was transferred to another tube and was twice extracted with benzene to eliminate benzaldehyde residues. The C-terminal amino acid was determined in the form of the DNS derivative.

Determination of Amino Acid Compositions. A mixture of 200 pmole of a protein or 500 pmole of a peptide and 50 μl of 5.7 N HCl was sealed under vacuum and hydrolysis was conducted at 110°C for 20 h. The hydrolysate was evaporated in a rotary evaporator, until the last traces of HCl had been eliminated and was analyzed on a D 550 amino acid analyzer (Durrum, USA).

Tryptic Hydrolysis of the Proteins. Trypsin (Sigma, USA) was added to a solution of 10 nmole of a protein in 0.1 N $NaHCO_3$, pH 8.0, in a substrate–enzyme ratio of 50:1. Hydrolysis was conducted at 37°C for 4 h, and the products were lyophilized.

Separation of the Tryptic Hydrolysates by the Peptide Map Method. A solution (20 μl) containing 10 nmole of the peptides of a tryptic hydrolysate was deposited in the middle of a 20 × 20 cm plate with a thin layer of Cellulose G1440 (Schleicher und Schull, FRG) 2 cm from the edge in the form of a 1 × 5 mm band perpendicular to the direction of electrophoresis. Pyridine acetate buffer, pH 5.6 (4 ml of pyridine, 1 ml of conc. CH_3COOH, water to 1 liter), was used to wet the layer of cellulose and as the electrode buffer. Horizontal electrophoresis was carried out after the cooling of the stage and the plate to 4°C in an instrument with a cooling stage (Desaga, FRG). Electrophoresis was conducted at 200 V for the first 15 min and 800 V for the following 60 min. After the end of electrophoresis, the plate was dried in a current of air at 40°C in a ventilation chamber. Chromatography was conducted in the perpendicular direction in the pyridine–butan-1-ol–CH_3COOH–water (10:15:3:12) system at room temperature. After drying, the peptide map was sprayed with a 0.2% solution of ninhydrin in acetone and was left in the dark. The colored spots were scraped off and the peptides were extracted with 50% pyridine. The extracts were evaporated and the residues were dried in a rotary evaporator until the last traces of pyridine had been eliminated. The homogeneity of the peptides isolated was shown by a determination of N-terminal amino acid residues.

Color Reaction for Arginine-containing Peptides. The following solutions were prepared: 1) a 0.01% solution of α-naphthol in 96% ethanol containing 5% of urea; and 2) a 5% solution of sodium hypobromite (2 g of bromine in 100 ml of 5% NaOH).

After the preparation of a peptide map, the plate was sprayed with the α-naphthol solution and was dried under a current of air. Then it was sprayed with the solution of bromine in caustic soda and dried. Arginine-containing peptides were colored pink.

Determination of the Primary Structures of Proteins and Peptides. The N-terminal amino acid sequences of proteins and the primary structures of peptides were determined by Edman's method on a gas-phase sequenator 470 (Applied Biosystems, USA), with detection of PTH derivatives by means of a 120 A analyzer (USA).

The authors express their gratitude to O. Yu. Chertov (M. M. Shemyakin–Yu. A. Ovchinnikov Institute of Biorganic Chemistry, Russian Academy of Sciences) for assistance in the performance of the structural investigations.

*No amounts given — Translator.

REFERENCES

1. A. M. Érmatov, D. A. Khashimov, and P. Kh. Yuldashev, Khim. Prir. Soedin., 736 (1995).
2. A. M. Érmatov, D. A. Khashimov, and P. Kh. Yuldashev, Khim. Prir. Soedin., 113 (1996).
3. U. K. Laemmli, Nature (London), **227**, 680 (1970).
4. P. Matsudaiza, J. Biol. Chem., **262**, No. 21, 10035 (1987).
5. B. G. Belen'kii, É. S. Gankina, and V. V. Nesterov, Dokl. Akad. Nauk SSSR, **172**, No. 1, 91 (1967).
6. R. P. Ambler, Methods Enzymol., **11**, 155 (1967).
7. A. Akabori, K. Ohno, T. Ikenaka, Y. Okada, H. Hanafusa, I. Haruna, K. Tsugita, K. Sugae, and T. Matsushima, Bull. Chem. Soc., Jpn., **29**, 507 (1956).
8. J. L. Bailey, Techniques in Protein Chemistry (1st ed.), Elsevier, Amsterdam (1962) [Russian translation (1965), p. 27].
9. A. A. Arutyunyan, E. C. Severin, and Ya. M. Varshavskii, Biokhimiya, **40**, 878 (1975).