Glycinin A3B4 mRNA

CLONING AND SEQUENCING OF DOUBLE-STRANDED cDNA COMPLEMENTARY TO A SOYBEAN STORAGE PROTEIN*

(Received for publication, November 14, 1984)

Chikafusa Fukazawa‡, Takayuki Momma‡, Hisashi Hirano§, Kyuya Harada¶, and Kyoko Udaka
From the Division of Applied Microbiology, National Food Research Institute, §National Institute of Agrobiological Resources, and ¶National Institute of Agro-Environmental Science, Ministry of Agriculture, Forestry, and Fisheries, Yatabe, Tsukuba, Ibaraki 305 Japan

The cDNA clones encoding the precursor form of glycinin A3B4 subunit have been identified from a library of soybean messenger RNA cDNA clones in the plasmid pBR322 by a combination of differential colony hybridizations, and then by immunoprecipitation of hybrid-selected translation product with A3-specific antisemum. A recombinant plasmid, designated pGA3B41425, from one of six clones covering codons for the NH2-terminal region of the subunit was sequenced, and the amino acid sequence was inferred from the nucleotide sequence, which showed that the mRNA codes for a precursor protein of 516 amino acids. Analysis of this cDNA also showed that it contained 1786 nucleotides of mRNA sequence with a 5'-terminal nontranslated region of 46 nucleotides, a signal peptide region corresponding to 24 amino acids, an A3 acidic subunit region corresponding to 320 amino acids followed by a B4 basic subunit region corresponding to 172 amino acids, and a 3'-terminal nontranslated region of 192 nucleotides, which contained two characteristic AAUAAA sequences that ended 110 amino acids followed by a B4 basic subunit region corresponding to 320 amino acids. Analysis of this cDNA also showed that it contained 1786 nucleotides of mRNA sequence with a 5'-terminal nontranslated region of 46 nucleotides, a signal peptide region corresponding to 24 amino acids, an A3 acidic subunit region corresponding to 320 amino acids followed by a B4 basic subunit region corresponding to 172 amino acids, and a 3'-terminal nontranslated region of 192 nucleotides, which contained two characteristic AAUAAA sequences that ended 110 amino acids followed by a B4 basic subunit region.

The inferred amino acid sequence of the mature basic subunit, B4, was compared to that of the basic subunit of pea legumin, Leg β, which contained 185 amino acids. Using an alignment that permitted a maximum homology of amino acids, it was found that overall 42% of the amino acid positions are identical in both proteins. These results led us to conclude that both storage proteins have a common ancestor.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
† To whom correspondence should be addressed.
‡ Carrier research fellow from Kirin Brewery Co., Plant Research Cester, Kita-ku, Chiba, 329-14, Japan, and recipient of a predoctoral trainee fellowship from National Food Research Institute.
§ The abbreviations used are: SDS, sodium dodecyl sulfate; dscDNA, double-stranded complementary DNA; bp, base pairs; Pipes, 1,4-piperazinediethanesulfonic acid.
LCI (pH 6.5) at 4 °C for 12 h. The resulting precipitate was dissolved in water, and 2 volumes of chilled ethanol were added. The resultant precipitate was rinsed 3 times with 80% chilled ethanol and dried in vacuo. This RNA pellet was dissolved in sterile water for further experiments.

Poly(A)-containing RNA was prepared by oligo(dT)-cellulose (Collaborative Research, Inc.) chromatography and tailed with oligo(dC) as described previously (10) with minor modifications as described previously (11). Translation products were immunoprecipitated by mixing 20 μl of ribonuclease A-treated lysate reaction mixture with 100 μg of monospecific antibody in the presence of 1% sodium deoxycholate, 0.1% SDS, 150 mM NaCl, 50 mM LiCl, 10 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, and 50 mM Tris-HCl, pH 7.2. The mixtures were incubated at 37 °C for 1 h, then at 4 °C for 12 h. Formaldehyde-fixed Staphylococcus aureus (IgG-SORB, The Enzyme Center, Boston, MA) was added in an amount sufficient to bind about 210 μg of IgG and mixed for 1 h at room temperature. The membrane filters with collected proteins were washed 6 times with 800 μl of the above Tris buffer by repeated resuspension (vortexing and sonication) and by centrifugation. Immunoprecipitates were dissolved by resuspending the pellets in 50 mM Tris-HCl, pH 7.2, containing 1% SDS and 4 mM dithiothreitol, and heating them at 100 °C for 5 min. Translation products were separated by electrophoresis on 12.5% polyacrylamide gels containing 0.1% SDS, according to Laemmli (12). The resultant gels were impregnated with commercial fluor (ENHANCE, New England Nuclear), according to the manufacturer's directions, and fluorographed at -75 °C.

**Purification of Glycinin Subunits**—Glycinin A and B subunits were purified from defatted soybean flour according to Kitamura et al. (1).

**Preparation and Characterization of Monospecific Antibodies to Glycinin A3 Subunit**—Antibodies against A3 subunit were elicited from random-bred albino rabbits, weighing 2.5 to 3.1 kg, according to the immunization schedule described previously (7). The antiserum obtained was absorbed with other glycinin acidic subunits such as A1, A2, A3, and A5, and the resultant antiserum was judged to be monospecific by protein-blotting experiments using a commercial kit of an enzyme immunoassay (The Bio-Rad Immuno-Blot Assay Kit), according to the manufacturer's directions. All the subunits of glycinin (about 600 ng of protein) were separated by SDS-polyacrylamide membrane filters and immunostained with the respective antibodies which were specific for AS subunit (1:120 dilution), washed to remove unbound antibodies, incubated with second antibodies which were conjugated with horseradish peroxidase by the above manufacturer, and immersed into developing solution.

**Construction of a Soybean Cotyledonary ds-cDNA Recombinant Plasmid Library**—Double-stranded cDNA was synthesized from poly(A)-containing RNA derived from the soybean cotyledonary tissue at the middle stage (38 days after flowering) of seed development essentially as described by Land et al. (13), with minor modifications. Briefly, 10 μg of poly(A)-containing RNA was used as a template for the synthesis of single-stranded cDNA with an oligo(dT) primer and avian myeloblastosis virus reverse transcriptase (Midwest Products Inc.). The 3'-terminus of the single-stranded cDNA was labeled with terminal transferase (P-L Biochemicals) with oligo(dC) primer and [γ-32P]ATP (3000 Ci/mmol, Amersham) to a specific activity of greater than 1010 cpm/μg. The labeled cDNA was fractionated by gel filtration through Sepharose CL-4B, and 20 μl of the fraction were used for the infection of E. coli (14). Recombinant work was performed under containment conditions in accordance with the Genetic Manipulation Advisory Group (Japan) guidelines.

**Identification of Transformants Containing the 5'-Coding Region of the Nucleotide Sequence Corresponding to the Glycinin A3 Subunit Family**—About 2200 tetracycline-resistant, ampicillin-sensitive transformants were selected and screened by colony DNA-labeling hybridization. The 5'-labeled probes were synthesized from cDNA sequences enriched or deficient in glycinin A3-type subunit mRNAs and to 32P-labeled oligonucleotide probe, which was constructed to correspond to all possible codons for the NH2-terminal amino acid residues 5-9 (Lys-Phe-Ase-Glu-Cys), which is a unique sequence of the A3 subunit family (3). Fig. 2 shows the sequence of mRNA from 5' to 3', specifying glycinin A3 subunit amino acids 5-9, as well as the cDNA sequence from 3' to 5'. This cDNA sequence was synthesized by a solid phase phosphatetriester method using the reaction conditions and procedures of Beaucage and Caruthers (15). After completion of the synthesis, the oligonucleotides were labeled at the 5'-OH end with T4 polynucleotide kinase (Toyobo Biochemicals) and [γ-32P]ATP (3000 Ci/mmol, Amersham) to a specific activity of greater than 1010 cpm/μg and then purified by a 20% polyacrylamide gel electrophoresis in 100 mM Tris, 100 mM boric acid, 2.5 mM EDTA, and 7 M urea (pH 8.3). Transformated bacterial colonies were allowed to grow on nitrocellulose membranes at 37 °C, and the nitrocellulose filters were collected and washed with 0.2 M sodium chloride and 1 mM phenylmethylsulfonyl fluoride. The resultant filters were impregnated with commercial fluor (ENHANCE, New England Nuclear), according to the manufacturer's directions, and fluorographed at -75 °C.

**Preparation and Characterization of Monospecific Antibodies to Glycinin A1 Subunit**—Antibodies against A1 subunit were purchased from Nippon Gene Inc. and New England Nuclear. Recombinant work was performed in accordance with the Genetic Manipulation Advisory Group (Japan) guidelines.
RNA Blot Analysis—To estimate the size of the mRNAs encoding the glycinin A3 subunit family, poly(A)-containing RNA was analyzed by RNA blot analysis essentially according to Thomas (18). Briefly, 10 μg of RNA were denatured with glyoxal and resolved by electrophoresis in 1% agarose gels in 10 mM sodium phosphate, pH 6.8, for 4 h at 5 V/cm. The RNA was transferred overnight to nitrocellulose by blotting, and the nitrocellulose was baked for 2 h at 80 °C under vacuum. Complete reversal of glyoxylation was achieved by soaking the baked blots in 0.3 M sodium citrate containing 3 M NaCl for 12 h. Blots were hybridized to 32P-labeled 162-bp PstI-MluI restriction endonuclease DNA fragment which contains a unique sequence for the glycinin A3 subunit family.

Nucleotide Sequencing—Nucleotide sequencing was performed according to the procedures of Maxam and Gilbert (19). Restriction endonuclease-digested DNA fragments were end-labeled at their 5'-termini with [γ-32P]ATP (Amersham), after treatment with calf intestine alkaline phosphatase (Boehringer Mannheim), and were end-labeled at their 3'-termini with [α-32P]dideoxy-ATP (Amersham) and terminal transferase (P-L Biochemicals). Nucleotide sequences were examined by computer analysis (Software Developing Company, Japan).

RESULTS AND DISCUSSION

As there is a strong relatedness among the glycinin subunits in amino acid sequences (2, 3) and in an immunological analysis (20), it has been difficult to identify plasmids containing the cDNA encoding the precursor form of a specific glycinin subunit using immunological techniques such as hybrid-selection and hybrid-arsenal translations. In order to identify the cDNA clones covering codons for the NH2-terminal region of the glycinin A3 subunit family (A3A5 and A3A5B3 subunits) from a library of soybean cotyledonary cDNA clones in the plasmid pBR322, we employed the following three procedures. 1) The A3 subunit monospecific antibodies which could be used for a positive hybridization selection of mRNA were prepared by an immunological adsorption technique.2 The antiserum elicited by hyperimmunization of the AB subunit was treated with purified A1, A1h, A5, A3, and A6 subunits of glycinin. Fig. 1 demonstrates that the antibodies absorbed are monospecific for the A3 subunit of glycinin in an enzyme-immunoblot assay system, whereas there is a strong cross-reactivity between A3 and A5 subunits in an immunological sense. Using this antiserum several recombinant plasmids selected as possible candidates for containing the A3 subunit sequence were used for the positive hybrid selection. 2) Total poly(A)-containing RNA isolated from soybean developing cotyledonary tissue at 38 days after flowering was fractionated by sedimentation through 5-29.9% isokinetic sucrose gradients, and the individual fractions of the sucrose gradient centrifugation were translated in the presence of [3H]leucine in a mRNA-dependent, protein-synthesizing system derived from rabbit reticulocyte lysates. The total translation products were immunoprecipitated with antibodies prepared against the A3 subunit of glycinin as described above and examined by gel electrophoresis and fluorography. Glycinin A1B4 subunit precursor was detected as Mr = 61,000 (data not shown). This value is in agreement with the previously reported value for the molecular weight of one of total immunoprecipitated products with anti-glycinin IgG (4). This result also showed that A3B4 subunit mRNA was enriched in three gradient fractions numbered 6–8. To screen the cDNA library by colony DNA-filter hybridization RNA fractions that were enriched (fraction numbers 6 and 8) and that were deficient in glycinin A1B4 mRNA (fraction 10) were employed, respectively. 3) To select the plasmids covering the nucleotide sequence corresponding to the NH2-terminal region of the A3 subunit family a mixed oligonucleotide probe was constructed to correspond to a unique sequence of the subunit as shown

Amino Acid Residues 5 6 7 8 9
(NH2-terminus) Lys Phe Asn Glu Cys

mRNA 5'-AA6 - UU6 - AA6 - GA6 - UG' 3'
cDNA 3'-TT6 - AA6 - TT6 - CT6 - AC' 5'

Fig. 2. The mixed oligonucleotide probe specific for the glycinin A3 subunit family. The figure shows amino acids 5–9 of the A3 subunit of glycinin (top, from left to right), the mRNA codons which may code for these amino acids, and the oligonucleotide probe designed to be complementary to all possible codons for this sequence.

Fig. 1. Demonstration of glycinin A3 subunit monospecificity of the antiserum, prepared by adsorption technique, using an immunoblot assay system. A crude glycinin preparation treated with 0.1% SDS and dithiothreitol was electrophoresed on a 12.5% acrylamide gel containing 0.1% SDS, followed by Amido Black staining (A) or by protein blotting experiments using rabbit anti-A3 subunit antibodies previously adsorbed with A1, A1h, A5, and A6 subunits (B), and with A1, A1h, A5, A6, and A3 subunits (C).

2 C. Fukazawa and K. Udaka, manuscript in preparation.
FIG. 4. Size of glycinin A3B4 subunit mRNA. Poly(A)-containing RNA (10 μg) from the developing cotyledonary tissue at 38 days after flowering was denatured, electrophoresed in agarose, transferred to nitrocellulose, and hybridized to a 32P-labeled 162-bp PstI-MluI restriction endonuclease DNA fragment which contains a unique sequence of the A3 subunit family. The molecular weight markers used are HindIII restriction endonuclease fragments of ϕX174 DNA that were labeled with [γ-32P]ATP at the 5'-ends.

FIG. 5. Restriction endonuclease map of the glycinin A3B4 subunit ds-cDNA insert of pGAsB4 1425. The top horizontal line represents the length of A3B4 mRNA corresponding to the relative positions of the cloned ds-cDNA. The cloned ds-cDNA is shown with only those restriction endonuclease sites employed in the sequence determination. Each of the ends of the ds-cDNA are bound by segments containing about 15 GC residues and also contain coding strand oligo(dA) segment of 5 nucleotides at the 3'-end. Arrows indicate the direction and length of sequencing of restriction endonuclease-digested DNA fragments. The solid circles represent the sites labeled at flush or recessed 5'-ends, and the open circles represent the sites labeled at protruding 3'-ends as indicated under "Experimental Procedures."

in Fig. 2. Using this 32P-labeled oligonucleotide probe plasmids from colonies that hybridized intensely with fraction 6 and 8 cDNAs, but not with fraction 10 cDNA were further characterized and the lengths of recombinant ds-cDNA were determined by restriction endonuclease digestion and electrophoresed.

A recombinant plasmid, pGAsB4 1425, containing a ds-cDNA insert of about 1800 bp, was selected as one of the longest possible candidates for containing NH2-terminal unique sequence of the A3 subunit family by a combination of differential colony hybridizations described above.

To establish the identity of the putative glycinin A3B4 subunit plasmid, pGAsB4 1425, DNA was immobilized on nitrocellulose and hybridized to total cotyledonary poly(A)-containing RNA of the developing soybean seed. The specific...
physically bound RNA was eluted and translated in the cell-free, protein-synthesizing system. As indicated in Fig. 3, lane 1, the pGAB,B, 1425 DNA specifically hybridized to a mRNA that directed the synthesis of the polypeptide of M ∼ 61,000. This translation product was immunoprecipitated with monospecific antibodies to the A3 subunit of glycinin (Fig. 3, lane 2), thus establishing the identity of the recombinant plasmid.

The size of glycinin A3B4 subunit mRNA was estimated by hybridizing 32P-labeled 162-bp PstI-MulI restriction endonuclease DNA fragment which contains a unique sequence of the A3 subunit family to nitrocellulose filter blots of poly(A)-containing RNA that had been denatured and partially resolved by agarose gel electrophoresis. As shown in Fig. 4, the restriction endonuclease DNA fragment probe hybridized to a RNA species of about 2000 nucleotides in length.

The complete nucleotide sequence of the cloned glycinin A3B4 subunit ds-cDNA was determined. The restriction endonuclease cleavage fragments and the sequencing strategy used in the nucleotide sequence determination are indicated in Fig. 5. The various restriction fragments overlap in several regions, generating essentially complete sequence information for each of the two strands of the ds-cDNA. Sites of labeling were independently sequenced within other restriction endonuclease fragments.

The complete nucleotide sequence for the coding strand of cloned glycinin A3B4 subunit ds-cDNA is shown in Fig. 6. The protein synthesis termination codon for this mRNA is UAA, which is followed by 192 untranslated nucleotides in the 3' region adjacent to a poly(A)-containing RNA that had been denatured and partially resolved by agarose gel electrophoresis. As shown in Fig. 4, the restriction endonuclease DNA fragment probe hybridized to a RNA species of about 2000 nucleotides in length.

The complete nucleotide sequence of the cloned glycinin A3B4 subunit ds-cDNA was determined. The restriction endonuclease cleavage fragments and the sequencing strategy used in the nucleotide sequence determination are indicated in Fig. 5. The various restriction fragments overlap in several regions, generating essentially complete sequence information for each of the two strands of the ds-cDNA. Sites of labeling were independently sequenced within other restriction endonuclease fragments.

The complete nucleotide sequence for the coding strand of cloned glycinin A3B4 subunit ds-cDNA is shown in Fig. 6. The protein synthesis termination codon for this mRNA is UAA, which is followed by 192 untranslated nucleotides in the 3' region adjacent to a poly(A)-containing RNA that had been denatured and partially resolved by agarose gel electrophoresis. As shown in Fig. 4, the restriction endonuclease DNA fragment probe hybridized to a RNA species of about 2000 nucleotides in length.
as α-helix and a relatively high degree of β-structure. These predicted secondary structures are in agreement with those deduced by optical rotatory dispersion and circular dichroism procedures (31-33).

The A₃ acidic component of the A₃B₄ subunit contained 17% acidic residues, 16% amide residues, and 10% basic residues, whereas the B₄ basic component contained 7, 15, and 11%, respectively. These acidic and basic amino acid residues were dispersed randomly throughout both molecules. To investigate the hydropathic character, both subunits were analyzed with a computer program and residues were dispersed randomly throughout both molecules. The predicted secondary structures are in agreement with those consecutive hydrophilic residues were found at positions 265 and 281 to 293.

It is of interest to know the mechanism of the post-translational cleavage of the glycinin A₃B₄ intermediary precursor into A₃ and B₄ subunits. The amino acid sequence around the cleavage site, Gly-Cys-Gln-Thr-Arg-Asn-Gly-Val-Glu-Glu-Asn-Ile (arrow indicates the cleavage site), was also applied to the analysis by the Chou and Fasman algorithm. The result suggests that the COOH-terminal region of the A₄ subunit may have β-structure, while the NH₂-terminal region of B₄ subunit is an α-helix. As the predicted A₃B₄ subunit precursor contained two additional Arg-Asn-Gly segments (Fig. 6), the predicted secondary structure of those amino acid sequences around the characteristic segment was analyzed by the same procedure described above. The results indicated that both amino acid sequences were predicted to have only β-structure. Therefore, it could be speculated that the maturation enzyme which cleaves the glycinin A₃B₄ subunit complex into acidic (A) and basic (B) subunits may recognize the specific β-structure - α-helix segment which has the unique Arg-Asn-Gly sequence, and cleave it at the site between Asn and Gly.

A comparison of the inferred amino acid sequence of glycinin A₃B₄ subunit with that of pea legumin, the pea storage protein which is analogous to glycinin, is shown in Fig. 7. In order to achieve maximum alignment of the amino acid positions between the two proteins, several hypothetical deletions were introduced. It was found that overall, 42% of the amino acid positions in the basic component of the two proteins are identical. Relatively long regions of non-homology could be identified at the COOH termini of the both subunits, whereas relatively short regions of non-homology could be identified near the middle of those proteins. These results indicate that both storage proteins have a common ancestor and gene rearrangement has occurred on evolution.

Acknowledgments—We are grateful to Dr. Akiyo Matsumoto for thoughtful discussion. We also appreciate the excellent technical assistance of Yukie Isoya and the manuscript preparation and editorial assistance of Noriko Kobayashi.

REFERENCES

1. Kitamura, K., Takagi, T., and Shibasaki, K. (1977) Agric. Biol. Chem. 41, 835-840
2. Morsira, M. A., Hermodson, M. A., Larkins, B. A., and Nielsen, N. C. (1981) Arch. Biochem. Biophys. 210, 633-642
3. Ereken-Tumer, N., Richer, J. D., and Nielsen, N. C. (1982) J. Biol. Chem. 257, 4016-4018
4. Tumer, N. E., Thanh, V. H., and Nielsen, N. C. (1981) J. Biol. Chem. 256, 8756-8760
5. Staswick, P. E., Hermodson, M. A., and Nielsen, N. C. (1981) J. Biol. Chem. 256, 8752-8755
6. Hirano, H., Fukazawa, C., and Harada, K. (1984) J. Biol. Chem. 259, 14371-14377
7. O-yanae, S., Fukazawa, C., and Harada, K. (1978) Agric. Biol. Chem. 42, 97-102
8. Brawerman, G., Mendelucki, J., and Lee, S. Y. (1972) Biochemistry 11, 637-641
9. Arov, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1408-1412
10. Felham, H. R. B., and Jackson, R. J. (1976) Eur. J. Biochem. 67, 247-256
11. McLean, J. W., Fukazawa, C., and Taylor, J. M. (1983) J. Biol. Chem. 258, 8993-9000
12. Laemmli, U. K. (1970) Nature (Lond.) 227, 680-685
13. Land, H., Grez, M., Hauser, H., Lindenmaier, W., and Schutze, G. (1981) Nucleic Acids Res. 9, 2251-2266
14. Dagert, M., and Ehrlich, S. D. (1979) Gene (Amst.) 6, 23-28
15. Beaucage, S. L., and Caruthers, M. H. (1981) Tetrahedron Lett. 22, 1859-1962
16. Grunstein, M., and Hogness, D. S. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 3961-3965
17. Birnboim, H. C., and Doly, J. (1979) Nucleic Acids Res. 5, 1513-1523
18. Thomas, P. S. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 5201-5205
19. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499-560
20. Moreira, M. A., Mahoney, W. C., Larkins, B. A., and Nielsen, N. C. (1981) Arch. Biochem. Biophys. 210, 643-646
21. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) Nucleic Acids Res. 8, 127-142
22. Kreil, G. (1981) Annu. Rev. Biochem. 50, 317-348
23. Lin-Lee, Y.-C., Bradley, W. A., and Chan, L. (1981) Biochem. Biophys. Res. Commun. 99, 654-661
24. Lycett, G. W., DeLaunay, A. J., Zhao, W., Gatehouse, J. A., Croy, R. R. D., and Boulter, D. (1984) Plant Mol. Biol. 3, 91-96
25. Chou, P. Y., and Fasman, G. D. (1974) Biochemistry 13, 211-222
26. Chou, P. Y., and Fasman, G. D. (1974) Biochemistry 13, 223-245
27. Weber, K., Pringle, J. R., and Osborn, M. (1972) Methods Enzymol. 26, 3-27
28. Dayhoff, M. O. (ed) (1972) Atlas of Protein Sequence and Structure, Vol. 5, pp. D2-D6, National Biochemical Research Foundation, Washington, D. C.
29. Catsimpoolas, N., Rogers, D. A., Circle, S. J., and Meyer, E. W. (1967) Cereal Chem. 44, 631-638
30. Kyte, J., and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132
31. Fukushima, D. (1968) Cereal Chem. 45, 203-224
32. Catsimpoolas, N., Wang, J., and Berg, T. (1971) Int. J. Protein Res. 3, 277-281
33. Koshiyama, I., and Fukushima, D. (1973) Cereal Chem. 50, 114-121