The primary structures of the A_4 and A_5 subunits are highly homologous to that of the A_3 subunit in the glycinin seed storage protein of soybean

Hisashi Hirano, Chikafusa Fukazawa[+] and Kyuya Harada[°]

Department of Molecular Biology, National Institute of Agrobiological Resources, [+]National Food Research Institute, and *[°]National Institute of Agro-Environmental Sciences, MAFF, Yatabe, Tsukuba, Ibaraki, 305 Japan*

Received 31 October 1984

The A_5 subunit of the glycinin seed storage protein of soybean was purified by ion-exchange chromatography followed by preparative SDS-gel electrophoresis and the complete amino acid sequence was determined. The A_5 subunit is composed of 97 amino acids which correspond to an M_r of ~ 10600. Approx. 86% of the sequence of the A_5 subunit was found to be identical to that of the NH_2-terminal region of the A_3 subunit of the glycinin determined previously. Also, the A_4 subunit of the glycinin was purified and the partial sequence was determined. The sequence of the A_4 subunit was estimated to be highly homologous to that of the COOH-terminal region of the A_3 subunit.

Soybean Glycine max *Seed storage protein* *Glycinin* *Amino acid sequence*

1. INTRODUCTION

Seeds of the soybean (*Glycine max* (L.) Merrill) contain two major storage proteins, glycinin (11 S globulin) and β-conglycinin (7 S globulin). The glycinin is the most abundant protein in the seeds and has a molecular mass of ~360 kDa [1]. It is considered to be composed of six pairs of subunits, each pair consisting of acidic (37–45 kDa) and basic (~22 kDa) subunits with different isoelectric points [2]. Staswick and Nielsen [3] have identified seven acidic (A_{1a}, A_{1b}, A_2, A_3, A_4, A_5 and A_6) and five basic (B_{1a}, B_{1b}, B_2, B_3 and B_4) subunits of the glycinin. Recently, we have determined the complete amino acid sequence of the A_3 subunit by protein-sequence analysis [4]. However, the complete amino acid sequences of the other subunits have not been established.

Here, we purified the A_4 and A_5 subunits and determined the complete amino acid sequence of the A_5 subunit and the partial sequence of the A_4 subunit. We compared the sequences of these subunits to that of the A_3 subunit and found the highly homologous sequence regions between these subunits.

2. MATERIALS AND METHODS

Glycinin was purified from dry mature seeds of the soybean (*G. max* (L.) Merr. cv. Bonminori) as described in [2]. Reduction and carboxymethylation were carried out as described in [5].

The A_5 subunit of glycinin was partially purified by ion-exchange chromatography on a DEAE-Sephadex A-50 (Pharmacia) column (2.6 cm diam., 200 ml) equilibrated with 0.1 M phosphate buffer (pH 6.3) containing 6 M urea and 0.2 M β-mercaptoethanol [6]. The partially purified A_5 subunit was purified by preparative SDS–polyacrylamide gel (16%) electrophoresis (SDS–PAGE) according to Laemmli [7]. After electrophoresis, the slice of the gel containing the A_5 subunit band was excised and the A_5 subunit was removed from the excised gel slices by electroelution [8]. Portions (10–100 μg) of the purified A_5 subunit were digested with cyanogen bromide

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

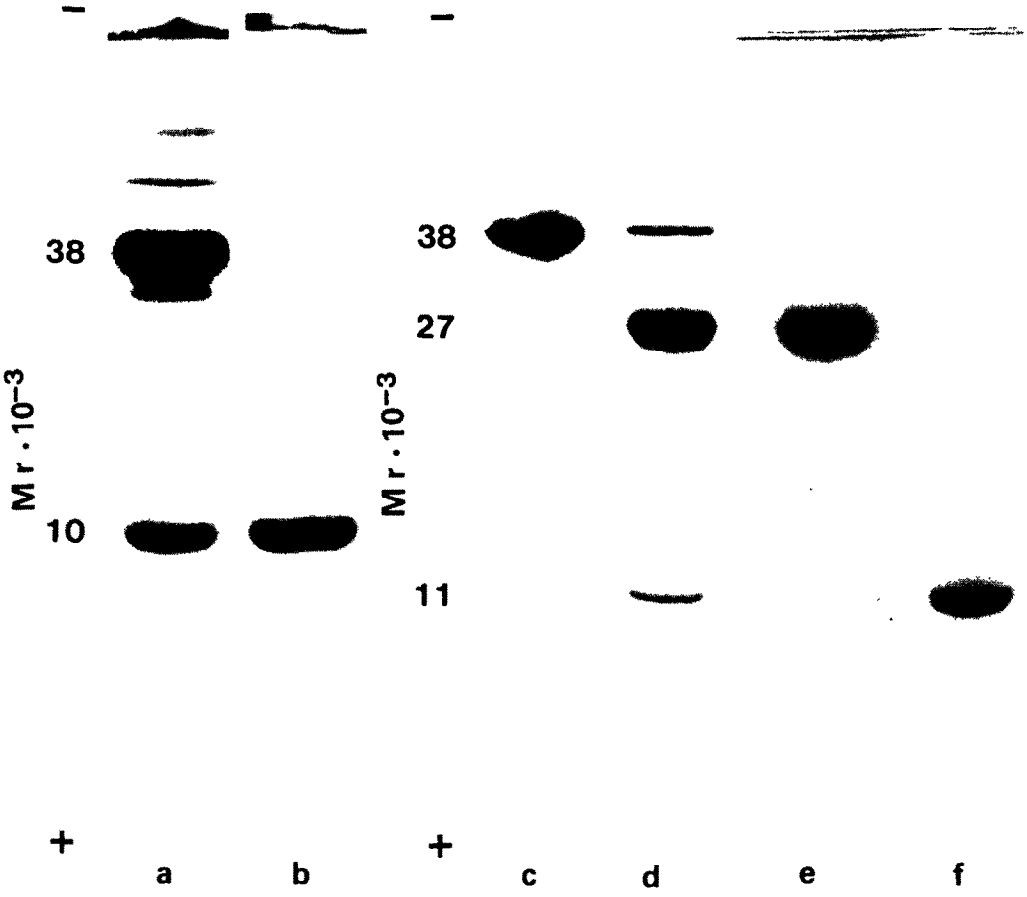

Fig.1. SDS–PAGE of the purified A_4 and A_5 subunits, and CNBr fragments of the A_4 subunit of glycinin. (a) A_5 subunit partially purified by ion-exchange chromatography; (b) A_5 subunit; (c) A_4 subunit; (d) whole CNBr digests of the A_4 subunit; (e) A_4CB2; (f) A_4CB1. M_r values were estimated as described previously [4].

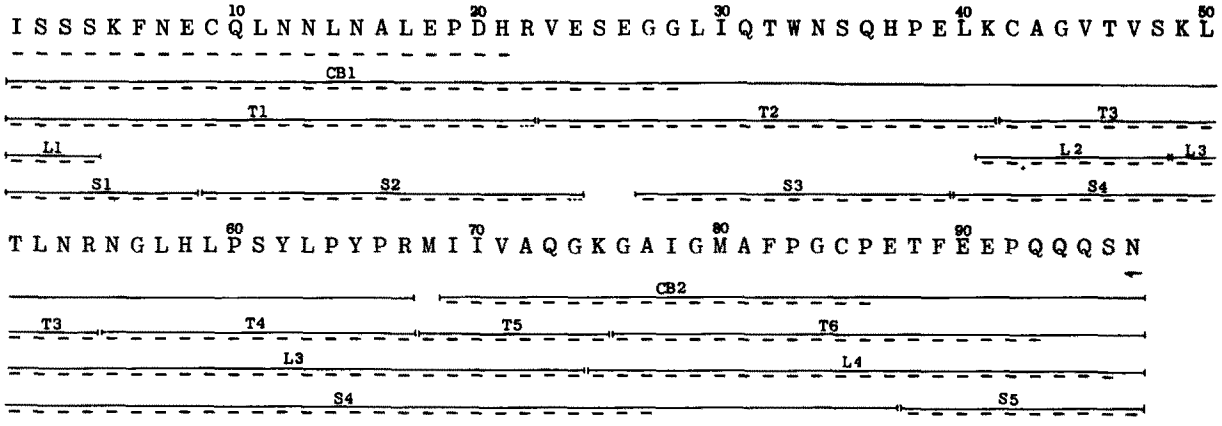

Fig.2. Amino acid sequence of the A_5 subunit of glycinin. CB, T, L and S represent CNBr, tryptic, metalloendopeptidase and Staphylococcal peptide fragments, respectively. Dashed lines represent the NH_2-terminal sequences of the A_5 subunit and the peptide fragments which were identified by DABITC analysis. Residues that were unambiguously identified are indicated with solid lines, others with broken lines. The COOH-terminal amino acid (—) was determined by carboxypeptidase A digestion followed by dansylation and TLC.

(CNBr) (Sigma), trypsin (Sigma) and *Staphylococcus aureus* V8 protease (Pierce) as described in [4], and metalloendopeptidase (Seikagaku Kogyo) (1:50, enzyme:protein in 0.2 M *N*-ethylmorpholine acetate buffer, pH 8.5, 4 h at 37°C) isolated from *Grifola frondosa*. The metalloendopeptidase cleaves specifically the peptide bonds on the NH_2-terminal side of lysine residues.

The A_4 subunit of glycinin was purified by ion-exchange chromatography on a DEAE-Sephadex A-50 column as described above. 50 mg of the purified A_4 subunit was digested with CNBr and the digest was purified by gel filtration on a Sephadex G-100 (Pharmacia) column (1.6 cm diam., 400 ml) as described in [4]. 2 mg of the purified CNBr fragment was digested with trypsin or metalloendopeptidase.

The resultant peptides were dissolved in 100 μl of 0.1% trifluoroacetic acid (TFA) and applied to a reverse-phase column (Nucleosil C18, 5 μ) for high performance liquid chromatography (HPLC) (Varian 5500), and eluted at a flow rate of 1.0 ml/min of 0.1% TFA with linear gradient of acetonitrile. The effluent was monitored by the absorbance at 216 nm.

The amino acid sequences were determined by manual sequence analysis using 4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate (DABITC) (Dojin) [9]. Leucine and isoleucine, which could not be identified by thin-layer chromatography (TLC) on polyamide sheets in DABITC analysis, were identified by TLC on silica gel [10]. The COOH-terminal amino acid of the A_5 subunit was determined by digestion with carboxypeptidase A (Sigma) followed by dansylation and TLC on polyamide sheets [11]. A blocked peptide was treated with pyroglutamate aminopeptidase (Boehringer) as described in [12] and the blocked NH_2-terminal amino acid was removed. The unblocked fragment was sequenced by a gas-phase sequenator (Applied Biosystems).

Amino acid composition of the A_5 subunit was determined as described in [4] with an amino acid analyser (Hitachi 835).

3. RESULTS AND DISCUSSION

The purified glycinin was dissociated and subjected to ion-exchange chromatography in 6 M urea. One of the two fractions eluted with the starting buffer was found to contain a mixture of a few acidic subunits including the A_5 subunit by SDS–PAGE (fig.1(a)). The A_5 subunit in the fraction was purified successfully by preparative SDS–PAGE followed by electroelution (fig.1(b)). On the other hand, one of the fractions eluted with a gradient of NaCl was found to contain a high proportion of the A_4 subunit by SDS–PAGE (fig.1(c)). Fig.1 shows that both of the A_4 and A_5 subunits purified were at least 90% pure.

Fig.2 shows the way in which the complete amino acid sequence of the A_5 subunit was deduced. The subunit contains 97 amino acids which correspond to an M_r of ~10600. The NH_2-terminal sequence of the A_5 subunit presented here is in good agreement with that described in [13]. The amino acid composition of the A_5 subunit is shown in table 1. The sequence shows reasonable agree-

Table 1

Amino acid composition of the A_5 subunit of glycinin

Asx	8.7 (9)
Thr	4.3 (4)
Ser	9.9 (8)
Glx	13.9 (15)
Pro	8.6 (8)
Gly	8.5 (8)
Ala	4.8 (5)
Val	4.4 (4)
Cys[a]	2.7 (3)
Met	1.4 (2)
Ile	4.9 (5)
Leu	10.2 (10)
Tyr	1.7 (2)
Phe	3.4 (3)
His	2.9 (3)
Lys	3.9 (4)
Arg	2.5 (3)
Trp	– (1)
Total	97

[a] Determined as carboxymethylcysteine

Acid hydrolysis was carried out at 105°C for 24, 48 and 72 h. Results are expressed as mol residue/mol, and the values for serine and threonine are corrected by extrapolation to zero time. Other values are average for the different times of hydrolysis. The values in parentheses are the numbers obtained from the sequence

ment with the composition derived by amino acid analysis.

Of the two CNBr fragments of the A_5 subunit purified by HPLC, the A_5CB2 fragment contained a methionine residue, but was not cleaved at the residue with CNBr. The reason is possibly that the A_5 subunit was incompletely cleaved with CNBr.

As compared in fig.3, the amino acid sequence of the A_5 subunit is closely similar to that of the NH_2-terminal region of the A_3 subunit determined previously [4]. In this region, only ~14% of the amino acid sequence of the A_5 subunit is different from the NH_2-terminal sequence of the A_3 subunit.

Two fragments, A_4CB1 and A_4CB2 were obtained by the CNBr digestion of the A_4 subunit (fig.1(d)). These CNBr fragments were purified by gel filtration (fig.1(e,f)) and their NH_2-terminal sequences were determined. The NH_2-terminal sequence of the A_4CB1 is consistent with that of the A_4 subunit [4]. Therefore, it is considered that the order of the CNBr fragments in the single chain of the A_4 subunit is as follows: H_2N-

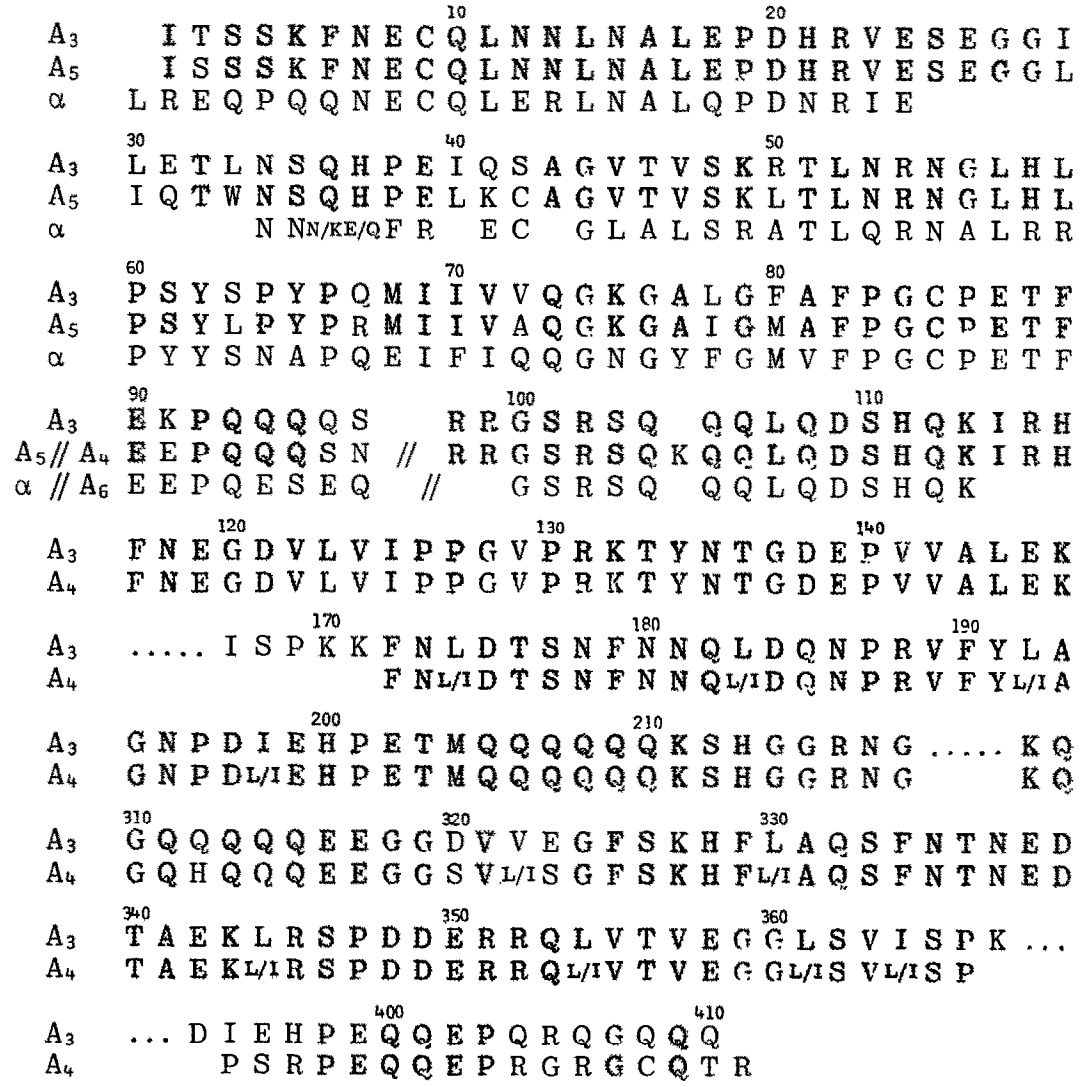

Fig.3. Structural homology among amino acid sequences of the A_3, A_4, A_5 and A_6 subunits of glycinin and the α subunit of pea legumin. A_3, A_3 subunit [4]: A_4, A_4 subunit A_5, A_5 subunit; A_6, A_6 subunit [3]; α,α subunit of legumin [14].

A_4CB1-A_4CB2-COOH.

The NH_2-terminal amino acid of the A_4CB2 isolated was blocked. The A_4CB2 was ultimately shown to be blocked by pyroglutamate formed in the course of cleavage with CNBr as well as the CNBr fragment CB3 of the A_3 subunit of glycinin as described previously [4], since the A_4CB2 was deblocked with pyroglutamate aminopeptidase.

We have found that the NH_2-terminal sequence of the A_4 subunit examined is closely similar to that of the internal region of the A_3 subunit and there is only one difference in the homologous regions: the amino acid residues in the positions of the A_3 subunit corresponding to the eighth residue (Lys) in the A_4 subunit are deleted [4]. Here, we determined the NH_2-terminal sequences of the A_4CB2, and some tryptic and metalloendopeptidase peptides of the A_4CB1 and A_4CB2, and found that not only the NH_2-terminal sequence of the A_4 subunit but the sequences of some peptides from the internal regions of the A_4 subunit are homologous to the internal sequence of the A_3 subunit (fig.3).

We have reported that there are long and highly conserved repeats in the sequence of the A_3 subunit [4]. However, we could not detect any repeated sequence in the A_4 subunit. Concerning the repeated sequence in the glycinin acidic subunits, more detailed study by both protein and DNA sequence analysis is now in progress.

Staswick and Nielsen [3] have determined the NH_2-terminal sequence of the A_6 subunit of glycinin, and found that the sequence is homologous to the NH_2-terminal sequence of the A_4 subunit. However, the amino acid residue in the position of the A_6 subunit corresponding to the eighth residue in the A_4 subunit is deleted. This is the same deletion as we have found in the A_3 subunit [4].

The sequence of the A_5 subunit and the partial sequence of the A_4 subunit were found to be closely similar to those of the NH_2-terminal region and the COOH-terminal region of the A_3 subunit, respectively, but not identical. This suggests that the A_4 and A_5 subunits do not arise by proteolysis of the A_3 subunit, but are possibly different gene products. However, it seems that the genes encoding the A_3, A_4, A_5 and possibly A_6 subunit are most strongly related in the glycinin genes.

We compared the sequence of the A_5 subunit determined to that of the pea legumin (11 S globulin) α subunit predicted from the DNA sequence [14] and found a considerable degree of homology between them. If optimal alignments are made, ~48% of the sequence of the A_5 subunit is identical to that of the α subunit (fig.3). The sequence homology among the 7 S storage proteins of peas [15], French beans [16] and soybeans [17] has suggested that all legume 7 S storage proteins share a common evolutionary origin. Similarly, the results obtained here suggest that all legume 11 S storage proteins share a common evolutionary origin.

REFERENCES

[1] Wolf, W.J. and Briggs, D.R. (1959) Arch. Biochem. Biophys. 85, 186–199.
[2] Kitamura, K., Takagi, T. and Shibasaki, K. (1977) Agr. Biol. Chem. 41, 833–840.
[3] Staswick, P.E. and Nielsen, N.C. (1983) Arch. Biochem. Biophys. 223, 1–8.
[4] Hirano, H., Fukazawa, C. and Harada, K. (1984) J. Biol. Chem., in press.
[5] Kitamura, K. and Shibasaki, K. (1975) Agr. Biol. Chem. 39, 945–951.
[6] Mori, T., Utsumi, S. and Inaba, H. (1979) Agr. Biol. Chem. 43, 2317–2322.
[7] Laemmli, U.K. (1970) Nature 227, 680–685.
[8] Bhown, A.S. and Bennet, J.C. (1983) Methods Enzymol., 450–455.
[9] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[10] Chang, J.Y., Creaser, E.H. and Hughes, G.J. (1977) J. Chromatog. 140, 125–128.
[11] Richardson, M. (1974) Biochem. J. 137, 101–112.
[12] Podell, D.N. and Abraham, G.N. (1978) Biochem. Biophys. Res. Commun. 81, 176–185.
[13] Moreira, M.A., Hermodson, M.A., Larkins, B.A. and Nielsen, N.C. (1979) J. Biol. Chem. 254, 9921–9926.
[14] Lycett, G.W., Delauney, A.J., Zhao, W., Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1984) Plant Mol. Biol. 3, 91–96.
[15] Hirano, H., Gatehouse, J.A. and Boulter, D. (1982) FEBS Lett. 145, 99–102.
[16] Sun, S.M., Slighton, J.L. and Hall, T.C. (1981) Nature 289, 37–41.
[17] Schuler, M.A., Schmitt, E.S. and Beachy, R.N. (1982) Nucl. Acids Res. 10, 8225–8244.