JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Molecular Analysis and Physicochemical Properties of Electrophoretic Variants of Wild Soybean *Glycine soja* Storage Proteins

TAKAKO FUKUDA,† NOBUYUKI MARUYAMA,† AKIRA KANAZAWA,‡ JUN ABE,‡ YOSHIYA SHIMAMOTO,‡ MIKI HIEMORI,§ HIDEAKI TSUJI,§ TAKATOSHI TANISAKA,⊥ AND SHIGERU UTSUMI*,†

Laboratory of Food Quality Design and Development, Graduate School of Agriculture, Kyoto University, Uji, Kyoto 611-0011, Japan, Laboratory of Plant Genetics and Evolution, Graduate School of Agriculture, Hokkaido University, Kitaku, Sapporo, Hokkaido 060-8589, Japan, Department of Nutritional Science, Faculty of Health and Welfare Science, Okayama Prefectural University, Kuboki 111, Soja, Okayama 719-1197, Japan, and Laboratory of Crop Breeding, Graduate School of Agriculture, Kyoto University, Sakyoku, Kyoto 606-8502, Japan

Cultivated soybeans (*Glycine max*) are derived from wild soybeans (*Glycine soja*) and can be crossed with them to produce fertile offspring. The latter exhibit greater genetic variation than the former, suggesting a possibility that wild soybeans contain storage proteins with properties different from and better than those of cultivated soybeans. To identify a wild soybean suitable for breeding a new soybean cultivar, we analyzed seed proteins from 390 lines of wild soybeans by electrophoresis. We found some lines containing electrophoretic variants of glycinin and β-conglycinin subunits: one line containing a small α' subunit of β-conglycinin and two and five lines containing small A3 and large A4 polypeptides of glycinin, respectively. β-Conglycinin and glycinin containing such variant subunits exhibited solubility and emulsifying ability similar to those of the predominant types of wild and cultivated soybeans. Glycinins containing small A3 and large A4 gave a shoulder derived from the start of denaturation at a temperature 4 °C lower than that of glycinin from the predominant types of wild and cultivated soybeans, although their thermal denaturation midpoint temperatures were very similar to each other. Cloning and sequencing of the predominant and variant subunit cDNAs revealed that the small α' and the small A3 lacked 24 amino acid residues in the extension region and four amino acid residues in the hypervariable region, respectively, and that the large A4 did not have an insert corresponding to the difference in the electrophoretic mobility but Arg279 and Gln305 were replaced by glutamine and histidine, respectively, in the hypervariable region. These suggest that small differences even in the hypervariable region can affect the thermal stability, as well as the electrophoretic mobilities, of the proteins.

KEYWORDS: Wild soybean; electrophoretic variant; storage protein; glycinin; *β*-conglycinin; thermal stability; physicochemical property; *Glycine soja*

INTRODUCTION

Wild soybeans (*Glycine soja*), in addition to cultivated soybeans (*Glycine max*), are members of the subgenus *Soja* in the genus *Glycine* and are distributed in East Asia including Japan, Korea, China, Russia, and Taiwan (*1*−*3*). These two species contain β-conglycinin (7S globulin) and glycinin (11S globulin) as major storage proteins and can be crossed to produce fertile offspring. Wild soybeans are thought to be a progenitor of cultivated soybeans and are known to have greater genetic variations than cultivated soybeans (*4*, *5*). Therefore, there is a possibility that some of the wild soybeans contain storage proteins with properties different from and better than those of cultivated soybeans and contain no Gly m Bd 30K, the primary allergen of soybean (*6*). Such a wild soybean can be a genetic resource suitable for breeding a new soybean cultivar with beneficial traits.

* Corresponding author. Tel: +81-774-38-3760. Fax: +81-774-38-3761. E-mail: sutsumi@kais.kyoto-u.ac.jp.
† Laboratory of Food Quality Design and Development, Graduate School of Agriculture, Kyoto University.
‡ Laboratory of Plant Genetics and Evolution, Graduate School of Agriculture, Hokkaido University.
§ Department of Nutritional Science, Faculty of Health and Welfare Science, Okayama Prefectural University.
⊥ Laboratory of Crop Breeding, Graduate School of Agriculture, Kyoto University.

10.1021/jf0479620 CCC: $30.25
Published on Web 03/31/2005

Glycinin is a hexameric protein with a molecular mass of around 320 kDa (*7*). Five constituent subunits of glycinin have been identified and are classified into two groups (group I—A1aB1b, A1bB2, A2B1a; group II—A3B4, A5A4B3) based on the identity of their amino acid sequences (*7*). The identities are around 82−86% within the groups and 42−45% between the groups.

β-Conglycinin is a trimeric protein with a molecular mass of 180 kDa (*7*). Three constituent subunits are identified α, α′, and β. The α and α′ subunits are composed of extension regions and core regions, whereas β consists of only the core region. The core regions of the three subunits exhibit high sequence identities with one another (90.4%, 76.2%, and 75.5% between α and α′, between α and β, and between α′ and β, respectively) (*8*). The extension regions of α and α′ exhibit lower identities (57.3%) than do the core regions and have a high acidic property (*8*).

Soybean is one of the representative allergenic foods in Japan, and 16 allergens in soybean seeds have been reported (*6*). Among them, Gly m Bd 30K, Gly m Bd 28K, and β-conglycinin α subunit are the major allergens; thus the IgE-binding frequency of these are 65.2%, 23.2%, and 23.2%, respectively (*6*). A mutant soybean line without the latter two allergens has been developed by γ-ray irradiation (*9*−*11*). On the other hand, a mutant soybean line without the major allergen, Gly m Bd 30K, has not been developed by natural breeding. Recently, such a soybean line has been developed by a recombinant technique (*12*). However, public acceptance of genetically modified organisms (GMOs) is currently low. If a wild soybean line without Gly m Bd 30K is identified, this can be used to develop a soybean line without this major allergen by traditional breeding.

In this study, we analyzed seed proteins from 390 lines of wild soybeans to screen for genetic variants of storage proteins by electrophoresis and to seek a line without Gly m Bd 30K by Western blotting using antibody against Gly m Bd 30K. While we were successful in identifying some lines containing electrophoretic variants of glycinin and β-conglycinin subunits, we did not find any line that had no Gly m Bd 30K. We compared the physicochemical properties of glycinins and β-conglycinins from mutant wild lines with the corresponding proteins from cultivated soybean to determine whether such mutant wild lines exhibit useful physicochemical properties for food production. We also cloned and sequenced the cDNAs encoding such variants and compared the sequences with their counterparts to further determine the molecular basis of their physicochemical differences.

MATERIALS AND METHODS

Plant Materials. Wild soybean lines widely collected from East Asia were used for the analysis.

Cultivation of Wild Soybean Lines. Wild soybean lines were cultivated in a field in Kyoto University, Japan, in 2003. Developing cotyledons and mature seeds were used for the preparation of total RNA and storage proteins, respectively.

Protein Extraction and SDS−PAGE Analysis. Seeds from 390 lines of wild soybean were ground in a mortar with a pestle, and proteins were extracted with buffer A (30 mM Tris-HCl, pH 8.0, 10 mM 2-mercaptoethanol, 1 mM EDTA, 0.1 mM (*p*-amidinophenyl)methanesulfonyl fluoride (*p*-APMSF), 1.2 μM leupeptin, 0.2 μM pepstatin A, 0.02% (w/v) NaN_3) at room temperature. After being stirred for 2 h, supernatants were collected by centrifugation at 4 °C (20 000 × *g*, 15 min). Supernatants were subjected to SDS−PAGE using 11% polyacrylamide gels according to the procedure of Laemmli (*13*).

Two-Dimensional (2D) Electrophoresis. Proteins extracted as described above were used. Gel strips (ReadyStrip IPG strip pH 3−10, Bio-Rad Laboratories, CA) for the first dimensional isoelectric focusing of the extracts were rehydrated overnight at room temperature with 8 M urea, 4% (w/v) 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonaic acid, 0.001% (w/v) bromophenol blue, 14 mM dithiothreitol, and the extracts (100 μg). Isoelectric focusing was performed in Protean IEF cell (Bio-Rad Laboratories) according to the manufacturer's instructions. After isoelectric focusing, gel strips equilibrated with 375 mM Tris-HCl, pH 8.8, 6 M urea, 2% (w/v) SDS, and 20% (v/v) glycerol were subjected to the second dimensional SDS−PAGE according to the procedure of Laemmli (*13*).

N-Terminal Amino Acid Sequence Analysis. The proteins in the bands of predominant polypeptides and electrophoretic variants excised from SDS−PAGE gels were extracted with SDS buffer (62.5 mM Tris-HCl, pH 6.8, 2% (w/v) SDS, 1% (v/v) 2-mercaptoethanol, 10% (v/v) glycerol), blotted onto a PVDF membrane by Prosorb cartridge (Applied Biosystems, CA), and subjected to N-terminal amino acid sequencing using a Procise 492 protein sequencer (Applied Biosystems).

Analysis of Wild Soybean Lines Using Gly m Bd 30K Antibody. The presence of the major soybean allergen, Gly m Bd 30K, was analyzed by sandwich ELISA and immunoblotting using two monoclonal antibodies according to the procedure of Tsuji et al. (*14*).

Purification of Glycinin and β-Conglycinin from Wild Soybean Seeds. The glycinin and β-conglycinin-rich fractions of wild soybean were prepared by the procedure of Nagano et al. (*15*). Glycinin was further purified by ammonium sulfate fractionation (45−65% saturation) to near homogeneity.

Protein Measurement. Protein concentrations of the samples were determined using a Protein Assay Rapid Kit (Wako Pure Chemical Industries, Japan) with bovine serum albumin as the standard.

Differential Scanning Calorimetry (DSC) Measurement. Protein samples were dialyzed against buffer B (35 mM sodium phosphate, pH 7.6, 0.4 M NaCl, 10 mM 2-mercaptoethanol, 1 mM EDTA, 0.1 mM *p*-APMSF, 1.2 μM leupeptin, 0.2 μM pepstatin A, 0.02% (w/v) NaN_3). In the case of glycinin samples, we used buffer B without 2-mercaptoethanol. After dialysis, the samples were used for the DSC measurement, which was carried out on a Microcal MC-2 ultrasensitive microcalorimeter (Micro Cal Inc., Northampton, MA) as described previously (*8*). All DSC experiments were performed with a protein concentration of 0.5 mg/mL. A DSC scan rate of 1 °C/min was used for all experiments.

Solubility as a Function of pH. The solubility of protein samples as a function of pH was measured as described previously (*16*). The protein solutions (0.8 mg/mL) were kept at 4 °C for 18 h at various pHs at high ionic strength ($\mu = 0.5$) and low ionic strength ($\mu = 0.08$). After centrifugation, protein concentrations in the supernatant were determined using a Protein Assay Rapid Kit (Wako Pure Chemical Industries). Solubility was expressed as percentage of the total protein content in the sample.

Emulsifying Ability. The emulsifying abilities of protein samples were measured as described previously (*16*). Protein samples were dialyzed against buffer B for $\mu = 0.5$ and buffer C (10 mM sodium phosphate, pH 7.6, 0.05 M NaCl, 10 mM 2-mercaptoethanol, 1 mM EDTA, 0.1 mM *p*-APMSF, 1.2 μM leupeptin, 0.2 μM pepstatin A, 0.02% (w/v) NaN_3) for $\mu = 0.08$. A portion (1.5 mL) of the dialyzed samples (1.0 mg/mL, $\mu = 0.5$ or 0.08) and 0.25 mL of soybean oil were homogenized and sonicated. Each sample was analyzed several times, and a representative pattern was presented.

Reverse Transcription−Polymerase Chain Reaction (RT−PCR). By using RT-PCR, we isolated cDNAs encoding all of predominant glycinin and β-conglycinin subunits and variant α′, A3B4, and A5A4B3. Total RNA was isolated from developing wild soybean seeds according to Shirzadegan et al. (*17*). The α′, A3B4, and A5A4B3 cDNAs were amplified using the RNA LA PCR kit (AMV), ver. 1.1 (Takara Bio, Japan). At first, mRNAs in total RNAs were reverse-transcribed into cDNAs by the primer (CGC GGATCC GGTACC CTGCAG GTCGAC TTTTTTTTTTTTTTTT), which is composed of the region complementary to poly(A) and sites of four restriction enzymes indicated by underlines. Next, the primers (GTGGAGAAAGAAGAATGTG for α; GTGGAGGAAGAAGAAGAATGC for α′; TTAAAGGTGAGAGAGGATG for β; TTCAGTTCCAGAGAGCAGCCTCAGC for A1aB1b; CGAGTGCCAGATCCAACGC for A1bB2; TTCGCTCTGAGAGAG-

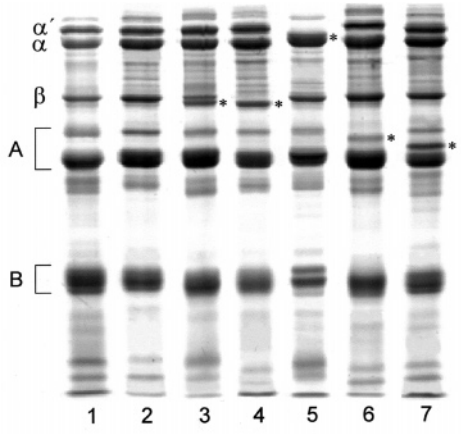

Figure 1. SDS−PAGE patterns of proteins from cultivated soybean, predominant type of wild soybean, and its variants. α′, α, and β indicate subunits of β-conglycinin. A and B indicate acidic and basic polypeptides of glycinin, respectively. Bands having unique mobilities are marked with asterisks: lane 1, cultivated soybean (var. Shirotsurunoko); lane 2, predominant type of wild soybean; lane 3, β•β*; lane 4, β*; lane 5, small α′; lane 6, small A3; lane 7, large A4.

CAGG for A2B1a; ATGGGGAAGCCCTTCTTCACTCTCTC for A3B4; ATGGGGAAGCCCTTCACTCTCTCTC for A5A4B3) corresponding to their N-termini were used. The primer for A1bB2 was designed to get cDNA lacking 29 base pairs for the N-terminus of the mature form to avoid the misannealing caused by the high homology of the nucleotides sequences of A1aB1b and A1bB2 (*18*). The primers for A3B4 and A5A4B3 correspond to the beginning of their signal peptides. The others correspond to the N-termini of the mature forms. At the same time, the primers (CGCGGATCCATACTTAT**TCA**GTAAAAAGCC for α and α′; CGCGGATCCATTTACGTAGTTAT**TCA**G for β; ACATACAAAGGATCCT**CTA**AGCCACAGCTCTCTTCTGAGACTC for A1aB1b; CTACGCACACTAACTAGTGC for A1bB2; TACGCGGATCC**CTA**AGCCACAGCTCTCCTCTGAG for A2B1a; CGCGGATCCTTGTTGTTAT**TTA**TGG for A3B4; CCGCGGATCC**TTA**TGCGACTTTAACACGGGGTGAGC for A5A4B3) corresponding to their C-termini were used. The primer for A1bB2 corresponds to its 3′-noncoding region because its C-terminus of the coding region exhibits high homology with those of A1aB1b and A2B1a. The others correspond to C-termini of the mature forms. Restriction enzyme sites (*Bam*HI) and stop codons are indicated by underlines and bold letters, respectively. The reactions were performed with 30 cycles at 95 °C for 30 s, 55 °C for 30 s for α, α′, β, A1bB2, A2B1a, and A3B4, 60 °C for A1aB1b, and 65 °C for A5A4B3, and 72 °C for 3 min. The amplified fragments having the expected sizes were blunted, phosphorylated, and treated with *Bam*HI, except A1bB2. Then, they were ligated with pBluescriptSK(−) (Stratagene, CA) that were previously treated with *Bam*HI, except A1bB2, and dephosphorylated, and their sequences were determined according to the dideoxy method using ABI Prism 3100 DNA analyzer (Applied Biosystems).

RESULTS AND DISCUSSION

Screening for Electrophoretic Variants. To identify mutant wild soybean lines giving electrophoretic variants of storage protein subunits, lacking a certain subunit, or both, protein extracts from 390 wild soybean lines and a predominant type of soybean cultivar (var. Shirotsurunoko) were analyzed by SDS−PAGE. All wild soybean lines except eight lines containing little or no β-conglycinin β gave bands corresponding to all subunits and polypeptides of glycinin and β-conglycinin. Most of them gave the same electrophoretic patterns as that of the cultivated soybean. However, some of them gave special or unique patterns. Twenty-one lines gave a band of β*, which migrated faster than the β (**Figure 1**, lane 3) in addition to that of β-conglycinin β, four lines contained only β* (**Figure 1**, lane 4), and eight lines contained little or no β. On the other hand, only 45 of 5828 soybean cultivars have been reported to contain the β* in addition to the β (*19*). We observed that three of 5828 soybean cultivars contained A3* instead of the A3 polypeptide of glycinin, which migrated slower than the A3 polypeptide (*19*). Among the 390 wild soybean lines, we did not observe a line containing A3*, but observed one, two, and five lines containing small α′ (accession number of the collection in Hokkaido University B09101), small A3 (B00171; B00193), or large A4 (B00159; B04077; B04105; B04119; B09090), respectively (**Figure 1**, lanes 5−7). These are consistent with the notion that wild soybeans have much more genetic variation than cultivated soybeans (*4*, *5*).

The 10 N-terminal amino acid sequences of these variants, as well as predominant ones, were found to be completely identical to those of the corresponding polypeptides. Furthermore, by 2D electrophoresis of protein extracts from these special lines (**Figure 2**), the isoelectric points of the small α′ and large A4 were found to be close to those of the predominant α′ and A4 (**Figure 2D,F**), respectively, but that of small A3 was more basic than that of the predominant A3 (**Figure 2E**).

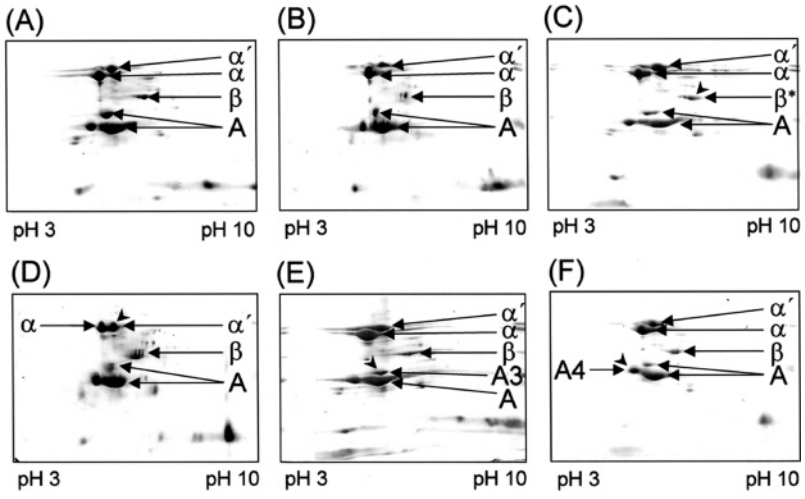

Figure 2. 2D electrophoresis patterns of proteins from cultivated soybean, predominant type of wild soybean, and its variants. Gels were stained with Coomassie Brilliant Blue R-250. α′, α, and β indicate subunits of β-conglycinin. A indicates acidic polypeptides of glycinin. An arrowhead indicates the unique spot. **Panel A** contains cultivated soybean (var. Shirotsurunoko); **panel B** contains predominant type of wild soybean; **panel C** contains β*; **panel D** contains small α′; **panel E** contains small A3; **panel F** contains large A4.

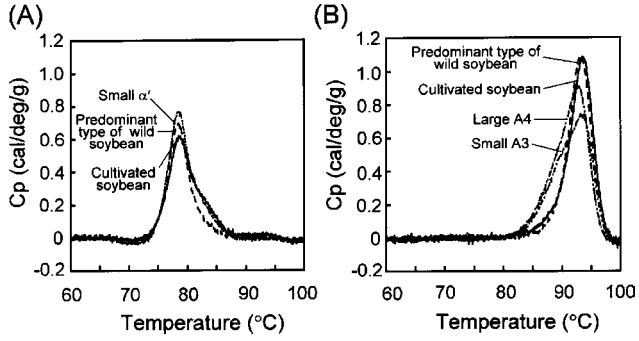

Figure 3. DSC scans of β-conglycinin (**A**) and glycinin (**B**) at $\mu = 0.5$.

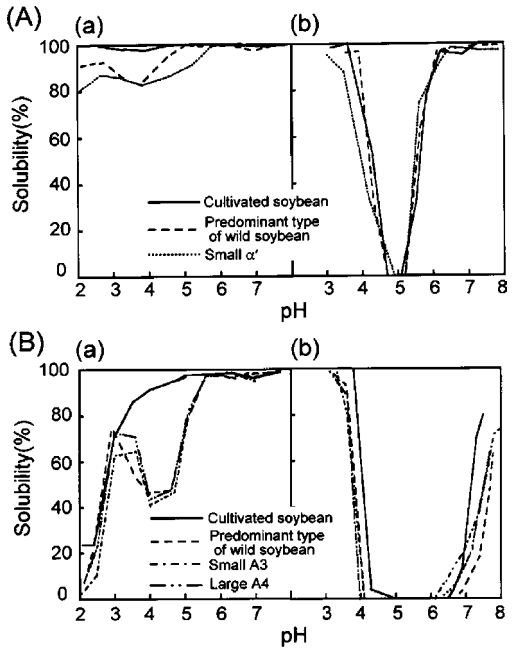

Figure 4. pH dependence of the β-conglycinin (**A**) and glycinin (**B**) solubility at $\mu = 0.5$ (**a**) and 0.08 (**b**).

Screening for a Line Lacking Gly m Bd 30K. Variations in the amount of Gly m Bd 30K was observed among lines, but no line lacking it or containing it at a very low level was found. So far, no soybean cultivar or wild soybean line without Gly m Bd 30K has been identified. Therefore, there was a possibility that Gly m Bd 30K is essential for the soybean plant. However, Herman et al. succeeded in developing a soybean containing no Gly m Bd 30K by genetic engineering (*12*). This indicates that Gly m Bd 30K is not essential, meaning that it may still be possible to identify a wild soybean without Gly m Bd 30K by screening a greater number of wild soybean lines.

Thermal Stability. Structural stability of proteins is an important factor for their gel forming and emulsifying abilities. Thermal denaturation is a prerequisite for the formation of the gel network structure. DSC profiles of β-conglycinins and glycinins from the predominant type of wild soybean and their variants together with those from a cultivated soybean are shown in **Figure 3**.

Although β-conglycinin containing small α' exhibited a very similar pattern to that of cultivated soybean β-conglycinin and a slightly different pattern from that of predominant type of wild soybean, their thermal denaturation midpoint temperatures (T_m) were the same, 78.5 °C (**Figure 3A**). Previously, we demonstrated that the order of T_m values of α, α', and β homotrimers is $\alpha < \alpha' < \beta$ (*20*) and that T_m values of heterotrimers are fundamentally determined by that of the subunit having the lowest T_m among the constituent subunits (*21*). These suggest that the proportion of heterotrimers containing no α subunit in the predominant type of wild soybean used here is lower than those in the cultivated soybean and the variant line.

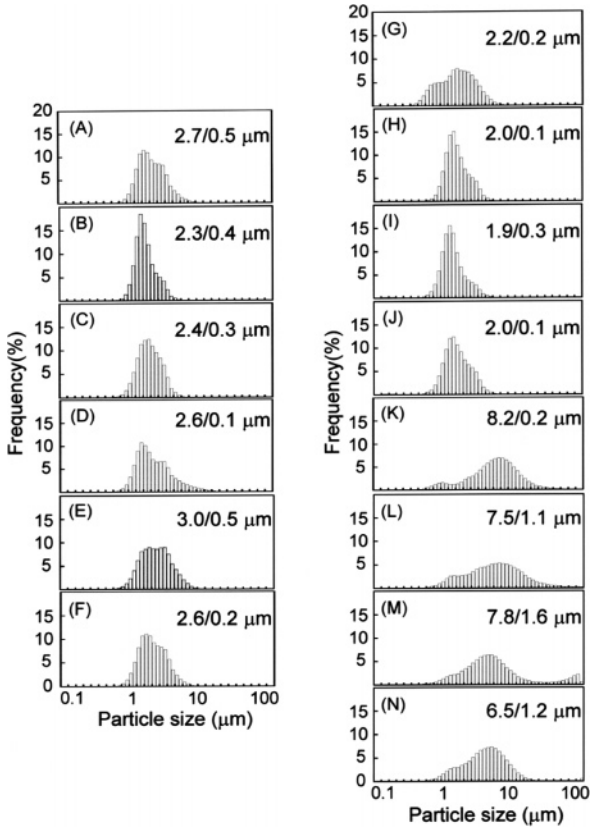

Figure 5. Particle size distributions of emulsions from β-conglycinin (**A**–**F**) and glycinin (**G**–**N**) at $\mu = 0.5$ (**A**–**C**, **G**–**J**) and $\mu = 0.08$ (**D**–**F**, **K**–**N**): (**A**, **D**, **G**, **K**) cultivated soybean (var. Shirotsurunoko); (**B**, **E**, **H**, **L**) predominant type of wild soybean; (**C**, **F**) small α'; (**I**, **M**) small A3; (**J**, **N**) large A4. Average particle size and standard error of each sample are shown in each panel.

Glycinin from predominant type of wild soybean exhibited a DSC profile with a T_m value of 93.1 °C very similar to that of glycinin from a cultivated soybean (**Figure 3B**). However, glycinins containing small A3 and large A4 gave a shoulder derived from the start of denaturation at a temperature 4 °C lower than did the others, although T_m values of these glycinins were very similar to each other (**Figure 3B**). We analyzed the thermal stability of glycinins composed of only group I, only group II, only A3B4, or only A5A4B3 and observed that the T_m value of glycinin composed of five subunits is not conferred by the subunit having the lowest T_m value (*22*), in contrast to β-conglycinin (*8*, *16*, *20*, *21*). Glycinin from the predominant type of wild soybean did not give any shoulder similar to that of a cultivated soybean, suggesting that this shoulder is due to the presence of small A3 or large A4. The lower denaturation-starting temperature denotes lower gelation temperature. This is coupled to reduction of costs for food production.

Solubility as a Function of pH. Generally, it is said that the higher the solubility of proteins, the higher the gel-forming, emulsifying, and forming abilities (*23*). We measured the solubilities of β-conglycinins and glycinins from the predominant type of wild soybean and its variants, together with those from a cultivated soybean at $\mu = 0.5$ and 0.08 (**Figure 4**). All β-conglycinin samples were soluble at all pH values examined here at $\mu = 0.5$ and exhibited isoelectric precipitation at pH

Figure 6. Alignment of the amino acid sequences of α′ subunit from cultivated soybean (var. Shirotsurunoko), predominant type of wild soybean, and small α′. Bold and dotted lines indicate the extension region and disorder regions, respectively. The regions marked by the diagonal pattern and the shade indicate amino acid deletions compared with the predominant type of wild soybean and cultivated soybean and predominant type of wild soybean, respectively.

around 5 at $\mu = 0.08$. Their profiles were very similar to each other (**Figure 4A**). In contrast to β-conglycinin, glycinins exhibited lower solubility at acidic pH at $\mu = 0.5$. Glycinin from a cultivated soybean became gradually insoluble with a lowering of pH, but all glycinins from a wild soybean and its variants exhibited different profiles. The solubility at pH 4−5 was lower than that at pH 3−4 and that at pH > 6 and < 3 was very similar to that of glycinin from a cultivated soybean (**Figure 4B**). We observed that soybean glycinin composed of only group I subunits exhibited a profile very close to that of glycinin containing both group I and group II subunits, and glycinin composed of only group II subunits was soluble at pH > 5, but very insoluble at pH < 5 (*22*). The profile of wild soybean glycinin is just between those of groups I and II. Therefore, the characteristic profile of wild soybean glycinin is probably due to higher proportion of group II subunits in wild soybean than in cultivated soybean. At ionic strength of 0.08, all glycinins from a cultivated soybean and wild soybeans were insoluble at pH 4−7. The profiles of β-conglycinins and glycinins from the predominant type of wild soybean and its variants were not identical in the same protein species, but the differences were within experimental error. Therefore, the variations do not seem to result in differences in solubility profiles.

Emulsifying Ability. Emulsifying ability of proteins is one of the useful physicochemical properties for food processing. We assessed the emulsifying abilities of samples by measuring the sizes of the particles after emulsification at $\mu = 0.5$ and 0.08 by homogenization and sonication. The smaller the particle sizes are, the better the emulsifying ability is.

All β-conglycinin and glycinin samples gave similar particle sizes in each protein species despite their origins (**Figure 5**). This indicates that their variations do not result in any difference in the emulsifying ability. The emulsifying ability of β-conglycinin at $\mu = 0.5$ was similar to that at $\mu = 0.08$, but that of glycinin was lower at $\mu = 0.08$ than at $\mu = 0.5$ (**Figure 5**). Measurement of the emulsifying ability was carried out at pH 7.6. The solubility of β-conglycinin at pH 7.6 was 100% at both ionic strengths of 0.5 and 0.08 (**Figure 4A**) but that of glycinin was 100% at $\mu = 0.5$ and only 30−40% at $\mu = 0.08$ (**Figure 4B**). Therefore, the emulsifying abilities of β-conglycinin and glycinin parallel their solubilities.

Cloning of the Predominant and Variant Subunit cDNAs. The N-terminal 10 amino acid sequences of the predominant and variant subunits of glycinins and β-conglycinins from the predominant and variant wild soybeans were identical to those of glycinin and β-conglycinin subunits from the cultivated soybean. Therefore, we used oligonucleotides synthesized based on the nucleotide sequences of soybean glycinin and β-conglycinin subunit genes for RT-PCR cloning. We cloned the RT-PCR products and sequenced two clones for each of predominant and variant subunits. The nucleotide sequence of each of the two clones was different from each other only at a few positions. However, we adopted the nucleotide sequences coinciding with those of the corresponding soybean because there is a possibility that PCR caused the mutations at these positions, although the enzyme used for PCR in this experiment has a relatively high fidelity. The differences in the amino acid sequences of glycinin and β-conglycinin subunits deduced from their nucleotide sequences between wild and cultivated soybeans are summarized in **Table 1**. Variations of the amino acid sequences among soybean cultivars deduced from nucleotide sequences were also observed. Four and one amino acid replacements were reported in glycinin A1aB1b and A2B1a subunit sequences between soybean cultivars Shirotsurunoko and Bonminori, respectively (*24*, *25*). However, the nucleotide sequences of these proteins from Bonminori were determined by the methods of Maxam and Gilbert (*26*, *27*), and some of the results could be artifacts. Therefore, the variations in the

Figure 7. Alignment of the amino acid sequences of A3B4 subunit from cultivated soybean (var. Shirotsurunoko), predominant type of wild soybean, and small A3. Dotted lines indicate the disordered regions. The region marked by the shade indicates the amino acid deletion.

Figure 8. Alignment of the amino acid sequences of A5A4B3 subunit from cultivated soybean (var. Shirotsurunoko), predominant type of wild soybean, and large A4. Dotted lines indicate the disordered regions. Boxes indicate substituted amino acids in comparison with those of predominant subunits.

amino acid sequences between cultivated and wild soybeans are higher than those between soybean cultivars.

The deduced amino acid sequences of the variants are shown with those of predominant types of wild and cultivated soybeans in **Figures 6**−**8**. In the case of β-conglycinin α', we sequenced four clones from two predominant lines. Both lines gave the same amino acid sequence, which contained two and four extra amino acid residues at two positions in the extension regions compared with that of a cultivated soybean (**Figure 6**). We can say that β-conglycinin α' of some wild soybean lines is six amino acid residues longer than that of a cultivated soybean, although it is unclear whether these lines are predominant or not. The small α' was lacking 18 amino acid residues in addition to the two and four amino acids in the extension region, thus, a total of 24 amino acids less than the α' of predominant types of wild soybeans. As a result of this deletion of 24 amino acid residues, the small α' had six fewer basic and seven fewer acidic amino acid residues than that of the predominant type, such that there was little change in the charge of the protein. This is consistent with the result of 2D electrophoresis (**Figure 2D**).

The A3B4 for the small A3 was lacking four amino acid residues in the hypervariable region (fourth disordered region) (**Figure 7**), which were all acidic, being consistent with the result of the 2D electrophoresis (**Figure 2E**). On the other hand, A5A4B3 for the large A4 did not have an insert corresponding to the difference in the electrophoretic mobility (**Figure 2F**), but Arg279 and Gln305 were replaced by glutamine and histidine, respectively, in the hypervariable region (**Figure 8**).

Table 1. Differences in the Amino Acid Sequences of Glycinin and β-Conglycinin Subunits between Wild and Cultivated Soybeans

subunit (Genbank accession no.)	wild soybean residue no.	wild soybean	cultivated soybean
A1aB1b	32	Phe	Leu
(SOYGLYBSU)	74	Val	Ile
	175	Gln	Lys
	194	Gly	Glu
	339	Arg	Ser
	381	Gly	Arg
	416	Arg	Lys
A1bB2	23	Asp	Gly
(AB030494)	137	Pro	Leu
	200	Ile	Met
	234	Thr	Ala
	338	Met	Val
A2B1a	115	Leu	Phe
(SOYGA2B1A)			
A5A4B3	129	Gly	Ser
(AB195712)	230	Gln	Glu
α	115	Gln	Arg
(AB008678)			
α′	18	Arg	*a*
(AB008680)	19	Pro	*a*
	98	Glu	*a*
	99	Gln	*a*
	100	Lys	*a*
	101	His	*a*
	371	His	Arg
β	13	Leu	Phe
(AB008679)	188	Met	Val
	333	Pro	Leu

[a] Amino acids in the corresponding positions are missing.

This can possibly explain the result of the 2D electrophoresis (**Figure 2F**) but not the difference in the electrophoretic mobility. Davis et al. (*28*) pointed out based on the report by Pacheco et al. (*29*) that structural changes that have effects on the electrophoretic mobility need not result from large alterations in the primary structure but may arise from a change as small as one amino acid substitution, despite no experimental data. We also demonstrated that the A3 polypeptide of which the Arg22 has been replaced with leucine in a β strand migrates more slowly than the predominant A3 (*19*). Similarly, the slower migration of large A4 is likely due to the two replacements. The results show that replacements in the disordered region are also able to affect the electrophoretic mobilities of proteins.

The variations in the amino acid sequences between the predominant A3 and the small A3 and between the predominant A4 and the large A4 are quite small. Therefore, we cannot explain well why glycinins from the predominant and variant lines exhibited different DSC profiles from each other. Recently, we created mutant procruciferins, which have replacements in the fourth variable region from the N-terminus with the corresponding regions of glycinin A1aB1b and A3B4. We observed that the replacements with those of A1aB1b and A3B4 resulted in no change and decrease of thermal stability at pH 7.6 and $\mu = 0.5$, respectively (*30*). These results indicate that the difference in the variable regions can affect the thermal stability. Therefore, the variations in the small A3 and the large A4 probably induce unfavorable interactions of the variable region with the core region of glycinins, resulting in the decrease of their thermal stability.

Screening of 390 wild soybean lines demonstrated that variations in amino acid sequences of glycinin and β-conglycinin subunits in wild soybeans are much higher than those in cultivated soybeans. We identified some lines containing glycinins with low thermal stability. It may be possible to identify wild soybean lines containing storage proteins with good physicochemical properties suitable for food production.

LITERATURE CITED

(1) Hymowitz, T.; Singh, R. J. Taxonomy and speciation. In *Soybeans: Improvement, Production, and Uses*, 2nd ed.; Wilcox, J. R. Ed.; Agronomy Monograph 16; American Society of Agronomy: Madison, WI, 1987; pp 23−48.

(2) Tateishi, Y.; Ohashi, H. Taxonomic studies on *Glycine* of Taiwan. *J. Jpn. Bot.* **1992**, *67*, 127−147.

(3) Tozuka, A.; Fukushi, H.; Hirata, T.; Ohara, M.; Kanazawa, A.; Mikami, T.; Abe, J.; Shimamoto, Y. Composite and clinal distribution of *Glycine soja* in Japan revealed by RFLP analysis of mitochondrial DNA. *Theor. Appl. Genet.* **1998**, *96*, 170−176.

(4) Hymowitz, T. On the domestication of the soybean. *Econ. Bot.* **1970**, *23*, 408−421.

(5) Kiang, Y. T.; Gorman, M. B. Soybeans. In *Isozymes in plant genetics and breeding. Part B.*; Tranksley, S. D., Orton, T. J., Eds.; Elsevier: Amsterdam, 1983; pp 295−328.

(6) Ogawa, T.; Bando, N.; Tsuji, H.; Okajima, H.; Nishikawa, K.; Sasaoka, K. Investigation of the IgE-binding proteins in soybeans by immunoblotting with the sera of the soybean-sensitive patients with atopic dermatitis. *J. Nutr. Sci. Vitaminol.* **1991**, *37*, 555−565.

(7) Utsumi, S.; Matsumura, Y.; Mori, T. Structure−function relationships of soy proteins. In *Food Proteins and Their Applications*; Damodaran, S., Paraf, A., Eds.; Marcel Dekker: New York, 1997; pp 257−291.

(8) Maruyama, N.; Katsube, T.; Wada, Y.; Oh, M. H.; Barba de la Rosa, A. P.; Okuda, E.; Nakagawa, S.; Utsumi, S. The roles of the N-linked glycans and extension regions of soybean β-conglycinin in folding, assembly and structural features. *Eur. J. Biochem.* **1998**, *258*, 854−862.

(9) Takahashi, K.; Banba, H.; Kikuchi, A.; Ito, M.; Nakamura, S. An induced mutant line lacking the α subunit of β-conglycinin in soybean [*Glycine max* (L.) Merrill]. *Breed. Sci.* **1994**, *44*, 65−66.

(10) Samoto, M.; Fukuda, Y.; Takahashi, K.; Tabuchi, K.; Hiemori, M.; Tsuji, H.; Ogawa, T.; Kawamura, Y. Substantially complete removal of three major allergenic soybean proteins (*Gly m* Bd 30K, *Gly m* Bd 28K, and the α subunit of conglycinin) from soy protein by using a mutant soybean, Tohoku 124. *Biosci. Biotechnol. Biochem.* **1997**, *61*, 2148−2150.

(11) Ogawa, T.; Samoto, M.; Takahashi, K. Soybean allergens and hypoallergenic soybean products. *J. Nutr. Sci. Vitaminol.* **2000**, *46*, 271−279.

(12) Herman, E. M.; Helm, R. M.; Jung, R.; Kinney, A. J. Genetic modification removes an immunodominant allergen from soybean. *Plant Physiol.* **2003**, *132*, 36−43.

(13) Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1979**, *227*, 680−685.

(14) Tsuji, H.; Bando, N.; Kimoto, M.; Okada, N.; Ogawa, T. Preparation and application of monoclonal antibodies for a sandwich enzyme-linked immunosorbent assay of the major soybean allergen, Gly m Bd 30K. *J. Nutr. Sci. Vitaminol.* **1993**, *39*, 389−397.

(15) Nagano, T.; Hirotsuka, M.; Mori, H.; Kohyama, K.; Nishinari, K. Dynamic viscoelastic study on the gelation of 7S globulin from soybeans. *J. Agric. Food Chem.* **1992**, *40*, 941−944.

(16) Maruyama, N.; Sato, R.; Wada, Y.; Matsumura, Y.; Goto, H.; Okuda, E.; Nakagawa, S.; Utsumi, S. Structure-physicochemical function relationships of soybean β-conglycinin constituent subunits. *J. Agric. Food Chem.* **1999**, *47*, 5278−5284.

(17) Shirzadegan, M.; Christie, P.; Seemann, J. R. An efficient method for isolation of RNA from tissue cultured plant cells. *Nucleic Acids Res.* **1991**, *19*, 6055.

(18) Prak, K.; Nakatani, K.; Katsube-Tanaka, T.; Adachi, M.; Maruyama, N.; Utsumi, S. Structure−function relationships of soybean proglycinins at subunit levels. *J. Agric. Food Chem.* **2005**, in press.

(19) Maruyama, N.; Fukuda, T.; Saka, S.; Inui, N.; Kotoh, J.; Miyagawa, M.; Hayashi, M.; Sawada, M.; Moriyama, T.; Utsumi, S. Molecular and structural analysis of electrophoretic variants of soybean seed storage proteins. *Phytochemistry* **2003**, *64*, 701−708.

(20) Maruyama, N.; Mohamad Ramlan, M. S.; Takahashi, K.; Yagasaki, K.; Goto, H.; Hontani, N.; Nakagawa, S.; Utsumi, S. The effect of the N-linked glycans on structural features and physicochemical functions of soybean β-conglycinin homotrimers. *J. Am. Oil Chem. Soc.* **2002**, *79*, 139−144.

(21) Maruyama, N.; Mohamad Ramlan, M. S.; Takahashi, K.; Yagasaki, K.; Goto, H.; Hontani, N.; Nakagawa, S.; Utsumi, S. Structure-physicochemical function relationships of soybean β-conglycinin heterotrimers. *J. Agric. Food Chem.* **2002**, *50*, 4323−4326.

(22) Maruyama, N.; Prak, K.; Motoyama, S.; Choi, S. K.; Yagasaki, K.; Ishimoto, M.; Utsumi, S. Structure-physicochemical function relationships of soybean glycinin at subunit level assessed by using mutant lines. *J. Agric. Food Chem.* **2004**, *52*, 8197−8201.

(23) Kinsella, J. E. Functional properties of soy proteins. *J. Am. Oil Chem. Soc.* **1979**, *56*, 242−258.

(24) Utsumi, S.; Kohno, M.; Mori, T.; Kito, M. An alternate cDNA encoding glycinin A1aBx subunit. *J. Agric. Food Chem.* **1987**, *35*, 210−214.

(25) Utsumi, S.; Kim, C.-S.; Kohno, M.; Kito, M. Polymorphism and expression of cDNAs encoding glycinin subunits. *Agric. Biol. Chem.* **1987**, *51*, 3267−3273.

(26) Negoro, T.; Momma, T.; Fukazawa, C. A cDNA clone encoding a glycinin A1a subunit precursor of soybean. *Nucleic Acids Res.* **1985**, *13*, 6719−6731.

(27) Momma, T.; Negoro, T.; Udaka, K.; Fukazawa, C. A complete cDNA coding for the sequence of glycinin A2B1a subunit precursor. *FEBS Lett.* **1985**, *188*, 117−122.

(28) Davies, C. S.; Coates, J. B.; Nielsen, N. C. Inheritance and biochemical analysis of four electrophoretic variants of β-conglycinin from soybean. *Theor. Appl. Genet.* **1985**, *71*, 351−358.

(29) Pacheco, E.; Barista, F.; Moreira, M. A.; Davies, C. S.; Nielsen, N. C. Urea-elicited changes in relative electrophoretic mobility of certain glycinin and β-conglycinin subunits. *Plant Physiol.* **1984**, *76*, 840−842.

(30) Tandang, M. R. G.; Adachi, M.; Inui, N.; Maruyama, N.; Utsumi, S. The effects of protein engineering of rapeseed procruciferin on its physicochemical and functional properties. *J. Agric. Food Chem.* **2004**, *52*, 6810−6817.

Received for review December 3, 2004. Revised manuscript received February 26, 2005. Accepted March 4, 2005. This work was supported in part by grants to S.U. and N.M. from the Ministry of Education, Science and Culture of Japan and to N.M. from Fuji Foundation for Protein Research.

JF0479620