Plant Mol Biol (2006) 62:111–125
DOI 10.1007/s11103-006-9007-0

The C-terminal region of α′ subunit of soybean β-conglycinin contains two types of vacuolar sorting determinants

Keito Nishizawa · Nobuyuki Maruyama · Shigeru Utsumi

Received: 10 July 2005 / Accepted: 21 April 2006 / Published online: 10 August 2006

Abstract In maturing seed cells, proteins that accumulate in the protein storage vacuoles (PSVs) are synthesized on the endoplasmic reticulum (ER) and transported by vesicles to the PSVs. Vacuolar sorting determinants (VSDs) which are usually amino acid sequences of short or moderate length direct the proteins to this pathway. VSDs identified so far are classified into two types: sequence specific VSDs (ssVSDs) and C-terminal VSDs (ctVSDs). We previously demonstrated that VSDs of α′ and β subunits of β-conglycinin, one of major storage proteins of soybean (*Glycine max*), reside in the C-terminal ten amino acids. Here we show that both types of VSDs coexist within this region of the α′ subunit. Although ctVSDs can function only at the very C-termini of proteins, the C-terminal ten amino acids of α′ subunit directed green fluorescent protein (GFP) to the PSVs even when they were placed at the N-terminus of GFP, indicating that an ssVSD resides in the sequence. By mutation analysis, it was found that the core sequence of the ssVSD is Ser-Ile-Leu (fifth to seventh residues counted from the C-terminus) which is conserved in the α and β subunits and some vicilin-like proteins. On the other hand, the sequence composed of the C-terminal three amino acids (AFY) directed GFP to the PSVs when it was placed at the C-terminus of GFP, though the function as a VSD was disrupted at the N-terminus of GFP, indicating that the AFY sequence is a ctVSD.

K. Nishizawa · N. Maruyama · S. Utsumi (✉)
Laboratory of Food Quality Design and Development, Graduate School of Agriculture, Kyoto University, Uji, Kyoto 611-0011, Japan
e-mail: sutsumi@kais.kyoto-u.ac.jp

Keywords β-Conglycinin · Protein sorting · Protein storage vacuole · Secretory pathway · Soybean · Vacuolar sorting determinant

Introduction

In plant cells there are at least two types of vacuoles: the lytic vacuoles (LV) and protein storage vacuoles (PSV). The LVs carrying the γ-tonoplast-intrinsic protein (TIP) in their tonoplast contain proteases active at low pH (Jauh et al. 1999). The PSVs in seed cells carrying both α- and δ-TIPs in their tonoplast contain storage proteins, defense proteins, proteases and the other enzymes (Herman and Larkins 1999; Jauh et al. 1999; Jiang et al. 2001). In pea cotyledon cells, the LVs disappear during development and become replaced by the PSVs (Hoh et al. 1995). Similarly, in soybean, during development, preexisting large central vacuoles (maybe the LVs) are replaced by numerous rounded PSVs whose tonoplasts contain α-TIP (Melroy and Herman 1991). Proteins destined for the LVs are transported by clathrin-coated vesicles from the Golgi apparatus, while proteins destined for the PSVs are transported by dense vesicles (Vitale and Raikhel 1999). It has been suggested that sorting of proteins to the respective vesicles occurs in the Golgi apparatus (Hillmer et al. 2001; Hinz et al. 1999; Hohl et al. 1996). In seeds of some plants such as soybean and pumpkin, storage proteins are also transported from the ER to the PSVs by ER-derived protein bodies or precursor-accumulating (PAC) vesicles bypassing the Golgi apparatus (Hara-Nishimura et al. 1998; Kinney et al. 2001; Mori et al. 2004).

A specific vacuolar sorting receptor (VSR) working in the pathway to the LVs has been found in pea (Kirsch et al. 1994; Paris et al. 1997). Two receptors that work in the pathway to the PSVs during seed maturation have been also identified (Shimada et al. 1997, 2003). One is PV72 probably working for recruiting storage proteins from the Golgi apparatus to PAC vesicles in pumpkin (Shimada et al. 1997, 2002). Another is VSR1 of *Arabidopsis thaliana* (AtVSR1), also referred to as AtELP. It was shown *in planta* that AtVSR1 functions to deliver storage proteins to the PSVs (Shimada et al. 2003). All of them are homologous proteins. However, it is still unclear if their ligand recognition-mechanisms are similar or not.

Secretory proteins destined to plant vacuoles must carry determinants for vacuolar sorting, termed vacuolar sorting determinants (VSDs). VSDs identified so far are classified into two types: sequence specific VSDs (ssVSDs) and C-terminal VSDs (ctVSDs). ssVSDs contain either essential Ile or Leu, and can be located in N-or C-terminal propeptides as well as within mature proteins. ctVSDs are present in C-terminal regions of polypeptides and often rich in hydrophobic amino acids, however, their length and sequences are very variable and neither Ile nor Leu is essential (Matsuoka and Neuhaus 1999; Neuhaus and Rogers 1998). VSDs of various PSV proteins such as barley lectin (Bednarek et al. 1990), common bean phaseolin (Frigerio et al. 1998a), Brazil nut 2S albumin (Saalbach et al. 1996), castor bean ricin (Frigerio et al. 1998b) and castor bean 2S albumin (Brown et al. 2003) have been reported. As for ricin, the VSD (LLIRP) is considered an ssVSD, since it is located in the internal propeptide, and the Ile, which is conserved in the internal propeptide regions of proricin and proricin-related proteins, is very important for vacuolar sorting and cannot be replaced by glycine (Frigerio et al. 1998b, 2001). In addition, the VSD of castor bean 2S albumin (STGEEVLRMPGDEN) is also considered an ssVSD, since it is located in the internal propeptide and Leu in the sequence cannot be replaced by glycine (Brown et al. 2003). On the other hand, the VSD of barley lectin is considered a ctVSD, since it functions only at the very C-terminus (Dombrowski et al. 1993). The VSD of phaseolin is also considered a ctVSD, since this VSD contains neither Ile nor Leu (Frigerio et al. 1998a).

Soybean β-conglycinin (7S globulin), one of the major storage proteins stored in the PSVs of seeds, is a trimeric protein composed of α (~67 kDa), α' (~71 kDa) and β (~50 kDa) subunits (referred to as α, α' and β, respectively). These are synthesized on the ER, assembled into trimers in the ER, and transported to the PSVs by the dense vesicles (Mori et al. 2004). We previously showed that the VSDs of α' and β exist in the C-terminal ten amino acids (Nishizawa et al. 2003, 2004). In this study, we examined the characteristics of these VSDs by transient expression assay using soybean-maturing seeds, and found that the C-terminal region of α' contains both an ssVSD and a ctVSD.

Materials and methods

Plant materials

Cotyledons at mid- to late-maturation stage of soybean (cv. Shirotsurunoko) were used in all experiments. Soybean was grown at 25°C under white fluorescent light with a day/night rhythm of 16 h light and 8 h dark.

Plasmid construction

Construction of plasmids for signal peptide-GFP or spGFP (pBSspGFP) and signal peptide-C-terminally blocked GFP or spcbGFP (pBSspcbGFP, previously denoted as pBSspmGFP) was previously reported (Nishizawa et al. 2003).

To construct the plasmid for cbGFP-α'CT6 (α'UTR), we performed PCR using pBSGFP-CT24 (Nishizawa et al. 2003), which is composed of pBluescript and the DNA insert, consisting of the promoter and the signal peptide-coding region of α', coding region of GFP, the sequence coding the C-terminal 24 amino acids of α', the stop codon and the 3′ untranslated region of α', as a template and the following primers: primer 1 (5′-**TCCTCCTCCTCC**CTTGTACAGCTCGTCCAT-3′, boldface indicates codons for glycines introduced at the C-terminus of GFP) and 2 (5′-ATTTTGAGGGCTTTTTAC*TGA*ATAAGTATGTAG-3′, italics indicates the stop codon). The amplified fragment was self-ligated to construct pBScbGFP-α'CT6 (α'UTR). To exchange the 3′ untranslated region of α' behind the stop codon with the nopaline synthase (nos) terminator, a DNA fragment from a *Bst*YI site (+572 from the initiation codon) in the GFP sequence to the stop codon was amplified using the following primers: primer 3 (5′-GGTGAACTTCAAGATCCGCCAC-3′, underline indicates a *Bst*YI site) and primer 4 (5′-ATAGTTTAGCGGCCGC*TCA*GTAAAAAGCCCT-3′; underline indicates an introduced *Not*I site, and italics indicate the stop codon). Since several *Bst*YI sites exist in the region other than multi-cloning site of pBluescript, a fragment from a *Sac*I site which resides upstream of the promoter to the *Bst*YI site in the GFP

sequence was prepared by restriction digestion of a mutated pBSGFP-CT24 in which *Bam*HI and *Not*I sites upstream of the promoter had been previously eliminated. Then this fragment and the fragment amplified by PCR were inserted together into pBSmGFP-CT24 (Nishizawa et al. 2003), which contains a *Sac*I site upstream of the promoter and a *Not*I site between the stop codon and the nos terminator, with *Sac*I/*Not*I sites to construct pBScbGFP-α′CT6. The *Sac*I-*Bst*YI fragment was used afterwards in the same way for the same reason. This process was followed in the construction of other plasmids and described as follows: fragments amplified by PCR were inserted into appropriate plasmids (pBSmGFP-CT24 in this case) with *Bst*YI (in GFP sequence)/appropriate restriction sites (in this case *Not*I) to construct appropriate plasmids (in this case pBScbGFP-α′CT6).

To construct pBScbGFP-PLSS and pBScbGFP-α′CT10-2G, we used pBSmGFP-CT10 (Nishizawa et al. 2003) as a template and the following primers: primer 3 and primer 5 (5′-ATAGTTTAGCGGCCGC*TT*-*A*TGAAGACAAAGGTCCTCC; underline indicates an introduced *Not*I site, and italics indicate the stop codon) for the former or primer 6 (5′-GCAG-CCGGGCGGCCGC*TTA***TCCTCC**GTAAAAAGCC-CTCAAAAT-3′; underline indicates a *Not*I site introduced behind the stop codon indicated in italics, boldface indicates codons for introduced glycines) for the latter. The amplified fragments were inserted into pBSmGFP-CT24 (Nishizawa et al. 2003) with *Bst*YI/*Not*I sites.

To construct the plasmids for cbGFP-α′CT10-6G, we used pBScbGFP-α′CT10-2G as a template and the following primers: primer 3 and primer 7 (5′-GCAGCCGGGCGGCCGC*TTA***TCCTCCTCCTCC**-TCCTCCGTAAAAAGCCCT-3′; underline indicates a *Not*I site introduced behind the stop codon indicated in italics, boldface indicates codons for newly introduced glycines). The amplified fragments were inserted into pBSmGFP-CT24 (Nishizawa et al. 2003) with *Bst*YI/*Not*I sites to construct pBScbGFP-α′CT10-6G.

To construct the plasmid for cbGFP-α′CT6-2G, we used pBScbGFP-α′CT6 as a template and the primers 3 and 6. The amplified fragments were inserted into pBSmGFP-CT24 (Nishizawa et al. 2003) with *Bst*YI/*Not*I sites to construct pBScbGFP-α′CT6-2G.

To construct the plasmid for cbGFP-α′CT6-6G, we used pBScbGFP-α′CT6-2G as a template and the primers 3 and 7. The amplified fragments were inserted into pBSmGFP-CT24 (Nishizawa et al. 2003) with *Bst*YI/*Not*I sites to construct pBScbGFP-α′CT6-6G.

To construct the plasmid for cbGFP-PLSS-2G, we used pBScbGFP-PLSS as a template and the following primers: primer 3 and primer 8 (5′-GCAGCCG-GGCGGCCGC*TTA***TCCTCC**TGAAGACAAAGG-3′; underline indicates a *Not*I site introduced behind the stop codon indicated in italics, boldface indicates codons for introduced glycines). The amplified fragments were inserted into pBSmGFP-CT24 (Nishizawa et al. 2003) with *Bst*YI/*Not*I sites to construct pBScbGFP-PLSS-2G.

To construct the plasmid for cbGFP-PLSS-6G, we used pBScbGFP-PLSS-2G as a template and the following primers: primer 3 and primer 9 (5′-GCAGCCGGGCGGCCGC*TTA***TCCTCCTCCTCC**-TCCTCCTGAAGACAAAGG-3′; underline indicates a *Not*I site introduced behind the stop codon indicated in italics, boldface indicates codons for newly introduced glycines). The amplified fragments were inserted into pBSmGFP-CT24 (Nishizawa et al. 2003) with *Bst*YI/*Not*I sites to construct pBScbGFP-PLSS-6G.

To construct the plasmids for cbGFP-α′CT10Δ1-6G, cbGFP-α′CT10Δ2-6G, cbGFP-α′CT10Δ3-6G, cbGFP-α′CT10Δ4-6G and cbGFP-α′CT10Δ5-6G, the DNA fragment between the *Bst*YI site in the GFP coding region and a *Kpn*I site downstream of the nos terminator in the vector region of pBScbGFP-α′CT10-6G was replaced by the following fragments. We prepared a DNA fragment containing a sequence coding six glycines, the stop codon, the nos terminator and the vector region up to the *Kpn*I site by PCR using pBScbGFP-α′CT10-6G as a template and the following primers: primer 10 (5′-**GGAGGAGGAGGAGGAG-GA***TAA*GCGGCCGCCCGGCT-3′; boldface indicates the sequence coding six glycines and italics indicate the stop codon) and T7 promoter primer. We also prepared a DNA fragment from the *Bst*YI site in the GFP coding region to the region coding the truncated C-terminus of the α′ by PCR using pBScbGFP-α′CT10-6G as a template and the following primers: primer 3 and each of primer 11 (5′-AAAAG-CCCTCAAAATTGAAGACAA-3′) for cbGFP-α′CT10Δ1-6G, primer 12 (5′-AGCCCTCAAAATTG-AAGACAAAGG-3′) for cbGFP-α′CT10Δ2-6G, primer 13 (5′-CCTCAAAATTGAAGACAAAGG-3′) for cbGFP-α′CT10Δ3-6G, primer 14 (5′-CAAAA-TTGAAGACAAAGGTCC-3′) for cbGFP-α′CT10Δ4-6G or primer 15 (5′-AATTGAAGACAAAGGTCC-TCC-3′) for cbGFP-α′CT10Δ5-6G. Both fragments were inserted together into pBScbGFP-α′CT10-6G with *Bst*YI/*Kpn*I sites.

To construct the plasmids for cbGFP-P612G-6G, cbGFP-L613G-6G, cbGFP-S614G-6G, cbGFP-S615G-6G, cbGFP-I616G-6G, cbGFP-L617G-6G, cbGFP-R618G-6G and cbGFP-A619G-6G, the DNA fragment from the *Bst*YI site in the GFP coding region to the *Kpn*I site in the vector region of pBScbGFP-α′CT10-6G was replaced by the following fragments. We

prepared a DNA fragment from the *Bst*YI site in the GFP coding region to the sequence coding the four glycines at the end of the cbGFP coding region by PCR using pBScbGFP-α′CT10-6G as a template, and the primers 1 and 3. We also prepared a DNA fragment containing the region coding the mutated C-terminus of the α′, a sequence coding six glycines, the stop codon, the nos terminator and the vector region up to the *Kpn*I site by PCR using pBScbGFP-α′CT10-6G as a template and the following primers: T7 promoter primer and each of primer 16 (5′-**GGA**TTGTCTTCAATTTTGAGGGCT-3′, boldface indicates a codon for Gly substituted for Pro 612), primer 17 (5′-CCT**GGG**TCTTCAATTTTGAGGGCT-3′, boldface indicates a codon for Gly substituted for Leu 613), primer 18 (5′-CCTTTG**GGA**TCAATTTTGAGGGCT-3′, boldface indicates a codon for Gly substituted for Ser 614), primer 19 (5′-CCTTTGTCT**GGA**ATTTTGAGGGCTTTT-3′, boldface indicates a codon for Gly substituted for Ser 615), primer 20 (5′-CCTTTGTCTTCA**GGA**TTGAGGGCTTTT-3′, boldface indicates a codon for Gly substituted for Ile 616), primer 21 (5′-CCTTTGTCTTCAATT**GGA**AGGGCTTTT-3′, boldface indicates a codon for Gly substituted for Leu 617), primer 22 (5′-CCTTTGTCTTCAATTTTG**GGA**GCTTTTTACGGA-3′, boldface indicates a codon for Gly substituted for Arg 618) or primer 23 (5′-CCTTTGTCTTCAATTTTGAGG**GGA**TTTTACGGAGGA-3′, boldface indicates a codon for Gly substituted for Ala 619). Both fragments were inserted together into pBScbGFP-α′CT10-6G with *Bst*YI/*Kpn*I sites.

To construct pBScbGFP-βCT10-6G, we used pBSmGFP-βCT10 (Nishizawa et al. 2004) as a template and the following primers: primer 3 and primer 24 GCAGCCGGGCGGCCGC*TTA***TCCTCCTCCTCCTCCTCC**GTAGAGAGCACCTAAGATTGA-3′; underline indicates a *Not*I site introduced behind the stop codon indicated in italics, boldface indicates sequence coding introduced Gly). The amplified fragments were inserted into pBSmGFP-βCT10 (Nishizawa et al. 2004) with *Bst*YI/*Not*I sites.

To construct the plasmid for α′CT10-cbGFP and I616G-cbGFP, the DNA region coding the C-teminal ten amino acids (PLSSILRAFY) or the mutated one (PLSSGLRAFY) were inserted between the DNA sequences coding the signal peptide of the α′ and GFP. We prepared a DNA fragment containing the promoter region, the signal peptide-coding region of α′ and the DNA sequence coding each of the first half of the C-terminal ten amino acids (PLSSI) or the mutated one (PLSSG) by PCR. As a template, pBSspcbGFP was used, which contains a *Sac*I site in the vector region followed by the promoter region, signal peptide-coding region of the α′, GFP-coding region, DNA coding four continuous glycines introduced prior to the stop codon, a *Not*I site and the nos terminator. The following primers were used: T3 promoter primer and each of primer 25 (5′-**AATTGAAGACAAAGG**AATGCCAAATGAGACAGA-3′, boldface indicates the sequence coding PLSSI of the C-terminal region of α′ introduced behind the signal peptide-coding region) for α′CT10-cbGFP or primer 26 (5′-**TCCTGAAGACAAAGG**AATGCCAAATGAGACAGA-3′, boldface indicates the sequence coding PLSSG of the mutated one) for I616G-cbGFP. We also prepared a DNA fragment containing the sequence coding the last half of the C-terminal ten amino acids of α′, the GFP-coding region, DNA sequence coding four continuous glycines and the stop codon by PCR using pBSspcbGFP as a template and the following primers: primer 27 (5′-**TTGAGGGCTTTTTAC**GTGAGCAAGGGCGAGGAG-3′, boldface indicates the DNA sequence coding LRAFY of the C-terminal region of α′ introduced prior to the GFP-coding sequence) and primer 28 (5′-GCAGCCGGGCGGCCGCT*TTA*TCCTCC-3′, underline indicates a *Not*I site introduced behind the stop codon indicated in italics). The amplified fragments were together inserted into pBSspcbGFP with *Sac*I/*Not*I sites.

To construct a plasmid for cbGFP-AFY, the DNA fragment between the *Bst*YI site in the GFP coding region and a *Kpn*I site downstream of the nos terminator in the vector region of pBScbGFP-α′CT6 was replaced by the following fragments. We prepared a DNA fragment containing a sequence coding AFY, the stop codon, the nos terminator and the vector region up to the *Kpn*I site by PCR using pBScbGFP-α′CT6 as a template and the following primers: primer 29 (5′-GCTTTTTAC*TAA*GCGGCCGCCCGGCT-3′; italics indicate the stop codon) and T7 promoter primer. We also prepared a DNA fragment from the *Bst*YI site in the GFP coding region to the region coding the continuous four glycines linked to the C-terminus of the GFP by PCR using pBScbGFP-α′CT6 as a template and following primers: primers 1 and 3. Both fragments were inserted together into pBScbGFP-α′CT6 with *Bst*YI/*Kpn*I sites.

To construct a plasmid for Aleu-cbGFP, we used pBSspcbGFP as a template and following primers: primer 30 (5′-**CCCGTCACCGACCGCGCCGCCTCGACG**GTGAGCAAGGGCGAGGAGCTGTTCACCGGG-3′, boldface indicates the DNA sequence coding PVTDRAAST that is the C-terminal half of the peptide from pro-aleurain, and underline indicates the GFP-coding sequence) and primer 31 (5′-**CCGGATCGGGTTGGAGTCGGCGAAGGAGGA**AATG-

CCAAATGAGACAGAAACTGATGCCAG-3′, boldface indicates the DNA sequence coding SSFADSNPIR that is the N-terminal half of the peptide from pro-aleurain, and underline indicates the coding sequence of signal peptide of α′ subunit). The amplified fragment was self-ligated to construct pBSAleu-cbGFP.

To construct a plasmid for cbGFP-Chit, we used pBSspcbGFP as a template and the following primers: primer 32 (5′-**GTCGATACAATG***TGA*GCGGCCGCCCGGCTGCAGATCGTTCAAAC-3′, boldface indicates the DNA sequence coding VDTM that is the C-terminal half of the peptide from the precursor of tobacco chitinase A, and italics indicate the stop codon) and primer 33 (5′-**AAGAAGACC**TCCTCCTCCTCCCTTGTACAGCTCGTCCATGCC-3′, boldface indicates the DNA sequence coding GLL that is the N-terminal half of the peptide from the precursor of tobacco chitinase A, and underline indicates the coding sequence of cbGFP). The amplified fragment was self-ligated to construct pBScbGFP-Chit.

The sequences of all PCR products were confirmed by DNA sequencing.

Transient expression assays

Transient expression assays were performed basically as described previously (Nishizawa et al. 2003, 2004). Briefly, the plasmids described above were introduced into maturing cotyledons of soybean by particle bombardment using Biolistic PDS-1000/He (Bio-Rad, Hercules, CA) as follows. Each bombardment used 750 μg of gold microcarriers (particle size of 1.0 μm) coated with 1.25 μg of the plasmids. Each sample was bombarded twice under the following conditions: vacuum, 26–28 in. Hg; target distance, 9 cm; and Helium pressure, 1,100 psi. After bombardment, cotyledons were incubated on the Murashige-Skoog agar plates at 25°C in the dark for 24–40 h. Fluorescent images were obtained using the MRC-1024 confocal laser-scanning microscope (Bio-Rad, Hercules, CA). GFP was excited with a laser wavelength of 488 nm and detected through a filter for wavelengths of 506–538 nm. We observed more than 20 cells with GFP fluorescence for each construct.

Results

Reevaluation of transient expression assay using maturing cotyledons of soybean

We previously showed that green fluorescent protein (GFP) with the signal peptide and the C-terminal sequence of α′ subunit of β-conglycinin (GFP-CT24) was transported to the PSVs in cotyledons of transgenic *Arabidopsis thaliana* (Nishizawa et al. 2003). We also confirmed that this GFP fusion protein was transported to the numerous rounded vacuoles (PSVs) in maturing soybean cotyledons in transient expression assay [unpublished data (but the fluorescent pattern was the same as that of mGFP-CT24 in which C-terminally blocked GFP was followed by the C-terminal sequence of α′, Nishizawa et al. 2003)]. In establishing the transient expression assay system, we used, at first, GFP with the signal peptide of α′ (spGFP, Fig. 1A) as a control of a secretory protein. However, in maturing cotyledons of soybean, spGFP was partially transported to the vacuoles as shown in Fig. 2A and B, though it was also secreted to the intercellular space (ICS). Two dot-like structures in Fig. 2A measuring ~2 μm are gold particles used in particle bombardment and only occasionally appeared under the condition used. Moreover, their fluorescent figures were always superimposed on the gold particles observed under transmitted light. There were cells, of course, in which spGFP was not transported to the vacuoles and was only secreted to the ICS (Figs. 2C, D; the dot near the ICS in Fig. 2C is also a gold particle used in particle bombardment). It is possible that the unexpected transport to the vacuoles was caused by the same mechanism as sorting by ctVSDs because the C-terminal sequence of GFP (MDELYK, underline indicates hydrophobic residues) is somewhat hydrophobic. It was reported that the ctVSD of barley lectin and the ctVSD of potato tuber 20 kDa protein lose their function when one or two glycines were linked to their C-termini (Dombrowski et al. 1993; Koide et al. 1999), although the mechanism of this inhibition by glycines at the C-termini remains to be elucidated. Both the C-terminal carboxylates of main chains and side chains of amino acids near the C-termini of ctVSDs might be recognized by receptors, and glycines at the C-termini might inhibit this recognition process. Similar to these reports, the unexpected transport of spGFP to the vacuoles was inhibited by adding four continuous glycines at its C-terminus (signal peptide-C-terminally blocked GFP or spcbGFP, which was referred to as spmGFP in our previous reports: Nishizawa et al. 2003, 2004; Fig. 1B), and this spcbGFP was efficiently secreted to the ICS (Fig. 2E, F; Nishizawa et al. 2003). We observed more than 20 cells with GFP fluorescence and found that the fluorescent pattern observed in these cells was the same as that observed when spGFP was efficiently secreted, indicating that the continuous glycines affected only the efficiency of

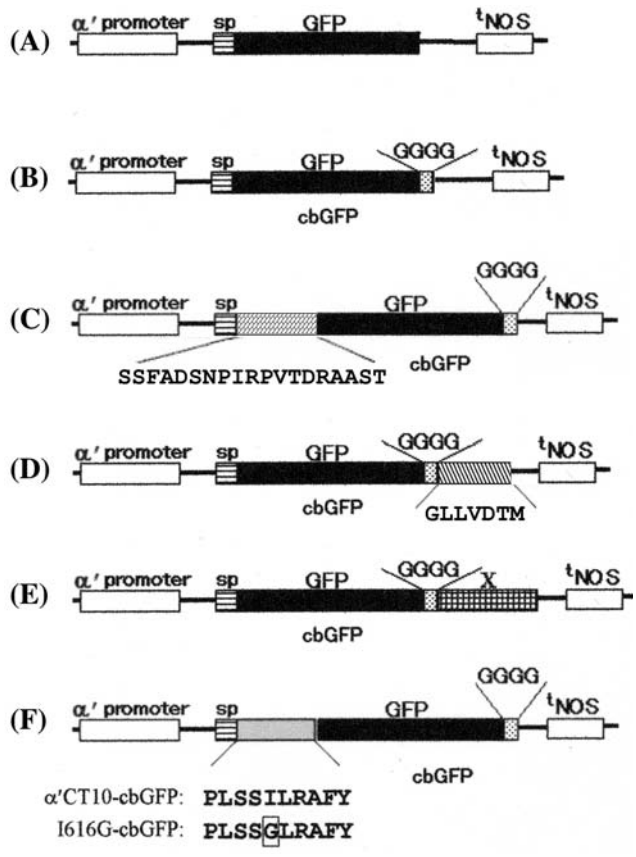

Fig. 1 Schematic diagrams of GFP derivatives used in this study. (**A**) Schematic diagram of spGFP. The signal peptide of α′ subunit of β-conglycinin was fused to the N-terminus of GFP. The genes coding the spGFP was driven by the promoter of α′. (**B**) Schematic diagram of spcbGFP. Four continuous glycines were fused to the C-terminus of spGFP. (**C**) Schematic diagram of Aleu-cbGFP. A portion of the propeptide of barley aleurain was fused to spcbGFP. (**D**) Schematic diagram of cbGFP-Chit. The C-terminal propeptide from tobacco chitinase A was fused to spcbGFP. (**E**) Schematic diagram of cbGFP-X mutants. Various sequences were fused to the C-terminus of spcbGFP. X represents mutated sequences listed in Table 1 derived from the C-terminal region of α′ or β. (**F**) Schematic diagrams of α′CT10-cbGFP and I616G-cbGFP. Either of the C-terminal ten amino acids of α′ or the mutated sequence in which Ile 616 was replaced by Gly was inserted between the signal peptide of α′ and cbGFP. sp, signal peptide of α′; tNOS, nopaline synthase terminator

secretion of GFP. Thus, we used cbGFP as a reporter protein in following experiments.

We examined if ssVSDs and ctVSDs can function on the pathway to the PSVs in soybean cotyledon cells. It was reported that the N-terminal propeptide of barley aleurain and the C-terminal propeptide of tobacco chitinase A contain an ssVSD and a ctVSD, respectively (Holwerda et al. 1992; Neuhaus et al. 1991). Plasmids containing the sequence coding spcbGFP fused to a portion of the propeptide (SSFADSNPIRPVTDRAAST, underline indicates the well-known motif of ssVSD) of barley aleurain (Aleu-cbGFP, Fig. 1C), and spcbGFP fused to a C-terminal propeptide (GLLVDTM) with ctVSD from tobacco chitinase A (cbGFP-Chit, Fig. 1D) were introduced into maturing soybean cotyledons by particle bombardment. Both genes were driven by the promoter of α′ to examine their transport under almost the same conditions in which α′ is transported during seed maturation. We observed more than 20 cells with GFP fluorescence for each construct, and found that both Aleu-cbGFP and cbGFP-Chit were transported to the numerous rounded vacuoles (diameter 2–10 μm) and were never secreted to the ICS (Fig. 2G, H). Several small fluorescent dots were sometimes observed in these cells. We could not identify them, but they might be the prevacuolar compartments. In soybean, during development, preexisting large central vacuoles (maybe the LVs) disappear and are replaced by numerous rounded PSVs accumulating storage proteins whose tonoplasts contain α-TIP (Melroy and Herman 1991). Therefore it is reasonable that the vacuoles where Aleu-cbGFP and cbGFP-Chit were transported are considered the PSVs based on their appearance. Thus the results indicated that both ssVSDs and ctVSDs can function on the pathway to the PSVs in soybean cotyledon cells.

In all the experiments shown below, we used this transient expression system using cbGFP as a reporter protein. We observed more than 20 cells with GFP fluorescence for each construct, and found that GFP fusions which were detected mainly in the PSVs were never detected in the ICS among all the observed cells, and GFP fusions which were detected in the ICS were never detected in the PSVs among all the observed cells unless specified. We have shown a typical fluorescent pattern for each construct in 'Figures'.

A partial sequence of the C-terminal region of α′ is enough to act as a VSD

We previously showed that the VSDs for the PSV matrix of α′ and β reside in the C-terminal ten amino acids (Table 1, underlines; Nishizawa et al. 2003, 2004), which are disordered in the surface of the molecules (Maruyama et al. 2001 2004). In these studies, the derivatives lacking the C-terminal ten amino acids of α′ and β (α′/βΔCT10) were secreted to the ICS in transgenic *Arabidopsis* seeds. In addition, it was elucidated by the transient expression system that the C-terminal ten amino acids of α′ and β were sufficient to direct spcbGFP (previously referred to as spmGFP) to the PSVs in maturing soybean cotyledons when they were fused to the C-terminus of spcbGFP (mGFP-CT10 and

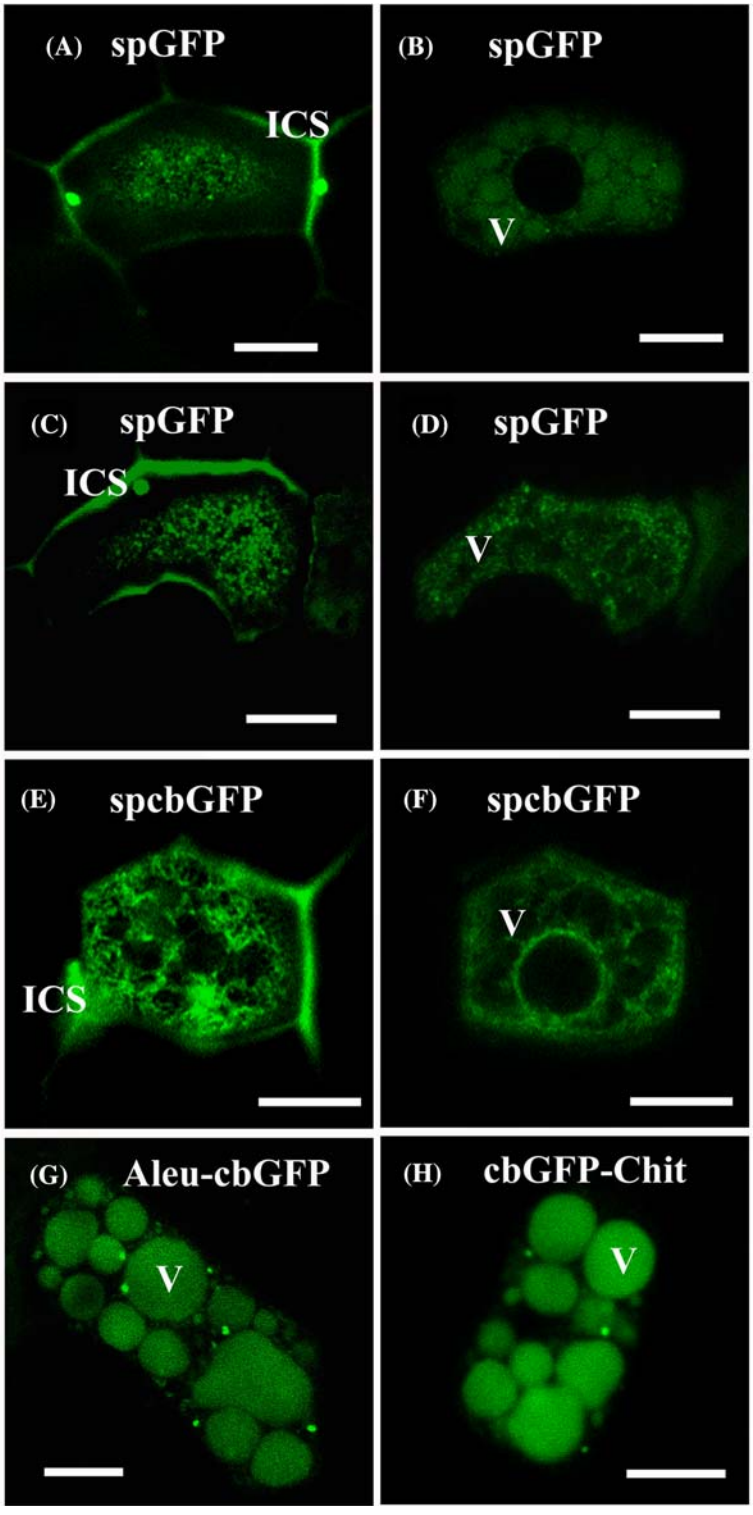

Fig. 2 Efficient secretion of spcbGFP compared to spGFP, and efficient transport of two standard GFP fusion proteins to the PSVs in cotyledons of maturing soybean. (**A**) Partial secretion of spGFP to the ICS. (**B**) The different section of the cell shown in (**A**). spGFP was also transported to the PSVs. Unexpected transportation to the PSVs was sometimes observed. (**C**) Secretion of spGFP to the ICS. In some cells, spGFP was mainly secreted to the ICS and was not transported to the PSVs. (**D**) The different section of the cell shown in (**C**). (**E**) Efficient secretion of spcbGFP to the ICS. (**F**) The different section of the cell shown in (**E**). spcbGFP was not transported to the PSVs. (**G**) Transportation of Aleu-cbGFP with one of the standard ssVSDs to the PSVs. (**H**) Transportation of cbGFP-Chit with one of the standard ctVSDs to the PSVs. V, PSV; ICS, intercellular space. Bars = 10 μm

mGFP-βCT10 in Nishizawa et al. 2003 2004; Table 1). Moreover, both the N-terminal portion containing four amino acids (PFPS) and the C-terminal portion containing six amino acids (ILGALY) of the C-terminal ten amino acids of β were sufficient to direct spcbGFP to the PSVs in maturing soybean cotyledons.

In this study, we examined if the N-terminal portion containing four amino acids (PLSS) and the C-terminal portion containing six (ILRAFY i.e., α'CT6) and three (AFY) amino acids, respectively, of the C-terminal ten amino acids of α' direct spcbGFP to the PSVs in soybean cotyledon cells. Plasmids containing the

Table 1 cbGFP-X mutants used in the study and their distribution

Name	X[a]	PSV[b]	ICS[b]
cbGFP-α′CT10[c]	^{612}PLSSILRAFY^{621}	++[c]	
cbGFP-α′CT10-2G	PLSSILRAFY*GG*	++	
cbGFP-α′CT10-6G	PLSSILRAFY*GGGGGG*	++	
cbGFP-α′CT10Δ1-6G	PLSSILRAFy*GGGGGG*	++	
cbGFP-α′CT10Δ2-6G	PLSSILRAfy*GGGGGG*	++	
cbGFP-α′CT10Δ3-6G	PLSSILRafy*GGGGGG*	+	+
cbGFP-α′CT10Δ4-6G	PLSSILrafy*GGGGGG*	+	++
cbGFP-α′CT10Δ5-6G	PLSSIlrafy*GGGGGG*		++
CbGFP-PLSS-6G	PLSS ilrafy*GGGGGG*		++
cbGFP-PLSS-2G	PLSSilrafy*GG*	++	
cbGFP-PLSS	PLSSilrafy	++	
cbGFP-AFY	plssilrAFY	++	
cbGFP-α′CT6	plssILRAFY	++	
cbGFP-α′CT6-2G	plssILRAFY*GG*	++	
cbGFP-α′CT6-6G	plssILRAFY*GGGGGG*	+	++
cbGFP-P612G-6G	GLSSILRAFY*GGGGGG*	++	
cbGFP-L613G-6G	PGSSILRAFY*GGGGGG*	(+)	(+)
cbGFP-S614G-6G	PLGSILRAFY*GGGGGG*	++	
cbGFP-S615G-6G	PLSGLRAFY*GGGGGG*	+	+
cbGFP-I616G-6G	PLSSGLRAFY*GGGGGG*		++
cbGFP-L617G-6G	PLSSIGRAFY*GGGGGG*		++
cbGFP-R618G-6G	PLSSILGAFY*GGGGGG*	++	
cbGFP-A619G-6G	PLSSILRGFY*GGGGGG*	++	
cbGFP-βCT10[d]	PFPSILGALY	++[d]	
cbGFP-βCT10-6G	PFPSILGALY*GGGGGG*	++	

[a]X represents the inserted sequences in the cbGFP-X mutants shown in Fig. 1E

Each sequence is represented using the single-letter amino acid codes. Underlines indicate the sequences containing the VSDs of α′ and β. '*G*'s in italics represent glycines used for blocking the C-termini. Lower cases represent deleted amino acids. 'G's enclosed by black lines represent glycines used for glycine scanning

[b]'++' represents the status that a large amount of GFP fluorescence was detected in the PSVs or the ICS. '+' represents the status that a part of the GFP fluorescence was detected. '(+)' represents the status that only slight fluorescence was detected. 'Empty' represents the status that almost no fluorescence was detected

[c]mGFP-CT10 (Nishizawa et al. 2003)

[d]mGFP-βCT10(Nishizawa et al. 2004)

sequence coding cbGFP-PLSS, cbGFP-α′CT6 and cbGFP-AFY (Fig. 1E and Table 1) were introduced into maturing soybean cotyledons by particle bombardment. As a result, cbGFP-PLSS, cbGFP-α′CT6 and cbGFP-AFY were transported to the PSVs (Fig. 3A–C). Thus, it was confirmed that the N-terminal and the C-terminal portions of the C-terminal ten amino acids of α′ in analogy with those of β are enough to work as VSDs.

Both ctVSD and ssVSD exist in the C-terminal ten amino acids of α′

ssVSDs function in any regions of proteins as long as they are exposed on surfaces of proteins. However, even if an essential Ile or Leu is replaced by Gly, they can function as ctVSDs only at the C-termini if the sequences have the characteristics of ctVSDs (Koide et al. 1997). It was also reported that the function of ctVSD of barley lectin and the ctVSD of potato tuber 20 kDa protein was inhibited by one or two glycines attached to their C-termini (Dombrowski et al. 1993; Koide et al. 1999), although the mechanism of this inhibition is not clear. To elucidate the type of the VSD in the C-terminal region of α′, we examined the effect of continuous glycines linked to the C-termini of cbGFP-α′CT6 and cbGFP-α′CT10 (Fig. 1E and Table 1). cbGFP-α′CT6-2G and cbGFP-α′CT10-2G with two continuous glycines were transported to the PSVs (Fig. 4A, B). However, a large amount of cbGFP-α′CT6-6G with continuous six glycines (Table 1) was secreted to the ICS although it was partially transported to the PSVs (Fig. 4C, D), indicating that a ctVSD exists in the C-terminal six amino acids of α′. From the results shown in these figures and in Fig. 3C, the sequence AFY (composed of the C-terminal three amino acids of α′) is considered a ctVSD. On the other hand, six glycines linked to the C-terminus of cbGFP-α′CT10 (Table 1, cbGFP-α′CT10-6G) did not inhibit the function of the C-terminal ten amino acids of α′ at

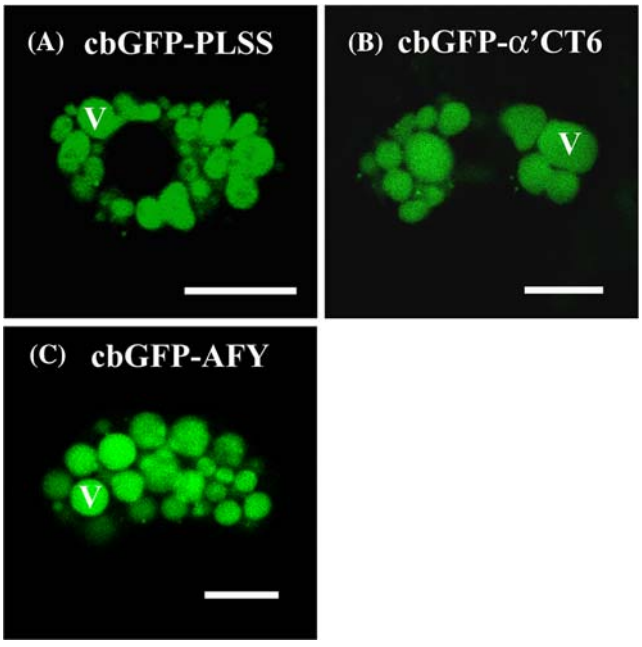

Fig. 3 Transport of cbGFP with the partial C-terminal sequences of α′ to the PSVs in maturing soybean cotyledons. (**A**) cbGFP-PLSS. (**B**) cbGFP-α′CT6. (**C**) cbGFP-AFY. V, PSV. Bars = 10 μm

all (Fig. 4E). This indicates that an ssVSD is contained in the C-terminal ten, not six, amino acids of α′. To further confirm this, the peptide composed of the C-terminal ten amino acids of α′ was placed at the N-terminus of cbGFP (Fig. 1F, α′CT10-cbGFP), and its destination in cotyledon cells was examined. α′CT10-cbGFP was never secreted to the ICS, and it was transported to the PSVs (Fig. 4F), indicating again that an ssVSD is contained in the C-terminal ten amino acids of α′. Thus these demonstrate that both an ssVSD and a ctVSD exist in the C-terminal ten amino acids of α′. Furthermore, cbGFP-PLSS-2G (Table 1) which has two glycines at the C-terminus was transported to the PSVs (Fig. 4G) in analogy with cbGFP-α′CT6-2G and cbGFP-α′CT10-2G, while cbGFP-PLSS-6G (Table 1) with continuous six glycines at the C-terminus was secreted to the ICS (Fig. 4H, I), indicating that an ssVSD does not exist in the sequence PLSS, and that the ssVSD in the C-terminal ten amino acids of α′ ranges over the two regions (PLSS and α′CT6).

A core sequence in the ssVSD is Ser-Ile-Leu

To determine the region where the ssVSD exists, it was examined if the derivatives having continuous six glycines at the sequentially truncated C-termini (Fig. 1E and Table 1) were transported to the PSVs. cbGFP-α′CT10Δ1-6G and cbGFP-α′CT10Δ2-6G (Table 1), which lack the C-terminal residue Tyr (621st residue of α′ counted from the first residue of signal peptide) and Phe and Tyr (620th–621st residues), respectively, were transported to the PSVs (Fig. 5A, B) in analogy with cbGFP-α′CT10-6G. However, cbGFP-α′CT10Δ3-6G and cbGFP-α′CT10Δ4-6G lacking the sequences AFY (619th–621st residues) and RAFY (618th–621st residues), respectively, were partially secreted to the ICS (Fig. 5C–E; the bright dots in Fig. 5C and E are gold particles used in particle bombardment). cbGFP-α′CT10Δ5-6G and cbGFP-PLSS-6G (Table 1) lacking the sequences LRAFY (617th–621st residues) and ILRAFY (616th–621st residues), respectively, were secreted to the ICS and were not transported to the PSVs (Figs. 5F, G and 4H, I). Thus, it was suggested that a core sequence for the function as an ssVSD exists in the sequence PLSSILRA (612th–619th residues of α′ counted from the first residue of signal peptide).

Next, we substituted Gly for each residue in the sequence PLSSILRA of cbGFP-α′CT10-6G (Fig. 1E and Table 1), and examined if the derivatives were transported to the PSVs. cbGFP-P612G-6G, cbGFP-S614G-6G, cbGFP-R618G-6G and cbGFP-A619G-6G (Table 1) in which Gly was substituted for each of Pro 612, Ser 614, Arg 618 and Ala 619 were transported to the PSVs (Fig. 6A–D). On the other hand, cbGFP-S615G-6G (Table 1) was secreted to the ICS though it was also transported to the PSVs (Fig. 6E, F). cbGFP-I616G-6G and cbGFP-L617G-6G (Table 1) were secreted to the ICS and were not transported to the PSVs at all (Fig. 6G–J; the dots in the cells in Fig. 6G and I are gold particles used in particle bombardment). Thus these indicate that the sequence SIL (615th to 617th residues of α′ counted from the first residue of signal peptide or fifth to seventh residues counted from the C-terminus) is a core sequence for the function as an ssVSD. In this sequence, the Ile and Leu are more important than the Ser, since cbGFP-S615G-6G was partly transported to the PSVs. This is largely consistent with the result obtained using cbGFP-α′CT6-6G lacking Ser 615 in which a large amount of cbGFP-α′CT6-6G was secreted to the ICS although it was partially transported to the PSVs (Fig. 4C, D). It is conceivable that Ser 615 has a supplementary role in function as an ssVSD. cbGFP-L613G-6G (Table 1), in which Gly was substituted for Leu 613, was barely detected in both the ICS and the PSVs, but was mainly detected in the ER-network (Fig. 6K). This result was unexpected, and we can only surmise that Leu 613 might play an important role different from the critical role played by Ile 616 and Leu 617. From the results shown in Fig. 5C–E, it seems as if Arg 618 and Ala 619

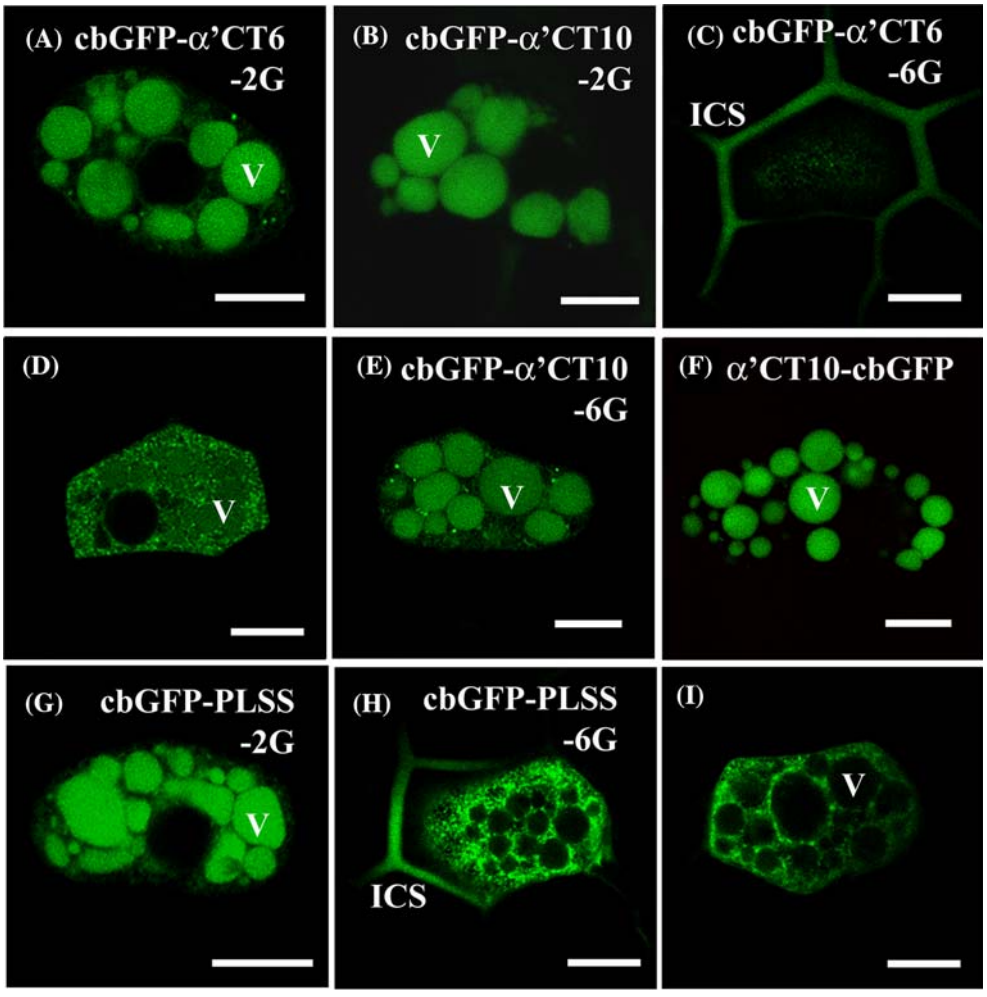

Fig. 4 Location of cbGFP with the partial C-terminal sequences of α′ followed by continuous glycines, and transport of cbGFP with the C-terminal ten amino acids of α′ at the N-terminus to the PSVs in maturing soybean cotyledons. (**A**) cbGFP-α′CT6-2G. (**B**) cbGFP-α′CT10-2G. (**C**) cbGFP-α′CT6-6G. (**D**) The cell in (**C**) was optically sectioned at the deeper position of *z*-axis. (**E**) cbGFP-α′CT10-6G. (**F**) α′CT10-cbGFP. (**G**) cbGFP-PLSS-2G. (**H**) cbGFP-PLSS-6G. (**I**) The cell in (**H**) was optically sectioned at the deeper position of *z*-axis. V, PSV; ICS, intercellular space. Bars = 10 μm

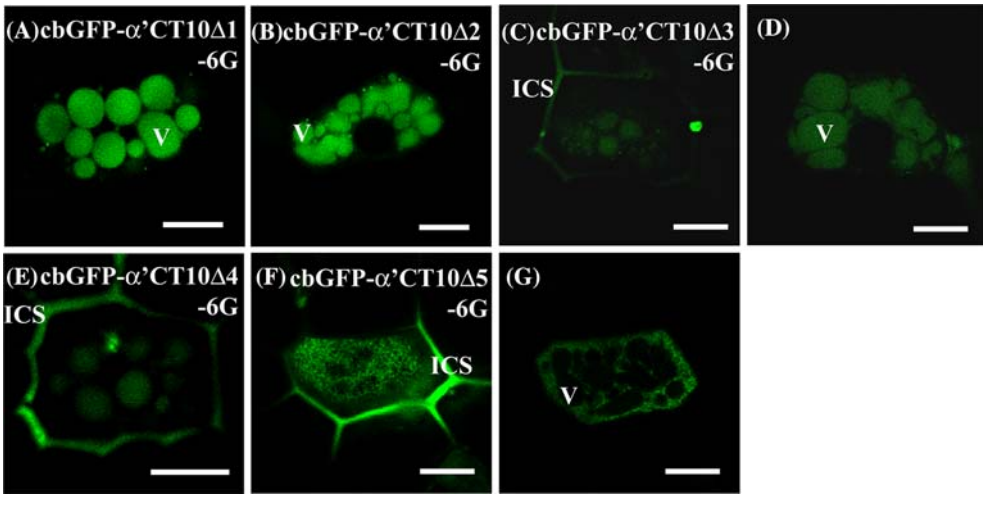

Fig. 5 Location of cbGFP with the truncated C-terminal sequences of α′ followed by continuous six glycines in maturing soybean cotyledons. (**A**) cbGFP-α′CT10Δ1-6G. (**B**) cbGFP-α′CT10Δ2-6G. (**C**) cbGFP-α′CT10Δ3-6G. (**D**) The cell in (**C**) was optically sectioned at the deeper position of *z*-axis. (**E**) cbGFP-α′CT10Δ4-6G. (**F**) cbGFP-α′CT10Δ5-6G. (**G**) The cell in (**F**) was optically sectioned at the deeper position of *z*-axis. V, PSV; ICS, intercellular space. Bars = 10 μm

had been necessary for transportation to the PSVs, since cbGFP-α′CT10Δ3-6G and cbGFP-α′CT10Δ4-6G were partly secreted to the ICS, while cbGFP-α′CT10Δ2-6G was not secreted (Fig. 5B). However, cbGFP-R618G-6G (Fig. 6C) and cbGFP-A619G-6G (Fig. 6D) in which glycines were substituted for Arg 618 and Ala 619, respectively, were transported to the PSVs and were never secreted. From these results,

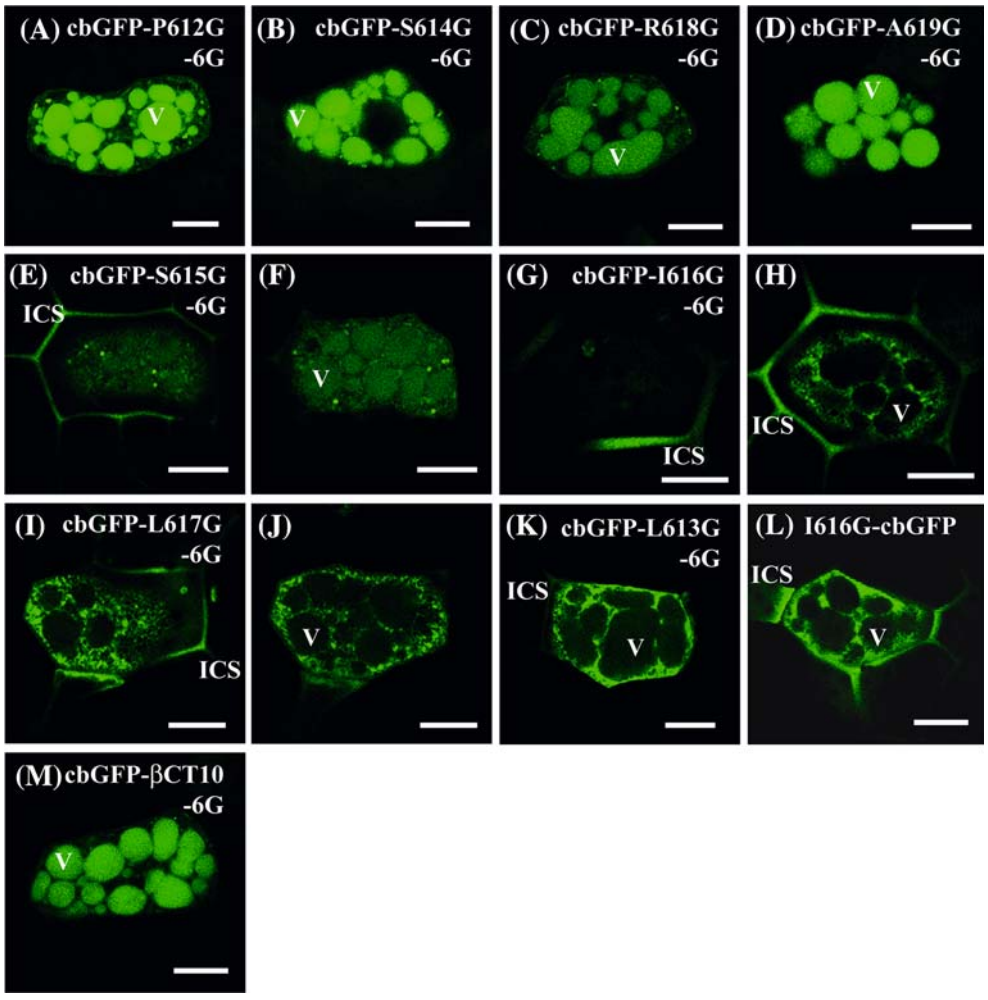

Fig. 6 Location of cbGFP with the mutated C-terminal sequences of α′ followed by continuous six glycines, with the mutated C-terminal sequence of α′ at the N-terminus, or with the C-terminal ten amino acids of β followed by continuous six glycines in maturing soybean cotyledons. (**A**) cbGFP-P612G-6G. (**B**) cbGFP-S614G-6G. (**C**) cbGFP-R618G-6G. (**D**) cbGFP-A619G-6G. (**E**) cbGFP-S615G-6G. (**F**) The cell in (**E**) was optically sectioned at the deeper position of *z*-axis. (**G**) cbGFP-I616G-6G. (**H**) The cell in (**G**) was optically sectioned at the deeper position of *z*-axis. (**I**) cbGFP-L617G-6G. (**J**) The cell in (**I**) was optically sectioned at the deeper position of *z*-axis. (**K**) cbGFP-L613G-6G. (**L**) I616G-cbGFP. (**M**) cbGFP-βCT10-6G V, PSV; ICS, intercellular space. Bars = 10 μm

neither Arg 618 nor Ala 619 was essential, and, alternatively, the SIL sequence requires being followed by at least two amino acids except for continuous glycines for complete transportation to the PSVs.

We examined if the core sequence described above are also significant at the N-terminus of GFP. I616G-cbGFP (Fig. 1F) in which Ile of α′CT10-cbGFP was replaced by Gly was secreted to the ICS (Fig. 6L). Thus this indicated that the SIL sequence is also critically important for the function as an ssVSD at the N-terminus. It was also indicated again that the C-terminal sequence composed of three amino acids (AFY) is a ctVSD, since it could not function as a VSD in the transport of I616G-cbGFP though it functions as a VSD at the C-terminus of the cbGFP (Fig. 3C).

The C-terminus of β subunit also contains an ssVSD

The sequence SIL also exists in the C-terminal ten amino acids of β. We examined if the C-terminal sequence of β subunit also contains an ssVSD. As shown in Fig. 6M, cbGFP-βCT10-6G (Fig. 1E and Table 1) was transported to the PSVs in analogy with cbGFP-α′CT10-6G, indicating that an ssVSD exists also in the C-terminal ten amino acids of β and that the sequence SIL is probably a core of the ssVSD.

Discussion

Both an ssVSD and a ctVSD exist in the C-terminal region of α′

It has been demonstrated that the PSVs are composed of (sub)compartments, since there are the globoid compartments and matrix in the PSVs of some plants. The matrix contains storage proteins, whereas the globoid compartments contain phytic acid crystals and are supposed to have characteristics of the lytic vacuoles (Jiang et al. 2001). We previously reported that the C-terminal ten amino acids of α′ and β of β-conglycinin (GenPept accessions AAB01374, BAA23361)

which accumulates in the PSVs function as VSDs in transportation to the PSV matrix, and that they are not involved in the transportation to the globoid compartments (Nishizawa et al. 2003, 2004). In this study, we showed that both an ssVSD and a ctVSD exist in proximity to each other in this region of α′, and that β also contains an ssVSD. We could not evaluate whether GFP fusion proteins used in this study were transported not only to the PSV matrix but also to the globoid compartments, since the globoid compartments in the PSVs of soybean is too small to be detected by confocal laser scanning microscope (Nishizawa et al. 2003). However, it is conceivable that both types of VSDs in the C-terminal region of α′ function in transportation to the PSV matrix (To avoid complexity, we are just using the word ''PSV'' instead of ''PSV matrix'' except for in this section).

Previously it was reported that the potato tuber 20 kDa protein precursor contains two VSDs in the N- and C-terminal regions (Koide et al. 1999). However, our study is the first one that shows distinct types of VSDs exist in proximity. Until recently, ssVSDs were thought to be specific for the LVs but not for the PSVs, and the VSRs such as BP-80 were thought to work for the trafficking to the LVs. However, it has been proposed that this vision must be revised because of several recent studies (Vitale and Hinz 2005). It has been reported that proricin and 2S albumin precursor of castor bean, which are stored in the PSVs, have ssVSDs (Frigerio et al. 2001; Brown et al. 2003). In addition, AtVSR1/AtELP, one of BP-80 homologs, was elucidated to act as a sorting receptor of storage proteins (Shimada et al. 2003). Also, in this study, not only both types of VSDs identified in the C-terminal region of α′ but also the well-known ssVSD of aleurain and ctVSD of chitinase transported cbGFP (Aleu-cbGFP, cbGFP-Chit) to the PSVs in soybean maturing cotyledons. We could not know whether these fusion proteins were transported to the PSV matrix through a same pathway or whether they were also transported to the globoid compartments. However, it is obvious that both types of VSDs, ssVSDs and ctVSDs, can function on the pathway to the PSV matrix in maturing cotyledons of soybean.

The sequence composed of C-terminal three amino acids (AFY) of α′, homologous to the ctVSD (AFVY) of phaseolin (Frigerio et al. 1998a), is considered a ctVSD, since it directed cbGFP to the PSVs (Fig. 3C), and I616G-cbGFP, which has a mutation in the core sequence of ssVSD, was not transported to the PSVs though they have the AFY sequence (Fig. 6L). ctVSDs can function only at the C-termini. This might be because both of the C-terminal carboxylates of main chains and side chains of amino acids near the C-termini must be recognized by receptors, like the interaction between peroxisomal targeting signal-1 (PTS1) and its receptor PEX5, the molecular basis for which was revealed by crystallographic analysis of the complex (Gatto et al. 2000). The function of ctVSD of α′ was not disrupted with two glycines but disrupted with six glycines in this study. In contrast, the function of ctVSDs of barley lectin and potato tuber 20 kDa protein was disrupted with one or two glycines in tobacco leaf protoplasts and suspension cultured cells, respectively (Dombrowski et al. 1993; Koide et al. 1999). This difference might be due to the difference in the sequence around the C-termini of the VSDs and the difference in the environment of ligand interaction sites of receptors (e.g., capacity of binding pockets) in distinct cell types derived from different plants. A thorough mechanism of interaction between ctVSDs and receptors, and the reason why several glycines at the C-termini disrupt the function of ctVSDs may be revealed by crystallographic analysis of the ctVSD-receptor complexes derived from various cell-types and plants in the future.

ssVSDs were at first found in the proteins destined for the LVs (Holwerda et al. 1992; Matsuoka and Nakamura 1991). However, it has been reported that proricin and 2S albumin precursor of castor bean, which are stored in the PSVs, also have ssVSDs (Frigerio et al. 2001; Brown et al. 2003). In this study, it was elucidated that the sequence Ser-Ile-Leu, which are fifth to seventh residues counted from the C-terminus, is a core sequence of the ssVSD of α′ since the C-terminal peptide containing this sequence transported cbGFP to the PSVs even when it was placed at the N-terminus of cbGFP. The Ile and Leu in the sequence play a critically important role in the function as an ssVSD since this function was completely disrupted when even one of them was replaced by Gly (Fig. 6G–J). The ssVSDs of barley aleurain and sweet potato sporamin precursors, which were found in the initial stage of the studies of transportation to the vacuole, have a consensus sequence (NPIRL/P) (Holwerda et al. 1992; Matsuoka and Nakamura 1991). However, Matsuoka and Nakamura (1999) showed by detailed mutation analysis of the VSD of sporamin that a strict conservation of these amino acids is not essential. According to their report, the requirements for NPIRL motif are as follows: the first amino acid should not be small and hydrophobic, the second position may not be acidic, the third position is essential Ile or Leu which cannot be replaced by any other amino acids, the fourth position can be any amino acid and the fifth position should contain a bulky and

preferably hydrophobic side-chain (Matsuoka and Nakamura 1999). The ssVSDs of castor bean ricin (LLIRP, underline indicates an essential residue though it was confirmed only by substitution of Gly but not confirmed by substitution of other amino acids) and 2S albumin precursors (EVLRM, underline indicates an essential residue though it also was confirmed only by substitution of Gly) fit this criteria (Brown et al. 2003; Frigerio et al. 2001). However, the sequence of the ssVSD of α′ (SSILRA, underline indicates essential Ile and Leu though we confirmed by substitution of only Gly but not of other amino acids for them) does not fit this very well. For example, the Ser prior to the essential Ile cannot be replaced by Gly while the Pro prior to the essential Ile in the NPIR motif can be replaced by Gly (Matsuoka and Nakamura 1999). This implies that many more various sequences than we expect are acceptable as ssVSDs.

The SIL sequence that is the core of the ssVSD of α′ exists also in the C-terminal regions of α (GenPept accession CAA35691) and β of β-conglycinin and some other vicilins, convicilins and vicilin-like proteins (Fig. 7) such as pea vicilin (swissprot accession P02855), pea convicilin (prf accession 1713472A), fava bean (broad bean) vicilin (pir accession A27288), *Vicia narbonensis* vicilin (GenPept accession CAA96514), *V. narbonensis* convicilin (GenPept accession CAA96513), black walnut vicilin (GenPept accession AAM54366), pumpkin PV100 (GenPept accession BAA34056), *Arabidopsis thaliana* cupin family protein (RefSeq accession NP566714), macadamia nut vicilin (GenPept accession AAD54246). Among these sequences, it was shown that the C-terminal region of β also contains an ssVSD (Fig. 6M). From comparison with the C-terminal regions of α′ and β, it is expected that the SIL sequence in all these proteins are exposed outside of the molecules (Fig. 7; Maruyama et al. 2001, 2004), and that the C-terminal regions of these proteins might also work as ctVSDs since they contain hydrophobic amino acids similar to the ctVSD (AFY) of α′. Pea convicilin (GenPept accession CAA29695) and jack bean canavalin (GenPept accession CAA42075)

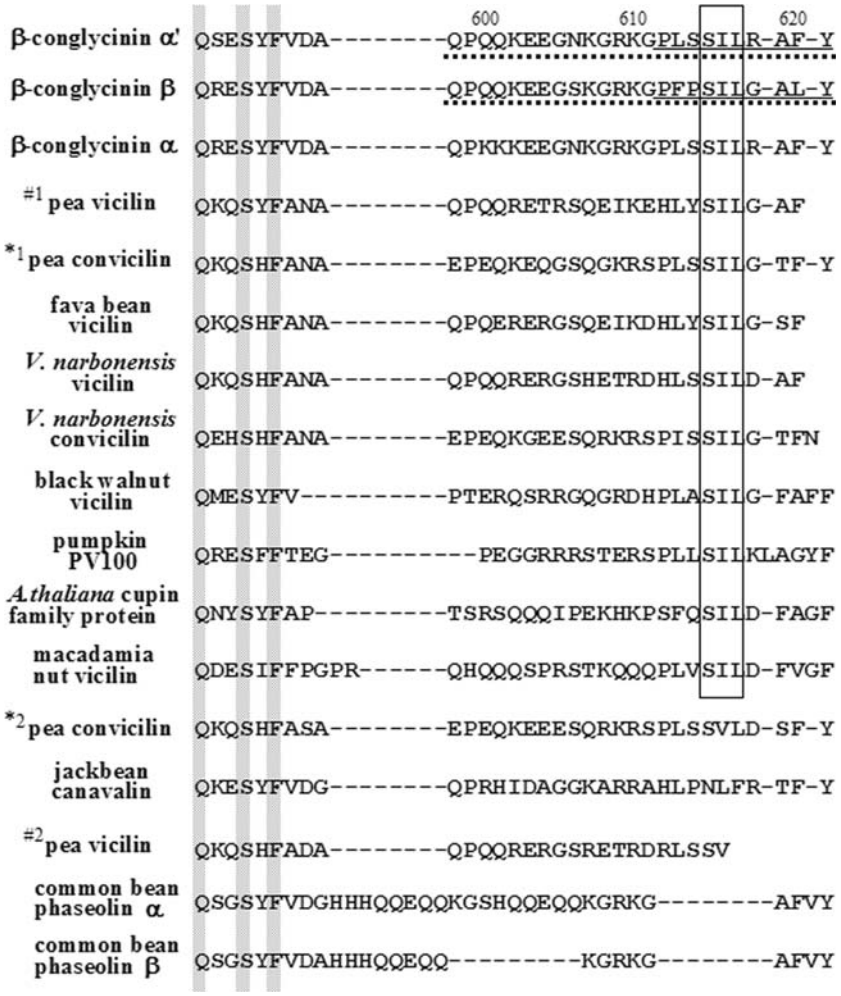

Fig. 7 Alignment of the C-terminal sequences of 7S globulins or related proteins from various plants. Dashed lines indicate gaps. Residue number of α′ counted from the first amino acid of the signal peptide are shown above the sequence. Underlines indicate the sequences containing the VSDs of α′ and β. Dotted underlines indicate the sequences, which are disordered on the surface of molecules elucidated previously by X-ray analysis. Shaded amino acids are conserved amino acids among these proteins. The conserved SIL sequence is enclosed by black line. #1: swissprot accession P02855, #2: GenPept accession CAA32239, *1: prf accession 1713272A, *2: GenPept accession CAA29695

contain similar sequences (SVL and NLF, respectively) in the C-terminal regions. They might also function as ssVSDs. However there are vicilins, which contain neither the SIL sequence nor similar sequences [e.g., pea vicilin (GenPept accession CAA32239) and α- and β-types of common bean phaseolin (GenPept accessions CAA26718, CAA26789)]. As for β-type phaseolin, it was elucidated that the sequence AFVY at the C-terminus is a ctVSD (Frigerio et al. 1998a).

Are both the ssVSD and ctVSD of α′ recognized by one vacuolar sorting receptor?

In this report we showed that the α′ subunit of β-conglycinin has two types of VSDs both of which function for trafficking to the PSVs. Are these VSDs recognized by one VSR? So far, VSRs which bind to the VSDs have been found in several plants, e.g., pea BP-80 (Kirsch et al. 1994; Paris et al. 1997), *Arabidopsis* AtVSR1/AtELP (Ahmed et al. 1997; Shimada et al. 2003) and pumpkin PV72 (Shimada et al. 1997). They possess high homology. Among these proteins, AtVSR1/AtELP was elucidated *in planta* to act as a sorting receptor of storage proteins (Shimada et al. 2003). It was also reported that AtVSR1/AtELP binds not only the VSD of sweet potato sporamin (Ahmed et al. 2000) and the N-terminal propeptide of *Arabidopsis* aleurain (AtAleu) which contain NPIR sequences but also the C-terminal sequence of *Arabidopsis* 12S globulin which contains neither Ile nor Leu (Shimada et al. 2003). Thus, it is conceivable that this (these) VSR(s) has (have) at least two distinct recognition mechanisms, which are for ssVSDs containing either Ile or Leu and for ctVSDs, respectively. Proteins homologous to the VSRs described above probably exist in soybean since expression of the homologous genes is detected in EST sequencing. Therefore it is possible that both of the ssVSD and ctVSD of α′ are recognized by one VSR in soybean maturing cotyledons. It was reported, however, that the 12S globulin and 2S albumin accumulated not only in the ICS but also in the PSVs in *atvsr1*mutant (Shimada et al. 2003), and that AtVSR1/AtELP did not bind the ctVSD of barley lectin (Ahmed et al. 2000). Recently, it was reported that *Arabidopsis thaliana* receptor homology region transmembrane domain ring H2 motif protein (AtRMR)1 functions as a receptor for ctVSD-mediated protein trafficking to the PSVs. By co-expression experiments using leaf protoplasts and *in vitro* binding assays, it was demonstrated that AtRMR1 binds the ctVSD (AFVY) of phaseolin and carries phaseolin to the PSVs (Park et al. 2005). Thus it is also possible that the ssVSD and ctVSD of α′ are recognized separately by different receptors, e.g., AtVSR1/AtELP and AtRMR1 homologs, respectively, in soybean.

Acknowledgements We thank Dr. Y. Niwa (University of Shizuoka) for providing the GFP (S65T) gene. This work was supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology (to N.M.), and Fuji Foundation for Protein Research (to N.M.).

References

Ahmed SU, Bar-Peled M, Raikhel NV (1997) Cloning and subcellular location of an Arabidopsis receptor-like protein that shares common features with protein-sorting receptors of eukaryotic cells. Plant Physiol 114:325–336

Ahmed SU, Rojo E, Kovaleva V, Venkataraman S, Dombrowski JE, Matsuoka K, Raikhel NV (2000) The plant vacuolar sorting receptor AtELP is involved in transport of NH_2-terminal propeptide-containing vacuolar proteins in *Arabidopsis thaliana*. J Cell Biol 149:1335–1344

Bednarek SY, Wilkins TA, Dombrowski JE, Raikhel NV (1990) A carboxyl-terminal propeptide is necessary for proper sorting of barley lectin to vacuoles of tobacco. Plant Cell 2:1145–1155

Brown JC, Jolliffe NA Frigerio L, Roberts LM (2003) Sequence-specific, Golgi-dependent vacuolar targeting of castor bean 2S albumin. Plant J 36:711–719

Dombrowski JE, Schroeder MR, Bendnarek SY, Raikhel NV (1993) Determination of the functional elements within the vacuolar targeting signal of barley lectin. Plant Cell 5:587–596

Frigerio L, Jolliffe NA, Cola AD, Felipe DG, Paris N, Neuhaus J-M, Lord JM, Ceriotti A, Roberts LM (2001) The internal propeptide of the ricin precursor carries a sequence-specific determinant for vacuolar sorting. Plant Physiol 126:167–175

Frigerio L, Virgilio M, Prada A, Faoro F, Vitale A (1998a) Sorting of phaseolin to the vacuole is saturable and requires a short C-terminal peptide. Plant Cell 10:1031–1042

Frigerio L, Vitale A, Lord JM, Ceriotti A, Roberts LM (1998b) Free ricin A chain, proricin and native toxin have different cellular fates when expressed in tobacco protoplasts. J Biol Chem 273:14194–14199

Gatto GJ, Geisbrecht BV, Gould SJ, Berg JM (2000) Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. Nat Struct Biol 7:1091–1095

Hara-Nisimura I, Shimada T, Hatano K, Takeuchi Y, Nishimura M (1998) Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicles. Plant Cell 10:825–836

Herman EM, Larkins BA (1999) Protein storage bodies and vacuoles. Plant Cell 11:601–613

Hillmer S, Movafeghi A, Robinson DG, Hinz G (2001) Vacuolar storage proteins are sorted in the *cis*-cisternae of the pea cotyledon Golgi apparatus. J Cell Biol 152:41–50

Hinz G, Hillmer S, Bäumer M, Hohl I (1999) Vacuolar storage proteins and the putative vacuolar sorting receptor BP-80 exit the Golgi apparatus of developing pea cotyledons in different transport vesicles. Plant Cell 11:1509–1524

Hoh B, Hinz G, Jeong B-K, Robinson DJ (1995) Protein storage vacuoles form de novo during pea cotyledon development. J Cell Sci 108:299–310

Hohl I, Robinson DG, Chrispeels MJ, Hinz G (1996) Transport of storage proteins to the vacuole is mediated by vesicles without a clathrin coat. J Cell Sci 109:2539–2550

Holwerda BC, Padgett HS, Rogers JC (1992) Proaleurain vacuolar targeting is mediated by short contiguous peptide interactions. Plant Cell 4:307–318

Jauh G-Y, Phillips T, Rogers JC 1999. Tonoplast intrinsic protein isoforms as markers for vacuole functions. Plant Cell 11:1867–1882

Jiang L, Phillips TE, Hamm CA, Drozdowicz YD, Rea PA, Maeshima M, Rogers SW, Rogers JC (2001) The protein storage vacuole: a unique compound organelle. J. Cell Biol 155:991–1002

Kinney AJ, Jung R, Herman EM (2001) Cosuppression of the alpha subunits of beta-conglycinin in transgenic soybean seeds induces the formation of endoplasmic reticulum-derived protein bodies. Plant Cell 13:1165–1178

Kirsch T, Paris N, Butler JM, Beevers L, Rogers JC (1994) Purification and initial characterization of a potential plant vacuolar targeting receptor. Proc Natl Acad Sci USA 91:3403–3407

Koide Y, Hirano H, Matsuoka K, Nakamura K (1997) The N-terminal propeptide of the precursor to sporamin acts as a vacuole-targeting signal even at the C terminus of the mature part in tobacco cells. Plant Physiol 114:863–870

Koide Y, Hirano H, Matsuoka K, Nakamura K (1999) The N-terminal propeptide of the precursor to sporamin acts as a vacuolar-targeting signal even at the C-terminus of the mature part in tobacco cells. Plant Physiol 114:863–870

Maruyama N, Adachi M, Takahashi K, Yagasaki K, Kohno M, Takenaka Y, Okuda E, Nakagawa S, Mikami B, Utsumi S (2001) Crystal structures of recombinant and native soybean β-conglycinin β homotrimers. Eur J Biochem 268:3595–3604

Maruyama Y, Maruyama N, Mikami B, Utsumi S (2004) Structure of the core region of the soybean β-conglycinin α' subunit. Acta Crystallogr D Biol Crystallogr 60:289–297

Matsuoka K, Nakamura K (1991) Propeptide of a precursor to a plant vacuolar protein required for vacuolar targeting. Proc Natl Acad Sci USA 88:834–838

Matsuoka K, Nakamura K (1999) Large alkyl side-chains of isoleucine and leucine in the NPIRL region constitute the core of the vacuolar sorting determinant of sporamin precursor. Plant Mol Biol 41:825–835

Matsuoka K, Neuhaus J-M (1999) *Cis*-elements of protein transport to the plant vacuoles. J Exp Bot 50:165–174

Melroy DL, Herman EM (1991) TIP, an integral membrane protein of the protein-storage vacuoles of the soybean cotyledon undergoes developmentally regulated membrane accumulation and removal. Planta 184:113–122

Mori T, Maruyama N, Nishizawa K, Higasa T, Yagasaki K, Ishimoto M, Utsumi S (2004) The composition of newly synthesized proteins in the endoplasmic reticulum determines the transport pathways of soybean seed storage proteins. Plant J 40:238–249

Neuhaus J-M, Rogers JC (1998) Sorting of proteins to vacuoles in plant cells. Plant Mol Biol 38:127–144

Neuhaus J-M, Sticher L, Meins FJ, Boller T (1991) A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole. Proc Natl Acad Sci USA 88:10362–10366

Nishizawa K, Maruyama N, Satoh R, Fuchikami T, Higasa T, Utsumi S (2003) A C-terminal sequence of soybean β-conglycinin α' subunit acts as a vacuolar sorting determinant in seed cells. Plant J 34:647–659

Nishizawa K, Maruyama N, Satoh R, Higasa T, Utsumi S (2004) A vacuolar sorting determinant of soybean β-conglycinin β subunit resides in a C-terminal sequence. Plant Sci 167:937–947

Paris N, Rogers SW, Jiang L, Kirsch T, Beevers L, Phillips TE, Rogers JC (1997) Molecular cloning and further characterization of a probable plant vacuolar sorting receptor. Plant Physiol 115:29–39

Park M, Lee D, Lee G-J, Hwang I (2005) AtRMR1 functions as a cargo receptor for protein trafficking to the protein storage vacuole. J Cell Biol 170:757–767

Saalbach G, Rosso M, Schumann U (1996) The vacuolar targeting signal of the 2S albumin from Brazil nut resides at the C terminus and involves the C-terminal propeptide as an essential element. Plant Physiol 112:975–985

Shimada T, Fuji K, Tamura K, Kondo M, Nishimura M, Hara-Nishimura I (2003) Vacuolar sorting receptor for seed storage proteins in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 100:16095–16100

Shimada T, Kuroyanagi M, Nishimura M, Hara-Nishimura I (1997) A pumpkin 72-kDa membrane protein of precursor-accumulating vesicles has characteristics of vacuolar sorting receptor. Plant Cell Physiol 38:1414–1420

Shimada T, Watanabe E, Tamura K, Hayashi Y, Nishimura M, Hara-Nishimura I (2002) A vacuolar sorting receptor PV72 on the membrane of vesicles that accumulate precursors of seed storage proteins (PAC vesicles). Plant Cell Physiol 43:1086–1095

Vitale A, Hinz G (2005) Sorting of proteins to storage vacuoles: how many mechanisms? Trends Plant Sci 10:316–323

Vitale A, Raikhel NV (1999) What do proteins need to reach different vacuoles? Trends Plant Sci 4:149–155