Properties and Distribution of Soybean Proglycinin Expressed in *Saccharomyces cerevisiae*

Chan-Shick Kim, Makoto Kito and Shigeru Utsumi *

Summary

Glycinin is a predominant storage protein of soybean. The high-level expression system of preproglycinin cDNA in yeast was established. The signal sequence of the expressed protein was cleaved at the correct site. However, most of the expressed proteins were insoluble due to their interaction with intracellular components at the acidic polypeptide region. The expressed proteins separated from the intracellular components by ion-exchange chromatography in the presence of urea were soluble without urea and self-assembled into trimers. The insolubility of the expressed proteins caused an accumulation of modified proteins with disturbed folding. Immunocytochemical distribution demonstrated that the expressed protein from the cDNA encoding preproglycinin accumulated in Golgi-like structure and organellas which may be derived from Golgi-like structure, and one from the cDNA encoding proglycinin homologue protein was found in the cytosol.

INTRODUCTION

Since soybean (*Glycine max* L.) proteins exhibit a higher nutritional value and better functional properties (e.g., heat-induced gelation and emulsification) than other plant foods, they have been widely utilized as food resources. Glycinin, one of the predominant storage proteins of soybean, plays important roles in the functional properties of soybean proteins. Therefore, it is of significance to elucidate the relationship between the structure and the functional properties of glycinin and to improve those properties by protein engineering for further application as a food source.

The constituent subunit of glycinin is synthesized as a single polypeptide precursor (preproglycinin) consisting of covalently linked acidic and basic polypeptides with a signal sequence. The signal sequence is removed cotranslationally in the endoplasmic reticulum, and the resultant proglycinin subunits assemble into trimers of about 8S (Barton et al., 1982; Chrispeels et al., 1982; Tumer et al., 1982). These complexes move from the endoplasmic reticulum into protein bodies, where a specific posttranslational cleavage occurs (Nielsen, 1984). The cleavage results in subunits that consist of an acidic and a basic polypeptide and assemble into hexamers of about 12S (Barton et al., 1982; Chrispeels et al., 1982; Tumer et al., 1982). The acidic polypeptide has an acidic isoelectric point and is hydrophilic, the basic polypeptide a basic isoelectric point and hydrophobic (Nielsen, 1984).

The polymorphism of cDNAs encoding glycinin subunits among cultivars and the occurrence of polymorphic $A_{1a}B_{1b}$ cDNAs in one cultivar have previously been demonstrated in this laboratory (Utsumi et al., 1987a,b). Therefore, the availability of a homogeneous polypeptide for each glycinin subunit affords a route for the precise comprehension of the relationship between the structure and the functional properties of glycinin. To attain this, we established a high-level expression system of cDNA encoding the $A_{1a}B_{1b}$ subunit precursor of glycinin in *Escherichia coli* (Utsumi et al., 1988a; Kim et al., 1990a). The deletion of the cDNA region encoding the signal sequence of the $A_{1a}B_{1b}$ preproglycinin was essential for the accumulation of the expressed protein in *E. coli* (Utsumi et al., 1988a). Fukazawa et al. (1987) observed low-level expression of preproglycinin cDNAs in *E. coli* and that the signal sequence of the expressed product was not processed. These findings suggest that the joining of the ATG codon under the control of the promoter in the expression vector to the first codon in the cDNA encoding proglycinin is required to produce the same protein as that of soybean glycinin in *E. coli*. However, since the N-terminal amino acids of the mature glycinin acidic polypeptides are phenylalanine, leucine, and isoleucine, the translation initiator methionine is not cleaved from the expressed protein in *E. coli* (Sherman et al., 1985; Utsumi et al., 1988a). In other words, we cannot express a protein with the same N-terminal amino acid as that of soybean glycinin in *E. coli*. On the other hand, in a preliminary paper, we revealed that the signal sequence of the soluble expressed protein (0.3% of total soluble proteins of yeast) from the cDNA encoding $A_{1a}B_{1b}$ preproglycinin in *Saccharomyces cerevisiae* was correctly processed (Utsumi et al., 1988b).

We established a relatively high level (~5% of the total yeast protein) expression system of $A_{1a}B_{1b}$ preproglycinin cDNA in yeast and observed that 85–90% of the expressed protein was insoluble. We investigated the properties of the insoluble expressed protein to evaluate the use of the yeast expression system for studying the relationship between the structure and the functional properties of glycinin using protein engineering. In addition, we studied the mechanism of the insolubilization and the cytological distribution of the expressed protein in yeast.

MATERIALS AND METHODS

Bacterial Strains, Medium, and Plasmids. Yeast strain AH22 (*a leu2 his4 can1 cir*$^{+}$) and a yeast–*E. coli* shuttle vector pAM82 with a *PHO5* promoter were the gifts of Dr. K. Matsubara (Miyanohara et al., 1983). Expression plasmids were constructed and propagated in *E. coli* strain JM105 (Yanisch-Perron et al., 1985). Plasmids employed here were pGST4-2-11-10 carrying the cDNA encoding preproglycinin $A_{1a}B_{1b}$ prepared according to the method of Okayama and Berg (1982) (Utsumi et al., 1987b), $pKGA_{1a}B_{1b}1$ (Utsumi et al., 1988a), $pKGA_{1a}B_{1b}$-DI+4Met, and pKK233-stop (Kim et al., 1990b). In the expression plasmid $pKGA_{1a}B_{1b}1$, the ATG codon was ligated to the last codon in the cDNA region encoding the signal sequence of the preproglycinin $A_{1a}B_{1b}$ and a *Nco*I site present in this

* Kyoto University, Uji, Kyoto 611, Japan

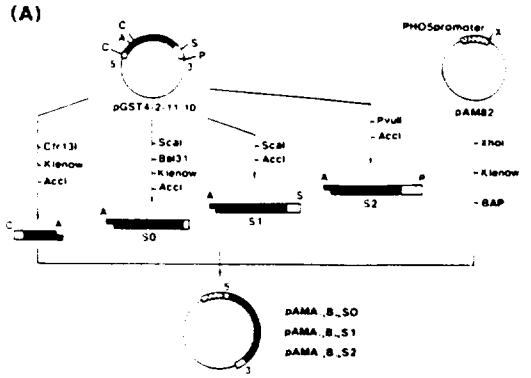

Figure 1. (A) Scheme for construction of expression plasmids for preproglycinin $A_{1a}B_{1b}$. Black, open, and stippled boxes represent preproglycinin coding sequence, 3′- and 5′-noncoding regions of $A_{1a}B_{1b}$ cDNA, and *PHO5* promoter, respectively. A, C, P, S, and X denote the *Acc*I, *Cfr*13I, *Pvu*II, *Sca*I, and *Xho*I sites. BAP is bacterial alkaline phosphatase. (B) Nucleotide and deduced amino acid sequences around the 3′-noncoding region of $A_{1a}B_{1b}$ cDNA. S0, S1, and S2 represent $pAMA_{1a}B_{1b}S0$, $pAMA_{1a}B_{1b}S1$, and $pAMA_{1a}B_{1b}S2$, respectively. The solid line following poly(A) indicates the DNA region derived from Okayama-Berg cloning vector (Okayama and Berg, 1982). Downward arrowheads indicate the 3′-end points of individual expression plasmids. The elements of homology with the consensus signals TTTTTAT (Henikoff et al., 1983) and TAG- - -TAGT- - -TTT or TAG- - -TATGT- - -TTT (Zaret and Sherman, 1982) are indicated.

junction (Utsumi et al., 1988a). The expressed protein from $pKGA_{1a}B_{1b}DI+4Met$ has Arg-Met-Met-Met-Met between Arg61 and Arg62 (Kim et al., 1990b). pKK233-stop is an expression vector with the universal transcription terminator [5′-d(GCT-TAATTAATTAAGC)-3′, Pharmacia] placed between filled-in *Nco*I and *Hin*dIII sites of pKK233-2 (Pharmacia) (Kim et al., 1990b). Burkholder minimal medium (Bostian et al., 1980) supplemented with histidine (20 μg/mg) was used for preparation of high-P_i (1.5 mg of KH_2PO_4/mL) or low-P_i medium (1.5 mg of KCl/mL, in place of P_i).

Construction of Expression Plasmids. Expression plasmids $pAMA_{1a}B_{1b}S0$, $pAMA_{1a}B_{1b}S1$, and $pAMA_{1a}B_{1b}S2$ were constructed as shown in Figure 1A. Briefly, to eliminate the oligo(dG) tail from the 5′ end of the cDNA, pGST4-2-11-10 was cut with *Cfr*13I, filled in with the large (Klenow) fragment of *E. coli* DNA polymerase I in the presence of all four deoxynucleotides, and then cut with *Acc*I, resulting in the *Cfr*13I-*Acc*I fragment. Fragment S0 was obtained by first cutting pGST4-2-11-10 with *Sca*I, digesting with BAL31, treating with Klenow, and cutting with *Acc*I. The 3′ end of this fragment was determined by sequencing. Fragments S1 and S2 were obtained by cutting pGST4-2-11-10 with *Sca*I-*Acc*I and *Pvu*II-*Acc*I, respectively. The *Cfr*13I (filled-in)-*Acc*I fragment was inserted into the *Xho*I (filled-in) site of pAM82 with fragment S0, S1, or S2 to generate $pAMA_{1a}B_{1b}S0$, $pAMA_{1a}B_{1b}S1$, or $pAMA_{1a}B_{1b}S2$, respectively. Expression plasmids having cDNA inserts in the proper orientation for expression were identified by restriction endonuclease analysis. The resultant expression plasmids have a *Xho*I site at the 5′ end of the cDNA.

Expression plasmids $pAMA_{1a}B_{1b}$-N, $pAMA_{1a}B_{1b}$-A, $pAMA_{1a}B_{1b}$-D, and $pAMA_{1a}B_{1b}$-X were constructed as shown in Figure 4A. Briefly, the *Nco*I (filled-in)-*Nde*I (filled-in), *Nco*I (filled-in)-*Acc*II, *Nco*I (filled-in)-*Dra*III (blunted with mung bean nuclease) or *Nco*I (filled-in)-*Xmn*I fragment from pGST4-2-11-10 was inserted into the *Nco*I site (blunted with mung bean nuclease) of pKK233-stop to place universal transcription terminater downstream of each fragment. The *Xma*I-*Xmn*I fragments from the resultant plasmids were obtained. The *Xma*I-*Pvu*II fragment from pGST4-2-11-10 was recirculated with the small *Xma*I-*Pvu*II fragment from $pAMA_{1a}B_{1b}S1$. The resultant plasmid was cut with *Eco*81I (filled-in) and *Xma*I. The *Xma*I-*Eco*81I (filled-in) fragment was ligated with each *Xma*I-*Xmn*I fragment. From the resultant plasmids the *Xma*I-*Pvu*II fragments were obtained. Each *Xma*I-*Pvu*II fragment was ligated with the large *Xma*I-*Pvu*II fragment from $pAMA_{1a}B_{1b}S1$ to generate $pAMA_{1a}B_{1b}$-N, $pAMA_{1a}B_{1b}$-A, $pAMA_{1a}B_{1b}$-D, and $pAMA_{1a}B_{1b}$-X.

Expression plasmid $pAMA_{1a}B_{1b}DI+4Met$ was constructed as shown in Figure 6. Briefly, the small *Pst*I-*Pvu*II fragment containing the whole cDNA region of pGST4-2-11-10 was inserted between the *Pst*I and *Sma*I sites of pUC19. The resultant plasmid was cut with *Acc*I and recirculated with synthetic DNA d(5′-AGAATGATGATGATG-3′) and d(3′-TTACTACTACTACTC-5′) to generate $pUA_{1a}B_{1b}DI+4Met$. The *Nco*I-*Xma*I fragment from $pUA_{1a}B_{1b}DI+4Met$ and the *Hin*dIII-*Nco*I fragment from $pAMA_{1a}B_{1b}S1$ were inserted between the *Hin*dIII and *Xma*I sites of $pAMA_{1a}B_{1b}S1$ to generate $pAMA_{1a}B_{1b}DI+4Met$.

Expression plasmid $pAMA_{1a}B_{1b}$-signal was constructed as shown in Figure 8A. The *Nco*I-*Xma*I fragment from $pKGA_{1a}B_{1b}1$ and the *Nco*I-*Hin*dIII fragment from $pAMA_{1a}B_{1b}S1$ were inserted between the *Xma*I and *Hin*dIII sites of $pAMA_{1a}B_{1b}S1$ to generate $pAMA_{1a}B_{1b}$-signal.

Transformation of Yeast Cells. Cells were transformed according to the method of Ito et al. (1983). Leu$^+$ transformants were selected on SD medium containing histidine (20 μg/mL) and 2% agar.

Expression and Detection of the Expressed Proteins. AH22 cells harboring individual expression plasmids were grown in 10 mL of high-P_i Burkholder minimal medium supplemented with histidine at 30 °C. At A_{610} = 0.3, yeast cells were collected and suspended in 10 mL of low-P_i medium for induction. At A_{610} = 1.0, yeast cells were centrifuged. The resultant cells were vortexed with 1-mm glass beads in 0.7 mL of 10 mM Tris-HCl buffer (pH 7.5) containing 1 mM EDTA and 1 mM phenylmethanesulfonyl fluoride (buffer A) for 5 × 1 min. The vortexed samples (10 μg) containing both the soluble and debris fractions were dialyzed against $NaDodSO_4$ sample buffer [62.5 mM Tris-HCl buffer (pH 6.8) containing 2% $NaDodSO_4$, 0.2 M 2-mercaptoethanol, and 10% glycerol] and then subjected to $NaDodSO_4$-polyacrylamide gel electrophoresis ($NaDodSO_4$-PAGE) according to the procedure of Laemmli (1970). The electrophoresed proteins were visualized in gels with Coomassie brilliant blue R-250 or transferred to nitrocellulose filters and analyzed with affinity-purified glycinin antibodies (Utsumi et al., 1987a, 1988b). The amount of the expressed protein was measured by densitometric analysis following Coomassie brilliant blue staining or immunoblotting of $NaDodSO_4$ gels.

Cultures (500 mL) of AH22 harboring $pAMA_{1a}B_{1b}S1$ induced as described above were harvested by centrifugation, and the cells were homogenized with a Braun homogenizer (B. Braun, Type 2876, Apparatebau Melsungers, Germany) as described previously (Utsumi et al., 1988b). The cell debris and the supernatant were fractionated by centrifugation at 10000g. The cell debris was suspended in buffer A. Aliquots (10 μg) of the cell debris and the supernatant fractions were analyzed by $NaDodSO_4$-PAGE as described above.

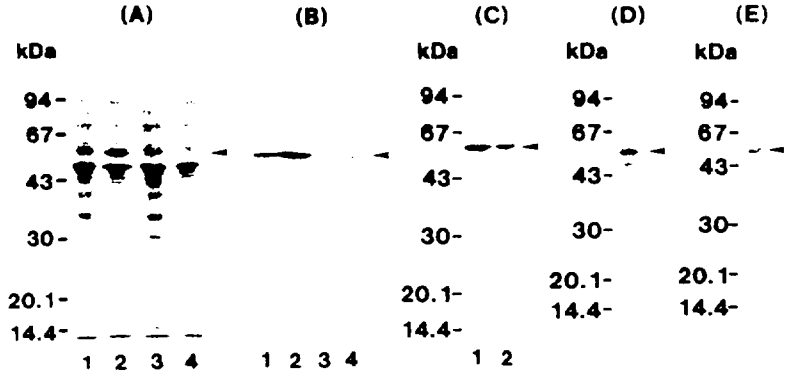

Figure 2. $NaDodSO_4$–PAGE analysis of the expressed proteins from individual expression plasmids. The total yeast cells and the soluble or insoluble fractions were analyzed by $NaDodSO_4$–PAGE using 11% gels. (A) The yeast cells harboring $pAMA_{1a}B_{1b}S2$ (lane 1), $pAMA_{1a}B_{1b}S1$ (lane 2), $pAMA_{1a}B_{1b}S0$ (lane 3), or pAM82 (lane 4) were applied to $NaDodSO_4$ gels and stained with Coomassie brilliant blue. (B) $NaDodSO_4$ gels identical with those in (A) were detected by immunoblotting. (C) $NaDodSO_4$ gels of the insoluble (lane 1) and soluble (lane 2) fractions of the extract from yeast harboring $pAMA_{1a}B_{1b}S1$ were detected by immunoblotting. (D) $NaDodSO_4$ gels of the insoluble fraction of the extract from yeast harboring $pAMA_{1a}B_{1b}S1$ were stained with Coomassie brilliant blue. (E) $NaDodSO_4$ gels of the pooled fractions of the expressed protein from $pAMA_{1a}B_{1b}S1$ fractionated by Q-Sepharose column chromatography were stained with Coomassie brilliant blue. Arrowheads indicate the position of the expressed protein. The numbers on the left denote MW.

Determination of N-Terminal Amino Acid Sequence. The N-terminal amino acid sequence of the expressed proteins was determined by using a gas-phase protein sequencer (Applied Biosystems, Model 477A) after the protein samples (12 μg) separated on $NaDodSO_4$ gels were blotted onto a PVDF membrane (Millipore) according to the procedure of Matsudaira (1987).

Analysis of N-Glycosylation. The expressed proteins was applied onto $NaDodSO_4$ gels for electrotransfer to nitrocellulose filters and then analyzed with peroxidase-conjugated Con A (Seikagaku Kogyo) as recommended by the manufacturer.

Q-Sepharose Column Chromatography. The cell debris fractions from 500-mL cultures were dissolved in 8 mL of 50 mM Tris-HCl buffer (pH 8.5) containing 6 M urea, 0.1 M 2-mercaptoethanol, and 1 mM EDTA (buffer B) and then dialyzed against buffer B. One-fifth of the dialyzed sample was applied to a Q-Sepharose column (1 × 6.4 cm) equilibrated with buffer B. The column was washed with buffer B until the A_{280} of the eluate had fallen below 0.01 and then eluted with a 200-mL linear gradient of 0–1.0 M NaCl in buffer B. Fractions (3 mL/fraction) were analyzed by $NaDodSO_4$–PAGE.

Analysis of Self-Assembly of the Expressed Protein. The expressed protein partially purified by Q-Sepharose column chromatography was dialyzed against 35 mM KP_i buffer (pH 7.6) containing 0.4 M NaCl and 40% glycerol for renaturation and then against 35 mM KP_i buffer (pH 7.6) containing 0.4 M NaCl (Utsumi et al., 1980). The dialysate was centrifuged in a sucrose density gradient as described previously (Utsumi et al., 1988a). The 2S, 7S, and 11S fractions purified from soybeans according to the method of Thanh and Shibasaki (1976) were run in parallel as size markers.

Protein Measurement. Proteins in the unfractionated and fractionated samples were measured according to the method of Bradford (1976) with bovine serum albumin as the standard.

Electron Microscope Immunocytochemistry. All subsequent stages in the technique, including antibody labeling, were carried out essentially as described by Griffiths et al. (1983). Yeast-expressed cells carrying $pAMA_{1a}B_{1b}S1$ or $pAMA_{1a}B_{1b}$-signal were fixed in a mixture of 4% formaldehyde and 1% glutaraldehyde in 0.1 M phosphate buffer (pH 7.4) for 24 h at 4 °C, infused with a mixture of 30% poly(vinylpyrrolidone) (PVP; MW 10 000; Sigma Chemical Co., St. Louis, MO) and 1.61 M sucrose, frozen in liquid nitrogen, and sectioned with the Reichert-Jung Ultracut-E with cryoattachment FC-4E at −100 to −125 °C into an estimated thickness of 50–100 nm. Sections were placed on Formvar-carbon-coated copper grids which were floated, section-side down, on 250 μL drops of reagents on a Parafilm sheet. All steps were performed at room temperature. Grids were incubated in 0.02 M glycine in PBG buffer [PBS (0.01 M phosphate buffer (pH 7.4) containing 0.15 M NaCl) supplemented with 0.5% BSA and 0.1% gelatin (E. Merck, art. no. 4070)] for 10 min and then were rinsed in PBG buffer for 5 min. The rinsed grids were placed on drops of 5% BSA (w/v) in PBG buffer for 10 min and then incubated for 1 h in primary antibody in PBG buffer. The grids were washed in PBG for 10 min, incubated for 1 h in 10 nm goat-anti-rabbit-IgG colloidal gold (Janssen Life Science Products) solution diluted in PBG, rinsed in PBG followed by PBS for 30 min, and postfixed in 2% glutaraldehyde in PBS. Sections were stained for 10 min in 4% (w/v) aqueous uranyl acetate and rinsed with distilled water. The grids were examined and photographed with a Hitachi H-700H electron microscope.

RESULTS

Expression of $A_{1a}B_{1b}$ Preproglycinin cDNA in Yeast and Properties of the Expressed Protein. To elucidate the effect of the length of the 3′-noncoding region of $A_{1a}B_{1b}$ cDNA on the level of the expression, we constructed the expression plasmids $pAMA_{1a}B_{1b}S0$, $pAMA_{1a}B_{1b}S1$, and $pAMA_{1a}B_{1b}S2$ which have different lengths of the 3′-noncoding region of $A_{1a}B_{1b}$ cDNA (Figure 1A). The 3′-noncoding regions of individual expression plasmids are shown in Figure 1B. Each expression plasmid was transformed into a yeast recipient strain, AH22. As shown in Figure 2, a protein of MW ~57 000 was stained by Coomassie brilliant blue (Figure 2A, lanes 1–3) and recognized by the glycinin antibodies (Figure 2B, lanes 1–3) in samples from yeast harboring each expression plasmid but not in those from control cells harboring pAM82 alone (Figure 2A,B, lane 4). $pAMA_{1a}B_{1b}S1$ (lane 2) was the most suitable expression plasmid, since the expressed protein comprised ~5% of the total yeast proteins, or 30–40 mg/L of culture. On the other hand, the expression levels of $pAMA_{1a}B_{1b}S0$ (lane 3) and $pAMA_{1a}B_{1b}S2$ (lane 1) were ~5% and ~30% of that of $pAMA_{1a}B_{1b}S1$ (lane 2), respectively (Figure 2B). Expressed proteins from individual expression plasmids were not detected in the media (data not shown).

The yeast cells harboring $pAMA_{1a}B_{1b}S1$ expressing $A_{1a}B_{1b}$ protein at the highest levels from large-scale culture were homogenized. The resultant soluble and insoluble fractions (the insoluble fraction corresponds to the debris fraction described under Materials and Methods) were electrophoresed on $NaDodSO_4$ gels. Ten to 15% and 85–

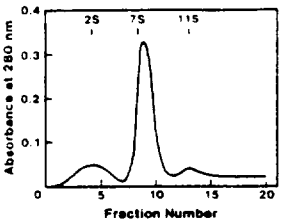

Figure 3. Self-assembly of the partially purified expressed protein. Assembly was assayed by centrifugation on a 12 mL 10–30% (w/v) linear sucrose density gradient. Sedimentation is from left to right. Sedimentation standards are given.

90% of the expressed proteins were observed in the soluble and the insoluble fractions, respectively (Figure 2C). In the insoluble fraction, the expressed protein accounted for ~50% of the total insoluble proteins (Figure 2D).

The sequence of the expressed protein in the insoluble fraction was determined to be Phe-Ser-Ser-(X)-Glu-Gln-Pro-Gln-Gln-Asn-Glu-. This is consistent with the N-terminal sequence of mature A_{1a} polypeptide predicted from the nucleotide sequence of $A_{1a}B_{1b}$ cDNA (Utsumi et al., 1987b). Thus, the signal sequence of the insoluble expressed protein accumulated at the level of ~5% of the total yeast proteins was cleaved at the same processing site as that in soybean. This together with the fact that the MW of the expressed protein is ~57 000 indicates that the expressed protein appears to be in the pro form.

The nucleotide sequence of $A_{1a}B_{1b}$ cDNA indicates that there is one potential site for N-glycosylation (Sharon and Lis, 1979): the asparagine residue at position 418, although glycinin is not a glycoprotein. We have observed that the soluble expressed protein is not N-glycosylated (Utsumi et al., 1988b). The blots of the insoluble expressed protein were allowed to interact with peroxidase-conjugated Con A. The protein did not interact with Con A, suggesting that the insoluble expressed protein is not N-glycosylated (data not shown).

We attempted to determine whether the insoluble expressed protein can assemble into trimers like the native proglycinin. The debris fraction could be solubilized in 35 mM KP_i buffer (pH 7.6) containing 6 M urea and 0.2 M 2-mercaptoethanol. However, when the solubilized sample was dialyzed against 35 mM KP_i buffer (pH 7.6) containing 0.4 M NaCl, the standard buffer for glycinin, the expressed protein precipitated. Since the expressed protein from $A_{1a}B_{1b}$ cDNA in *E. coli* is soluble, the insolubility of the expressed protein in yeast may be due to its interaction with some intracellular component. Therefore, the debris fraction from the yeast cells harboring $pAMA_{1a}B_{1b}S1$ was applied to a Q-Sepharose column in the presence of urea. The $A_{1a}B_{1b}$ protein was eluted as a single peak (data not shown). The pooled fractions with a purity of ~80% (Figure 2E) were renatured by dialysis. The dialyzed expressed protein was soluble. It was centrifuged on a sucrose density gradient (Figure 3). The $A_{1a}B_{1b}$ protein predominantly sedimented in fraction 9, which corresponds to a size of 7–9 S. This indicates that the $A_{1a}B_{1b}$ protein expressed in yeast can self-assemble into trimers.

Interaction Site of the Expressed Protein with an Insolubility Forming Component. It is likely that the component which induces the insolubilization of the expressed protein in yeast interacts with the acidic or the basic polypeptide region of the expressed protein, since they have different characteristics. To investigate this hypothesis, we deleted stepwise the cDNA region corresponding to the basic polypeptide and constructed expression plasmids $pAMA_{1a}B_{1b}$-N, $pAMA_{1a}B_{1b}$-A, $pAMA_{1a}B_{1b}$-D, and $pAMA_{1a}B_{1b}$-X (Figure 4). After the disruption of the cells harboring individual expression plasmids by a Braun homogenizer, the debris and the soluble fractions were applied to $NaDodSO_4$-PAGE and the proteins were visualized by Coomassie brilliant blue staining (Figure 5A) and immunoblotting (Figure 5B). All deleted proteins were predominantly observed in the debris fractions (Figure 5). This suggests that the region corresponding to the acidic polypeptide of the $A_{1a}B_{1b}$ protein interacts with the component.

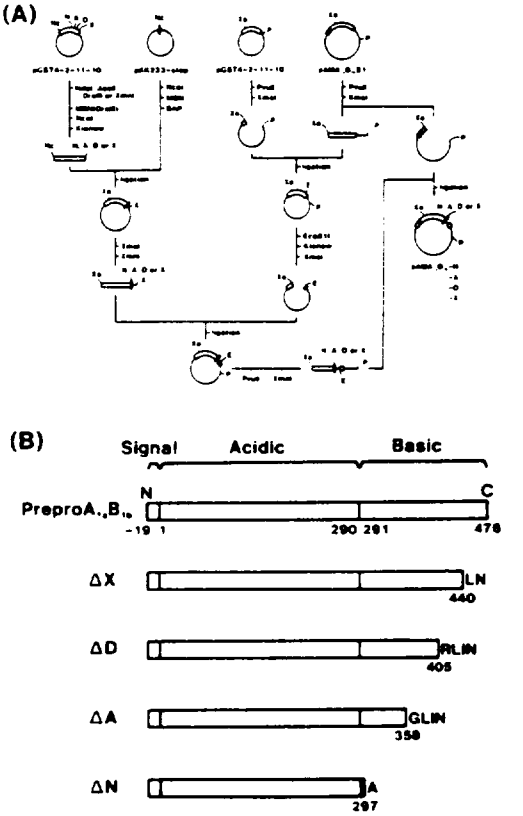

Figure 4. (A) Scheme for construction of expression plasmids $pAMA_{1a}B_{1b}$-N, -A, -D, and -X. Open, stippled, and thick boxes represent $A_{1a}B_{1b}$ cDNA, *PHO5* promoter, and the universal translation terminator [5'-d(GCTTAATTAATTAAGC)-3', Pharmacia], respectively. A, D, E, N, Nc, P, X, and Xa denote the *Acc*II, *Dra*III, *Eco*81I, *Nde*I, *Nco*I, *Puv*II, *Xmn*I, and *Xma*I sites, respectively. MBN is mung bean nuclease; BAP is bacterial alkaline phosphatase. (B) Construction of the deleted proteins. ΔX, ΔD, ΔA, and ΔN lack from the 441st, 406th, 359th, and 298th to the C terminus, respectively. Extra C-terminal amino acids in each deleted protein are indicated.

Advantage of the Insolubilization of the Expressed Protein in Yeast. There was very low accumulation of expressed protein ($A_{1a}B_{1b}$DI+4Met) in *E. coli* (Kim et al., 1990b) from the cDNA to which a synthetic DNA encoding four continuous methionines was inserted into the *Acc*I site in the region corresponding to domain I (Argos et al., 1985). However, we would expect that the modified protein would accumulate in yeast cells, because the expressed protein from the glycinin cDNA was insoluble in yeast cells. We constructed an expression plasmid $pAMA_{1a}B_{1b}$-

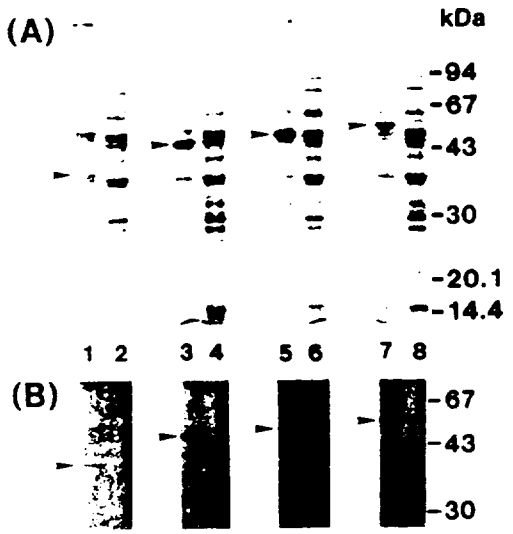

Figure 5. $NaDodSO_4$-PAGE analysis of the expressed protein from individual deleted expression plasmids. The soluble (lanes 2, 4, 6, and 8) and insoluble (lanes 1, 3, 5, and 7) fractions of the extract from the yeast harboring individual expression plasmids were applied to $NaDodSO_4$ gels and stained with Coomassie brilliant blue (A) or detected by immunoblotting (B). (Lanes 1 and 2) $pAMA_{1a}B_{1b}$-N; (lanes 3 and 4) $pAMA_{1a}B_{1b}$-A; (lanes 5 and 6) $pAMA_{1a}B_{1b}$-D; (lanes 7 and 8) $pAMA_{1a}B_{1b}$-X. Arrowheads indicate the position of the expressed protein. The numbers on the right denote MW.

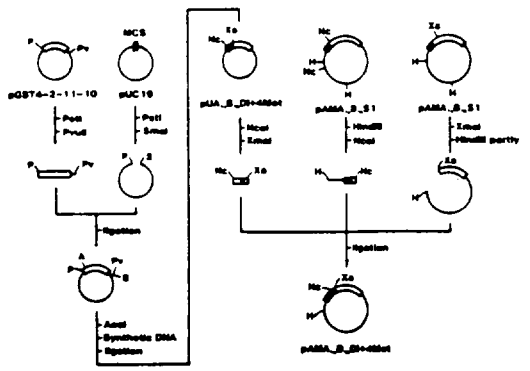

Figure 6. Scheme for construction of expression plasmid $pAMA_{1a}B_{1b}$DI+4Met. Synthetic DNAs used for construction of $pAMA_{1a}B_{1b}$DI+4Met were 5′-AGAATGATGATGATG-3′ and 3′-TTACTACTACTACTC-5′. Open, black, and stippled boxes represent $A_{1a}B_{1b}$cDNA, inserted synthetic DNA, and *PHO5* promoter, respectively. A, H, Nc, P, Pv, S, and Xa denote the *Acc*I, *Hin*dIII, *Nco*I, *Pst*I, *Pvu*II, *Sma*I, and *Xma*I sites, respectively. MCS is multicloning site.

DI+4Met (Figure 6). The expressed protein from this plasmid was predominantly insoluble (Figure 7), and the expression level was similar to that from $pAMA_{1a}B_{1b}$S1 (Figure 7A; compare with Figure 2D). This indicates that yeast accumulates a modified protein which cannot be accumulated in *E. coli* (Kim et al., 1990b). The immunoreactivity of the expressed protein from $pAMA_{1a}B_{1b}$-DI+4Met was much lower than that of the expressed protein from $pAMA_{1a}B_{1b}$S1 (Figure 7B, lanes 1 and 2).

Cytological Distribution of the Expressed Protein in Yeast. In soybean seeds, the signal sequences of the preproglycinins synthesized on polysomes are removed cotranslationally in the endoplasmic reticulums and the

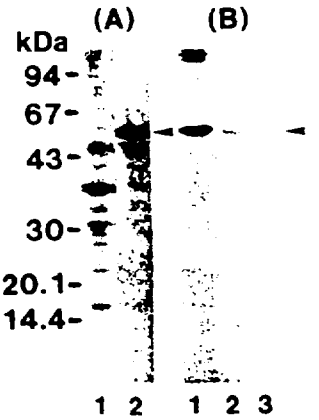

Figure 7. $NaDodSO_4$-PAGE analysis of the expressed protein from $pAMA_{1a}B_{1b}$DI+4Met. (A) The soluble (lane 1) and insoluble (lane 2) fractions from the yeast cells harboring $pAMA_{1a}B_{1b}$-DI+4Met were applied to $NaDodSO_4$ gels and stained with Coomassie brilliant blue. (B) Detection by immunoblotting. (Lane 1) Insoluble fraction of $pAMA_{1a}B_{1b}$S1; (lane 2) insoluble fraction of $pAMA_{1a}B_{1b}$DI+4Met; (lane 3) soluble fraction of $pAMA_{1a}B_{1b}$DI+4Met. Arrowheads indicate the position of the expressed protein. The numbers on the left denote MW.

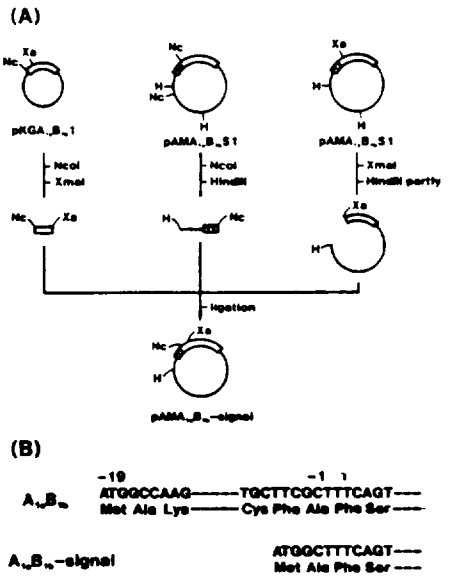

Figure 8. (A) Scheme for construction of expression plasmid $pAMA_{1a}B_{1b}$-signal. Open and stippled boxes represent $A_{1a}B_{1b}$ cDNA and *PHO5* promoter, respectively. H, Nc, and Xa denote the *Hin*dIII, *Nco*I, and *Xma*I sites, respectively. (B) Construction of the expressed protein from $pAMA_{1a}B_{1b}$-signal.

resultant proglycinins move into the protein bodies (Barton et al., 1982; Chrispeels et al., 1982; Turner et al., 1982). It is of interest to determine if preproglycinins are processed similarly in yeast. To observe also the expressed protein from the cDNA-deleted region corresponding to the signal sequence, we constructed an expression plasmid $pAMA_{1a}B_{1b}$-signal of which the expressed protein lacks 18 residues of the signal sequence consisting of 19 residues (Figure 8). The expression level of this plasmid was 20–30% of that of $pAMA_{1a}B_{1b}$S1, and the expressed protein was insoluble (data not shown). The N-terminal amino acid sequence of this protein could not be determined.

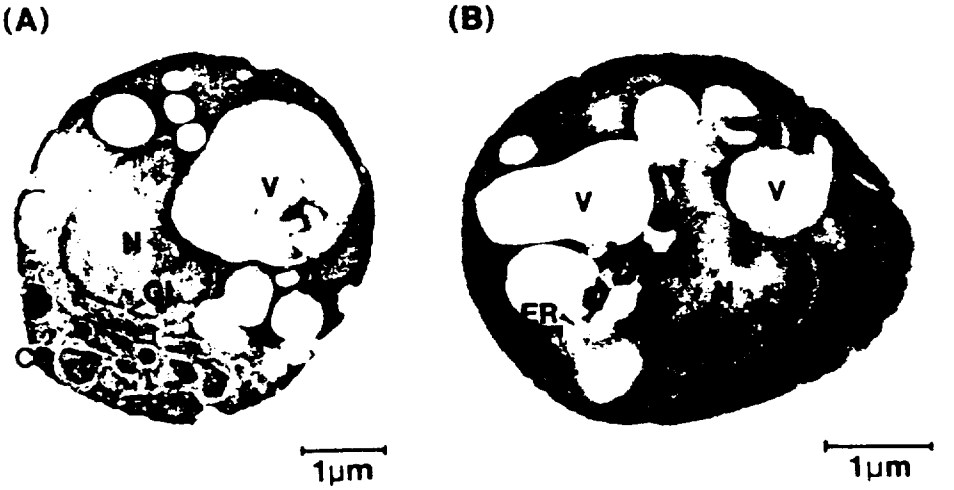

Figure 9. Immunogold labeling of yeast harboring pAMA$_{1a}$B$_{1b}$S1 (A) and pAMA$_{1a}$B$_{1b}$-signal (B). N, nucleus; V, vacuole; GL, Golgi-like structure; ER, endoplasmic reticulum-like structure; CO, a certain organella.

This may be due to acylation of the N terminus, since the penultimate residue of the initiator methionine is alanine (Huang et al., 1987).

Distribution of the expressed proteins from pAMA$_{1a}$B$_{1b}$-S1 and pAMA$_{1a}$B$_{1b}$-signal was demonstrated by immunocytochemical localization in sections cut by cryoultramicrotomy. In pAMA$_{1a}$B$_{1b}$S1, most of the amorphous materials were associated with Golgi-like structure, which is similar to that observed by Semenza et al. (1990), and a certain organella which seems to be derived from the Golgi-like structure. Some additional staining was also associated with the endoplasmic reticulum-like structure, which resembles that observed by Yarwood et al. (1987) (Figure 9A). On the other hand, in pAMA$_{1a}$B$_{1b}$-signal, most label was observed in the cytosol, most of them associated with the surrounding of the vacuole (Figure 9B).

DISCUSSION

Among the expression plasmids pAMA$_{1a}$B$_{1b}$S0, pAMA$_{1a}$B$_{1b}$S1, and pAMA$_{1a}$B$_{1b}$S2 where 3′-noncoding regions of the cDNAs differ in length from each other (Figure 1B), pAMA$_{1a}$B$_{1b}$S1 containing approximately two-thirds of the 3′-noncoding region was the most efficient expression plasmid. The expression levels of pAMA$_{1a}$B$_{1b}$-S2 containing the whole 3′-noncoding region, 125 nucleotides of poly(A), and a short DNA fragment derived from the Okayama-Berg cloning vector and pAMA$_{1a}$B$_{1b}$S0 containing approximately one-fifth of the 3′-noncoding region were ~30% and ~5% of that of pAMA$_{1a}$B$_{1b}$S1, respectively (Figure 2A,B). The 3′-noncoding region of the cDNA has several sequences of homology with the consensus signals TTTTTAT (Henikoff et al., 1983) and TAG- - -TAGT- - -TTT or TAG- - -TATGT- - -TTT (Zaret and Sherman, 1982), thought to be important in yeast transcription termination (Figure 1B). If we assume that the element, similar to the consensus sequence proposed by Zaret and Sherman (1982), occurs in pAMA$_{1a}$B$_{1b}$S1 and pAMA$_{1a}$B$_{1b}$S2 but not in pAMA$_{1a}$B$_{1b}$S0 as a transcription terminator and that the function of this element is inhibited by the poly(A) region of A$_{1a}$B$_{1b}$ cDNA and/or the short DNA fragment derived from the Okayama-Berg cloning vector, the difference in the expression level among individual expression plasmids can be explained. However, identification of the sizes of transcripts from individual expression plasmids is necessary to determine whether this assumption is feasible. The expression level (~5% of the total yeast proteins) of pAMA$_{1a}$B$_{1b}$S1 is comparable with those of other heterologous proteins, 1-2% of pea legumin (Yarwood et al., 1987), 5% of pea vicilin (Watson et al., 1988), 3% of phaseolin (Cramer et al., 1987), and 4-5% of zein (Coraggio et al., 1988).

Most of the expressed proteins from the cDNAs encoding the preproglycinin with a signal sequence and the proglycinin homologue protein without the signal sequence were insoluble. However, the expressed protein from the cDNA encoding the proglycinin homologue protein in *E. coli* is absolutely soluble and is able to self-assemble correctly (Utsumi et al., 1988a; Kim et al., 1990a). Moreover, the glycinin from soybean can refold correctly from a reduced unfolded state (Utsumi et al., 1980). These facts suggest that interaction of the yeast-expressed protein with some component derived from the yeast cells causes the insolubilization. Cytological distributions of the expressed proteins from pAMA$_{1a}$B$_{1b}$S1 and pAMA$_{1a}$B$_{1b}$-signal are different from each other (Figure 9). This indicates that the component which interacts with the expressed protein is ubiquitous in the yeast cells or that the expressed proteins interact with the component during extraction from the yeast cells.

The expressed acidic polypeptide from pAMA$_{1a}$B$_{1b}$-N was also insoluble. This indicates that the interaction site of the expressed protein with the insolubility forming partner is located in the acidic polypeptide region. The acidic polypeptide is rich in acidic and hydrophilic amino acids and has an acidic isoelectric point (Nielsen, 1984). Therefore, the molecular forces involved in the interaction between the acidic polypeptide and the partner may be electrostatic interactions and/or hydrogen bonds. Ten to 15% of the total expressed protein was soluble after extraction. Furthermore, additional expressed proteins became soluble after the debris fraction was diluted with 35 mM KP$_i$ buffer (pH 7.6) containing 0.4 M NaCl (data not shown), suggesting that electrostatic interaction is involved. Polyamines or basic proteins are possible candidates for binding the acidic portion and precipitating

the protein. Derby and Creighton (1990) recently suggested that one reason for the insolubility of bovine pancreatic trypsin inhibitor (BPTI) expressed in *E. coli* is the formation of a complex with some other tightly bound substance. They thought that the complexing partner is a polyanion since BPTI is a very basic protein (Derby and Creighton, 1990).

The signal sequence of the preproglycinin expressed in yeast was cleaved at the same processing site as that in soybean, although the expression level was relatively high (~5% of the total yeast protein). This indicates that the yeast processing system was not saturated even under high expression levels. After cotranslational cleavage of the signal sequence, the resultant proglycinin was transported into the endoplasmic reticulum and then moved into a certain organella (seems to be derived from Golgi-like structure) via Golgi-like structure (Figure 9A) but was not secreted. The organeilas would not be related to inclusion body, because they were not observed in the yeast harboring $pAMA_{1a}B_{1b}$-signal. Yarwood et al. (1987) observed that pea legumin expressed in yeast was accumulated in the Golgi apparatus and the endoplasmic reticulum. On the other hand, the proglycinin homologue protein expressed in yeast was found in the cytosol (Figure 9B).

The signal peptide of the preproglycinin expressed in yeast was processed at the orrect site. The resultant proglycinin self-assembled into trimers (Figure 3) and was not N-glycosylated, similar to that of soybean. On the other hand, modified proglycinin in which folding was partly disturbed due to the introduced alteration accumulated in yeast because of its insolubility (Figure 7). In *E. coli* expression system, we indicated that the expressed protein containing a modification which disturbs the correct folding is susceptible to proteinase digestion and is then scarcely accumulated in the *E. coli* cells (Kim et al., 1990b). The availability of large quantities of modified glycinin is crucial to the study of the protein engineering of glycinin. Therefore, it is concluded that the yeast expression system of glycinin cDNA is useful for studying the relationship between the structure and function of protein-engineered glycinins, although it is necessary to solubilize the expressed proteins.

ACKNOWLEDGMENT

We thank Professor K. Matsubara for his kind gifts of pAM82 and AH22 and Ms. S. Nakamura (Hitachi) for thin sectioning. This research was partially supported by grants from the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED

Argos, P.; Narayana, S. V. L.; Nielsen, N. C. Structural Similarity between Legumin and Vicilin Storage Proteins from Legumes. *EMBO J.* **1985**, *4*, 1111–1117.

Barton, K. A.; Thompson, J. F.; Madison, J. T.; Rosenthal, R.; Jarvis, N. P.; Beachy, R. M. The Biosynthesis and Processing of High Molecular Weight Precursors of Soybean Glycinin Subunits. *J. Biol. Chem.* **1982**, *257*, 6089–6095.

Bostian, K.; Lemire, J. M.; Cannon, L. E.; Halvorson, H. O. *In vitro* Synthesis of Repressible Yeast Acid Phosphatase: Identification of Multiple mRNAs and Products. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 4504–4508.

Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72*, 248–254.

Chrispeels, M. J.; Higgins, T. J. V.; Spencer, D. Assembly of Storage Protein Oligomers in the Endoplasmic Reticulum and Processing of the Polypeptides in the Protein Bodies of Developing Pea Cotyledons. *J. Cell Biol.* **1982**, *93*, 306–313.

Coraggio, I.; Martegani, E.; Compagno, C.; Porro, D.; Alberghina, L.; Bernard, L.; Faoro, F.; Viotti, A. Differential Targeting and Accumulation of Normal and Modified Zein Polypeptides in Transformed Yeast. *Eur. J. Cell Biol.* **1988**, *47*, 165–172.

Cramer, J. H.; Lea, K.; Schaber, M. D.; Kramer, R. A. Signal Peptide Specificity in Posttranslational Processing of the Plant Protein Phaseolin in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **1987**, *7*, 121–128.

Derby, N. J.; Creighton, T. E. Folding Proteins. *Nature* **1990**, *344*, 715–716.

Fukazawa, C.; Udaka, K.; Murayama, A.; Higuchi, W.; Totsuka, A. Expression of Soybean Glycinin Subunit Precursor cDNAs in *Escherichia coli*. *FEBS Lett.* **1987**, *224*, 125–127.

Griffiths, G.; Simons, K.; Warren, G.; Tokuyasu, K. T. Immunoelectron Microscopy Using Thin, Frozen Sections: Application to Studies of the Intracellular Transport of Semliki Forest Virus Spike Glycoproteins. *Methods Enzymol.* **1983**, *96*, 466–485.

Henikoff, S.; Kelly, J. D.; Cohen, E. H. Transcription Terminates in Yeast Distal to a Control Sequence. *Cell* **1983**, *33*, 607–614.

Huang, S.; Elliott, R. C.; Liu, P.-S.; Kodur, R. K.; Weickmann, J. L.; Lee, J.-H.; Blair, L. C.; Ghosh-Dastidar, P.; Bradshaw, R. A.; Bryan, K. M.; Einarson, B.; Kendall, R. L.; Kolacz, K. H.; Saito, K. Specificity of Cotranslational Amino-Terminal Processing of Proteins in Yeast. *Biochemistry* **1987**, *26*, 8242–8246.

Ito, H.; Fukuda, Y.; Murata, K.; Kimura, A. Transformation of Intact Yeast Cells Treated with Alkali Cations. *J. Bacteriol.* **1983**, *153*, 163–168.

Kim, C.-S.; Kamiya, S.; Kanamori, J.; Utsumi, S.; Kito, M. High Level Expression, Purification and Functional Properties of Soybean Proglycinin from *Escherichia coli*. *Agric. Biol. Chem.* **1990a**, *54*, 1543–1550.

Kim, C.-S.; Kamiya, S.; Sato, T.; Utsumi, S.; Kito, M. Improvement of Nutritional Value and Functional Properties of Soybean Glycinin by Protein Engineering. *Protein Eng.* **1990b**, *3*, 725–731.

Laemmli, U. K. Cleavage of Structural Proteins During Assembly of the Head of Bacteriophage T4. *Nature* **1970**, *227*, 680–685.

Matsudaira, P. Sequence from Picomole Quantities of Proteins Electroblotted onto Polyvinylidene Difluoride Membranes. *J. Biol. Chem.* **1987**, *262*, 10035–10038.

Miyanohara, A.; Toh-e, A.; Nozaki, C.; Hamada, F.; Ohtomo, H.; Matsubara, K. Expression of Hepatitis B Surface Antigen Gene in Yeast. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 1–5.

Nielsen, N. C. The Chemistry of Legume Storage Proteins. *Philos. Trans. R. Soc. London B* **1984**, *304*, 287–296.

Okayama, H.; Berg, D. High-Efficiency Cloning of Full-Length cDNA. *Mol. Cell. Biol.* **1982**, *2*, 161–170.

Semenza, J. C.; Hardwick, K. G.; Dean, N.; Pelham, H. R. B. *ERD2*, a Yeast Gene Required for the Receptor-Mediated Retrieval of Luminal ER Proteins from the Secretory Pathway. *Cell* **1990**, *61*, 1349–1357.

Sharon, N.; Lis, H. Comparative Biochemistry of Plant Glycoproteins. *Biochem. Soc. Trans.* **1979**, *7*, 783–799.

Sherman, F.; Stewart, J. W.; Tsunasawa, S. Methionine or Not Methionine at the Beginning of a Protein. *BioEssays* **1985**, *3*, 27–31.

Thanh, V. H.; Shibasaki, K. Major Proteins of Soybean Seeds. A Straightforward Fractionation and Their Characterization. *J. Agric. Food Chem.* **1976**, *24*, 1117–1121.

Tumer, N. E.; Richter, J. D.; Nielsen, N. C. Structural Characterization of the Glycinin Precursors. *J. Biol. Chem.* **1982**, *257*, 4016–4018.

Utsumi, S.; Inaba, H.; Mori, T. Formation of Pseudo- and Hybrid-11S Globulins from Subunits of Soybean and Broad Bean 11S Globulins. *Agric. Biol. Chem.* **1980**, *44*, 1891–1896.

Utsumi, S.; Kim, C.-S.; Kohno, M.; Kito, M. Polymorphism and Expression of cDNAs Encoding Glycinin Subunits. *Agric. Biol. Chem.* **1987a**, *51*, 3267–3273.

Utsumi, S.; Kohno, M.; Mori, T.; Kito, M. An Alternate cDNA Encoding Glycinin $A_{1a}B_{5}$ Subunit. *J. Agric. Food Chem.* **1987b**, *35*, 210–214.

Utsumi, S.; Kim, C.-S.; Sato, T.; Kito, M. Signal Sequence of Preproglycinin Affects Production of the Expressed Protein in *Escherichia coli*. *Gene* **1988a**, *71*, 349–358.

Utsumi, S.; Sato, T.; Kim, C.-S.; Kito, M. Processing of Preproglycinin Expressed from cDNA-Encoding $A_{1a}B_{1b}$ Subunit in *Saccharomyces cerevisiae*. *FEBS Lett.* **1988b**, *233*, 273–276.

Watson, M. D.; Lambert, N.; Delauney, A.; Yarwood, J. N.; Croy, R. R. D.; Gatehouse, J. A.; Wright, D. J.; Boulter, D. Isolation and Expression of a Pea Vicillin cDNA in the Yeast *Saccharomyces cerevisiae*. *Biochem J.* **1988**, *251*, 857–864.

Yanisch-Perron, C.; Vieira, J.; Messing, J. Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mp18 and pUC19 Vectors. *Gene* **1985**, *33*, 103–119.

Yarwood, J. N.; Harris, N.; Delauney, A.; Croy, R. R. D.; Gatehouse, J. A.; Watson, M. D.; Boulter, D. Construction of a Hybrid cDNA Encoding a Major Legumin Precursor Polypeptide and Its Expression and Localization in *Saccharomyces cerevisiae*. *FEBS Lett.* **1987**, *222*, 175–180.

Zaret, K. S.; Sherman, F. DNA Sequence Required for Efficient Transcription Termination in Yeast. *Cell* **1982**, *28*, 563–573.