

Limited proteolysis of β -conglycinin and glycinin, the 7S and 11S storage globulins from soybean [*Glycine max* (L.) Merr.] Structural and evolutionary implications

Andrei D. SHUTOV¹, Irina A. KAKHOVSKAYA¹, Angela S. BASTRYGINA¹, Valentina P. BULMAGA¹, Christian HORSTMANN² and Klaus MÜNTZ²

¹ State University of Moldova, Kishinev, Moldova

² Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

(Received 31 May/2 August 1996) – EJB 96 0794/3

The G2 (A2B1a) glycinin subunit from soybean (*Glycine max* L. Merr.) was purified and renatured to the homohexameric holoprotein. This protein along with purified β -conglycinin were subjected to limited proteolysis by trypsin. The generated polypeptide fragments were separated via SDS/PAGE and the amino acid sequence of the N-terminals was determined. Four cleavage points were detected in the α -chain A2 of glycinin as well as in the α' -chain of β -conglycinin. From the known three-dimensional structure of 7S globulin and the hypothetical model of 7S globulin-like 11S globulin structure, it was possible to draw the conclusion that two distinct types of susceptible sites for proteolytic cleavage are characteristic of the subunits of both globulins. The first includes the sequences linking N- and C-terminal domains of both globulins and the sequence of N-terminal extensions of 70-kDa subunits from the vicilin-like 7S globulins. The second type includes the loop between β -strands E and F of the N-terminal domain of 11S globulins and of the C-terminal domain of 7S globulins. A statistically significant similarity was found between the N-terminal extension of the α' -chain of β -conglycinin and the interdomain linker regions of soybean glycinin and pea legumin. It is proposed that the three sequence regions which form the first type of susceptible sites are of similar structural function and might have evolved from the N-terminal segment of a putative single-domain ancestor.

Keywords: protein structure; seed storage globulins; gene evolution; limited proteolysis; *Glycine max* (L.) Merr.

Glycinin and β -conglycinin from soybean seeds belong to the legumin-like 11S and vicilin-like 7S storage globulin families, respectively. Secondary structure predictions and multiple alignment of amino acid sequences [1] suggest that structural similarities should exist between 11S and 7S seed storage globulins. The trimer which is formed from 11S globulin subunit precursors before their cleavage into α - and β -chains [2, 3] as well as the trimeric half-molecule from the mature hexameric 11S globulin holoprotein [4] appear to be structurally similar to the trimeric 7S globulin molecule. X-ray-crystallographic analysis of two 7S globulins, canavalin from jack bean [5] and phaseolin from garden bean [6, 7], revealed that their subunits are composed of two similar domains. Each domain is composed of a β -barrel followed by a region of α -helices (Fig. 1, A1 and B1). A canonical model for the domain structure of 7S globulins, including both 50-kDa and 70-kDa subunits, was postulated [7]. However, the structural role of the extended N-terminal sequence upstream from the N-terminal domain of the 70-kDa subunits remains unclear. Different alignment methods that have been employed to compare the domains of the 7S globulin subunits with corresponding regions from the 11S globulin subunits suggest that similar N- and C-terminal domains should exist in the legumin-like storage globulins [7, 8] (Fig. 1, A2 and B2). It

has been predicted that all four structural domains, the two domains of 11S and the two domains of 7S globulin subunits, are similar and may have evolved from a common single-domain ancestor [8]. This model of storage protein evolution is supported by similarities of the storage globulin domains to the single-domain proteins germin and spherulin involved in desiccation/hydration processes in plant seeds and fungal spores, respectively [9]. According to this model of seed-storage globulin evolution [8], the sequence elements upstream from the β -barrel of the N-terminal domain and those elements linking N- and C-terminal domains could have evolved from the N-terminal segment of a common legumin/vicilin-like single-domain ancestor.

In vivo as well as *in vitro* limited proteolysis produces several high-molecular-mass fragments from both the 7S and 11S globulin subunits (see review [10] and [11, 12]). The position of the proteolytic cleavage sites must be determined by the tertiary structure of globulin subunits in the assembled holoproteins. Determining the exact location of the cleavage sites by amino acid sequencing would allow the projection of the sites onto the three-dimensional structure models and thereby provide an additional test of the model for 7S and 11S globulin subunit structure. Conformational similarity between the two globulins has already been proposed based upon the pattern of limited proteolysis of glycinin and β -conglycinin [11].

Recently, we studied high-molecular-mass fragments formed by limited proteolysis of glycinin by trypsin. The SDS/PAGE patterns of the hydrolyzates suggest that three sites are suscepti-

Correspondence to K. Müntz, Institute of Plant Genetics and Crop Plant Research, Corrensstr. 3, D-06466 Gatersleben, Germany

Abbreviation. SBP, 62-kDa sucrose binding protein from *Glycine max* (L.) Merr.

Definition. SD, similarity degree.

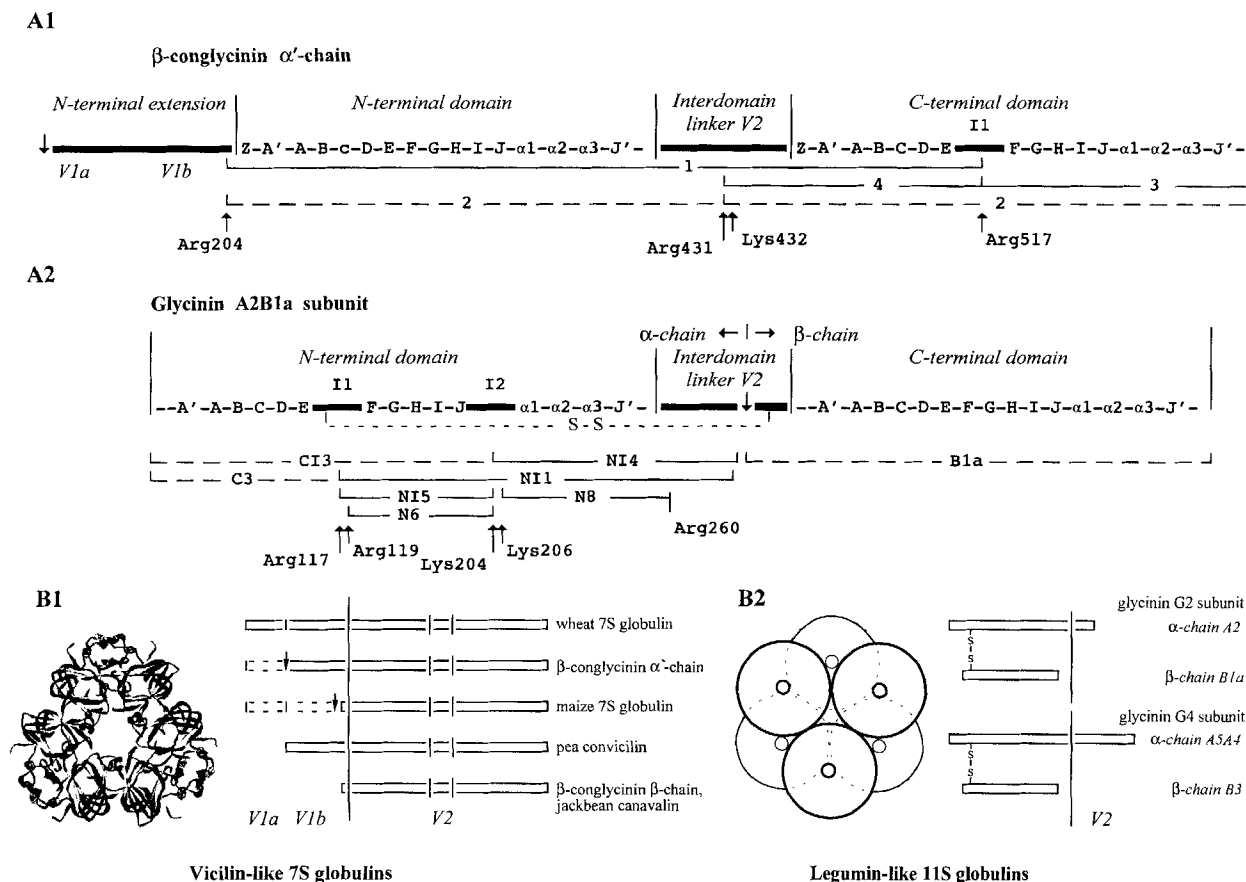


Fig. 1. Principles of organization of 7S and 11S storage globulin molecules. (A) Topography of secondary structures and sites susceptible to limited proteolysis in (A1) β -conglycinin α' -chain and (A2) glycinin G2 (A2B1a) subunit. β -Strands Z, A', A ... J forming β -barrels and α -helices $\alpha 1$, $\alpha 2$, $\alpha 3$ are indicated according to the canonical model of 7S globulin structure and to the vicilin-like model of 11S globulin structure [7, 8]. The sites susceptible to limited proteolysis are highlighted by thick solid lines. (---S---) disulfide bridge between α -chain A2 and β -chain B1a. Cleavage points and trypsinolytic fragments (compare Fig. 2) are shown below the schemes of subunits: (I) points of proteolytic processing of β -conglycinin α' -chain and glycinin A2B1a subunit precursors; (f) trypsinolytic cleavage points according to the data of the N-terminal sequencing of corresponding fragments. The nomenclature of fragments corresponds to our previous publication [12]. C, fragments covalently linked with β -chain B1a; N, non-dissociable fragments retained in holoprotein molecule only due to non-covalent interactions; I, cleavage intermediates. C-terminal limits of the N-terminal fragments have been positioned according to their apparent molecular masses [12]. (B) Polymorphic subunits and quaternary structure models of 7S globulin (diagram of the canavalin trimer from Fig. 13 of [5]) and 11S globulin (scheme of the model of the helianthin hexamer from Fig. 5 of [19]). Both holoprotein molecules of 7S and 11S globulins are composed of subunits differently sized because of hypervariability of the segments V1a and V1b of the N-terminal extensions in 7S globulin subunits and of the interdomain linkers V2 in 11S globulin subunits, respectively.

ble to proteolytic attack in the α -chain A2 [12]. The sites appeared to lie within the three variable hydrophilic regions characteristic of 11S globulins sequences and probably lacking structurally ordered elements. Here we report the exact localization of the cleavage sites by N-terminal amino acid sequencing of trypsinolytic fragments of glycinin and β -conglycinin subunits. In both globulins two types of cleavage sites were found. The first is located in the linker region between N- and C-terminal domains, and the second is located in the loop between β -strands E and F. Sequence alignment revealed similarity between the variable N/C-domain linker region of 11S globulins and the N-terminal extension in 70-kDa subunits of 7S globulins which also represents the first type of cleavage site. The structural and evolutionary implications of the similarities of 11S and 7S storage globulins are discussed.

MATERIALS AND METHODS

Purification of storage globulins. Glycinin and β -conglycinin were isolated from mature soybean seeds [*Glycine max* (L.)

Merr., cv. Kishinevskaya-16] according to Thanh and Shibasaki [13] and purified by phenyl-Sepharose CL-4B chromatography [12]. β -Conglycinin was additionally purified by Sephacryl S-300 gel filtration. N-ethylmaleyl-glycinin was prepared according to Kitamura et al. [14]. All the purification steps were performed in the presence of 0.02% NaN_3 .

Reconstitution of glycinin G2. After denaturation by 6 M urea, N-ethylmaleyl-glycinin subunits were fractionated by DEAE-Sepharose CL-6B chromatography according to Staswick et al. [15]. The isolated A2B1a subunits were reconstituted into oligomers using the procedure of Kitamura et al. [14]. Analysis by sucrose density gradient centrifugation revealed that approximately 50% of the applied protein appeared in the hexameric 11S globulin peak whereas the other 50% corresponded to trimers in the 7S peak. This agreed with previous findings by analytical ultracentrifugation [12]. Reconstituted hexameric glycinin G2 was isolated by Sephacryl S-400 gel filtration.

Proteolysis of glycinin G2 and β -conglycinin and fragment separation. Protein substrates (5 mg/ml in 0.05 M Tris/HCl pH 8.0 containing 0.45 M NaCl and 0.04% NaN_3) were

incubated with trypsin (Spofa, Czech Republic) at 30°C, the active enzyme/substrate ratio being 1:600 (by mass). The reaction was stopped by adding trichloroacetic acid to a final concentration of 6% (mass/vol.). The resulting precipitates were analyzed by SDS/PAGE under non-reducing conditions using the Laemmli buffer system [16] in either 10–25% gradient gels (glycinin G2) or 12.5% gels (β -conglycinin).

N-terminal amino acid sequencing. After electrophoresis glycinin G2 and β -conglycinin fragments were blotted onto poly(vinylidene difluoride) membranes. Subsequently, N-terminal Edman degradation was performed according to the manufacturer's instructions using a Beckman Sequencer LF3400.

Alignment of amino acid sequences. The computer program CLUSTAL was used for multiple alignments. Similarity between the sequences at the amino acid as well as at the nucleic acid level was maximized by first aligning the amino acid sequences and using these results to align the nucleic acid sequences. Alignment was improved by limited manual correction to obtain the best consensus between amino acid and nucleic acid alignments. The statistical significance of amino acid sequence similarity [17] was calculated for 200 randomizations.

Sequence sources and numbering. 11S and 7S globulin sequences were extracted from the EMBL Data Library (database accession codes in parentheses): glycinin A2B1a subunit (X15122) and A5A4B3 subunit (X02626); pea legumin (M16890); β -conglycinin α -chain (X17698), α' -chain (M13759) and β -chain (P25974); pea vicilin (P02854) and convicilin (X06398); jackbean canavalin (X06733); 7S globulins from cacao (X62625), cotton (M16891), maize (M24845) and wheat (M81719); soybean sucrose binding protein (L06038). Numbering of the amino acid residues corresponds to that of the data base.

RESULTS

Proteolysis of glycinin G2. Because of the complexity of fragment composition of hydrolyzates of the native polymorphic glycinin [12], it is expedient to use as a substrate a homohexameric 11S globulin. Such a protein can be prepared by the reconstitution of the isolated 11S globulin subunit [18]. The reconstituted 11S globulin hexamer was found to be identical with the native protein when judged from several data, including circular dichroic analyses [14], electron microscopy [18] and limited trypsinolysis [12, 14].

Glycinin G2 is a homo-oligomeric 11S globulin composed of A2B1a subunits. Its limited trypsinolysis gave a relatively simple pattern of bands in SDS/PAGE (Fig. 2A) suitable for the N-terminal sequencing of α -chain A2 fragments. Since α - and β -chains are linked by a disulfide bridge [20], under non-reducing conditions some of intermediate and final A2 chain fragments remain covalently linked with B1a chain. Other fragments are retained in the holoprotein molecule under non-denaturing conditions due to non-covalent interactions (nomenclature and positions of fragments see Fig. 1, A2).

Two susceptible sites in the vicinity of Arg117 and Lys204 were detected by N-terminal sequencing of the two groups of non-covalently linked fragments (Table 1). Both sites with the presumed N- and C-terminal borders of the corresponding fragments (Fig. 1, A2) have already been previously deduced on the basis of their apparent molecular masses and several other indirect data [12]. Obviously, the intermediate fragment CI3 and the final fragment C3, which are covalently bound to the B1a β -chain, correspond to A2 chain fragments upstream from these cleavage points. According to the apparent molecular masses of

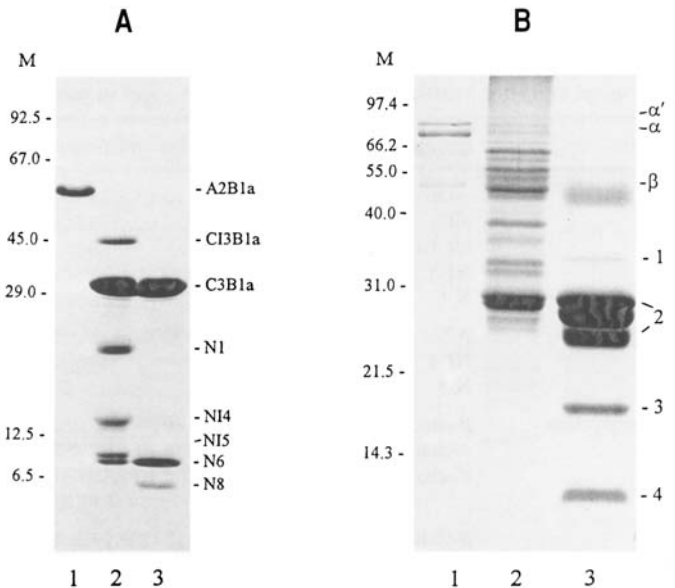


Fig. 2. SDS/PAGE patterns of glycinin G2 and β -conglycinin hydrolyzates. (A) Trypsinolysis of glycinin G2. For nomenclature of the fragments see Fig. 1. M, Serva molecular mass markers (values on left in kDa); glycinin chain A5 (10.6 kDa) was also used as a standard. Lane 1, intact glycinin G2; lanes 2 and 3, 3-min and 24-h hydrolyzates, respectively. The samples were treated by SDS in the absence of 2-mercaptoethanol. (B) Proteolysis of β -conglycinin. M, Promega molecular mass markers (values on left in kDa). Lane 1, intact β -conglycinin; lanes 2 and 3, 3-min and 2-h hydrolyzates, respectively.

the fragments CI3 (22 kDa) and C3 (12 kDa) [12], the N-termini of both fragments probably correspond to the N-terminus of the non-cleaved A2 α -chain. The low apparent molecular mass of the fragment N8 (5.9 kDa) indicates that a relatively large C-terminal fragment was cleaved off from the A2 α -chain, probably downstream from Arg260. Indirect data suggest that this segment is digested into peptides [12].

In accordance with the presumed structural similarity between 7S and 11S globulins [8], the cleavage points detected in the A2 α -chain are located in the EF loop and in the β -barrel/ α -helix linking region (Fig. 1, A2).

Proteolysis of α' -chain of β -conglycinin. The β -conglycinin preparation was enriched in α - and α' -chains (Fig. 2B). This could be the result of a partial fractionation of the native mixture of polymorphic 7S globulin holoproteins. Even after the gel filtration step some contaminating bands remained in the undigested β -conglycinin preparation. They could result from partially degraded α - and α' -chains. No specific degradation products of the contaminating bands were apparent in SDS gels after trypsinolysis of β -conglycinin (Fig. 2B).

More than 10 fragments with the apparent molecular masses between 70 kDa to 35 kDa transiently appeared on SDS gels shortly after the β -conglycinin preparation was incubated with trypsin (Fig. 2B). After 1 h of proteolysis a relatively simple pattern of several minor bands and one group of major bands was generated. Three of the minor fragments, the intermediate fragment 1 and the stable fragments 3 and 4 were subjected to N-terminal amino acid sequencing (Table 1). By comparing these sequences with the known sequences of β -conglycinin chains, fragments 1 and 3 can only be derived from the α' -chain whereas fragment 4 can be assigned to the α' - as well as α -chains (Table 1). Products of trypsinolysis of the β -chain of β -

Table 1. N-terminal sequences of the fragments of glycinin α -chain A2 and β -conglycinin α' -chain. Compare this table with Fig. 1. (m) indicates N-terminal sequence of a minor component present in the electrophoretic band. Splitting points predicted earlier [12] are designated by asterisks. Sequence numbering of α -chain A2 of glycinin subunit G2 (X15122) and β -conglycinin α -chain (X17698), α' -chain (M13759) and β -chain (P25974) correspond to that of EMBL Data Library (accession codes in parentheses).

Protein	Fragment	N-terminal sequence detected	Splitting point
Glycinin	α -chain A2	115 QQRGRSQRPDHRHQVHRFREG...	
	NI 1	GRSQRPDHRHQVHR...	R117–G118*
	NI 1 (m)	SQXPXDRHQKVXRF...	R119–S 120*
	NI 5	GRSQRPDHR...	R117–G 118*
	N 6	SQRPDHRHQVHRF...	R119–S 120*
α -chain	A2	202 SQKGKQEEENECSNILSGFAPE...	
	NI 4	GKQEEENECS...	K204–G 205*
	N 8	QXEENECSNILSGF...	K206–Q 207*
β -Conglycinin	β -chain	26 VREDENNPFFYFRSSNSFQT...	
	α -chain	186 ELRRHKKNKPNFLFGSNRFETL	
	α' -chain	202 EPRRHKNKPNFFHNSKRFQTL...	
	1	RHKKNKPNFFHNSXRF...	R204–R205
	β -chain	315 QQKQKQEEEPLEVRKYRAELSEDDVF...	
	α -chain	481 QQEQKQEEQPLEVRKYRAELSEQDIF...	
	α' -chain	515 QQRQKQEEQPLEVRKYRAELSEQDIF...	
	3	QQEQKQPLEVRKYRAELSYQ...	R517–Q518
	β -chain	228 SSRKTISSEDEPFNLRSRNPISNNFGKFFE...	
	α -chain	394 SSRKTISSEDKPFNLRSRDPISNKLGLKFFE...	
	α' -chain	429 SSRKTISSEDKPFNLGSRDPISNKLGLKFE...	
	4	KTISSEDKPFNLXSXDPIYSNXLGX...	R431–K432
	4(m)	TISXEDKPFNLXSXXPIYXNX...	K432–T433

conglycinin could not be detected since it was only a minor component of our β -conglycinin preparation.

According to the three-dimensional model of 7S globulin structure [7], the three detected cleavage points lie within the following regions of β -conglycinin chains (Fig. 1, A1): (a) a segment just upstream from the beginning of the N-terminal domain of the α' -chain; (b) the linker region between N- and C-terminal domains of α' -chain (or of α -chain); and (c) the EF loop of the C-terminal domain of α' -chain. The sequences of α - and α' -chains are highly similar to each other. This suggests that EF loops in the C-terminal domain as well as interdomain linking regions are potentially susceptible to proteolytic attack in both chains. However, the region of α -chain which is similar to the cleavage site Arg517/Gln518 in the EF loop of α' -chain (Table 1) does not contain cleavage sites accessible to trypsin. Consequently, fragment 4 as well as fragment 3 must have originated from the C-terminal domain of α' -chains (Fig. 1, A1). Thus, the stable fragments 3 and 4 together should correspond to the complete sequence of the α' -chain C-terminal domain. It seems likely that the intermediate fragment 1, with an apparent molecular mass of 35 kDa, corresponds to the α' -chain sequence between the first and the third cleavage points (Arg204 and Arg517, respectively).

A number of transient fragments, including those located between α' -, α - and β -chains on the SDS gels (Fig. 2B, lane 2), were probably generated due to several cleavages of the unstable N-terminal extension of both, the α - and α' -chains, the area where they differ from the β -chain of β -conglycinin (Fig. 1, B1). Probably, group 2 of major fragments (Fig. 2B) corresponds to the inaccessible N-terminal domains of both the α' - and α -chains (without N-terminal extensions) and to the C-terminal domain of the α -chain which is inaccessible as well. The α' -chain C-terminal domain as an intermediate before its cleavage into the fragments 3 and 4 also might be a constituent of the group 2 fragments.

Sequence similarity between the N-terminal extension of the α' -chain and the interdomain linker of 11S globulin subunits.

The sequences of the N-terminal segment of 70-kDa 7S globulin subunits and the interdomain linker region of 11S globulin subunits are highly variable. Nevertheless it appears that both sequence segments could have evolved from the N-terminal V region of the putative common single-domain ancestor of 7S and 11S globulins [8]. A large portion of the N-terminal extension region of the α' -chain of β -conglycinin can be aligned with the extended interdomain linker of glyciner A5A4B3 subunit as well as with the similar region of pea legumin subunit (Fig. 3). The statistical significance of similarity between these sequence regions of glyciner and β -conglycinin is high (8.0 SD units) for the entire compared segment. It still remains relatively high (5.9 SD units) when the first 40 residues are aligned without gaps. The compared segments of both the pea legumin and β -conglycinin subunits contain extremely high numbers of Asp and Glu residues. Therefore, the statistical significance of their similarity is less reliable (5.6 SD from random for the entire compared region which contains three long deletions). However, similarity of the internal repeats is readily apparent if one compares the β -conglycinin and pea legumin sequences. As shown in Fig. 3, the four segments corresponding to these repeats can be aligned simultaneously and exhibit similarity in cross-combinations as well. In addition, the similarity between α' -chain of β -conglycinin and glyciner as well as legumin was discernable in both polypeptide and DNA sequence comparisons (Table 2).

DISCUSSION

The structural similarity between 7S and 11S seed storage globulins has been noted before, and it has been proposed that this may be because they share a common single-domain ancestor [7–9]. Potential proteolytic cleavage sites could be predicted

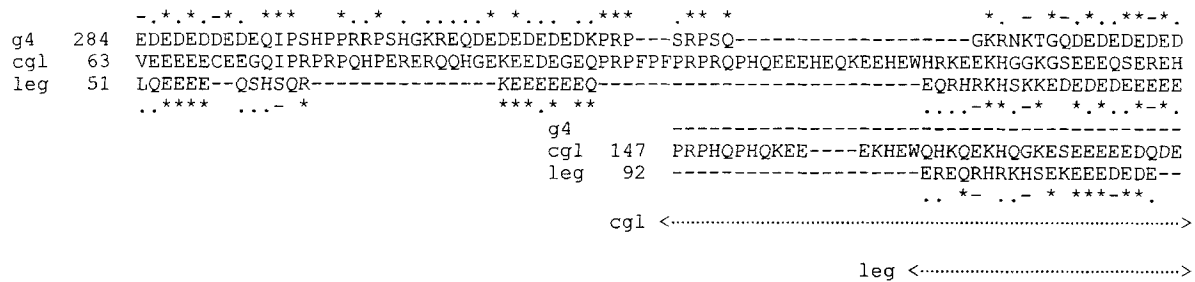


Fig. 3. Sequence similarity between the N-terminal extension of β -conglycinin α' -chain and the interdomain linkers of glycine G4 and pea legumin subunits. G4, glycine A5A4B3 subunit; leg, pea legumin from cDNA probe pCD32 [21]; cgl, β -conglycinin α' -chain. (*) Identical residues; (●) conserved substitutions which correspond to matrix elements $m(i,j) > 0$ of MDM78 according to Dayhoff [22]; (—) non-conserved substitutions [matrix elements $m(i,j) < 0$]. The regions of internal repeats of cgl and leg sequences (dotted lines) are shown in combination according to the multiple alignment of those four segments together with the segment of G4.

Table 2. Similarity between amino acid and nucleic acid sequences of the segment V1b of β -conglycinin α' -chain N-terminal extension (cgl) and interdomain linkers V2 of glycine A5A4B3 (Gy4) and pea legumin (leg) subunits.

Indications of similarity	Pairwise comparisons of sequences	
	gy4/cgl	leg/cgl
Total length of gy4 and leg segments	62	60
Identical amino acid residues	23 (37%)	25 (42%)
Conserved amino acid substitutions [$m(i,j) > 1$]	22 (35%)	21 (35%)
Non-conserved amino acid substitutions [$m(i,j) < 0$]	4 (6%)	6 (10%)
Average $m(i,j)$ of positions matched without gaps	2.4	2.3
Identical base pairs	106 (57%)	110 (61%)

from the structure models for 7S and 11S globulin subunits as well as from our previous experiments [12]. Here we have examined in detail several of the proteolytic fragments that are produced by limited proteolysis and compared amino acid sequences in the vicinity of proteolytic cleavage sites. The results confirm the three-dimensional structural model of these classes of seed storage proteins, are in accord with our triplication model of globulin evolution [8] and have revealed several new details.

The most striking difference between the two 7S globulin domains, as well as between the two 11S globulin domains, is that one of them (the N-terminal domain in 7S and C-terminal domain in 11S globulin subunits, respectively) never tolerates large hydrophilic insertions inside the structurally ordered sequence regions forming a specific β -barrel/ α -helix [7, 8]. This conformation, which looks like a core of the subunit structure of both storage globulins, is inaccessible to proteolytic attack in assembled holoproteins of both storage globulins. In contrast, the other domains of the storage globulin subunits (the C-terminal domain in 7S and N-terminal domain in 11S globulin subunits) tolerate extended hydrophilic insertions and are susceptible to proteolytic attack. Such an insert I1 is wedged between the β -strands E and F in C-terminal and N-terminal domains of 7S and 11S globulin subunits, respectively (Fig. 4A). Thus, the susceptibility to proteolytic attack as well as the structural fea-

tures reveal that the domains of 7S and 11S globulin subunits are most similar if their domains are compared cross-wisely. This argues in favor of the triplication model of storage globulin evolution [8].

Another type of sites that are susceptible to limited proteolysis in both storage globulin holoproteins belongs to sequence regions located outside of the β -barrel/ α -helix structures (Fig. 4A): V1 corresponds to the N-terminal hypervariable extension where 70-kDa subunits differ from 50-kDa subunits of 7S globulins, e.g. the α' -chain of β -conglycinin and V2, the sequence linking N-terminal and C-terminal domains in 7S and 11S globulin subunits, respectively. In 11S globulin subunits this region includes the hypervariable segment at the C-terminus of the α -chain and the α/β -chain cleavage site.

The susceptibility of both types of the sites mentioned above has already been observed *in vitro* as well as *in vivo* (Table 3) during the molecular maturation of 11S and 7S globulin precursor polypeptides [2, 23–27] (and Jung et al., unpublished results). Circumstantial evidence from the literature supports this conclusion (see reviews [5, 10, 12, 28]). Obviously, the segments V1 and V2 as well as the inserts I1 correspond to regions exposed on the surface of the trimeric 7S and hexameric 11S holoproteins, respectively.

In the light of predicted structural similarity between 7S and 11S globulin subunits [7], we drafted a simplified model of their tertiary structures (Fig. 4B). This is in accordance with the triplication model of storage globulin evolution [8], i.e. in conformity with cross-wise relations between N-terminal and C-terminal domains of 7S and 11S globulins. The structural resemblance between the I1 inserts of 7S and 11S globulin subunits is obvious. In contrast, the structural function of the extended insert I2 of 11S globulin subunits cannot be derived from our scheme. However, it is highly probable that this segment, which is both hydrophilic and susceptible to proteolytic cleavage, like insert I1 forms a loop on the surface of the hexameric 11S globulin holoprotein.

In the X-ray structure of 7S globulin subunits, the precise conformation of the interdomain linker has remained unclear. It only can be depicted on the basis of its conjunctive function. The V2 segments are susceptible for cleavage and occupy similar positions in the rotation symmetry of 7S and 11S globulin subunits (Fig. 4B). At least the part of the N-terminal extension of the 70-kDa subunit of 7S globulins that forms the susceptible segment V1 in our scheme can be positioned like the 11S globulin subunit interdomain linker. The similarity of the primary structures of the segments V1 and V2 of 7S and 11S globulin subunits, respectively (Fig. 3), favours the idea of them sharing some secondary and higher levels of structural similarity.

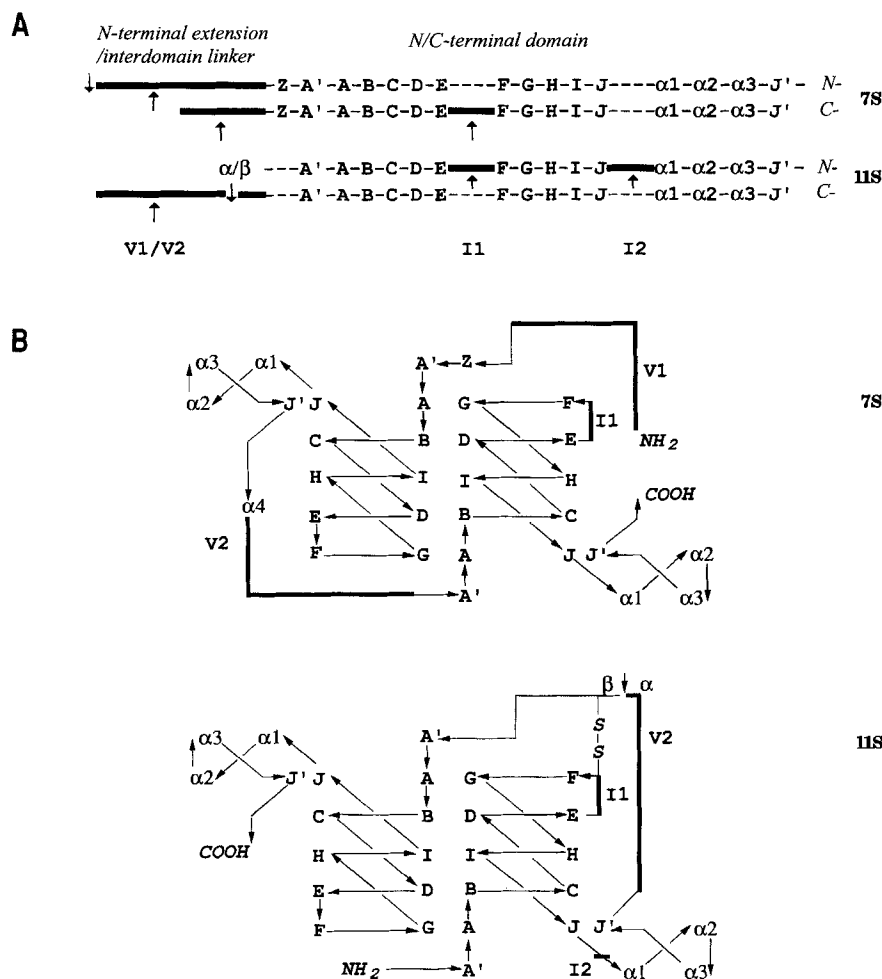


Fig. 4. The sites of 7S and 11S globulin subunits that are susceptible to limited proteolysis. (A) Schematic presentation of susceptible sites projected into aligned sequences of 7S and 11S globulin domains [8] and V1/V2 segments (see Fig. 1). Cleavage points of the susceptible sites V1/V2, I1 and I2 are documented in Fig. 1 and in the Tables 1 and 3. (B) A simplified scheme of storage globulin structures based on the structure models of the phaseolin and canavalin subunits [5, 7] (a view along pseudo-dyad axis). The putative 11S globulin structure is shown in accordance with the triplication scheme of storage globulin evolution [8]. The sites susceptible to limited proteolysis are highlighted by a thick solid line.

Table 3. Cleavage points inside susceptible sites V1/2 and I1 of 11S and 7S globulin subunits. The N-terminal amino acid residue in P1 position of a cleaved peptide bond is given. Sequence numbering corresponds to that of the EMBL Data Library. V1, the N-terminal extension of 70-kDa vicilin subunits. V2, sequence segment linking 7S/11S globulin subunit N- and C-terminal domains. I1, the insert specific of EF loop of vicilin C-terminal and legumin N-terminal domains. P.W., the data of the present work. The assignment of detected cleavage points to V1, V2 and I1 susceptible sites and their position relative to structurally ordered elements of storage globulins are based on the multiple alignment of the sequences of 11S/7S globulin domains [8] and on the vicilin-like model of legumin tertiary structure [7].

Protein	Enzyme	Susceptible site and cleavage point			Reference
		V1	V2	I1	
Glycinin					
A2B1a subunit	trypsin			R117	P.W.
	trypsin			S119	P.W.
A5A4B3 subunit	processing		N377	N120	[19]
β-Conglycinin					
α'-chain	processing	K62			[8]
α'-chain	trypsin	R204	R431	R517	P.W.
	trypsin		K432		P.W.
α-chain	C1	E138			[23]
α-chain	C1	E169			[23]
Pea vicilin	processing			N321	[10]
Cacao 7S globulin	processing			D436	[28]

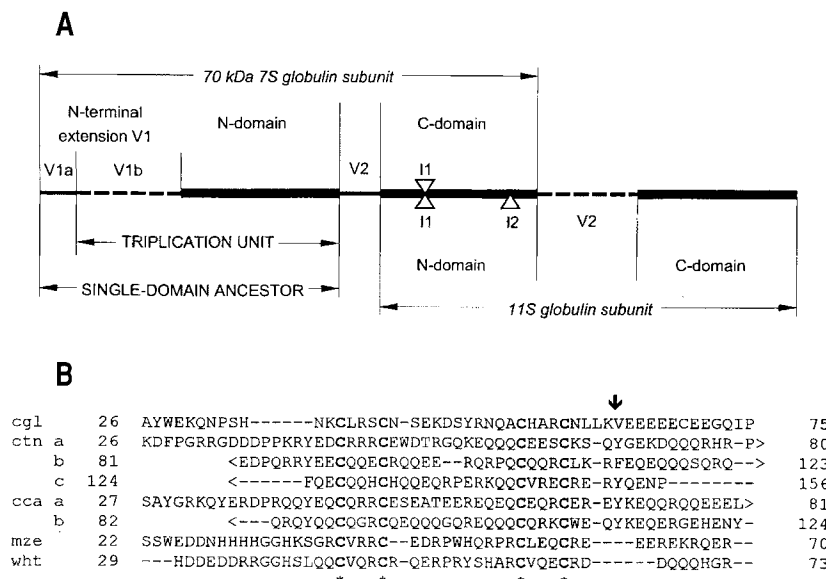


Fig. 5. Relationships between the N-terminal extension V1 of 70-kDa 7S globulin and 11S globulin subunits. (A) A hypothetical scheme of the composition of a putative three-domain ancestor according to the triplication model of storage globulin evolution [8]. An ancient single-domain ancestor was composed of V1 sequence consisting of V1a and V1b segments and the domain segment. The unit composed of V1b and domain segments was triplicated, thus generating the three-domain ancestor common to 7S and 11S globulins. During the evolution of 7S globulins portions of the triplicated ancestor were lost but the N-terminal and central triplication units as well as the V1a segment were retained. On the other hand, the 11S globulins retained only the central and the C-terminal triplication units. According to this scheme, I1 was inserted into the central triplication linker of 11S globulins retained sequence similarity (dotted line). The V2 interdomain linker of 7S globulin (solid line) was truncated during evolution and lost any reliable indication of sequence similarity of either the V1b segment of 7S globulins or the V2 segment of 11S globulins. (B) Sequence similarity between V1a segments of the 7S globulin N-terminal extensions. cgl, β -conglycinin α' -chain; ctn, *Gossypium hirsutum* vicilin; cca, *Theobroma cacao* vicilin; mze, *Zea mays* vicilin; wht, *Triticum aestivum* vicilin. During the multiple alignment the sequence repeats have been regarded as additional independent sequences. Brackets (<>) indicate limits of the internal repeats of ctn and cca sequences. Cys residues of CXXXC motifs are indicated by asterisks. The processing site of β -conglycinin α' -chain [24] is labelled by an arrow.

The sequence similarity between these segments could be either a reflection of some shared structural constraint or of their common origin from the just mentioned N-terminal V segment predicted for the common single-domain ancestor [8]. However, the structural impact of the highly variable N-terminal extension segment in 70-kDa subunits of 7S globulins cannot be strong. On the other hand, the interdomain linker of 11S globulin subunits exhibits a high degree of variability as well. This makes it difficult to evaluate the probability of common ancestral origin of V1 and V2 segments of 7S and 11S globulin subunits, respectively. Nevertheless, a certain sequence similarity is recognizable among the interdomain linkers of 11S globulin subunits from even distantly related dicotyledons (for instance, see the multiple alignment described by Lawrence et al. [7]). Thus, it is justified to assume an evolutionary relation between the β -conglycinin V1 segment and the interdomain linkers not only of soybean glycinin and pea legumin but also of other 11S globulin subunits.

The situation with the hypervariable N-terminal extensions of 7S globulins, including that of β -conglycinin, appears to be more complex. These segments were regarded initially as a result of independent recent insertions in the 70-kDa 7S globulins of different species, but its origin from an ancient ancestor common to all of 7S globulins seems more probable [29].

The N-terminal extension V1 of 7S globulins can be divided into distinct N-terminal and C-terminal parts, V1a and V1b segments, respectively, as shown at Fig. 5A. In some cases either only the V1a segment (the subunits of pea convicilin and β -conglycinin α' -chain) or both the V1a and the V1b segments (50-kDa 7S globulin subunits) are missing either in genes or in mature protein (Fig. 1, B1). In spite of their variability, at least

the V1a segments of the N-terminal extension even of distantly related 7S globulins exhibit similarity including the segment of 70-kDa chains of β -conglycinin that is detached during maturation of its propeptide (Fig. 5B). These related V1a segments are composed of one, two or three sequence repeats containing two CXXXC motifs. Remarkably, the same two CXXXC motifs are characteristic of the N-terminal region of 62-kDa sucrose binding protein (SBP) from soybean cotyledons and leaves [30]. This two-domain protein obviously does not have any storage function but it displays an extremely high degree of similarity to the entire sequences of all known 7S storage globulins (22.6 SD from random for SBP/cacao 7S globulin comparison [31]). Undoubtedly, SBP is an additional member of 7S/11S globulin superfamily. Since SBP is involved in such a basic process as sucrose transport [32], it seems likely that its evolutionary origin is even more deeply rooted than that of a separate storage 7S globulin family. Multiple alignment of the amino acid sequences of only their N-terminal extensions, including both V1a and V1b segments (data not shown) reveal similarity between SBP and several 7S globulins. This similarity between SBP and at least the β -conglycinin α' -chain and the cacao 7S globulin subunit was found to be statistically significant (8.1 SD and 7.6 SD from random, respectively), as well as that between α' -chain and cacao 7S globulin (6.5 SD from random). Evidently, both the V1a and V1b segments of 7S globulin extensions like the domains of both 7S and 11S globulins are rooted in the presumed common ancestral polypeptide. Possibly, the formation of those highly divergent segments which are dissimilar to the V1b segment of β -conglycinin and 11S globulin interdomain linkers took place after the evolution of 7S and 11S globulin subunits had already begun to diverge.

There are no indications of even a vestigial sequence similarity between specific V1a segments and either the C-terminal half of 7S globulins or both halves of 11S globulins. According to the triplication model of globulin subunit evolution [8], separate 7S and 11S globulin ancestors evolved from a putative three-domain ancestor as shown in Fig. 5A. It can be assumed that the V1a part of the V segment of the ancient single-domain ancestor did not participate in the triplication but was only retained in the separate 7S globulin ancestor.

A. D. Shutov and I. A. Kakhovskaya thank the *Deutsche Forschungsgemeinschaft* (grants 436 Mol 17/2/95 and Mu 925/4-1) and the *Deutscher Akademischer Austauschdienst* for financial support. In addition the authors want to express their immense gratitude to Prof McPherson (Riverside, USA) and Dr Reichelt for permission to reproduce figures from their publications.

REFERENCES

- 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.
- Müntz, K., Jung, R. & Saalbach, G. (1993) Synthesis, processing, and targeting of legume seed proteins, *Proc. Phytochem. Soc. Eur.* **35**, 128–146.
- Nielsen, N. C., Nam, Y.-W., Beaman, T., Oliviera, L. O. & Bassiner, R. (1995) Synthesis and assembly of 11S globulins, *J. Plant Physiol.* **145**, 641–647.
- Dickinson, C. D., Hussein, E. H. A. & Nielsen, N. C. (1989) Role of posttranslational cleavage in glycinin assembly, *Plant Cell* **1**, 459–469.
- 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.
- 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.
- 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.
- Shutov, A. D., Kakhovskaya, I. A., Braun, H., Bäumlein, H. & Müntz, K. (1995) Legumin and vicilin-like seed storage proteins: evidence for a common single-domain ancestral gene, *J. Mol. Evol.* **41**, 1057–1069.
- Bäumlein, H., Braun, H., Kakhovskaya, I. A. & Shutov, A. D. (1995) Seed storage proteins of spermatophytes share a common ancestor with desiccation proteins of fungi, *J. Mol. Evol.* **41**, 1070–1075.
- Shutov, A. D. & Vaintraub, I. A. (1987) Degradation of storage proteins in germinating seeds, *Phytochemistry* **26**, 1557–1566.
- Kamata, Y., Otsuka, S., Sato, M. & Shibasaki, K. (1982) Limited proteolysis of soybean beta-conglycinin, *Agric. Biol. Chem.* **46**, 2829–2834.
- Shutov, A. D., Senyuk, V. I., Kakhovskaya, I. A. & Pineda, J. (1993) High molecular mass products of hydrolysis of soybean glycinin by trypsin, *Biochemistry Engl. Transl. Biokhimiya* **58**, 174–182.
- Thanh, V. H. & Shibasaki, K. (1976) Major proteins of soybean seeds. A straightforward fractionation and their characterization, *J. Agric. Food Chem.* **24**, 1117–1121.
- Kitamura, K., Takagi, T. & Shibasaki, K. (1977) Renaturation of soybean 11S globulin, *Agric. Biol. Chem.* **41**, 833–840.
- Staswick, P. E., Hermodson, M. A. & Nielsen, N. C. (1981) Identification of the acidic and basic subunit complexes of glycinin, *J. Biol. Chem.* **256**, 8752–8755.
- Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* **227**, 680–685.
- Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) Establishing homologies in protein sequences, *Methods Enzymol.* **91**, 524–545.
- Adler, K. & Horstmann, C. (1984) Electron microscopic investigations of reconstituted products of isolated subunit pairs of the plant storage protein legumin from *Vicia faba*, *Kulturpflanze* **32**, S227–S228.
- Reichelt, R., Schwenke, K. D., König, T., Pähitz, W. & Wangermann, G. (1980) Electron microscopic studies for estimation of the quaternary structure of the 11S globulin (Helianthinin) from sunflower seed (*Helianthus annuus* L.), *Biochem. Physiol. Pflanzen* **175**, 653–663.
- Staswick, P. E., Hermodson, M. A. & Nielsen, N. C. (1984) The amino acid sequence of the A₂B_{1a} subunit of glycinin, *J. Biol. Chem.* **259**, 13424–13430.
- Domoney, C., Barker, D. & Casey, R. (1986) The complete deduced amino acid sequences of legumin β -polypeptides from different genetic loci in *Pisum*, *Plant Mol. Biol.* **7**, 467–474.
- Dayhoff, M. O. (1978) A model of evolutionary change in proteins: Matrices for detecting distant relationships, in *Atlas of protein sequences and structures* (Dayhoff, M. O., ed.) vol. 5, suppl. 3, pp. 345–358, Academic Press, New York.
- Momma, T., Negoro, T., Hirano, H., Matsumoto, A., Uda, K. & Fukazawa, C. (1985) Glycinin A5A4B3 mRNA: cDNA cloning and nucleotide sequencing of a splitting storage protein subunit of soybean, *Eur. J. Biochem.* **149**, 491–496.
- Doyle, J. J., Schuler, M. A., Godette, W. D., Zenger, V., Beachy, R. N. & Slightom, J. L. (1986) The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. Structural homologies of genes and proteins, *J. Biol. Chem.* **261**, 9228–9238.
- Qi, X., Chen, R., Wilson, K. A. & Tan-Wilson, A. L. (1994) Characterization of a soybean β -conglycinin-degrading protease cleavage site, *Plant Physiol.* **104**, 127–133.
- Gatehouse, J. A., Lycett, G. W., Delaney, A. J., Croy, R. R. D. & Boulter, D. (1983) Sequence specificity of the post-translational proteolytic cleavage of vicilin, a seed storage protein of pea (*Pisum sativum* L.), *Biochem. J.* **212**, 427–432.
- Scott, M. P., Jung, R., Müntz, K. & Nielsen, N. C. (1992) A protease responsible for post-translational cleavage of a conserved Asn-Gly linkage in glycinin, the major seed protein in soybean, *Proc. Natl Acad. Sci. USA* **89**, 658–662.
- Müntz, K., Bassiner, R., Lichtenfeld, C., Scholz, G. & Weber, E. (1985) Proteolytic cleavage of storage proteins during embryogenesis and germination of legume seeds, *Physiol. Vég.* **23**, 75–94.
- Dure, L. III (1990) An unstable domain of the vicilin genes of higher plants, *New Biol.* **2**, 487–493.
- Grimes, H. D., Overvoorde, P. J., Ripp, K., Franceschi, R. & Hitz, W. D. (1992) A 62-kDa sucrose binding protein is expressed and localized in tissues actively engaged in sucrose transport, *Plant Cell* **4**, 1561–1574.
- Braun, H., Czihal, A., Shutov, A. D. & Bäumlein, H. (1996) A vicilin-like seed protein of cycads: Similarity to sucrose binding proteins, *Plant Mol. Biol.*, in the press.
- Overvoorde, P. J., Frommer, W. B. & Grimes, H. D. (1996) A soybean sucrose binding protein independently mediates nonsaturable sucrose uptake in yeast, *Plant Cell* **8**, 271–280.