A complete cDNA coding for the sequence of glycinin
$A2B_{1a}$ subunit precursor

Takayuki Momma, Takaharu Negoro, Kyoko Udaka and Chikafusa Fukazawa*

Genetic Engineering Laboratory, Division of Applied Microbiology, National Food Research Institute, MAFF, Yatabe, Tsukuba, Ibaraki 305, Japan

Received 13 May 1985; revised version received 17 June 1985

Analysis of the $A2B_{1a}$ subunit precursor, one of the $A_2$-subunit family of glycinin, the main storage protein of soybean, revealed that it is composed of a signal peptide segment (18 amino acids), the $A_2$ acidic polypeptide (282 amino acids), followed by the $B_{1a}$ basic polypeptide (185 amino acids). There was overall 63% homology between this subunit complex and pea legumin, which is an analogous protein to glycinin. As this degree of homology is rather higher than that in the $A_3B_4$ subunit, one of the $A_3$ subunit family, it seems that the genes encoding the $A_2$ subunit family are phylogenetically more strongly related to the legumin genes.

1. INTRODUCTION

Glycinin, the most abundant storage protein of soybean (*Glycine max* (L.) Merr.) seeds, is a hexameric protein which has an $M_r$ of $\sim 360\,000$ and a sedimentation coefficient ($s_{20,w}$) of $\sim 12$ [1,2]. Each of the 6 subunits in a glycinin molecule is composed of an acidic ($M_r$, 35 000–42 000) polypeptide and a basic ($M_r$, 22 000) polypeptide [2–4]. These subunit pairs are not random but heterogeneous and are composed of at least 5 distinct subunit complexes, which are denoted $A_1A_2B_2$, $A_1B_1B_2$, $A_2B_{1a}$, $A_3B_4$ and $A_5B_3$ that are covalently linked by disulfide bonds [5,6]. Recent reports [7–9] have established that these specific subunit complexes are synthesized as a single precursor polypeptide which may be cleaved to form a specific subunit pairing during a post-translational processing as speculated previously [10], and that glycinin subunits have a strong relatedness to legumin subunits in their amino acid sequences. To elucidate the evolutionary process of the glycinin gene structure and the regulatory mechanism of the gene expression during development, it is of importance to obtain a better understanding of the gene structure for each subunit precursor. Along this line, we have already determined the predicted amino acid sequence of the $A_3B_4$ subunit precursor, one of the $A_3$ subunit family [8]. Here we describe the complete amino acid sequence of the $A_2B_{1a}$ subunit precursor, one of the $A_2$ subunit family, deduced from the cloned cDNA, although the partial genomic structure for this glycinin precursor has been determined [7].

2. EXPERIMENTAL

Glycinin was purified from defatted flour of the soybean (*G. max* (L.) Merr. cv. Bonnimir) by a combination of a fractional acid precipitation procedure and gel chromatography [11]. Glycinin $A_2$ and $B_{1a}$ subunits were purified from the above glycinin preparation according to the procedures of Moreira et al. [12]. The COOH-terminal V8 peptide of the $A_2$ subunit which was isolated and identified by its NH$_2$-terminal sequence [13] was ana-
lyzed with carboxypeptidase A digestion, while the COOH-terminal metalloendoproteinase (Grifola frondosa, Seikagaku Kogyo) peptide of the B1a subunit [14] was also analyzed by the same procedure.

The preparation of a cDNA library from soybean cotyledon tissue at the middle stage of seed development was performed as described in [8]. To select the plasmids covering the nucleotide sequence corresponding to the NH2-terminal region of the A2 subunit family a mixed oligonucleotide probe was constructed to correspond to a unique sequence of the A2 subunit as shown in fig. 1. The procedures of labeling and hybridization of this synthetic probe were performed as described [8]. Nucleotide sequences were determined by the methods of Maxam and Gilbert [15] and examined by computer analysis (Software Development Corporation, Japan).

3. RESULTS

3.1. Identification and DNA sequencing of the cloned glycinin A2B1a cDNA

Using the mixed oligonucleotide probe as shown in fig.1, 13 colonies that showed an intensive hybridization signal were selected to determine the length and the partial nucleotide sequence of those recombinant DNAs. A recombinant plasmid, designated as pGA2B1a 521, containing a ds-cDNA insert of about 1700 base pairs was identified as the plasmid containing the corresponding nucleotide sequences for both the A2 and B1a subunits. The restriction endonuclease map and the sequencing strategy used in the sequence determination are indicated in fig. 2. As shown in fig. 3, the overall length of the sequence in this cloned DNA insert is 1712 nucleotides and it covers the whole protein coding region with the nontranslated regions of both 5'- and 3'-termini. At the 5'-terminus, a nontranslated region of 44 nucleotides preceding the AUG translation start codon was found, whereas the protein synthesis termination codon for this mRNA is UAG, which is followed by 210 untranslated nucleotides in the 3'-region adjacent to a poly(A) segment. The sequence AAUAAA, usually found near the 3'-end of eukaryotic mRNAs [16], is located in the glycinin A2B1a mRNA, ending 16 nucleotides upstream from poly(A) segment. This unique sequence is also found in the unexpected position 127 nucleotides upstream from the segment. Analysis of the coding region of this mRNA indicates that the glycinin A2B1a subunit precursor is synthesized from the mRNA encoding a signal peptide segment (18 amino acids), the A2 acidic subunit, followed by the B1a basic subunit, as the NH2-terminal residues of those mature subunits were identified by comparison to the results reported previously [5,13].

3.2. The predicted protein sequences

To determine the accurate size of the mature A2 and B1a subunits, the COOH-termini of both polypeptides were analyzed according to the carboxypeptidase A digestion of their subfragments derived from Staphylococus aureus V8 protease or from Grifola frondosa metalloendopeptidase. The results indicate that the COOH-termini of the mature A2 and B1a subunits are asparagine and alanine, respectively. Therefore, the mature A2 subunit is encoded by 846 nucleotides (282 amino acids) and the mature B1a subunit is encoded by 555 nucleotides (185 amino acids). The Mε calculated from the inferred amino acid sequences of

![Fig.2. Restriction endonuclease map and nucleotide sequencing strategy for glycinin A2B1a subunit cDNA. The arrows indicate the direction and length of the sequence, (●) the sites labeled at flush or recessed 5'-ends, and (○) the sites labeled at protruding 3'-ends.](image-url)
Fig. 3. Nucleotide sequence of glycinin A2B14 subunit cDNA. The complete nucleotide sequence for the coding strand of the cloned ds-cDNA is shown with the predicted amino acid sequence for the primary translation product of glycinin A2B14 subunit mRNA. (*) and (**) indicate the NH2-terminal residues of the mature A2 and B14 subunits, respectively; the cleavage site of the subunit precursor is indicated by the arrow; the termination codon by (**); a putative polyadenylation signal sequence is underlined.
the A₂ and the B₁₈ subunits are 32,078 and 20,340, respectively.

3.3. Comparison of amino acid sequences of the glycinin A₂B₁₈ and the legumin αβ

A comparison of the predicted amino acid sequence of the glycinin A₂B₁₈ subunit with that of pea legumin is shown in fig. 4. Using an alignment that permitted maximum homology of amino acids, it was found that overall 59% of the amino acid positions were identical in the 2 acidic polypeptide regions, while there was 67% homology.

Fig. 4. Comparison of the amino acid sequence of glycinin A₂B₁₈ subunit with that of the legumin αβ. The dashed boxes enclose amino acids that are identical in the signal peptide segments of the two proteins, while the solid boxes enclose the identical amino acids of those precursor polypeptides. The dots indicate the hypothetical deletion spaces which were introduced to optimize alignment. The arrow indicates the cleavage site, whereas the asterisk indicates the NH₂-terminal residue of the basic polypeptide, and (●) cysteine residues which form a disulfide bond during the post-translational processing.
between the 2 basic polypeptide regions. The extent of homology in the NH₂-terminal regions of both acidic and basic subunits is greater than that in the COOH-terminal regions. Remarkable differences in amino acid sequence occurred in the COOH-terminal regions of both acidic polypeptides.

When the entire amino acid sequences of both precursor proteins are plotted as a function of hydropathic index [17], similar patterns were observed (not shown). In these alternating patterns, the sequence spanning residues 244-254 of the A₂B₁₈ subunit precursor shows a relatively striking hydrophilicity, suggesting that it may be localized near the surface of the protein molecule. Although these hydrophilic sequences contained repeated units in glycinin A₃B₄ [8] and legumin αβ [9,18], there are none in the A₂B₁₈ subunit precursor. However, these hydrophilic regions which are located immediately upstream of the post-translational cleavage sites of glycinin and legumin precursors appear to be functionally similar. By the Chou and Fasman algorithm [19], the secondary structure of the amino acid sequence around the cleavage site was predicted. The result suggested that the predicted structure has a similar domain segment as speculated previously [8].

4. DISCUSSION

Recently, Staswick et al. [13] have determined the amino acid sequences of the A₂ and B₁₈ subunits of glycinin using a protein sequencing technique, and Marco et al. [7] have also determined the partial genomic DNA sequence for the subunit precursor. To further understand the molecular structure and the evolutionary process of the glycinin A₂ subunit family we have deduced the complete amino acid sequence of the A₂B₁₈ subunit precursor from the nucleotide sequence of a cloned cDNA and have also determined the accurate size of the mature A₂ and B₁₈ subunits. The determination of the COOH-terminal residues for both mature subunits revealed that there was no evidence for the existence of a linker sequence [7,13]. This is in good agreement with those from glycinin A₃B₄ [8] and legumin αβ [9,18].

In comparison with the primary structure of pea legumin, the extent of homology to the glycinin A₂B₁₈ is greater than that to the A₃B₄ subunit precursor. This suggests that the glycinin A₂ subunit family appears to be evolutionarily more closely related to the pea legumin that the A₃ subunit family is. Although there was a remarkable sequence divergence in the COOH-terminal region of the acidic polypeptide, the secondary structure predicted by the Chou and Fasman algorithm [19] and the hydrophilic character [17] of the boundary region between the A₂ and B₁₈ subunit of the precursor were quite similar to those of the A₃B₄ and legumin precursors. Therefore, the hypothetical cleavage rule [8] will also apply to the A₂B₁₈ subunit precursor.

ACKNOWLEDGEMENTS

We are very grateful to Drs Kyuya Harada and Hisashi Hirano for their thoughtful discussions. We also appreciate the excellent technical assistance from Yukie Isoyama, and the manuscript preparation and editorial assistance of Michiko Murakami.

REFERENCES


