

## Crystal Structure of Soybean Proglycinin A1aB1b Homotrimer

Motoyasu Adachi, Yasuyuki Takenaka, Andrew B. Gidamis  
Bunzo Mikami and Shigeru Utsumi\*

*Research Institute for Food  
Science, Kyoto University, Uji  
Kyoto 611-0011, Japan*

Soybean glycinin is a member of the 11 S globulin family. The crystal structure of proglycinin was determined by X-ray crystallography at 2.8 Å resolution with an *R*-factor of 0.199 and a free *R*-factor of 0.250. A trimer molecule was found in an asymmetric unit of crystals. The trimer model contains three A1aB1b subunits and comprises 1128 amino acid residues and 34 water molecules. The constituent protomers of the homo-trimeric protein are arranged around a 3-fold symmetry axis with dimensions of 95 Å × 95 Å × 40 Å. The protomer model is composed of five fragments which correspond roughly to conserved regions based on the sequence alignment of various 11 S globulins. The core of the protomer consists of two jelly-roll β-barrels and two extended helix domains. This structure of proglycinin is similar to those of canavalin and phaseolin belonging to the 7 S globulin family, strongly supporting the hypothesis that both 7 S and 11 S globulins are derived from a common ancestor. The inter and intra-chain disulfide bonds conserved in the 11 S globulin family are clearly observed. It is found that the face with the inter-chain disulfide bond (IE face) contains more hydrophobic residues than that with the intra-chain disulfide bond. This suggests that a mature hexamer is formed by the interaction between the IE faces after processing.

© 2001 Academic Press

\*Corresponding author

**Keywords:** crystal structure; proglycinin; glycinin; legumin; seed storage protein

### Introduction

Soybean is one of the major dietary seeds and its proteins have better amino acid contents, in terms of nutrition, than other plant proteins. Glycinin, 11 S globulin, is one of the major storage proteins in soybean and accounts for about 40 % of the total seed proteins. Thus, glycinin plays an important role in the properties of food made from soybean. The glycinin extracted from seeds has a hexameric structure with a molecular mass of 300–380 kDa. Five major subunits of glycinin have been identified and classified into two groups according to

their amino acid sequences (group I, A1aB1b, A2B1a and A1bB2; group II, A3B4 and A5A4B3). Each subunit is composed of an acidic polypeptide with a molecular mass of about 32 kDa and a basic polypeptide with a molecular mass of about 20 kDa. The acidic and basic polypeptides are linked by a disulfide bond.<sup>1</sup> The sequence identity in each group is about 80 %, and that between groups is about 45 %.

In the cell, the constituent subunits are synthesized as a single polypeptide precursor, pre-proglycinin. The signal sequence is removed co-translationally in the endoplasmic reticulum (ER) and the resultant proglycinin assembles into trimers of about 8 S in size. As the proglycinin is sorted to protein storage vacuoles (PSV), a specific post-translational cleavage occurs between asparagine and glycine, resulting in a mature subunit consisting of the acidic and basic polypeptides. Finally, glycinin assembles into a hexamer of about 11 S. It has been demonstrated that the cleavage is a trigger for the formation of hexamers.<sup>2</sup> However,

Present address: A. B. Gidamis, Department of Food Science and Technology, Faculty of Agriculture, Sokoine University of Agriculture, PO Box 3006, Morogoro, Tanzania.

Abbreviations used: ER, endoplasmic reticulum; PSV, protein storage vacuoles; rmsd, root-mean-square deviation; NCS, non-crystallographic symmetry.

E-mail address of the corresponding author: [utsumi@food2.food.kyoto-u.ac.jp](mailto:utsumi@food2.food.kyoto-u.ac.jp)

it is not known how the trimers stack to form hexamers. Moreover, the signal for sorting from the ER to the PSV and the mechanism for the highly packed accumulation remain unclear. For understanding the cellular mechanisms, it is very important to elucidate the protein structures of pro and mature glycinins.

Many seeds, including soybean, contain another type of storage protein, the 7 S globulins. On the basis of amino acid sequence alignment, it is suggested that the two families are evolutionarily related.<sup>3–7</sup> Until now, the crystal structures of only two 7S globulins, canavalin from jack bean<sup>8,9</sup> and phaseolin from kidney bean,<sup>7,10</sup> have been reported. Their subunits are composed of two similar domains containing a  $\beta$ -barrel followed by a region of helices. Lawrence *et al.*<sup>7</sup> described the detailed analysis of structural elements of phaseolin, and predicted that 11 S globulins also contain two copies of the conserved Pro/Gly motif, as do 7 S globulins. In addition, from comparison of sequences between the N and C-terminal domains, the known three-dimensional structures of 7 S globulins, and limited proteolysis analysis, Shutov *et al.*<sup>11,12</sup> proposed a mechanism for 11 S and 7 S gene evolution in which triplication of a single-domain ancestor occurred. In a recent report, they suggest gene duplication mechanism instead.<sup>13</sup> To understand the evolutionary relationship between the two protein families, it is necessary to compare their structures. However, there has been no report on the high-resolution crystal structure for an 11 S globulin. Because of the heterogeneity of molecular species, it is difficult to crystallize mature glycinin prepared from seeds.<sup>14</sup> Therefore, we previously expressed cDNA coding for the proglycinin A1aB1b subunit in *Escherichia coli* to obtain a homogeneous proglycinin homotrimer<sup>15</sup> and succeeded in its crystallization.<sup>16,17</sup>

Here, we have determined the crystal structure of the proglycinin A1aB1b homotrimer and revealed the core structure of the proglycinin protomer. The core structure consists of two jelly-roll barrels and two extended helix domains in analogy with 7 S globulins.

## Results and Discussion

### Quality of the final model

The final statistics for the refinement and data collection are shown in Table 1. The proglycinin model was refined at 2.8 Å resolution with an *R*-factor of 19.9% and a free *R*-factor of 25.0% (PDB:1FXZ). The final  $2|F_o| - |F_c|$  map contoured at the  $1\sigma$  level provided good traces, with the exception of the six disordered regions described below. The final model includes 1128 amino acid residues and 34 water molecules. From the Luzzati<sup>18</sup> plot, the mean absolute positional error was estimated to be 0.3 Å. A Ramachandran plot (Figure 1) of the main-chain conformation angles showed that about 84.6% lie within the core

**Table 1.** Statistics of data collection and refinement

|                                                       |                              |
|-------------------------------------------------------|------------------------------|
| Space group                                           | $P4_1$                       |
| Cell dimensions<br><i>a</i> , <i>b</i> , <i>c</i> (Å) | 115.524, 115.524,<br>146.998 |
| Data collection                                       |                              |
| No. of crystals used                                  | 1                            |
| Resolution range (Å)                                  | $\infty$ –2.8                |
| No. of measured reflections                           | 142,631                      |
| No. of unique reflections                             | 46,010                       |
| <i>R</i> <sub>sym</sub> (%)                           | 8.9                          |
| Completeness (%)                                      | 99.5                         |
| Refinement                                            |                              |
| Resolution range (Å)                                  | 8.0–2.8                      |
| No. of reflections used ( $F > 2\sigma F$ )           | 35,286 [3162] <sup>a</sup>   |
| Completeness (%)                                      | 77.6 [55.6]                  |
| No. of residues                                       | 1128                         |
| No. of water molecules                                | 34                           |
| Average <i>B</i> -factor (Å <sup>2</sup> )            |                              |
| Protein atoms                                         | 22.5                         |
| Water molecules                                       | 23.9                         |
| rmsd                                                  |                              |
| Bond length (Å)                                       | 0.013                        |
| Bond angle (deg.)                                     | 3.27                         |
| <i>R</i> -factor                                      | 0.199 [0.271]                |
| Free <i>R</i> -factor                                 | 0.250 [0.341]                |

<sup>a</sup> Statistics for outer shells from 2.92 to 2.80 Å and  $F > 2\sigma F$ .

region, with 99.7% lying within the allowed region.<sup>19,20</sup> Three Phe463 residues in the model were found in a disallowed region ( $\phi = 57$  to  $59^\circ$ ,  $\psi = -54$  to  $-65^\circ$ ). The Phe463 residues have well-defined density and are located in a sharp bend between the barrel and extended helix domains. The residue Phe463 corresponds to Tyr376 in phaseolin<sup>7</sup> and Tyr417 in canavalin,<sup>9</sup> which in turn have a structure similar to that of proglycinin as will be discussed. The Tyr376 residue in phaseolin and Tyr417 in canavalin have  $\phi, \psi$  angles similar to those of Phe463 in proglycinin. The average *B*-factors were calculated to be 22.5 Å<sup>2</sup> and 23.9 Å<sup>2</sup> for protein atoms and water molecules, respectively. Figure 2 shows the plot of *B*-factors averaged among three protomers along the amino acid sequence. The average *B*-factors were relatively low in continuous strand regions and slightly high in tandem helix regions.

### Structure of proglycinin

The crystal structure indicates that the protein is a trimer composed of three protomers in each asymmetric unit, and the trimeric protein arranges the constituent protomers around a 3-fold symmetry axis with dimensions of 95 Å × 95 Å × 40 Å (Figure 3). The dimensions are consistent with those previously reported from preliminary crystallographic analysis of the proglycinin A1aB1b homotrimer<sup>17</sup> and electron microscopic studies of various 11 S globulins,<sup>21,22,23</sup> but smaller than those of hemp 11 S globulin,<sup>24</sup> and very similar to those of 7 S globulins.<sup>7,8,10</sup> Superimposition of protomers (A, B and C) gave a root-mean-square deviation (rmsd) of 0.30 Å, 0.39 Å and 0.34 Å for the C $\alpha$  atoms between A and B, B and C, and C and A, respectively. The differences in the deviations

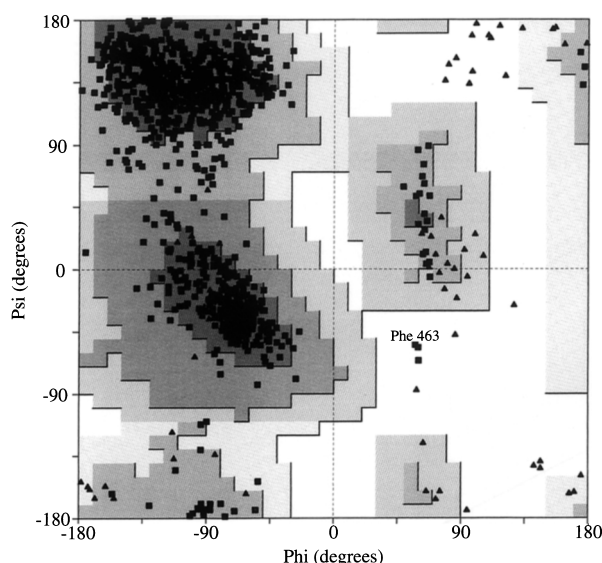


**Figure 1.** Ramachandran plot for a proglycinin trimer. Glycine and other residues are indicated by filled triangles and filled squares, respectively. The Figure was drawn with the program PROCHECK.<sup>20</sup>

among the subunits are mainly caused by differences in the structure formed by residues from Gln308 to Ser310, because the Gln308 in protomer C contacts with a symmetric molecule in the crystals. The removal of the three residues resulted in decrease of deviations between B and C, and between C and A by 0.08 and 0.09 Å, respectively, while the deviation did not change between A and B. The average *B*-factors of protomers A, B and C were 21.1 Å<sup>2</sup>, 24.2 Å<sup>2</sup> and 21.8 Å<sup>2</sup>, suggesting that the structures of protomers A and C could be better fixed than that of protomer B in crystals.

The protomer structure contains 25 strands and five helices, and is folded into two β-barrel (jelly-roll) domains and two extended helix domains containing two helices. These domains are superimposed by a pseudo-dyad symmetry. Here, we divide the protomer into N and C-terminal mod-

ules according to the pseudo-dyad symmetry axis, and elements of the secondary structure in both terminal modules are numbered sequentially, as shown in Figure 4. N and C-terminal modules are composed of 204 and 172 amino acid residues. Figure 5 shows a comparison between the structures of the two terminal modules. A least-squares fit of 132 C<sup>α</sup> atoms in a distance of 2.0 Å produced an rmsd of 0.97 Å. Twenty-nine residues of the superimposed 132 residues were identical between the two terminal modules. Although the structures of the N and C-terminal modules in the core of proglycinin are very similar to each other, the C-terminal module of the protomer does not contain any helices in the barrel domain and contains one strand less than the N-terminal module. There is one more significant difference in the size of the loop between strand I and J (IJ loop). In the N-terminal module, eight residues from Asn150 to Asp157 are inserted, and make a larger loop compared with that of the C-terminal module, which is located near the center of the trimer.

### Disordered regions

We observed ten termini and constructed the model containing five visible regions on the basis of electron density maps. The electron density maps of the ten terminal residues of the five visible regions, with the exception of the side-chain of Glu227, were clearly observed, and the averaged *B*-factors at both termini of the five visible regions are relatively higher (Figure 2). Since the same molecular size was indicated by SDS-PAGE for both the freshly purified protein and the protein in the crystals (data is now shown), the model of proglycinin A1aB1b was thought to have six disordered regions (Figure 6). The four internal disordered regions are concentrated around the N-terminal module (Table 2). In the crystals, since the trimer interacts with symmetric molecules, the space of disordered regions 2 and 4 can be limited, while there are sufficient open spaces around disordered regions 1, 3, 3' and 5 on all asymmetric

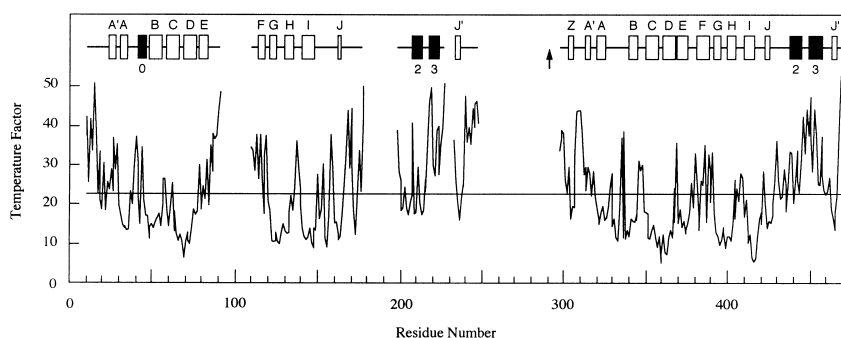


**Figure 2.** Average isotropic temperature factor as a function of residue number in the final model. Each point is averaged over three protomers. Secondary structures are indicated at the top. Helix and strand elements are shown by filled and open boxes, respectively. The middle line indicates the average *B*-factor (22.5 Å<sup>2</sup>). The arrow indicates the processing site between Asn291 and Gly292.

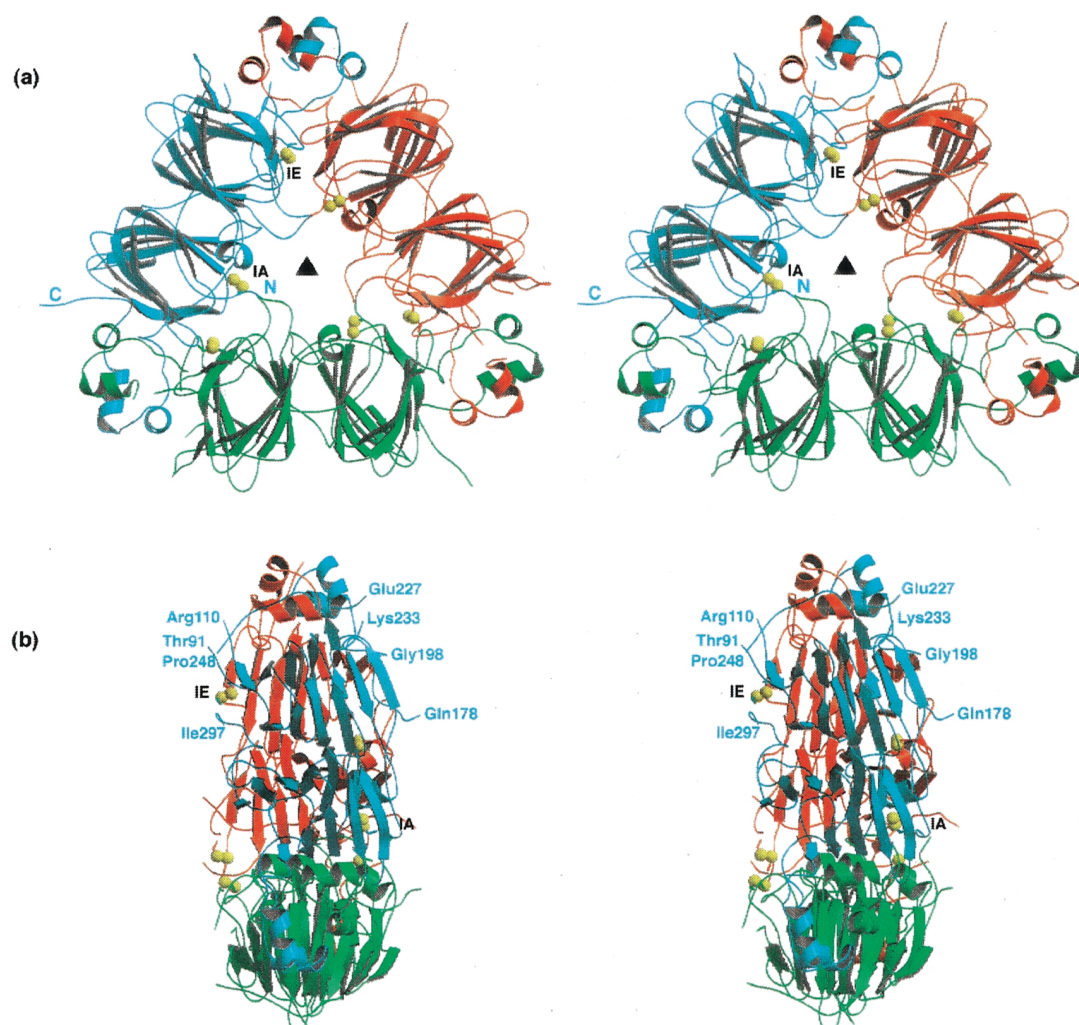


**Figure 3.** Stereo views of the ribbon diagram. Three protomers in the proglycinin trimer are shown by red, cyan and green, respectively. The disulfide bonds in the trimer molecule are represented by yellow spheres with a radius of 1.2 Å. IA and IE indicate the disulfide bonds between Cys12 and Cys45, and between Cys88 and Cys298, respectively. (a) The 3-fold symmetry axis runs perpendicular to the paper and is shown by a filled triangle. N and C indicate the N and C termini, respectively. (b) Viewed after 180° rotation around the vertical axis. The residues just after and before the internal disordered regions are labeled by blue letters. The model was generated using the programs MOLSCRIPT<sup>45</sup> and Raster3D.<sup>46</sup>

protomers (Table 2). All protomers in the trimer are packed in between the sides of the two  $\beta$ -barrels. The biggest disordered region, 4, is comprised of 48 residues. It is likely that there is not enough space for the disordered region 4 to exist on the side of the trimer due to the crystal packing. This indicates that disordered region 4 may be located on the face perpendicular to the 3-fold axis. Besides, it is possible that disordered region 2 could interact with disordered region 4, since it exists near Pro248 (Figure 3), or disordered region 2 might be extended over the helix domain of the next protomer in the trimer.

On the other hand, protein sequences have previously been aligned among various 11 S globulins from legumes and non-legumes, and five variable regions have been proposed.<sup>4,5</sup> Here, we examined how these six disordered regions correspond to the

five variable regions. Five of the six disordered regions corresponded roughly to the variable regions (Figure 6), although there were three significant differences between the disordered and the variable regions. First, there is no variable region corresponding to disordered region 3'. The disordered region 3' is composed of only five residues and is rich in acidic residues. In glycinin group II subunits, two residues are deleted in this region compared with those of group I. This region could be primarily variable. Secondly, disordered regions 3 and 5 are much smaller than the corresponding variable regions III and V, respectively. The structure of 11 residues at the variable region III was visible by crystallography. Further, we were able to construct a model for 23 residues in variable region V. This region contains a helix and is responsible for interaction between subunits, as

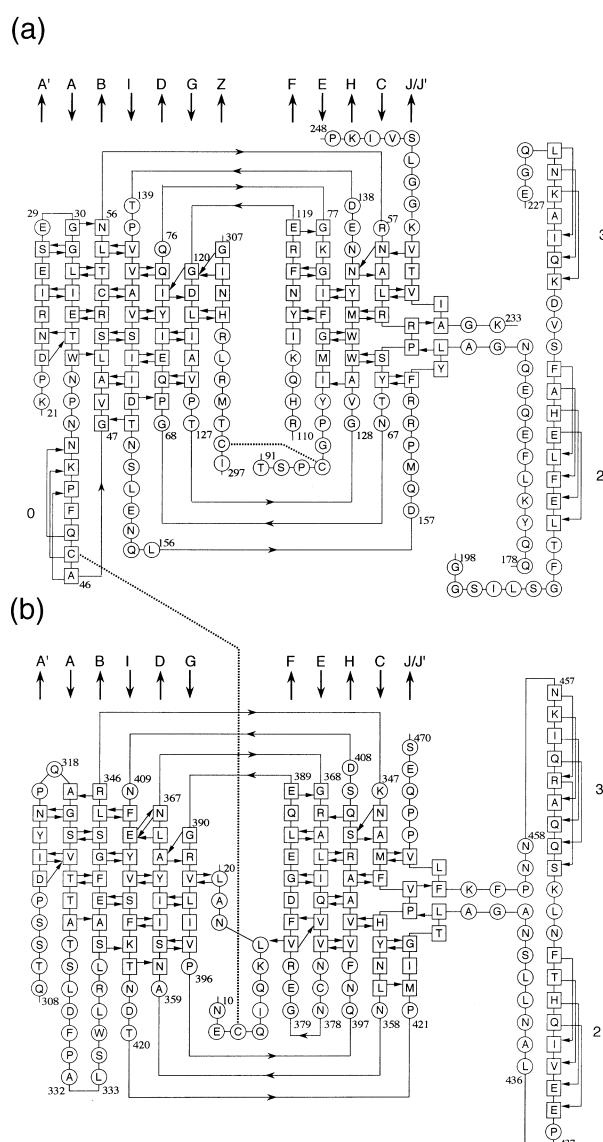


**Figure 4.** Schematic diagram of the hydrogen bonding pattern found in the proglycinin protomer. (a) N-terminal module; (b) C-terminal module. Boxes indicate residues comprising secondary structural elements, while circles indicate those in interconnection and terminal regions. Hydrogen bonds between main chains are represented by means of directed arrows with heads corresponding to the carbonyl oxygen atom. Helix and strand elements are labeled as defined previously.<sup>7,9</sup> The arrows at the top show the direction of the antiparallel strands. Broken lines represent disulfide bridges.

described below. Finally, disordered region 4 is six residues longer at its C terminus than variable region IV. The conservation in the six residues is high among 11 S globulins (Figure 7). This conserved sequence contains the processing site Asn291-Gly292 for the maturation. The processing site would protrude into the solvent to be susceptible to the attack of asparaginyl endopeptidase,

although we could not observe the electron density map of the site. Jung *et al.*<sup>25</sup> suggested that the three conserved residues after site Asn291-Gly292 are not necessary for the processing. The role of the three conserved residues remains unclear. The A5A4B3 subunit is processed between Asn96 and Arg97 in addition to the processing site, corresponding to site Asn291-Gly292 in A1aB1b subunit. As a result, an acidic chain of the A5A4B3 subunit is separated into A5 and A4 chains. The site Asn96-Arg97 is located near the center of disordered region 2. Any other asparagine residues are not found in the disordered regions far from the visible regions of the five glycinin subunits.

### Disulfide bonds

Two disulfide bonds are highly conserved among 11 S globulins. Here, we were able to clearly confirm the two disulfide bonds in an electron density map of the proglycinin protomer. One is an intra-chain bond formed between Cys12 and Cys45, and the other is an inter-chain between Cys88 and Cys298 (between the acidic and basic chains) (Figure 8). The dihedral angles of the intra and inter-chain bonds are  $102^\circ$  (right-handed) and  $-93^\circ$  (left-handed), respectively, and the distances between C $\alpha$  atoms in the two cysteine residues are 4.6 Å and 5.9 Å, respectively. The two bonds are located near the two symmetry-related terminal modules. It is found that the intra-chain bond is positioned in a slightly buried region compared with the inter-chain bond. Since three cysteine residues forming disulfide bonds, i.e. all those except Cys45, exist just after or before disordered regions, we cannot rule out the possibility that these disordered regions completely cover the two disulfide bonds. In addition, the intra-chain bond is located nearer the 3-fold axis in the trimer than the inter-chain bond and is located on the interface between protomers. These disulfide bonds could contribute to structural stability and folding of the protein. However, Utsumi *et al.*<sup>26</sup> disrupted inter and intra-chain disulfide bridges of glycinin using site-directed mutagenesis, and showed that these mutants are able to assemble into a trimer and form crystals.<sup>27</sup> Jung *et al.*<sup>25</sup> also constructed a mutant of legumin lacking an inter-chain disulfide bond, and demonstrated that the mutant can assemble into a trimer.

### Interactions between acidic and basic chains

Proglycinin is divided into two chains after processing by an endopeptidase. They are linked by an inter-chain disulfide bond in mature form, as described above. In general, they are called acidic and basic chains, respectively. For example, the putative isoelectric points of N and C-terminal chains are 5.24 and 8.22, respectively, in the case of the A1aB1b subunit (proglycinin; 5.72).

Table 3 lists hydrogen bonds between the acidic and basic chains in the intra-protomer. A hydro-

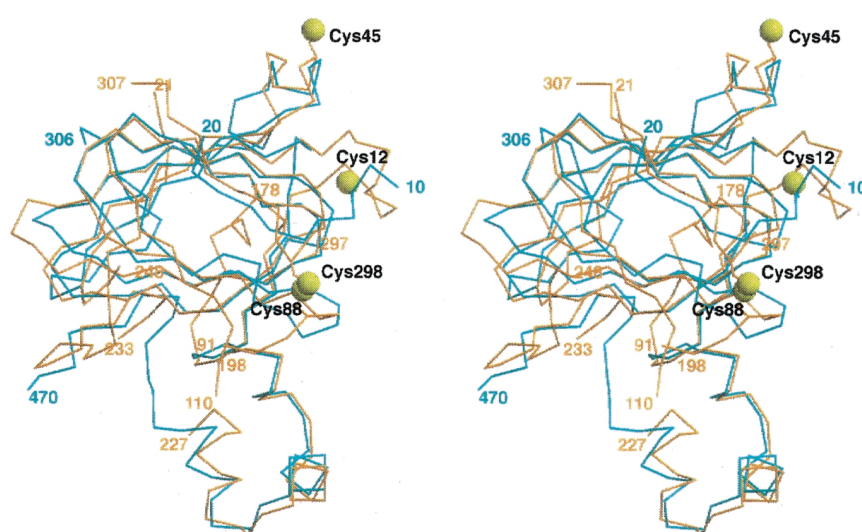


**Figure 5.** Comparison of structures between the N and C-terminal modules of proglycinin. Orange and cyan lines indicate C $\alpha$  traces of the N and C-terminal modules, respectively. Sulfur atoms forming disulfide bonds are represented by a yellow sphere with a radius of 1.2 Å.

gen-bonded salt bridge was found between Asp121 and Arg303 on the molecular surface. Thirteen residues of 23 contributing hydrogen bonds between acidic and basic chains are conserved among the five glycinin subunits. Most of the hydrogen bonds are found around the segments that link the  $\beta$ -barrels (from Asn10 to Leu20 in C-terminal module and from Ile297 to Gly307 in N-terminal module). The segments should play an important role in the binding of acidic and basic chains. Five other hydrogen bonds were found outside of the segments (Table 3). Although three bonds are formed near the solvent, two strong hydrogen bonds were observed between Glu71 and Ser339 in a completely buried region. When the carboxyl group of Glu71 is ionized, the residue has no partner for the salt bridge. If the proglycinin is destabilized by the buried ionizable group of Glu71, the desolvation energy may be compensated by the two hydrogen bonds between Glu71 and Ser339.<sup>28</sup>

To account for the contribution of hydrophobic interaction between the acidic and basic chains in the intra-protomer, we calculated the accessible

surface area. Binding between the two chains results in the decrease of approximately 1800 Å<sup>2</sup>, each in area of the surface of the acidic and basic regions, which corresponds to 15% and 18% of the whole surfaces of the acidic and basic regions, respectively. The hydrophobic residues occupy 65% of the decreased area and are surrounded by hydrophilic residues in the buried surface. These results suggest that hydrophobic interaction plays an important role in the interaction between the acidic and basic chains. Seven residues of Leu17, Tyr73, Leu122, Leu302, Leu336, Ile394 and Phe415 contribute to the decrease in the surface area of 750 Å<sup>2</sup>, and occupy 32% of the hydrophobic area. Among these residues, Leu17, Leu302 and Phe415 are strictly conserved in glycinin subunits, and other residues are conserved in group I and conservatively mutated in group II.

### Interactions between protomers

As mentioned above, a protomer has two helix domains extended from the barrels. The helix domains of a protomer stack on those of the adja-

**Table 2.** Putative location of disordered regions

| Region | Residue numbers                   | Modules | Location             | Steric constraints <sup>a</sup> |
|--------|-----------------------------------|---------|----------------------|---------------------------------|
| 1      | 1-9 <sup>b</sup> (9) <sup>c</sup> | C       | IA face <sup>d</sup> | No                              |
| 2      | 92-109(18)                        | N       | IE face              | Yes                             |
| 3      | 179-197(19)                       | N       | IA face              | No                              |
| 3'     | 228-232(5)                        | N       | IA face              | No                              |
| 4      | 249-296(48)                       | N       | IE face              | Yes                             |
| 5      | 471-476(6)                        | C       | Side                 | No                              |

<sup>a</sup> Steric constraints owing to crystal packing.

<sup>b</sup> Three lacking residues are included.

<sup>c</sup> The number of residues in disordered regions are indicated in parentheses.

<sup>d</sup> IA and IE faces are defined in section Disulfide bonds.

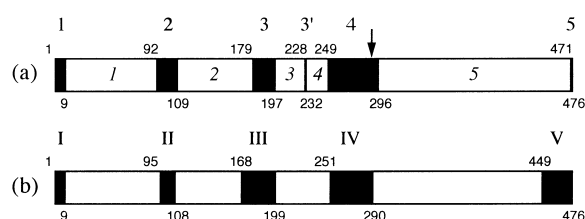


**Figure 6.** Comparison of variable regions with disordered regions. (a) Disordered regions obtained by crystallography. Open and filled boxes represent modeled and disordered parts, respectively. Numerals 1-5 and italicized numerals 1-5 indicate modeled and disordered parts, respectively. The disordered region 1 includes the lacking three residues.<sup>15</sup> The arrow indicates the processing site. (b) Variable regions based on sequence alignment. Open and closed boxes represent conserved and variable regions, respectively. Roman numerals I to V indicate variable regions.

cent protomers. The protomers interact with the interfaces between helix domains, between helix and barrel domains, and between barrel domains. A part of the  $\beta$ -barrel including strands C, E, H and J is responsible for these interactions in addition to the helix domains. Table 4 lists common hydrogen bonds less than 3.5 Å in all of the three interfaces. About half of the hydrogen bonds are found on the interface between barrel and barrel domains. While almost all of the hydrogen bonds are positioned near the solvent in the trimer, two bonds, those between Tyr65 and Asn398 and between Gly128 and Tyr355, are buried and located at the center of the whole interfaces. A hydrogen-bonded salt bridge is formed between Arg110 and Glu438 in the interaction area of the helix and barrel domains, but Arg110 is not con-

served in group II, indicating that the salt-bridge does not play an important role in the formation of the trimer.

To estimate hydrophobic interactions, we calculated accessible surface area. One protomer shares an interface of about 2700 Å<sup>2</sup> with two adjacent protomers in the trimer, which occupies about 30% of the whole surface of one protomer. The hydrophobic residues comprise 48% of the total residues in the interface and were distributed in positions far from the solvent. These hydrophobic residues were evenly distributed in all interfaces between helix domains, between barrel domains, and between helix and barrel domains. Forty-one important hydrophobic residues on the interfaces are shown in Figure 7, and occupy 87% of the hydrophobic area. Thirty-eight residues of the 41 (all those except for Met159, Ala213 and Ile220) are conserved or conservatively mutated among glycinin subunits.

In addition to these interactions, an off-set stacking interaction by three aromatic residues of Tyr65, Phe163 and Tyr355 was observed. Tyr65 and Phe163 are strictly conserved in five glycinin subunits, and Tyr355 is conserved as tyrosine or tryptophan residues. Phe163 is strictly conserved, and Tyr65 and Tyr355 are conserved as tyrosine/phenylalanine and tyrosine/tryptophan, respectively, in all 11 S globulins.<sup>7</sup> Tyr65 is located in the center of this stack and interacts with Tyr355 derived from the adjacent protomer. In addition, OH groups of Tyr65 and Tyr355 form buried hydrogen bonds between subunits. These stacking residues are not found in canavalin and phaseolin. It seems that the stacking interaction plays an important role in binding between monomers in 11 S globulins.

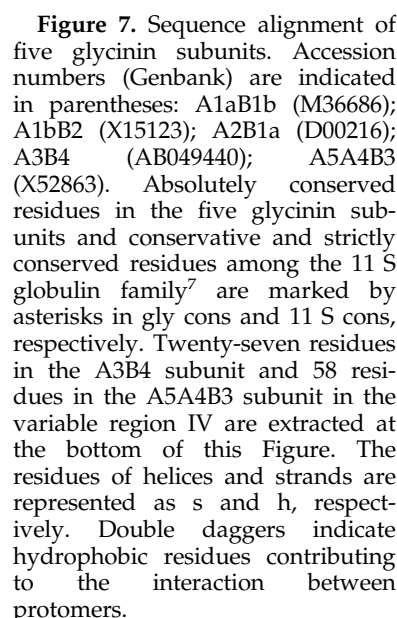
**Table 3.** Hydrogen bonds between acidic and basic chains

| Residue              | Atom | Residue              | Atom | Distance (Å)       |
|----------------------|------|----------------------|------|--------------------|
| 17 Leu <sup>a</sup>  | N    | 382 Val <sup>*</sup> | O    | 2.9                |
| 20 Leu <sup>*</sup>  | N    | 392 Val              | O    | 2.8                |
| 20 Leu <sup>*</sup>  | O    | 392 Val              | N    | 2.8                |
| 46 Ala <sup>*</sup>  | O    | 360 Asn <sup>*</sup> | ND2  | 3.0 † <sup>b</sup> |
| 52 Arg               | NH1  | 364 Tyr <sup>*</sup> | OH   | 2.8 †              |
| 71 Glu               | OE1  | 339 Ser <sup>*</sup> | O    | 2.7 †              |
| 71 Glu               | OE2  | 339 Ser <sup>*</sup> | O    | 2.9 †              |
| 75 Gln               | OE1  | 364 Tyr <sup>*</sup> | OH   | 3.0 †              |
| 116 Asn              | O    | 303 Arg              | NE   | 3.4                |
| 119 Glu <sup>*</sup> | O    | 305 Asn <sup>*</sup> | ND2  | 3.0                |
| 120 Gly <sup>*</sup> | O    | 306 Ile <sup>*</sup> | N    | 2.9                |
| 120 Gly <sup>*</sup> | O    | 307 Gly              | N    | 3.2                |
| 121 Asp <sup>*</sup> | OD1  | 303 Arg              | NH2  | 2.9                |
| 121 Asp <sup>*</sup> | OD1  | 305 Asn <sup>*</sup> | ND2  | 3.1                |
| 121 Asp <sup>*</sup> | OD2  | 303 Arg              | NE   | 2.8                |
| 121 Asp <sup>*</sup> | OD2  | 303 Arg              | NH2  | 3.2                |
| 122 Leu              | N    | 304 His              | O    | 2.8                |
| 122 Leu              | O    | 304 His              | N    | 3.0                |
| 124 Ala              | N    | 302 Leu <sup>*</sup> | O    | 3.5                |

Hydrogen bonds observed in protomer A of proglycinin are shown.

<sup>a</sup> Asterisk indicates residues conserved among five glycinin subunits.

<sup>b</sup> Dagger indicates bond observed in the region except for segments that link the  $\beta$ -barrels.



To discuss glycinin maturation, we term the two faces perpendicular to the 3-fold axis in a proglycinin trimer the IE and IA faces. The IE and IA faces contain an inter-chain and an intra-chain disulfide bond, respectively, which correspond to the left and right faces in Figure 3(b), respectively. To form a hexamer after processing, two trimers must interact with each other between the same faces, because interaction between the different faces should result in oligomeric rod-like assemblies larger than a hexamer. The dimensions of hexamers observed from primary crystallographic analysis<sup>24</sup> and electron microscopic studies<sup>21-23</sup> are consistent with the hypothesis that a hexamer is formed by interactions between the same faces of two trimers. Here, the question is raised of which face is responsible for the interaction, the IA or IE face. Comparison of the two faces indicates that the IA face has a larger curvature than the IE face. The

number of hydrophobic residues that are exposed on the IE face is larger than that on the IA face, while there is no remarkable difference in the distribution of electrostatic potentials between the two faces (Figure 9). These facts suggest that the hexamer is formed by the interaction between the IE faces. Another question is that of what interferes with the hexamer formation of proglycinin. Dickinson *et al.*<sup>2</sup> reported that processing is required for the hexamer formation. The processing site is located on the IE face, just beyond the disordered region 4. Since Pro248, the C-terminal residue of the conserved region 4, also exists on the IE face, the disordered region 4 including 48 amino acid residues must be located on the IE face, as mentioned above. Therefore, it is possible that the disordered region 4 interferes with hexamer formation by steric hindrance. The disordered region 4 must be able to move from the IE face to the side of the trimer by the processing.

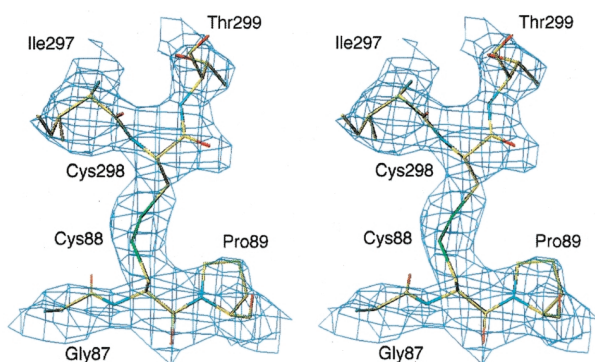


**Figure 8.** Stereo view of the schematic drawing around the inter-chain disulfide bond between Cys88 and Cys298. Representation of the protein model in a  $2|F_o| - |F_c|$  electronic density map contoured at  $1.2\sigma$  generated using the program TURBO-FRODO (Bio-Graph). The map shown in blue was superimposed on the refined 2.8 Å resolution coordinates of proglycinin. Carbon, oxygen, nitrogen and sulfur atoms are shown by yellow, red, cyan and green, respectively.

### Comparison with the structure of 7S globulins

It is found that the overall structure of proglycinin is very close to those of canavalin<sup>8,9</sup> and phaseolin.<sup>7,10</sup> Lawrence *et al.*<sup>7</sup> addressed the structural relationship between 11 S and 7 S globulins by 11 S/7 S alignment based on the canonical 7 S element, and pointed out that Pro/Gly motifs in internal structural repeats (Pro67/Gly81 and Pro254/Gly269 in phaseolin) are the most conserved key residues. Our analysis showed that the corresponding residues of Pro63 and Pro353 in proglycinin A1aB1b are positioned on strand C in both terminal modules and form a  $\beta$ -bulge structure, and Gly77 and Gly368 residues are positioned in a loop between the strands D and E in analogy with those of 7 S globulins. This indicates that the Pro/Gly motif is certainly conserved between the 11 S and 7 S globulins as predicted.<sup>7</sup> Thus, we compared the backbone structure of the proglycinin protomer with those of phaseolin and canavalin based on the Pro/Gly motif. A least-squares fit produced an rmsd of 1.35 Å between proglycinin and phaseolin using 179 C $\alpha$  atoms, and 191 C $\alpha$  atoms were superimposed with an rmsd of 1.30 Å between proglycinin and canavalin. These results indicate that the backbone structure of glycinin is similar to those of canavalin and phaseolin, and is more similar to that of canavalin than that of phaseolin. In addition, the results strongly support the hypothesis that both 7 S and 11 S globulins are derived from a common ancestor. Figure 10 shows the superimposition of the C $\alpha$  models of the proglycinin with those of 7 S globulins. The two barrel domains in the proglycinin are well superimposed on those of 7 S globulins, whereas the position of helix domain appears to be slightly different. How-

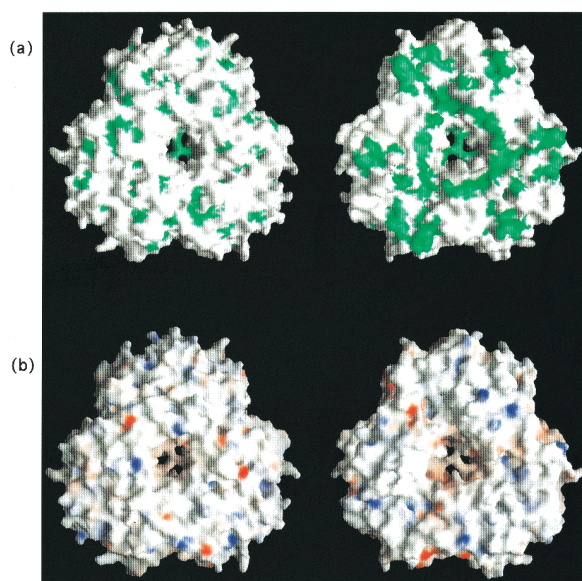


**Figure 9.** Comparison of IA and IE faces. (a) Diagrams of the hydrophobic surface of the proglycinin trimer. Alanine, valine, leucine, isoleucine, methionine, proline, phenylalanine, tyrosine and tryptophan are colored green. The left and right diagrams show the IA and IE faces, respectively. (b) Electrostatic surface potential calculated using the program GRASP.<sup>47</sup> Positive potential is colored blue and negative potential is colored red (values range from -33 (red) to 32 (blue) in units of  $k_B T$ , where  $k_B$  is the Boltzmann constant and  $T$  is temperature). Orientation is the same as in (a).

ever, the overall structure of the helix domains of the proglycinin seems to be similar to those of canavalin and phaseolin, although the extended helix domains contain one less helix than canavalin and phaseolin. When the N and C-terminal modules of the proglycinin were compared separately with those of 7 S globulins, the rmsd values decreased to 1.15 and 1.0 Å, respectively. These results indicate that the two terminal modules of the 11 S and 7 S globulin protomers are similar to each other, but that their configurations are slightly different.

Based on the three-dimensional structures, we aligned protein sequences of canavalin, phaseolin, and proglycinin, as shown in Figure 11. Here, we found 46 identical residues among the three proteins, and 27 identical residues in C-terminal module, suggesting that the C-terminal module is more conservative than N-terminal module between 11 S and 7 S globulins. The alignment between phaseolin and proglycinin is basically similar to that described.<sup>7</sup> There is, however, a deletion of three residues between helix 0 and strand B, while they introduced a gap in strand A of the N-terminal module. This difference is probably due to the fact that proglycinin has a helix structure in this region. Additionally, a deletion of one residue is present between strands H and I in the N-terminal module

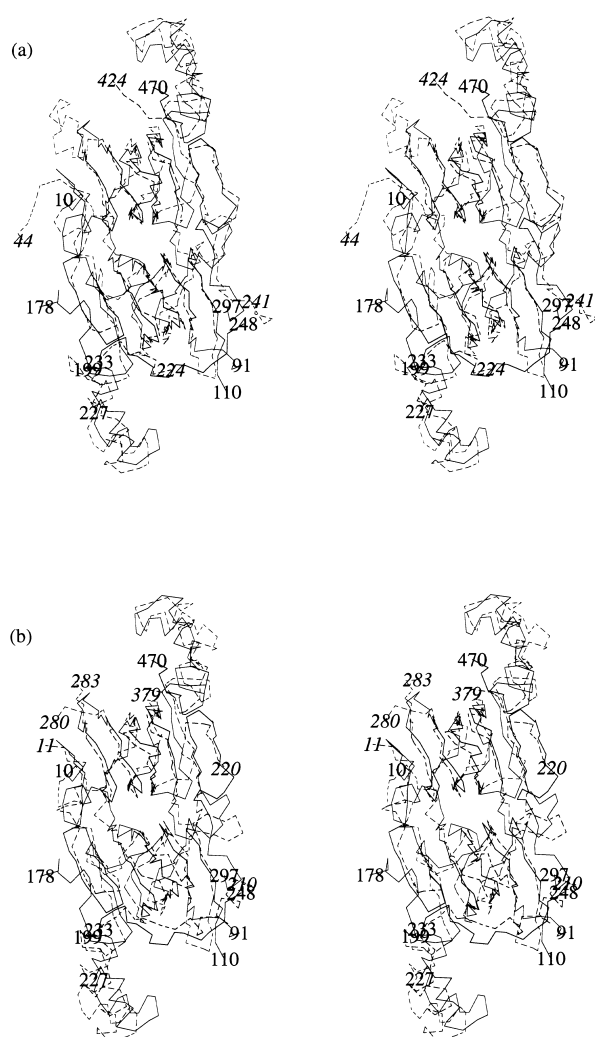


**Figure 10.** Comparison of the structure of proglycinin with those of canavalin (PDB; 1CAU) and phaseolin (PDB; 2PHL). (a) C $\alpha$  traces showing the superimposition of proglycinin with canavalin. Thin and broken lines indicate proglycinin and canavalin, respectively. (b) C $\alpha$  traces showing the superimposition of proglycinin with phaseolin. Thin and broken lines indicate proglycinin and phaseolin, respectively.

of proglycinin compared to their alignment. Since an insertion of five residues exists just after strand I, it may have been difficult to find the deletion. In their 7 S/11 S alignment, 29 globally conserved residues were shown. Only one residue between strands H and I is inconsistent with our result. This is probably due to the deletion between strands H and I. These results indicate that the 28 globally conserved residues could be applied to other related proteins including the structurally conserved module.<sup>29</sup>

In the case of canavalin and phaseolin, the reduced accessible surface area following the formation of a trimeric structure was about 6000 Å<sup>2</sup>, of which hydrophobic residues comprised 48 and

|         | Z                    | A'                            | A                     | 0   |   |
|---------|----------------------|-------------------------------|-----------------------|-----|---|
| Gly 1   | FSSREQPQNECQIQKLNALK | PDNR IESEGGLIETWNP            | NKPFQCA               | 46  |   |
| Can 41  | ESRAQNNPYLFRSNK      | FLTLFKNQHGSLRLLRQFNEDTEKLE    | 81                    |     |   |
| Pha 1   | TSLREEEESQDNPFYFNSD  | NSWNTLFKNQYGHIRVLQRFDDQSKRLQ  | 47                    |     |   |
|         | B                    | C                             | D                     | E   |   |
| Gly 47  | GVALSRCTLNRRALRR     | PSYTNCPQBIYIQGKG              | IFGMIYFGCPST          | 91  |   |
| Can 82  | NLRDYRVLEYCSKPNTLL   | PHHSDSDLLVLVLEGG              | QAILVLVNP             | 125 |   |
| Pha 48  | NLEDYRLVEFRSKPETLL   | PQQADAEILLVVRSGS              | AILVLVKPDD            | 93  |   |
|         | F                    | G                             | H                     |     |   |
| Gly 92  | FEEPQQPQQRQSSRPQDR   | RHOKIYNFRE                    | GDLIAVPTGVAMWY        | 134 |   |
| Can 126 | GRDVTYKLDQ           | GDIAIKIAGTFFYLI               | 149                   |     |   |
| Pha 94  | RREYFFLTSDNPIFS      | DHOKIPAGTIFVLV                | 121                   |     |   |
|         | I                    | J                             |                       |     |   |
| Gly 135 | NNE DTPVVAVSIIDTNS   | LENQLDQMPRRFYLAGNQEQEFLKYQQBQ | 181                   |     |   |
| Can 150 | NPNNQNLRLKFAIT       | FRRPGTVEDFLSSTRKL             | 183                   |     |   |
| Pha 122 | NPDPKEDLRITQLAMP     | VNNPQIHEFFLSSTEAQ             | 154                   |     |   |
|         | 1                    | 2                             | 3                     |     |   |
| Gly 182 | GHQSQKQKQEEENEGGS    | ILSGFTLEFLEHAFSVDKQIAKNLQGENE | 229                   |     |   |
| Can 184 | PSYLSAFSKNFLEASVD    | SPYDEIEQTLLQE                 | 214                   |     |   |
| Pha 155 | QSYLQEFSSKHILEASFN   | SKEFEINRVLFEED                | 185                   |     |   |
|         | J'                   | 4                             |                       |     |   |
| Gly 230 | GEDKGATVTVKGGLSVIK   | PTDEQQQRQPEEEEEDEKPCQCKGDKH   | 277                   |     |   |
| Can 214 | EQEGVTVKMPKDKIQE     | ISKHAQSSSRKTL                 | 240                   |     |   |
| Pha 186 | GQEGVTVNIDSEQIKEL    | SKHAKSSSRKSLSKQD              | 216                   |     |   |
|         | Z                    | A'                            | A                     |     |   |
| Gly 278 | CQRPRGSQSKSRNGIDET   | ICTMRLRHNIQTSSPD              | IYNPQAGSVTTA          | 325 |   |
| Can 241 | SSQDKPPNLR           | SRDPTYSNN YGKLYE              | 268                   |     |   |
| Pha 217 | NTIGNE               | FGNLTER                       | 232                   |     |   |
|         | 0                    | B                             | C                     | D   | E |
| Gly 326 | TSLDFPALSWLRLSAEFG   | SLRKNAMFVPHYNLANS             | IYALNGRALIQ           | 373 |   |
| Can 269 | TPEKNSQLRDLILLNCL    | QMEGALFVPHYNSTRAT             | IVLVANEGRAEVE         | 316 |   |
| Pha 233 | TD                   | NSLVLISSIEMEEGALF             | VPHYYSRAIVILVNEGEAIVE | 274 |   |
|         | F                    | G                             | H                     |     |   |
| Gly 374 | VVNC                 | NGERVFDGELQEGRV               | LIVPQNFVVAARSQSDNF    | 410 |   |
| Can 317 | LVGLEQQQQGLESMLRR    | YAATLSEGDIIIV                 | SSFPVALKAAASDLN       | 363 |   |
| Pha 275 | LVGPKGN              | KETLEIES YRAELSKDDV           | FVIPAAYPVAIKATSNVN    | 317 |   |
|         | I                    | J                             | 1                     | 2   | 3 |
| Gly 411 | EYVSFKTNDTPMIGTLAG   | ANSLNAL                       | PEEVIOHTFNLSQQA       | 452 |   |
| Can 364 | MVGIGVNAENNERNF      | LGHKENVIRQI                   | PROVSDLTFFPGSGEEV     | 416 |   |
| Pha 318 | FTGFGINANNNNRNL      | LAKTDNVISSIGRALD              | GKDVGLGLTFSGSGDEV     | 365 |   |
|         | J'                   |                               |                       |     |   |
| Gly 453 | ROIKNNNPFKFLV        | PPQESQKRAVA                   | 476                   |     |   |
| Can 407 | BEILENKQESYFVDG      | OPRHIDAGGKARRAHL              | PNLFRFTFY             | 445 |   |
| Pha 366 | MKLINKQSGSYFVDA      | HHHQEQQKGRKGAFVY              | 397                   |     |   |

**Figure 11.** Alignment of protein sequences among the glycinin A1aB1b subunit, canavalin and phaseolin based on the three-dimensional structures. Gly, Can and Pha indicate glycinin, canavalin<sup>9</sup> and phaseolin,<sup>7</sup> respectively. The helices and strands are numbered above the alignment. The identical residues are represented by red. The green and black underlines indicate helices and strands, respectively. The disordered or lacking residues are shown by blue letters.

53%, respectively, comparable with 48% of the proglycinin. In addition, the distribution of the hydrophobic residues in the interacting interfaces of the proglycinin is similar to those of both 7 S globulins. The hydrophobic residues of both 7 S globulins are also located far from the solvent, and are surrounded by hydrophilic residues. The distribution and the number of hydrogen bonds are also similar between the proglycinin and 7 S globulins, and buried hydrogen bonds are observed in the center of the interface between the barrel domains. However, the residues contributing to the hydrogen bonds are not well conserved among the three globulins compared with the hydrophobic residues.

Shutov *et al.*<sup>12</sup> described the gene triplication hypothesis between  $\beta$ -conglycinin (7 S) and glyci-

**Table 4.** Hydrogen bonds between protomers

| Residue | Atom | Domain         | Residue | Atom | Domain | Distance (Å)    |     |                  |
|---------|------|----------------|---------|------|--------|-----------------|-----|------------------|
|         |      |                |         |      |        | AB <sup>a</sup> | BC  | CA               |
| 62 Arg  | NH1  | B <sup>b</sup> | 444 Thr | O    | H      | 2.8             | 2.7 | 3.0              |
| 65 Tyr  | OH   | B              | 398 Asn | O    | B      | 2.9             | 2.9 | 2.8 <sup>c</sup> |
| 110 Arg | O    | B              | 456 Lys | NZ   | H      | 2.8             | 3.1 | 2.6 *            |
| 112 Gln | N    | B              | 435 Ala | O    | H      | 2.9             | 2.9 | 3.1 *            |
| 128 Gly | O    | B              | 355 Tyr | OH   | B      | 2.9             | 3.0 | 2.8              |
| 153 Glu | O    | B              | 397 Gln | NE2  | B      | 3.4             | 3.0 | 3.2 *            |
| 155 Gln | N    | B              | 397 Gln | OE1  | B      | 2.7             | 3.0 | 3.0 *            |
| 155 Gln | OE1  | B              | 417 Thr | OG1  | B      | 3.0             | 2.6 | 2.7 *            |
| 155 Gln | OE1  | B              | 417 Thr | O    | B      | 3.4             | 2.9 | 3.1 *            |
| 155 Gln | NE2  | B              | 358 Asn | O    | B      | 3.0             | 3.2 | 3.0 *            |
| 157 Asp | OD2  | B              | 12 Cys  | N    | B      | 2.8             | 2.8 | 2.9 *            |
| 171 Gln | OE1  | B              | 378 Asn | ND2  | B      | 2.9             | 2.8 | 2.8              |
| 177 Gln | NE2  | B              | 378 Asn | OD1  | B      | 2.9             | 3.4 | 2.9              |
| 204 Gly | O    | H              | 381 Arg | NH1  | B      | 2.9             | 2.7 | 2.8              |
| 206 Thr | N    | H              | 373 Gln | OE1  | B      | 3.1             | 3.1 | 3.0              |
| 213 Ala | O    | H              | 459 Asn | ND2  | H      | 2.6             | 2.7 | 2.7              |
| 215 Ser | O    | H              | 458 Asn | ND2  | H      | 2.7             | 2.9 | 2.9              |
| 223 Asn | ND2  | H              | 446 Asn | O    | H      | 2.7             | 2.8 | 2.9              |
| 244 Ser | N    | B              | 439 Glu | OE1  | H      | 2.9             | 2.9 | 3.2 *            |
| 244 Ser | N    | B              | 439 Glu | OE2  | H      | 3.2             | 3.0 | 3.1 *            |

<sup>a</sup> Distance between protomers A and B in a trimer.<sup>b</sup> B and H indicate barrel and helix, respectively.<sup>c</sup> Asterisk indicates hydrogen bonds which may be conserved among five glycinin subunits.

nin based on sequence comparison and limited proteolysis analysis. However, they have suggested the gene duplication mechanism recently.<sup>13</sup> Table 5 shows mutual and cross-comparisons of the N and C-terminal modules of canavalin, phaseolin and proglycinin based on superimposition of their C<sup>α</sup> models. The N-terminal module of proglycinin is most similar to the C-terminal module of the proglycinin, and the comparison of the N-terminal module to the four terminal modules of canavalin and phaseolin shows that the N-terminal module of proglycinin is evenly similar to the four terminal modules, since there is no significant difference of rmsd value and the number of superimposed atoms among them. On the other hand, the C-terminal module of the proglycinin is more similar to the C-terminal modules than to the N-terminal modules of both canavalin and phaseolin. These results suggest that the evolutionary conservation between

11 S and 7 S globulins in the structure of the C-terminal modules in a subunit of the proglycinin is greater than that of the N-terminal module, supporting the gene duplication hypothesis. In addition, the result of conservation is consistent with the number of identical residues among glycinin, canavalin and phaseolin, as shown in Figure 11. The N-terminal module of the proglycinin may be most diverged from the other terminal modules. The rmsd values show that the N and C-terminal modules of canavalin are most similar to the N and C-terminal modules of phaseolin, respectively, indicating the conservation between the N and the C-terminal module structures of 7 S globulins and the validity of this structure-based comparison.

### The center of the trimer

Ko *et al.*<sup>8</sup> found a hydrophilic channel with a diameter of around 18 Å along the 3-fold axis in

**Table 5.** rms distance of C<sup>α</sup> atoms among six modules of proglycinin, canavalin and phaseolin (Å)

| Module            | GlyN                                 | GlyC      | CanN      | CanC      | PhaN      |
|-------------------|--------------------------------------|-----------|-----------|-----------|-----------|
| GlyC              | 0.97 <sup>a</sup> (132) <sup>b</sup> |           |           |           |           |
| CanN <sup>c</sup> | 1.10(119)                            | 1.00(108) |           |           |           |
| CanC              | 1.11(115)                            | 1.05(125) | 0.93(108) |           |           |
| PhaN <sup>d</sup> | 1.14(110)                            | 1.08(102) | 0.76(154) | 0.99(106) |           |
| PhaC              | 1.08(112)                            | 0.95(120) | 0.93(106) | 0.72(150) | 1.01(105) |

<sup>a</sup> The rmsd values were calculated using distances between superimposed C<sup>α</sup> atoms in a distance less than 2.0 Å.<sup>b</sup> The number of C<sup>α</sup> atoms superimposed in a distance less than 2.0 Å are represented in parentheses.<sup>c</sup> N and C-terminal modules of canavalin contain 183 (55 to 224 and 241 to 253) and 182 (44 to 54 and 254 to 424) residues, respectively.<sup>d</sup> N and C-terminal modules of phaseolin contain 193 (18 to 210) and 169 (11 to 17 and 220 to 381) residues, respectively.

a canavalin trimer. A similar hydrophilic channel was also observed in the phaseolin trimer.<sup>7</sup> However, we could not find any hydrophilic channels in the center of the trimer, because the channel is closed by the IJ loop. In particular, a side-chain of Leu152 protrudes toward the 3-fold axis from each protomer, because the side-chains of three Leu152 residues made a hydrophobic cluster in the center. The length of the IJ loop of the proglycinin was longer than those of phaseolin and canavalin by more than four residues (Figure 11). This loop is conserved among 11 S globulins. In glycinin, Leu152 was conserved as leucine and phenylalanine. The reason for the structural difference in the center of the trimer between 11 S and 7 S globulins is not clear.

### Potential N-glycosylation sites

Among the five glycinin subunits, the A1aB1b and A2B1a subunits have one potential N-glycosylation site of Asn418 and Asn300, respectively.<sup>30,31</sup> However, they are not glycosylated as  $\beta$ -conglycinin is in soybean. To investigate why the two potential sites are not glycosylated, we analyzed the structures in the vicinities of the two asparagine residues. The structure of Asn418 in the A1aB1b subunit was determined in this study. The residue is positioned just after strand I in the basic chain and near the 3-fold axis of the trimer. The side-chain of this residue is not very accessible to solvent molecules, indicating that the glycosylation at this site does not occur due to steric hindrance. On the other hand, we hypothesized the position of Asn300 in A2B1b on the basis of the crystal structure of the A1aB1b subunit and sequence alignment (Figure 7). The results suggested that Asn300 may be located on strand C and on the side of the trimer, and that the side-chain of the residue may face the solvent. Therefore, it appears that the disordered regions cannot cover the residue. Further, we cannot rule out the possibility that the glycosyl chain is trimmed off from the A2B1a subunit before maturation.

### Sorting of glycinin to the vacuole

Proglycinin is assembled into a trimer in the ER and is sorted to PSV during seed maturation. It is believed that a vacuolar sorting determinant is present on the molecular surface of the proglycinin trimer.<sup>32,33</sup> The protein sequence of glycinin, with the exception of the signal sequence, must play a role in the sorting and accumulation. Saalbach *et al.*<sup>34</sup> investigated the intracellular localization of bacterial chloramphenicol acetyltransferase fused with the fragment of *Vicia faba* prolegumin B4 in tobacco, and suggested that a prolegumin contains multiple targeting information, probably formed by higher structures of relatively long peptide sequences. Recently, based on the secretion granule for-

mation in mammals,<sup>35,36</sup> a mechanism of aggregation sorting was proposed for legumin in developing pea cotyledons.<sup>37–40</sup> The more hydrophobic prolegumin participates in aggregate formation and in the binding of prolegumin to membranes of dense vesicle or multivesicular bodies. Moreover, the processing of prolegumin to mature form and the resultant hexamer formation cause a decrease in hydrophobicity that leads to the dissociation of legumin aggregates from the membrane.<sup>32</sup> The crystal structure of proglycinin suggests that the hexamer may be formed by the interaction between the IE faces of the trimer. This suggestion is consistent with the result that the prolegumin is much more hydrophobic than mature legumin,<sup>37</sup> since the IE face of proglycinin is more hydrophobic than the IA face (Figure 9). However, the formation of a dense vesicle containing storage proteins in the densely packed state cannot be explained by hydrophobic interactions alone, and thus other factors will need to be investigated.

## Materials and Methods

### Crystallization and data collection

The crystals of proglycinin A1aB1b homotrimer were grown at 281 K by the dialysis method<sup>16</sup> and by hanging-drop vapor diffusion method employing 5  $\mu$ l of a 10 mg/ml protein solution mixed with 5  $\mu$ l of precipitant solution containing 8% PEG6000, 0.1 M Mes (pH 5.5) and 0.4 M NaCl. The crystals made by the dialysis method were used for collection of native data. The data collections for the native and derivative crystals were carried out at 293 K on a Bruker multiwire HI-STAR detector mounted on a MAC Science M18XHF rotating anode generator at the Research Institute for Food Science, Kyoto University, or a Siemens X-1000 multiwire detector at the Institute for Protein Research, Osaka University.

### MIR phasing

The multiple isomorphous replacement (MIR) method was applied to determine the protein structure. The native crystals were soaked in two different pH solutions containing heavy-atom compounds. The soaked crystals were checked for isomorphism by measuring changes in their cell dimensions and diffraction intensity. The isomorphous difference Patterson functions at 3.5 Å resolution were calculated for each derivative. The refinement of the heavy-atom parameters and phase determination were carried out by the conventional MIR method using the program package PHASES.<sup>41</sup> Additional heavy-atom sites were identified using difference Fourier techniques. It was possible to locate the 3-fold axis of the trimer in the asymmetric unit by comparing the identified heavy-atom sites. The initial MIR phases were calculated to a resolution of 3.2 Å from five derivatives with a mean figure of merit = 0.58 (Table 6). After solvent flattening, a mean figure of merit of 0.70 was obtained at 3.2 Å resolution. The phases were improved and extended to a resolution of 2.9 Å by applying 3-fold symmetry averaging, solvent-flattening and partial modeling. This resulted in a mean figure of merit of 0.62 at this higher

**Table 6.** Summary of statistics for heavy-atom derivatives

|                           | PCMB-1   | CH <sub>3</sub> HgOH | KAu(Cl) <sub>4</sub> | KAu(CN) <sub>2</sub> | PCMB-2       |
|---------------------------|----------|----------------------|----------------------|----------------------|--------------|
| Method of crystallization | Dialysis | Dialysis             | Dialysis             | Hanging drop         | Hanging drop |
| Soaking conditions        |          |                      |                      |                      |              |
| Concentration (mM)        | 1        | 10                   | 2                    | 1                    | 10           |
| Period (hour)             | 18       | 15                   | 18                   | 15                   | 1.5          |
| pH <sup>a</sup>           | 7.6      | 7.6                  | 7.6                  | 5.5                  | 5.5          |
| Detector                  | X-1000   | X-1000               | X-1000               | High Star            | High Star    |
| Wavelength (Å)            | 1.54     | 1.54                 | 1.54                 | 1.54                 | 1.54         |
| Resolution (Å)            | 3.4      | 3.2                  | 5.0                  | 3.2                  | 3.2          |
| No of observations        | 68,998   | 127,263              | 50,480               | 152,342              | 147,621      |
| No of unique reflections  | 23,245   | 29,156               | 8,470                | 51,491               | 44,651       |
| Completeness (%)          | 74.2     | 86.7                 | 99.7                 | 88.4                 | 94.0         |
| R-merge (%)               | 14.9     | 14.0                 | 10.4                 | 10.5                 | 14.1         |
| MIR phasing statistics    |          |                      |                      |                      |              |
| R-Cullis                  | 0.673    | 0.697                | 0.634                | 0.698                | 0.567        |
| R-Kraut                   | 0.102    | 0.098                | 0.073                | 0.120                | 0.135        |
| Phasing power             | 1.31     | 1.34                 | 1.59                 | 0.95                 | 1.60         |
| Heavy-atom sites          | 3        | 3                    | 3                    | 6                    | 12           |

<sup>a</sup> 0.1 M Tris (pH 7.6), 0.1 M Mes (pH 5.5).

resolution. The solvent flattening and phase improvement were performed by using the program PHASES.

### Model building and refinement

The initial structure model of glycinin was built against the electron density map at 2.9 Å referring to the structures of canavalin and phaseolin<sup>7,8,10</sup> using the program TURBO-FRODO (Bio-Graphics). Crystallographic refinement was carried out against 2.8 Å data with  $F > 2\sigma F$ . A randomly selected 5% of the data was removed prior to refinement for analysis of the free  $R$  factor.<sup>42</sup> Starting with this model, the refinement was performed with the program X-PLOR.<sup>43</sup> The initial model was subject to one round of simulated annealing, followed by a round of Powell minimization. The refinement was performed at 2.8 Å resolution with the three monomers restrained by the non-crystallographic symmetry. The progress of the refinement was guided by the decrease in both the conventional and free  $R$  factors. Non-crystallographic symmetry (NCS) restraints were imposed with a weight of 20 kcal/mol per Å<sup>2</sup>, based on the analysis of the free  $R$ -factor where the NCS weight was varied. A number of rounds of manual rebuilding of poorly fitting residues followed. Each round of manual rebuilding was followed by conjugate-gradient energy minimization and individual  $B$ -factor refinement. NCS was not included in the final cycle of refinement, and 34 solvent molecules were built.

### Calculation of accessible surface

The accessible surface was calculated using program NACCESS.<sup>44</sup> The probe is taken to be a water molecule as a sphere with a radius of 1.4 Å.

### Acknowledgements

Computation time was provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University. We thank Dr E. M. T. Mendoza for her critical reading of the manuscript. We are grateful to

Drs M. Kusunoki, and Y. Katsube of the Institute for Protein Research, Osaka University for their technical support.

This work was supported in part by grants from the Ministry of Education, Science, and Culture of Japan (to M. A., Y. T. and S. U.), the Program for Promotion of Basic Research Activities for Innovative Biosciences (to B. M. and S. U.) and the Fuji Foundation for Protein Research (to S. U.).

### References

1. Staswick, P. E., Hermodson, M. A. & Nielsen, N. C. (1984). Identification of the cysteines which link the acidic and basic components of the glycinin subunits. *J. Biol. Chem.* **259**, 13431-13435.
2. Dickinson, S. D., Hussein, E. H. A. & Nielsen, N. C. (1989). Role of posttranslational cleavage in glycinin assembly. *Plant Cell*, **1**, 459-469.
3. Argos, P., Narayana, S. V. L. & Nielsen, N. C. (1985). Structural similarity between legumin and vicilin storage proteins from legumes. *EMBO J.* **2**, 1111-1117.
4. Wright, D. J. (1987). The seed globulins. In *Developments in Food Proteins* (Hudson, B. J. F., ed.), vol. 5, pp. 81-157, Elsevier, London.
5. Wright, D. J. (1988). The seed globulins. In *Developments in Food Proteins* (Hudson, B. J. F., ed.), vol. 6, pp. 119-178, Elsevier, London.
6. Gibbs, P. E. M., Strongin, K. B. & McPherson, A. (1989). Evolution of legume seed storage proteins - a domain common to legumines and vicilins is duplicated in vicilins. *Mol. Biol. Evol.* **6**, 614-623.
7. Lawrence, M. C., Izard, T., Beuchat, M., Blagrove, R. J. & Colman, P. M. (1994). Structure of phaseolin at 2.2 Å resolution. *J. Mol. Biol.* **238**, 748-770.
8. Ko, T. P., Ng, J. D. & McPherson, A. (1993). The three-dimensional structure of canavalin from jack bean (*Canavalia ensiformis*). *Plant Physiol.* **101**, 729-744.
9. Ko, T. P., Ng, J. D. & McPherson, A. (2000). The refined structure of canavalin from jack bean in two crystal forms at 2.1 and 2.0 Å. *Acta Crystallog. sect. D*, **56**, 411-420.

10. Lawrence, M. C., Suzuki, E., Varghese, J. N., Davis, P. C., Van Donkelaar, A., Tulloch, P. A. & Colman, P. M. (1990). The three-dimensional structure of the seed storage protein phaseolin at 3 Å resolution. *EMBO J.* **9**, 9-15.
11. Shutov, A. D., Kakhovskaya, I. A., Braun, H., Bäumlein, H. & Müntz, K. (1995). Legumin-like and vicilin-like seed storage proteins: Evidence for a common single-domain ancestral gene. *J. Mol. Evol.* **41**, 1057-1069.
12. Shutov, A. D., Kakhovskaya, I. A., Bastrygina, A. S., Bulmaga, V. P., Horstmann, C. & Müntz, K. (1996). Limited proteolysis of  $\beta$ -conglycinin and glycinin, the 7S and 11S storage globulins from soybean. *Eur. J. Biochem.* **241**, 221-228.
13. Shutov, A. D. & Bäumlein, H. (1999). Origin & evolution of seed storage globulins. In *Seed Proteins* (Shewry, P. R. & Casey, R., eds), pp. 543-561, Kluwer, London.
14. Utsumi, S. (1992). Plant food protein engineering. *Advan. Food Nutr. Res.* **36**, 89-208.
15. Utsumi, S., Kim, C. S., Sato, T. & Kito, M. (1988). Signal sequence of preproglycinin affects production of the expressed protein in *Escherichia coli*. *Gene*, **71**, 349-358.
16. Utsumi, S., Gidamis, A. B., Mikami, B. & Kito, M. (1993). Crystallization and preliminary X-ray crystallographic analysis of the soybean proglycinin expressed in *Escherichia coli*. *J. Mol. Biol.* **233**, 177-178.
17. Utsumi, S., Gidamis, A. B., Takenaka, Y., Maruyama, N., Adachi, M. & Mikami, B. (1996). Crystallization and X-ray analysis of normal and modified recombinant soybean proglycinins. Three-dimensional structure of normal proglycinin at 6 Å resolution. In *Macromolecular Interactions in Food Technology, ACS Symposium Series 650* (Parris, N., Kato, A., Creamer, L. K. & Pearce, J., eds), pp. 257-270, ACS, Washington, DC.
18. Luzzati, V. (1952). Traitements statistiques des erreurs dans la détermination des structures cristallines. *Acta Crystallog.* **5**, 802-810.
19. Ramakrishnan, C. & Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformation. *Biophys. J.* **5**, 909-933.
20. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
21. Badley, R. A., Atkinson, D., Hauser, H., Oldani, D., Green, J. P. & Stubbs, J. M. (1975). The structure, physical and chemical properties of the soybean protein glycinin. *Biochim. Biophys. Acta*, **412**, 214-228.
22. Reichelt, R., Schwenke, K. D., König, T., Pöhtz, W. & Wangermann, G. (1980). Electron microscopic studies for estimation of the quaternary structure of the 11S globulin (Helianthinin) from sunflower seed. *Biochem. Physiol. Pflanzen*, **175**, 653-663.
23. Tulloch, P. A. & Blagrove, R. A. (1985). Electron microscopy of seed-storage globulins. *Arch. Biochem. Biophys.* **241**, 533-539.
24. Patel, S., Cudney, R. & McPherson, A. (1994). Crystallographic characterization and molecular symmetry of edestin, a legumin from hemp. *J. Mol. Biol.* **235**, 361-363.
25. Jung, R., Scott, M. P., Nam, Y. W., Beaman, T. W., Bassüner, R. & Saalbach, I. *et al.* (1998). The role of proteolysis in the processing and assembly of 11S seed globulins. *Plant Cell*, **10**, 343-357.
26. Utsumi, S., Gidamis, A. B., Kanamori, J., Kang, I. J. & Kito, M. (1993). Effects of deletion of disulfide bonds by protein engineering on the conformation and functional properties of soybean proglycinin. *J. Agric. Food. Chem.* **41**, 1417-1444.
27. Gidamis, A. B., Mikami, B., Katsube, T., Utsumi, S. & Kito, M. (1994). Crystallization and preliminary X-ray analysis of soybean proglycinins modified by protein engineering. *Biosci. Biotech. Biochem.* **58**, 703-706.
28. Giletto, A. & Pace, N. (1999). Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T<sub>1</sub>. *Biochemistry*, **38**, 13379-13384.
29. Shutov, A. D., Blattner, F. R. & Bäumlein, H. (1999). Evolution of a conserved protein module from *Archaea* to plants. *Trends Genet.* **15**, 348-9.
30. Utsumi, S., Kim, C. S., Kohno, M. & Kito, M. (1987). Polymorphism and expression of cDNAs encoding glycinin subunits. *Agric. Biol. Chem.* **51**, 3267-3273.
31. Utsumi, S., Kohno, T., Mori, T. & Kito, M. (1987). An alternate cDNA encoding glycinin A<sub>1a</sub>B<sub>x</sub> subunit. *J. Agric. Food Chem.* **35**, 210-214.
32. Müntz, K. (1998). Deposition of storage proteins. *Plant Mol. Biol.* **38**, 77-99.
33. Neuhaus, J. M. & Rogers, J. C. (1998). Sorting of proteins to vacuoles in plant cells. *Plant Mol. Biol.* **38**, 127-44.
34. Saalbach, G., Jung, R., Kunze, G., Saalbach, I., Adler, K. & Müntz, K. (1991). Different legumin protein domains act as vacuolar targeting signals. *Plant Cell*, **3**, 695-708.
35. Torrent, M., Geli, M. J., Ruiz-Avila, L., Canales, J. M., Puigdomenech, P. & Ludevid, D. (1994). Role of structural domains for maize  $\gamma$ -zein retention in *Xenopus oocytes*. *Planta*, **192**, 512-518.
36. Thiele, C., Gerdes, H. H. & Huttner, W. B. (1997). Protein secretion: puzzling receptors. *Curr. Biol.* **7**, R496-R500.
37. Hinz, G., Menze, A., Hohl, I. & Vaux, D. (1997). Isolation of prolegumin from developing pea seeds: its binding to endomembranes and assembly into prolegumin hexamers in the protein storage vacuole. *J. Exp. Bot.* **48**, 139-149.
38. Robinson, D. G., Bäumer, M., Hinz, G. & Hohl, I. (1997). Ultrastructure of the pea Golgi apparatus: origin of dense vesicles and the action of brefeldin A. *Protoplasma*, **200**, 198-209.
39. Robinson, D. G., Bäumer, M., Hinz, G. & Hohl, I. (1998). Vesicle transfer of storage proteins to the vacuole: the role of the Golgi apparatus and multi-vesicular bodies. *J. Plant Physiol.* **152**, 659-667.
40. Robinson, D. G., Hinz, G. & Holstein, S. E. H. (1998). The molecular characterization of transport vesicles. *Plant Mol. Biol.* **38**, 49-76.
41. Furey, W. & Swaminathan, S. (1997). PHASE-95, a program package for processing and analysing diffraction data from macromolecules. *Methods Enzymol.* **277**, 590-620.
42. Brünger, A. T. (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallog. sect. D*, **49**, 24-36.
43. Brünger, A. T. (1992). *X-PLOR version 3.1 Manual: a system for crystallography and NMR*, Yale University, New Haven, CT.

- 
44. Hubbard, S. J. & Thornton, J. M. (1993). ACCESS computer program, Department of Biochemistry and Molecular Biology, University College, London.
45. Kraulis, P. (1991). MOLSCRIPT - a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
46. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0, a program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869-873.
47. Nicholls, A., Sharp, K. A. & Hoing, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.

*Edited by R. Huber*

*(Received 19 July 2000; received in revised form 2 November 2000; accepted 2 November 2000)*