Categories **b. Food Science**

Title

Cloning of a cDNA encoding the Gly m Bd 28K precursor and its vacuole transport in tobacco BY2 suspension-cultured cells.

Names of authors

Hitomi YUMIOKA-ITO[1,2], Ryo MISAKI[1], Miyuki YOKORO[3], Makiko SUZUKI[4], Hiromi YAMASHITA[4], Miki HIEMORI-KONDO[5], Masumi KIMOTO[4], Ko KATO[6], Kazuhito FUJIYAMA[1] and Hideaki TSUJI[2]

Names of institutions

[1]International Center for Biotechnology, Osaka University, Yamada-oka 2-1, Suita, Osaka 565-0871, Japan.

[2]Okayama Prefectural University, Kuboki 111, Soja 719-1197, Japan.

[3]Division of Nephrology, Department of Medicine, Kurume University School of Medicine, Asahi-machi67, Kurume, Fukuoka 830-0011, Japan.

[4]Department of Nutritional Science, Faculty of Health and Welfare Science, Okayama Prefectural University, Kuboki 111, Soja 719-1197, Japan.

[5]Department of Food and Nutritional Sciences, Faculty of Human Life Sciences, Tokushima Bunri University, Nishihama 180, Yamashiro, Tokushima 770-8514, Japan

[6]Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan.

Number of tables and figures: 8 figures and 1 table

A brief running head: Vacuole transport of Gly m Bd 28K precursor in tobacco BY2 cells.

Name and address of corresponding author: Hitomi YUMIOKA-ITO, International Center for Biotechnology, Osaka University, Yamada-oka 2-1, Suita, Osaka 565-0871, Japan.

Key words: soybean allergen, Gly m Bd28K, 23-kDa peptide, signal peptide, vacuole

Summary Gly m Bd 28K (Gm28K), a soybean allergen, is formed as a preproprotein consisting of a predicted signal peptide, Gm28K, and the 23-kDa peptide (Gm23K). Gm28K and Gm23K are found in the protein-storage vacuoles (PSVs) of developing soybean seeds. However, the complete structure of Gm28K has not yet been identified and its processing and transport to the vacuoles has never been clarified. In the present study, we elucidated the 5'-nucleotide sequence of cDNA encoding the Gm28K precursor and identified a putative signal peptide (SP) with 24 N-terminal amino acid residues. We expressed peptides from the Gm28K precursor as fusion proteins with enhanced green fluorescent protein (EGFP) in tobacco BY2 suspension-cultured cells. BY2 cells transformed by an expression vector for SP-EGFP-Gm28-Gm23K (SP-EGFP-Gm28-Gm23K/BY2 cells) and SP-Gm28-Gm23K-EGFP/BY2 cells produced the EGFP fused-Gm28K precursor, and the EGFP-fluorescence in their vacuoles were recorded. In the experiments with SP-EGFP/BY2 and SP-EGFP-Gm28K/BY2 cells, large amounts of the EGFP segments were secreted into the medium. On the other hand, the fluorescence of EGFP in SP-EGFP-Gm23K/BY2 cells was shown to accumulate only in the endoplasmic reticulum without secretion into the medium. These findings show that the SP signals the precursor to enter the lumen of the endoplasmic reticulum and that both the Gm28K and Gm23K components may be involved in the transport from the endoplasmic reticulum (ER) lumen via the Golgi to the vacuoles in a proprotein form.

Footnote: The new nucleotide sequence of the 5'-nucleotide sequence region of cDNA encoding the Gly m Bd 28K precursor elucidated in the present study was registered under accession number AB046874.2 on DDBJ.

INTRODUCTION

In Japan, the incidence of food allergy has been steadily increasing, and now constitutes a serious problem that urgently warrants new treatments. Eggs, dairy milk, rice, wheat and soybeans are all considered major allergenic foodstuffs in Japan. Soybean, of course, is a major source of protein and lipid, and is utilized as a supplement for the preparation of many food products. In an investigation in which the sera of soybean-sensitive patients were examined with an immunoblotting technique in order to identify the allergens in soybeans, Ogawa *et al*. (*1*) found that soybean has 16 IgE-binding components. Among them, Gly m Bd 68K, Gly m Bd 30K and Gly m Bd 28K (Gm28K) were found to be the major allergens (*1*). Gly m Bd 68K was identified as the α-subunit of β-conglycinin (*2*). Furthermore, the N-terminal amino acid residues of Gly m Bd 30K were found to be identical to those of the 34-kDa soybean seed storage vacuole protein (*3*), which is a papain superfamily protein that defends against *Pseudomonas* (*4*).

In another study, Gm28K was purified from defatted soybean flakes and characterized (*5*). A cDNA that encodes Gm28K and encompasses 473 amino acid residues, including 220 amino acid residues of the allergen, was cloned (*6*). However, its N-terminal amino acid sequence has not been fully characterized because the cloned cDNA did not contain any start codon. By homology search of the amino acid sequences deduced from a comparison of the cDNA with the MP27/MP32 proteins in pumpkin seeds, Gm28K was suggested to be synthesized in the form of a preproprotein consisting of a signal peptide (SP), Gm28K, and an additional peptide with a molecular mass of 23 kDa (*6*). Furthermore, Hiemori *et al*. showed that the 23-kDa peptide (Gm23K) exists in the 7S globulin fraction of soybean proteins by immunoblotting with a monoclonal antibody specific for the peptide. Both Gm28K and Gm23K locate in the crystalloids of protein storage vacuoles (PSVs) in growing cotyledons (*7*).

However, how the Gm28K and Gm23K are processed and transported to the vacuoles is unknown. Therefore, study of the production of recombinant Gm28K in plant cells would be useful for clarifying the characteristics of the Gm28K precursor.

Tobacco BY2 suspension-cultured cells are a useful host for the production of recombinant proteins. We attempted to construct the sequence of cDNA encoding the Gm28K connecting to a signal peptide of the sweet potato sporamin. However, we failed to express Gm28K protein in tobacco BY2 suspension-cultured cells in our preliminary study. This failure may have been due to incorrect expression or folding of Gm28K in tobacco BY2 cells, because its N-terminal cDNA sequence including the region encoding for its signal peptide has not been fully characterized as mentioned above. This led us to clone the cDNA encoding the Gm28K precursor with the full-length signal peptide portion and to examine the transport of Gm28K and Gm23K to vacuoles in tobacco BY2 cells.

The seed storage proteins in plants are known to be transported into PSVs by two routes. In one route, the proteins are transported to the PSVs via the endoplasmic reticulum (ER) and the Golgi apparatus. In the other, they are transported to the PSVs directly via the ER. The proteins transported to the PSVs depend on the vacuolar-sorting determinants (VSDs), sequence-specific VSDs (ssVSDs) (*8, 9*), C-terminal VSDs (ctVSDs) (*10-13*) and protein-structure-related VSDs (*14*). However, it is not clear how the Gm28K precursor is transported into the PSVs in soybean.

In the present study, we investigated the role in vacuole transport of the newly identified SP and the region of Gm28K and Gm23K in tobacco BY2 cells.

MATERIALS AND METHODS

Preparation of total RNA and cDNA. Total RNA was extracted from developing soybean seeds (*Glycine max*, L. cv. Wasesuzunari) using an RNeasy Plant

Mini Kit (QIAGEN, North Rhine-Westphalia, Germany). 5' RNA-ligand-mediated rapid amplification of cDNA ends (5' RLM-RACE) was carried out using a First Choice RLM-RACE kit (Ambion, Austin, TX) according to the manufacturer's instructions. Two RNA 5'-adaptor primers were supplied with the RLM-RACE kit and the following three gene-specific reverse primers were used to obtain the Gm28K precursor gene (*6*): 5'-TCCTACCACCATGGCTTTTCA-3' (outer primer), 5'-GTCTTGAAAACCCTCGTGGA-3' (first inner primer) and 5'-GCCACTCCATGGCAAAGAACAA-3' (second inner primer). These PCR products were purified and subsequently cloned into the pGEM®-T Easy vector (Promega, Madison, WI), and their sequences were elucidated using an ABI PRISM™ 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA).

Plasmid construction. First, the gene encoding the *Nicotiana tabacum* alcohol dehydrogenase 5'-untranslated region (*NtADH* 5'-UTR) was amplified using a plasmid containing the *NtADH* 5'-UTR region (*15*) that was constructed using a vector as the template and primers 1 and 2 (Table 1). The signal peptide of the Gm28K precursor was amplified using cDNA prepared by the 5'RLM-RACE method and primers 3 and 4. The Gm28K- and Gm23K-coding regions were amplified using cDNA encoding the Gm28K precursor and primers 5 and 6. These amplified fragments were then used to construct *Xba*I-*NtADH* 5'-UTR-SP-Gm 28-Gm23K-*Sac*I with primers 1 and 6.

Each construct shown in Fig. 1 was prepared using PCR. SP-Gm28-Gm23K-EGFP and SP-EGFP-Gm28-Gm23K were constructed using the coding regions of the *Xba*I-*NtADH* 5'-UTR-signal peptide-Gm28-Gm23K-*Sac*I, EGFP, and the following primer sets: primers 1, 9, 18, and 19 for SP-Gm28-Gm23K-EGFP; and primers 1, 6, 7, 8, 10, and 11 for SP-EGFP-Gm28-Gm23K. EGFP-Gm28-Gm23K was constructed using the coding region of SP-Gm28-Gm23K-EGFP and primers 1, 6, 16, and 17. EGFP-Gm28K and enhanced green fluorescent protein (EGFP) were

prepared using the coding region of EGFP-Gm28-Gm23K and the following primer sets: primers 1 and 12 for EGFP-Gm28K; and primers 1 and 9 for EGFP. SP-EGFP-Gm28K, SP-EGFP-Gm23K and SP-EGFP were constructed using the coding region of SP-EGFP-Gm28-Gm23K and the following primer sets: primers 1 and 12 for SP-EGFP-28K; primers 1 and 6 for SP-EGFP-Gm23K; and primers 1 and 9 for SP-EGFP. This resulted in, respectively, pBI121-SP-EGFP-Gm28-Gm23K, pBI121-EGFP-Gm28-Gm23K, pBI121-SP-Gm28-Gm23K-EGFP, pBI121-SP-EGFP-Gm28K, pBI121-EGFP-Gm28K, pBI121-SP-EGFP-Gm23K and pBI121-SP-EGFP.

These PCR products were digested with *Xba*I and *Sac*I, and were ligated to plant expression vector pBI121. The vectors were then introduced into the *Agrobacterium tumefaciens* strain LBA4404 by electroporation using a Bio-Rad Gene Pulser (Bio-Rad, Hercules, CA).

The location of Table 1.

The location of Figure 1.

Transformation of tobacco BY2 cells. Agrobacterium carrying a plant expression vector were inoculated onto 2×YT medium containing antibiotics (50 mg/l kanamycin, 200 mg/l streptomycin, and 50 mg/l rifampicin) to transform tobacco BY2 cells by the Agrobacterium-mediated transformation method (*16*). The transformants were selected on mLS medium (*17, 18*) containing antibiotics (250 mg/l carbenicillin sodium salt and 150 mg/l kanamycin).

Fluorescent microscopy and laser-scanning confocal microscopy.

The transformed BY2 cells were examined on an inverted Leica TCS-SPE confocal laser scanning microscope on a DMI 4000 fluorescent microscope, version 2.3 (Leica Microsystems, Mannheim, Germany). The laser was focused on cells through a 40×NA 0.75 or 63×NA 1.4 oil objective. The 488-nm excitation wavelength and a 500-531-nm emission filter (Green) were used to detect EGFP.

To visualize the tonoplast in BY2 cells, 50 µM *N*-(3-triethylammoniumpropyl)-4-(6-(4-(diethylamino) phenyl) hexatrienyl)-pyridinium dibromide (FM4-64; Molecular Probes Inc., Eugene, OR) was added to the cultured-cell medium. The cells were incubated at room temperature for 1 h and washed three times with fresh mLS medium. A 555-700-nm emission filter (red) was used to detect tonoplast stained with FM4-64 at an excitation wavelength of 532 nm (*19*).

To visualize the Golgi apparatus in BY2 cells, the cells were stained by incubation in medium containing 5 µM BODIPY TR ceramide (Molecular Probes Inc.) for 30 min on ice and 30 min at 37°C, then washed three times with fresh mLS medium. A 530-592-nm emission filter (red) was used to detect the Golgi apparatus stained with BODIPY TR ceramide at an excitation wavelength of 532 nm (*20*).

To visualize the ER in BY2 cells, the cells were also stained for 5 min by incubation in a medium containing 1 µM ER-tracker blue-white DPX (Molecular Probes Inc.) transferred to fresh medium without the marker dyes (*21*). Differential Interference Contrast (DIC) and fluorescence images were taken with an AxioCam MRc digital camera controlled by Axio Vision software release 4.8 (Carl Zeiss, Oberkochen, Germany) on an upright microscope Axioskop 2 (Carl Zeiss) fitted for a 40×NA 0.75 or 100×NA 1.30 oil objective. Fluorescence measurements were performed with Image J software (http://rsb.info.nih.gov/ij/index.html).

Protein extraction. Cells were suspended in 10 mM Tris-HCl, pH 7.5, containing Complete Protease Inhibitor Cocktail (Roche Diagnostics GmbH, Mannheim, Germany) and homogenized on ice using a Handy Sonic model UR-20P (Tomy Seiko Co., Ltd., Tokyo, Japan). The cell lysate was centrifuged at 12,000 ×*g* for 20 min at 4°C and its supernatant was used as the cell extract.

The medium was filtered through a filter paper to remove the cells, and then the filtrate was centrifuged at 5,000 ×*g* for 15 min at 4°C. The obtained supernatant was

treated with four volumes of acetone for 1 h, followed by centrifugation. The precipitate was suspended in 5 mM Tris-HCl, pH 7.0, and then centrifuged at 12,000 ×*g* for 20 min to obtain a solution containing the secreted proteins. The protein concentration in the supernatant was determined using a Bio-Rad Bradford assay (Bio-Rad).

Immunoprecipitation of expressed proteins. The cell extracts were centrifuged at 12,000 ×*g* for 20 min at 4°C. The proteins in the supernatants were immunoprecipitated with a rabbit anti-GFP polyclonal antibody (Medical & Biological Laboratories Co., Nagoya, Japan), and protein G Sepharose^{TM} 4 Fast Flow (GE Healthcare Bio-Science AB, Uppsala, Sweden). The immunoprecipitates linked with protein G Sepharose were washed 5 times with cold 20 mM sodium phosphate buffer, pH 7.0, resuspended in Laemmli's sample buffer, and subjected to sodium dodecyl sulfate-poly acrylamide gel electrophoresis (SDS-PAGE) and Western blotting.

SDS-PAGE and Western blotting. Total protein extracts from suspension-cultured tobacco BY2 cells and medium were subjected to SDS-PAGE according to the method of Laemmli (*23*) and transferred to an Immobilon-P polyvinylidene fluoride membrane (Millipore Co., Bedford, MA) at 1 mA/cm^2 for 1 h in a semi-dry transfer system (transfer buffer: 25 mM Tris, 190 mM glycine, 20% methanol). The membrane was blocked with 3% skim milk in PBS containing 0.05% Tween 20 (PBS-T) at room temperature for 1 h. After washing with PBS-T, the membranes were incubated with the primary antibodies against either EGFP (anti-GFP polyclonal antibody) (Medical & Biological Laboratories Co.) or Gm23K (monoclonal antibody 1G4) (*7*) for 1 h at room temperature. After washing with PBS-T, the membranes were incubated with the secondary antibody horseradish peroxidase (HRP)-conjugated anti-rabbit IgG (GE Healthcare UK, Ltd., Buckinghamshire, UK) or

HRP-conjugated anti-mouse IgG (GE Healthcare UK) for 1 h at room temperature followed by visualization using Luminata Forte Western HRP substrate (Millipore Co.) according to the manufacturer's instructions. For detection of the proteins immunoprecipitated with protein G Sepharose, SDS-PAGE and Western blotting were performed in the same manner as described for the experiments with the total proteins, except that anti-GFP monoclonal antibody (mAb) (Medical & Biological Laboratories Co.) was used as a primary antibody and HRP-conjugated anti-mouse IgG was used as a secondary antibody.

RESULTS

Determination of the full-length sequence of the 5'-nucleotide sequence region of cDNA encoding the Gm28K precursor. To determine the cDNA sequence encoding the 5'-untranslated region and the Gm28K precursor, 5'RLM-RACE was performed using the total RNA derived from developing soybean cotyledons. Consequently, the new nucleotide sequence identified in the present study was revealed to consist of a nucleotide sequence corresponding to 24 amino acids and a 5' untranslated nucleotide sequence with 17 nucleotides, as shown in Fig. 2. The Gm28K would be produced as a precursor with a signal peptide of 24 amino acid residues including MGN. Furthermore, a stop codon, TAA, was located 45 base-area upstream of the methionine residue, and there was no methionine within this 45 bases in the genomic sequence from soybean.

The location of Figure 2.

Functional analysis of the putative SP of the Gm28K precursor in tobacco BY2 suspension-cultured cells. We next examined the function of the putative N-terminal SP of the Gm28K precursor. EGFP was used as a reporter protein and was fused to SP, Gm28K, Gm23K, and Gm28-Gm23K (Fig. 1). Each construct was stably

expressed in tobacco BY2 suspension-cultured cells and the detailed localization was imaged by a fluorescent microscope.

When the EGFP gene was introduced into tobacco BY2 suspension-cultured cells, the EGFP was localized in the cytoplasm and the nucleus (Fig. 3A). On the other hand, in the case of the 4-d-old SP-EGFP/BY2 cells, EGFP was mainly localized in the ER (Fig. 3B). However, in the 8-d-old SP-EGFP/BY2 cells, the EGFP was partially transported to the vacuoles, although most of the EGFP fluorescence was still found in the ER (Fig. 3B). Figure 3C also shows that EGFP was predominantly localized in both the ER and vacuoles of 8-d-old SP-EGFP/BY2 cells whose tonoplasts were stained with FM4-64. In a Western blot of total soluble proteins in 4-d-old SP-EGFP/BY2 cells, the anti-EGFP antibody-positive band was detected at the same position (27 kDa) as that obtained from 4-d-old EGFP/BY2 cells (Fig. 4A). The protein with a molecular mass of 27 kDa was EGFP, and thus these findings indicate that the co-translational cleavage of the SP of SP-EGFP results in EGFP. No band was detected in the Western blot obtained with empty vector-transformed BY2 cells and the 7S globulin fraction containing Gm28K and Gm23K (Fig. 4A). However, the Western blot of the culture medium of 10-d-old SP-EGFP/BY2 cells showed that a large amount of EGFP (27 kDa) was secreted into the medium, although the medium from 4-d-old SP-EGFP/BY2 cells did not contain any EGFP (data not shown). These data suggest that the SP region of the Gm28K precursor functions in intracellular trafficking and secretion of EGFP.

The location of Figure 3.

The location of Figure 4.

Effect of the Gm28K precursor on transport of EGFP to the vacuoles in tobacco BY2 suspension-cultured cells. We expressed a Gm28-Gm23K peptide fused to SP-EGFP (SP-EGFP-Gm28-Gm23K) to investigate the behaviors of the Gm28K and Gm23K regions in tobacco BY2 cells. Figure 5A shows the localization patterns of EGFP in SP-EGFP-Gm28-Gm23K/BY2 cells over the indicated days after subculture.

In the SP-EGFP-Gm28-Gm23K/BY2 cells, the produced protein was mainly found in the ER from the 1st to the 3rd day but was not observed in vacuoles. In contrast, a remarkable accumulation of EGFP-Gm28-Gm23K was observed in vacuoles of 4-d-old SP-EGFP-Gm28-Gm23K/BY2 cells, and the main localization site of EGFP-Gm28-Gm23K shifted from the ER to the ER/vacuoles after 1 wk. The vacuolar localization of EGFP-Gm28-Gm23K in SP-EGFP-Gm28-Gm23K/BY2 cells was confirmed by visualization of the vacuole membrane with FM4-64 (Fig. 5B). The localization pattern of EGFP in EGFP-Gm28-Gm23K/BY2 cells was similar to that in EGFP/BY2 cells (Fig. 5C). Western blotting with an anti-GFP antibody and the total soluble proteins from 4-d-old SP-EGFP-Gm28-Gm23K/BY2 cells showed that a 27-kDa protein was detected but that the EGFP-Gm28-Gm23K with a molecular mass of 78 kDa could not be detected (Fig. 4A). Therefore, the immunoprecipitation with an anti-GFP antibody was applied to concentrate EGFP-Gm28-Gm23K in the cells. A protein band with a molecular mass of 78 kDa corresponding to that of EGFP-Gm28-Gm23K was observed in the Western blot with the immunoprecipitated sample (Fig. 6A). The putative EGFP-Gm28-Gm23K was also detected with the anti-Gm23K antibody prepared in our previous study (*7*) (Fig. 6B). On the other hand, Gm23K was also detected in the cell extracts with anti-Gm23K antibody (Fig. 6C). This finding suggests that the EGFP-Gm28-Gm23K was proteolytically cleaved between Gm28K and Gm23K in SP-EGFP-Gm28-Gm23K/BY2 cells. No EGFP-Gm28-Gm23K was found in the culture medium (Fig. 4B).

Further, in order to investigate the influence of EGFP bound to the N-terminus of Gm28K on the transport of the precursor, we tried to express SP-Gm28-Gm23K-EGFP, in which EGFP was fused to the C-terminal of Gm23K. The fluorescence of EGFP was accumulated in the vacuoles of SP-Gm28-Gm23K-EGFP/BY2 cells, which was similar to the localization observed with 10-d-old SP-EGFP-Gm28-Gm23K/BY2 cells (data

not shown). In SP-Gm28-Gm23K-EGFP/BY2 cells, Western blot analysis using an anti-EGFP and an anti-Gm23K antibody showed that EGFP was not secreted into the medium. These results suggest that EGFP is not involved in vacuolar transport of the precursor.

The above observations clearly demonstrate that both EGFP-Gm28-Gm23K and Gm28-Gm23K-EGFP are transported to the vacuoles. The location of Figure 6.

Effect of the Gm28K precursor sequence on vacuolar transport in tobacco BY2 suspension-cultured cells. Next we attempted to prepare SP-EGFP-Gm28K/BY2 cells and SP-EGFP-Gm23K/BY2 cells to examine the contribution of the Gm28K or Gm23K region to vacuolar sorting of the Gm28K precursor in tobacco BY2 cells. The fluorescence of EGFP was very weak in the vacuoles of the SP-EGFP-Gm28K/BY2 cells as compared to that in the vacuoles of SP-EGFP-Gm28-Gm23K/BY2 cells (Fig. 5A, Fig. 7). A Western blot of the total soluble proteins from 4-d-old SP-EGFP-Gm28K/BY2 cells revealed the presence of EGFP-Gm28K (55 kDa) (Fig. 4A). However, the SP-EGFP-Gm28K/BY2 cells secreted a large amount of EGFP into the 10-d-old culture medium, which was similar to the result for SP-EGFP/BY2 cells (Fig. 4B). The location of Figure 7.

An NPIRL motif derived from sweet potato sporamin has been shown to function as a sequence-specific VSD, and its detailed mechanisms have been analyzed in tobacco BY2 cells (*23*). The Gm28K precursor has an NPIRL-like motif, NPISD, around residue 330 of Gm23K. Therefore, we examined the function of Gm23K as a VSD. We expressed an SP-EGFP-Gm23K without Gm28K in tobacco BY2 cells. The strong and stable fluorescence of EGFP was localized in the ER lumen of the SP-EGFP-Gm23K/BY2, but not in the vacuoles, Golgi, endosomes or cell wall (Fig. 8). This distribution remained constant over the 10-d culture (data not shown). A Western blot of total soluble proteins from the 4-d-old SP-EGFP-Gm23K/BY2 cells

demonstrates that EGFP-Gm23K (50 kDa) was detected with anti-GFP antibody and anti-Gm23K antibody (Fig. 4A). Moreover, in a Western blot with the anti-Gm23K antibody, a degraded protein derived from the EGFP-Gm23K region with a molecular mass of approximately 30 kDa was observed (Fig. 6B). The EGFP was not secreted into the medium of SP-EGFP-Gm23K/BY2 cells (Fig. 4B).

These findings show that EGFP cannot be transported from the ER to the Golgi under the conditions examined here. Thus, it is suggested that the Gm28K-Gm23K region of the Gm28K precursor plays an important role in vacuolar-sorting in tobacco BY2 cells.

The location of Figure 8

DISCUSSION

Previous reports have suggested that the precursor Gm28K is synthesized as a preproprotein which would be composed of an SP, and Gm28K and Gm23K (*7*), the latter two of which are plant-specific *N*-glycosylated proteins in soybean (*7, 24*). Both proteins are localized in the protein bodies derived from PSVs (*7*). This suggests that Gm28K and Gm23K are sorted to the vacuoles through the ER and the Golgi. However, the N-terminal sequence of the Gm28K precursor has not been fully characterized. In the present study, we demonstrated that the Gm28K precursor has an SP of 24 amino acids. A typical SP for transport into the ER lumen has positively charged amino acids followed by 7-15 hydrophobic amino acids and polar amino acids (*25*). The SP from the Gm28K precursor also has similar characteristics. This suggests that the peptide of 24 amino acids functions as a signal peptide in soybean to transport the Gm28K precursor to the ER lumen.

What is the function of the SP of the Gm28K precursor with regard to its transport to vacuoles? To address this question, we expressed SP-EGFP-fused proteins in tobacco BY2 suspension-cultured cells. The main location site of EGFP in the SP-EGFP/BY2

cells was the ER. Moreover, Western blotting data showed that EGFP is formed by cleavage from SP-EGFP after removal of the SP (Fig. 4A), and that a large amount of EGFP is secreted into the medium of 10-d-cultured cells (Fig. 4B). These findings suggest that the SP acts as a signal peptide for the translocation of the Gm28K precursor into the ER lumen, and this translocation results in the default secretory pathway. In contrast, SP-EGFP-Gm28-Gm23K/BY2 cells gave a steady-state fluorescence in the vacuoles (Fig. 5A) and no secretion of the proteins, including EGFP, into the 10-d culture medium (Fig. 4B). On the other hand, the fluorescence of EGFP in the EGFP-Gm28-Gm23K/BY2 cells was observed in the cytoplasm and nucleus due to the absence of SP (Fig. 5C). These results suggest that the VSDs exist in the Gm28-Gm23K region.

The SP-EGFP-Gm28K/BY2 cells and SP-EGFP-Gm23K/BY2 cells were expressed to study the role of the Gm28K or Gm23K segment of the precursor in its transport to the vacuole. The fluorescence of EGFP in SP-EGFP-Gm28K/BY2 cells showed a significant accumulation in the ER and a large amount of EGFP was secreted into the culture medium (Figs. 4 and 7). The above observations indicate that the Gm28K itself is not sufficient for transport of the Gm28K precursor to the vacuoles.

The C-terminal amino acid sequence often participates in an important signal for protein trafficking, such as the ER retention signal (*26*, *27*), membrane association via lipidation signal (*28*) or vacuolar sorting signal (*10*-*13*). However, the Gm28K has neither specific signal nor a similar vacuolar-sorting motif at its C-terminus. It was reported that a ctVSD exhibited only weak transport ability when GFP was fused to its C-terminal region in plant cells (*10*, *29*). This report suggests that the Gm28K precursor might not have a ctVSD, because SP-Gm28-Gm23K-EGFP/BY2 cells can transport Gm28-Gm23K-EGFP to the vacuoles. An amino acid motif as a vacuolar-sorting signal was found on the vacuole-resident protein sporamin in sweet

potatoes. This motif consists of 5 amino acids, NPIRL, and even if the arginine and leucine residues are replaced by any amino acid residue and any high hydrophobic amino acid residue, respectively, its vacuolar-sorting function is not lost in tobacco BY2 cells (*23*). The Gm28K precursor contains an NPIRL-like motif, NPISD, on the Gm23K segment. If the NPISD acts as a vacuolar-sorting signal, EGFP-Gm23K should be transported to the vacuoles of the SP-EGFP-Gm23K/BY2 cells. However, confocal images and a Western blotting analysis revealed an accumulation of EGFP-Gm23K in the ER and no secretion into the culture medium. In SP-EGFP/BY2 cells and SP-EGFP-Gm28K/BY2 cells, a large amount of EGFP was secreted into the medium. The alteration of the SP-EGFP-Gm23K structure by fusion of Gm23K may cause its incorrect folding and accumulation in the ER. The detection of the 30-kDa protein in Fig. 6C shows that an unpredicted cleavage of SP-EGFP-Gm23K occurred on the Gm23K part in the ER. It was suggested that the ER-accumulation of EGFP-Gm23K inhibited its transport to the Golgi followed by trafficking to the vacuoles or secretion to the extracellular compartment. The seed storage proteins in the 7S globulin, common bean phaseolin and jack bean canavalin form their multimers (*30, 31*). For example, when the common bean phaseolin was expressed as a monomer, it bound to the binding protein (a member of the heat-shock 70 family) and was degraded by ER-associated degradation (ERAD). On the other hand, it was shown that the wild-type formed trimers in the ER to transport them to the vacuole (*32*). Some newly synthesized proteins that cannot fold correctly are degraded by ERAD (*33*, *34*). Gm28K and Gm23K may be required for the multimerization of the Gm28K precursor in the ER, and then Gm28K and Gm23K may also be transported to PSVs by VSDs through the Golgi apparatus.

In the present study, we revealed that a previously identified 24-amino-acid segment on the N-terminus of the Gm28K precursor functions as a signal peptide to

enter the ER lumen on its vacuolar-sorting pathway. EGFP is secreted into the medium of SP-EGFP-Gm28K/BY2 cells, suggesting that Gm23K is required for transport of the Gm28K precursor to the vacuoles. Although, in the SP-EGFP-Gm23K/BY2 cells, the fluorescence was shown to accumulate in the ER but not in the vacuoles, some types of VSDs on the Gm23K segment may be recognized by the vacuolar sorting receptor for VSDs in the Golgi apparatus after leaving the ER. Our findings suggest that the Gm28K and Gm23K segments are transported to the soybean vacuoles from the ER lumen via the Golgi in the Gm28-23K proprotein form.

REFERENCES

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

2) Ogawa T, Bando N, Tsuji H, Nishikawa K, Kitamura K. 1995. α-subunit of β-conglycinin, an allergenic protein recognized by IgE antibodies of soybean-sensitive patients with atopic dermatitis. *Biosci Biotech Biochem* **59**: 831-833.

3) Ogawa T, Tsuji H, Bando N, Kitamura K, Zhu YL, Hirano H, Nishikawa K. 1993. Identification of the soybean allergenic protein, Gly m Bd 30K, with the soybean seed 34-kDa oil-body-associated protein. *Biosci Biotechnol Biochem* **57**: 1030-1033.

4) Ji C, Boyd C, Slaymaker D, Okinaka Y, Takeuchi Y, Midland SL, Sims JJ, Herman E, Keen N. 1998. Characterization of a 34-kDa soybean binding protein for the syringolide elicitors. *Proc Nat Acad Sci USA* **95**: 3306–3311.

5) Tsuji H, Bando N, Hiemori M, Yamanishi R, Kimoto M, Nishikawa K, Ogawa T. 1997. Purification of characterization of soybean allergen Gly m Bd 28K. *Biosci Biotechnol Biochem* **61**: 942-947.

6) Tsuji H, Hiemori M, Kimoto M, Yamashita H, Kobatake R, Adachi M, Fukuda T, Bando N, Okita M, Utsumi S. 2001. Cloning of cDNA encoding a soybean allergen, Gly m Bd 28K. *Biochim Biophys Acta* **1518**: 178-182.

7) Hiemori M, Ito H, Kimoto M, Yamashita H, Nishizawa K, Maruyama N, Utsumi S, Tsuji H. 2004. Identification of the 23-kDa peptide derived from the precursor of Gly m Bd 28K, a major soybean allergen, as a new allergen. *Biochim Biophys Acta* **1675**: 174-183.

8) 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.

9) Frigerio L, Jolliffe NA, Di Cola A, Felipe DH, Paris N, Neuhaus JM, 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.

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

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

12) Nishizawa K, Maruyama N, Utsumi S. 2006. The C-terminal region of α' subunit of soybean β-conglycinin contains two types of vacuolar sorting determinants. *Plant Mol Biol* **62**: 111-125.

13) Saint-Jore-Dupas C, Gilbert MA, Ramis C, Paris N, Kiefer-Meyer MC, Neuhaus JM, Faye L, Gomord V. 2005. Targeting of proConA to the plant vacuole depends on its nine amino-acid C-terminal propeptide. *Plant Cell Physiol* **46**: 1603-1612.

14) Saalbach G, Jung R, Kunze G, Saalbach I, Adler K, Muntz K. 1991. Different legumin protein domains act as vacuolar targeting signals. *Plant Cell* **3**: 695-708.

15) Satoh J, Kato K, Shinmyo A. 2004. The 5'-untranslated region of the tobacco alcohol dehydrogenase gene functions as an effective translational enhancer in plant. *J Biosci Bioeng* **98**: 1-8.

16) Rempel HC, Nelson LM. 1995. Analysis of conditions for Agrobacterium-mediated transformation of tobacco cells in suspension. *Transgenic Res* **4**: 199-207.

17) Murashige, Toshio, Skoog, Folke. 1962. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum **15***: 473–497.

18) Misaki R, Sakai Y, Omasa T, Fujiyama K, Seki T. 2011. N-terminal vacuolar sorting signal at the mouse antibody alters the *N*-linked glycosylation pattern in suspension-cultured tobacco BY2 cells. *J Biosci Bioeng* **112**: 476-484.

19) Bolte S, Talbot C, Boutte Y, Catrice O, Read ND, Satiat-Jeunemaitre B. 2004. FM-dyes as experimental probes for dissecting vesicle trafficking in living plant cells. *J Microsc* **214**: 159-173.

20) Xie B, Deng Y, Kanaoka MM, Okada K, Hong Z. 2012. Expression of Arabidopsis callose synthase 5 results in callose accumulation and cell wall permeability alteration. *Plant Sci* **183**: 1-8.

21) Jun L, Saiki R, Tatsumi K, Nakagawa T, Kawamukai M. 2004. Identification and subcellular localization of two solanesyl diphosphate synthases from Arabidopsis thaliana. *Plant Cell Physiol* **45**: 1882-1888.

22) Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

23) 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.

24) Hiemori M, Bando N, Ogawa T, Shimada H, Tsuji H, Yamanishi R, Terao J. 2000. Occurrence of a soybean allergen, Gly m Bd 28K. *Int Arch Allergy Immunol* **122**: 238-245.

25) von Heijne G. 1990. The Signal Peptide. *J.Membrane Biol* **155**: 195-201.

26) Munro S, Pelham HR. 1987. A C-terminal signal prevents secretion of luminal ER proteins. *Cell* **48**: 899-907.

27) Pelham HR. 1988. Evidence that luminal ER proteins are sorted from secreted proteins in a post-ER compartment. *EMBO J* **7**: 913-918.

28) Lane KT, Beese LS. 2006. Thematic review series: lipid posttranslational modifications. Structural biology of protein farnesyltransferase and geranylgeranyltransferase type I. *J Lipid Res* **47**: 681-699.

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

30) Ceriotti A, Pedrazzini E, Bielli A, Giovinazzo G, Bollini R, Vitale A. 1995. Assembly and intracellular transport of phaseolin, the major storage protein of *Phaseolus vulgaris* L. *J Plant Physiol* **145**: 648-653.

31) Ko TP, Ng JD, McPherson A. 1993. The three-dimensional structure of canavalin from jack bean (*Canavalia ensiformis*). *Plant Physiol* **101**: 729-744.

32) Foresti O, Frigerio L, Holkeri H, de Virgilio M, Vavassori S, Vitale A. 2003. A phaseolin domain involved directly in trimer assembly is a determinant for binding by the chaperone BiP. *Plant Cell* **15**: 2464-2475.

33) Pedrazzini E, Giovinazzo G, Bielli A, de Virgilio M, Frigerio L, Pesca M, Faoro F, Bollini R, Ceriotti A, Vitale A. 1997. Protein quality control along the route to the plant vacuole. *Plant Cell* **9**: 1869-1880.

34) Meusser B, Hirsch C, Jarosch E, Sommer T. 2005. ERAD: the long road to destruction. *Nat Cell Biol.* 7: 766-772.

FIGURE LEGENDS

Fig. 1. Schematic models of proteins expressed in tobacco BY2 suspension-cultured cells: EGFP, Gm28K and/or 23K with or without signal peptide. Each construct was controlled under the cauliflower mosaic virus 35S promoter, NtADH 5'UTR, and the nopaline synthase terminator in plant expression vectors.

Fig. 2. Identified 5'-nucleotide sequence encoding the Gm28K precursor and its deduced amino acid sequence. Underlining and an arrowhead show a predicted signal peptide and the N-terminal site of matured Gm28K, respectively.

Fig. 3. Localization of EGFP in EGFP/BY2 cells and SP-EGFP/BY2 cells. A: Confocal images of 8-day-old EGFP/BY2 cells stably expressing EGFP. B: Confocal images of 4-day-old and 8-day-old SP-EGFP/BY2 cells stably expressing EGFP. C: Confocal images of 8-day-old SP-EGFP/BY2 cells stained with FM4-64 stably expressing EGFP. Bars represent a length of 10 μm.

Fig. 4. Western blotting analysis of recombinant proteins in the cells and the culture media vectors using anti-GFP antibody. A: Four-day-old cell extracts from the indicated transformants. B: Ten-day-culture media from the indicated transformants. Mock: BY2 cells transformed with the pBI121-empty vector; 7S-globulin: 7S-globulin

fraction from soybean seeds; Cell extracts: Cell extracts of EGFP/BY2 cells in panel A; Ce: Blot of the proteins in the cell extracts; Me: Blot of the proteins in the media.

Fig. 5. Localization of EGFP-Gm28-Gm23K with or without SP in BY2 cells. A: The time-dependent accumulation of EGFP-Gm28-Gm23K in SP-EGFP-Gm28-Gm23K/BY2 cells cultured for the indicated number of days. B: Colocalization image of SP-EGFP-Gm28-Gm23K/BY2 cells with FM4-64 in 8-day-old cells. C: EGFP-Gm28-Gm23K/BY2 cells. Bars represent a length of 10 μm.

Fig. 6. Western blotting of recombinant proteins and their derivatives in the indicated transformants. A: Immunoprecipitates using anti-GFP antibody. B: Immunoprecipitates using anti-Gm23K antibody. C: Cell extracts from the indicated cell lines stained with anti-Gm23K antibody.

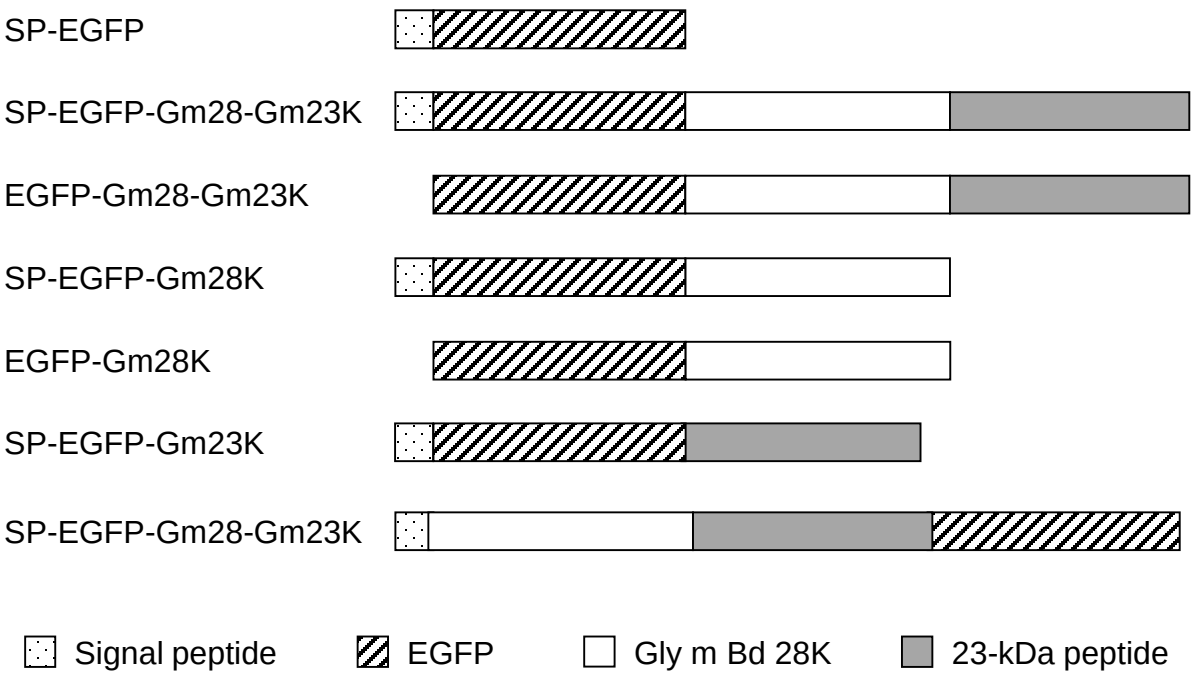
Fig. 7. Localization of EGFP-Gm28K with or without SP in BY2 cells. A: The time-dependent accumulation of EGFP-Gm28K in SP-EGFP-Gm28K/BY2 cells cultured for the indicated number of days. B: Colocalization image of SP-EGFP-Gm28K/BY2 cells with FM4-64 in 8-day-old cells. C: EGFP-Gm28K/BY2 cells. Bars represent a length of 10 μm.

Fig. 8. Fluorescence images of SP-EGFP-Gm23K cells with an ER tracker to stain the ER (A), FM4-64 to stain the tonoplast (B), and BODIPY to stain the Golgi (C) in 8-day-old SP-EGFP-Gm23K/BY2 cells. Bars represent a length of 10 μm.

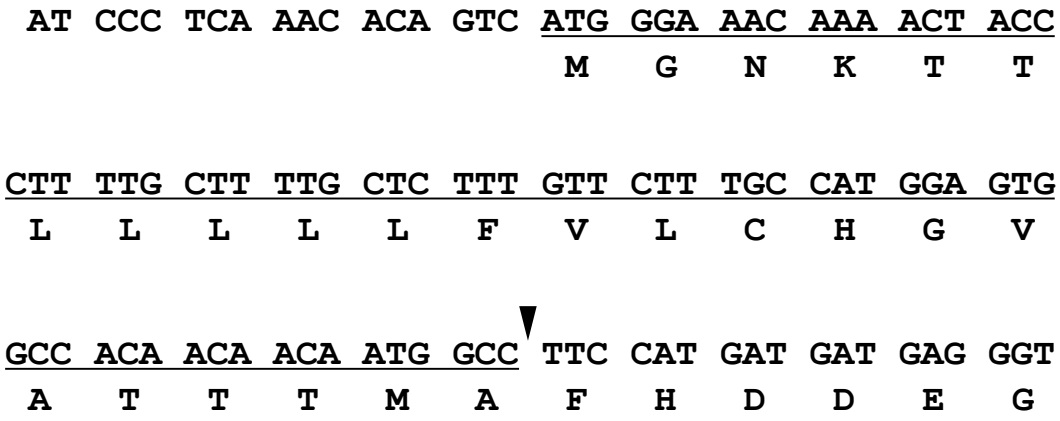
TABLE 1. Sequences of oligonucleotides used as PCR primers in the present study

Primer	Sequence*
1	5’-TTATCTAGAGTCTATTTAACTCAGTATTCA-3’
2	5’-AGTTTTGTTTCCCATTTATTTTTCTTGATT-3’
3	5’-AATCAAGAAAAATAAATGGGAAACAAAACT-3’
4	5’-ATGGAAGGCCATTGTTGTTGTGGCCACTCC-3’
5	5’-ACAACAACAATGGCCTTCCATGATGATGAG-3’
6	5’-TAAGAGCTCAAAAAACATCCATAACCACAT-3’
7	5’-GCCCTTGCTCACCATGGCCATTGTTGTTGT-3’
8	5’-ACAACAACAATGGCCATGGTGAGCAAGGGC-3’
9	5’-TAAGAGCTCTTACTTGTACAGCTCGTCCAT-3’
10	5’-CTCATCATCATGGAACTTGTACAGCTCGTC-3’
11	5’-GACGAGCTGTACAAGTTCCATGATGATGAG-3’
12	5’-TAAGAGCTCAGTTCTCTATCTTCTCATT-3’
13	5’-ACCAGCAGTGTCTTTCTTGTACAGCTCGTC-3’
14	5’-GACGAGCTGTACAAGAAAGACACTGCTGGT-3’
15	5’-TAAGAGCTCTTAAAAAACATCCATAACCAC-3’
16	5’-GCCCTTGCTCACCATTTATTTTTCTTGATT-3’
17	5’-AATCAAGAAAAATAAATGGTGAGCAAGGGC-3’
18	5’-GTTATGGATGTTTTTATGGTGAGCAAGGGC-3’
19	5’-GCCCTTGCTCACCATAAAAACATCCATAAC-3’
20	5’-TAATCTAGAATGATTCACACCAACCTGAAG-3’
21	5’-CTCGGTGTTGTCCATCCACACACAGATGAC-3’
22	5’-GTCATCTGTGTGTGGATGGACAACACCGAG-3’
23	5’-TAAGAGCTCCTACTGGGAGCCGGAGTGGCG-3’

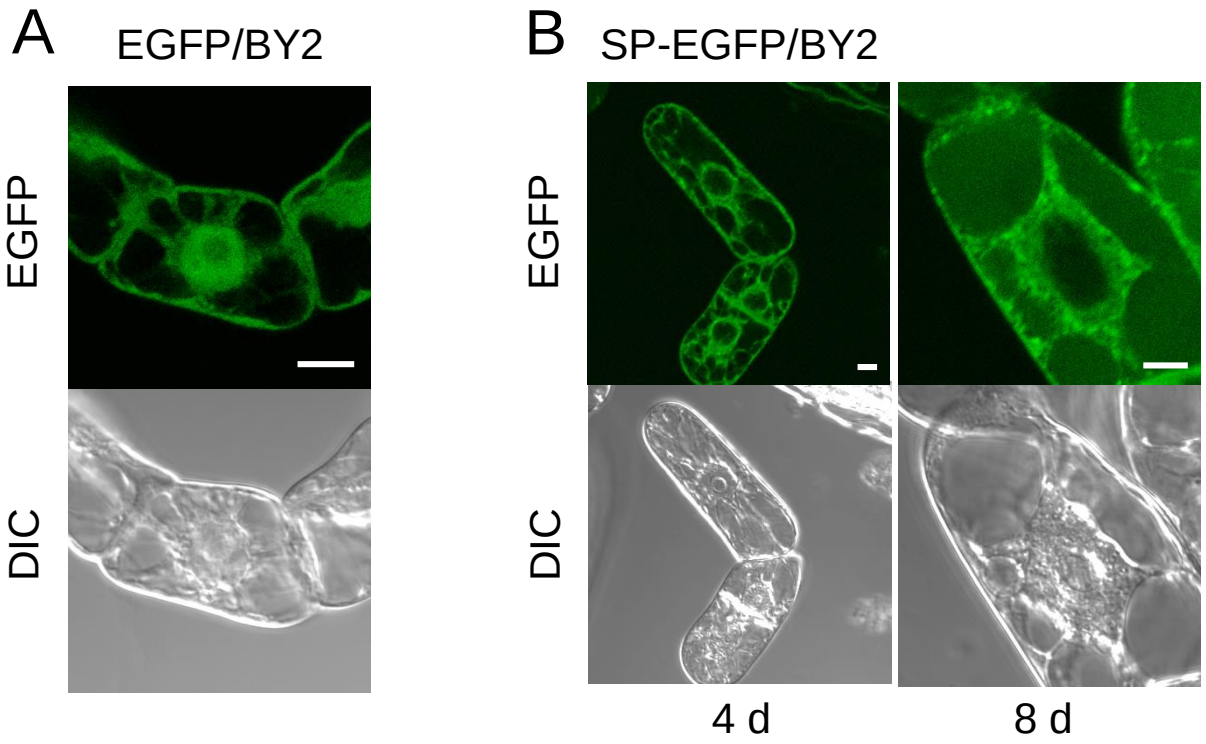
*The oliconucleotides used as PCR primers are listed in 5’-3’ orientation.
*Underlines mean the restriction sites of *Xba*I and *Sac*I.

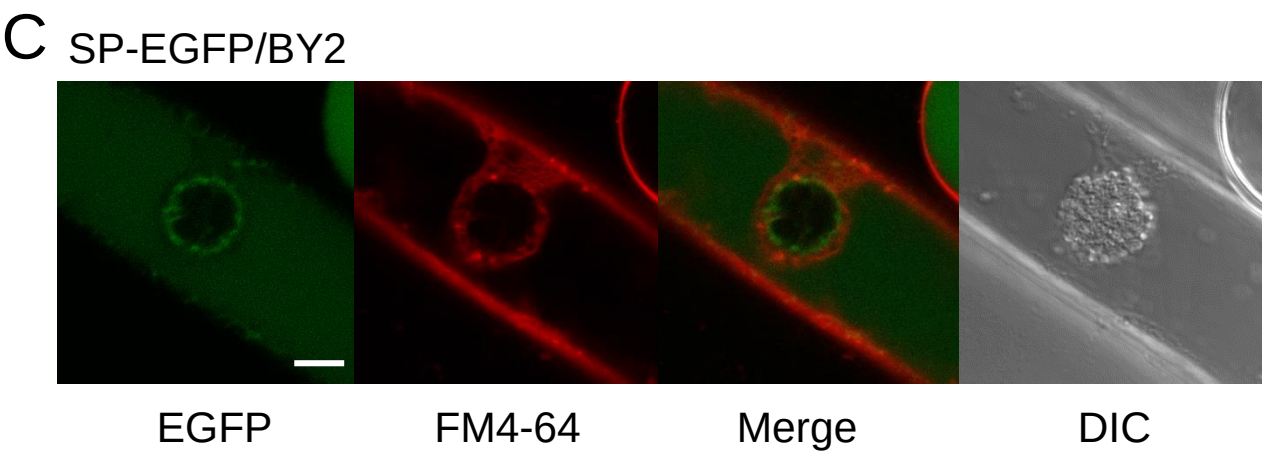


Yumioka-Figure 1

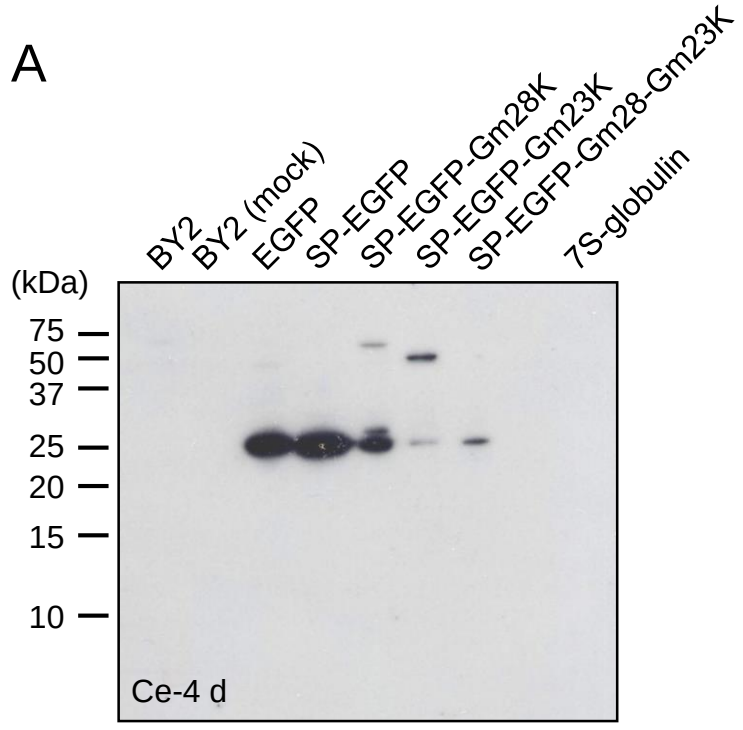


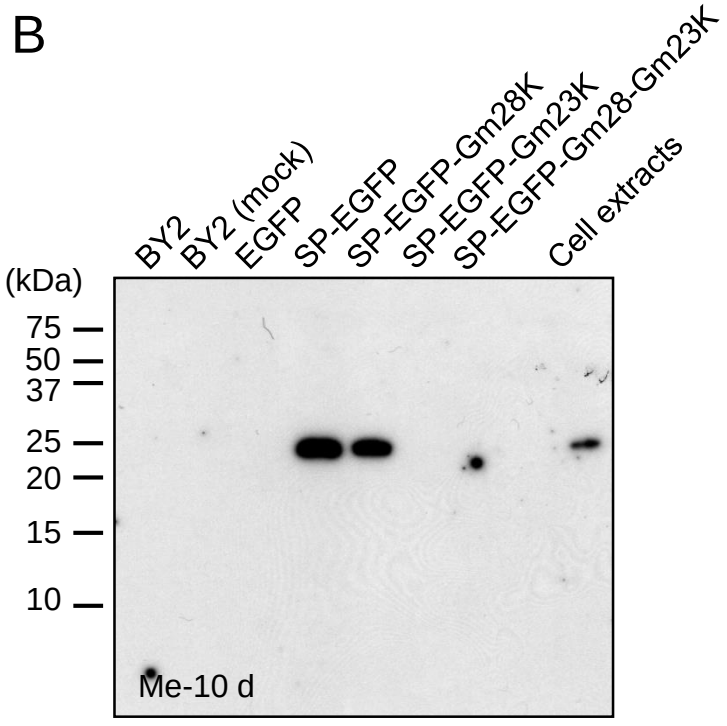
Yumioka-Figure 2



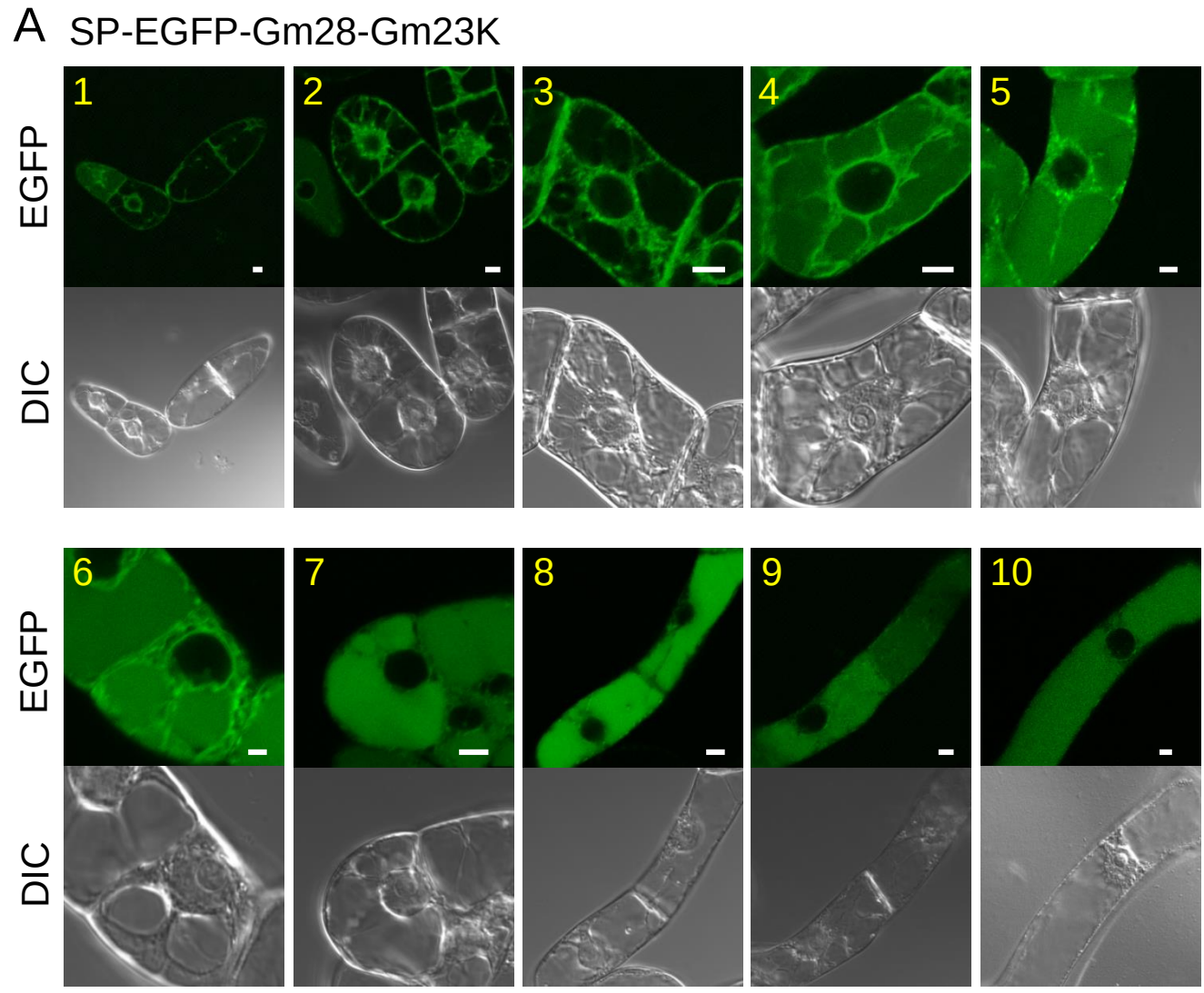


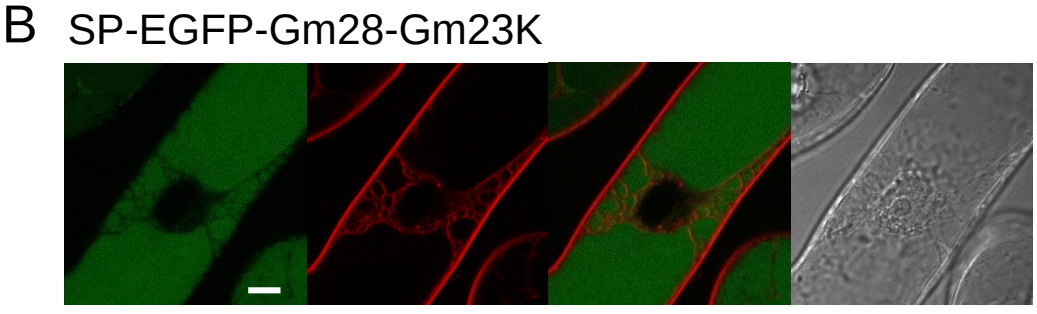
Yumioka-Figure 3

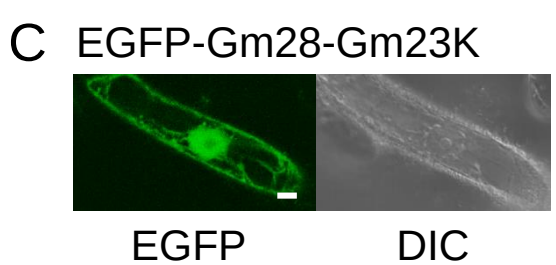




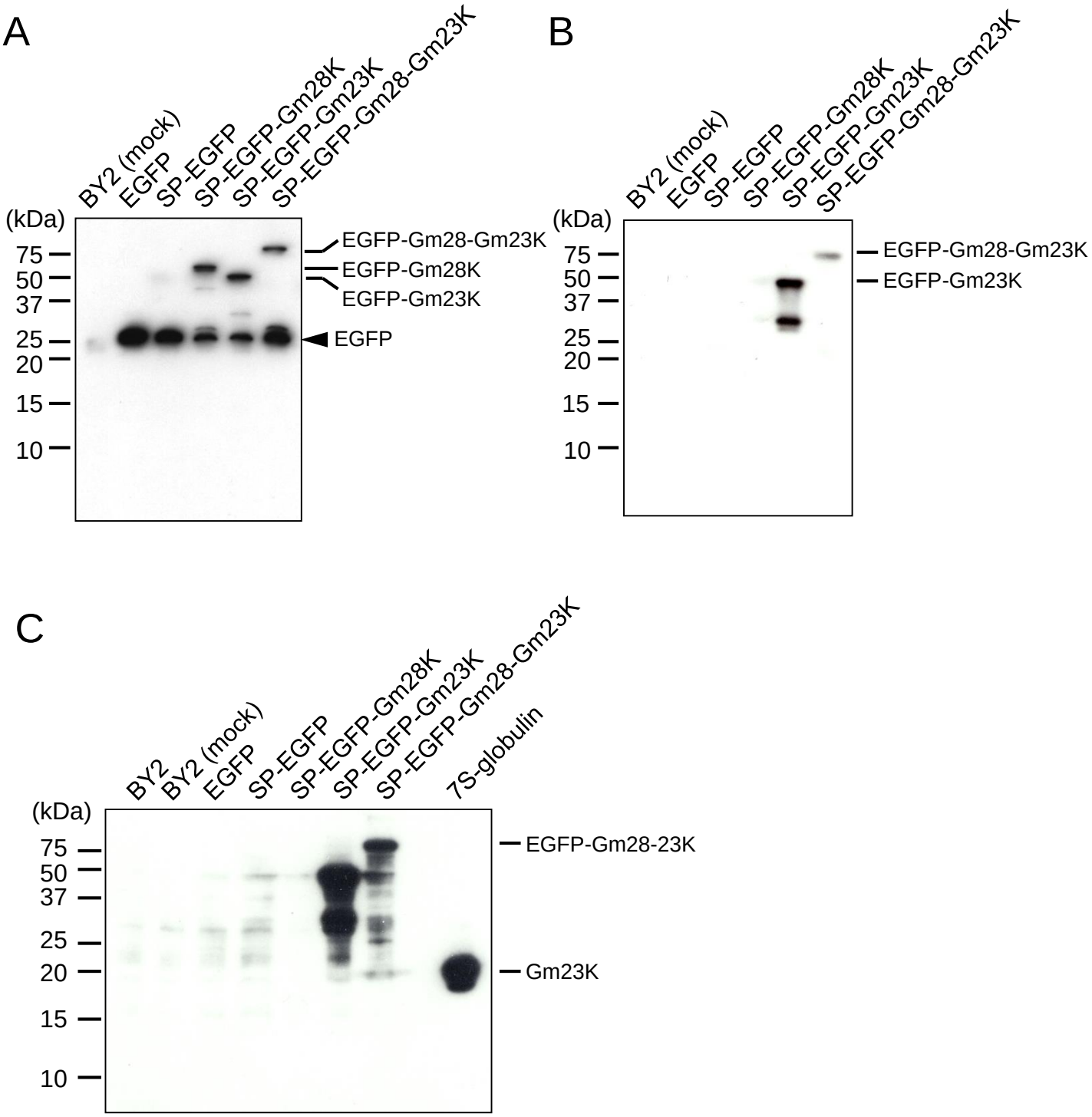
Yumioka-Figure 4



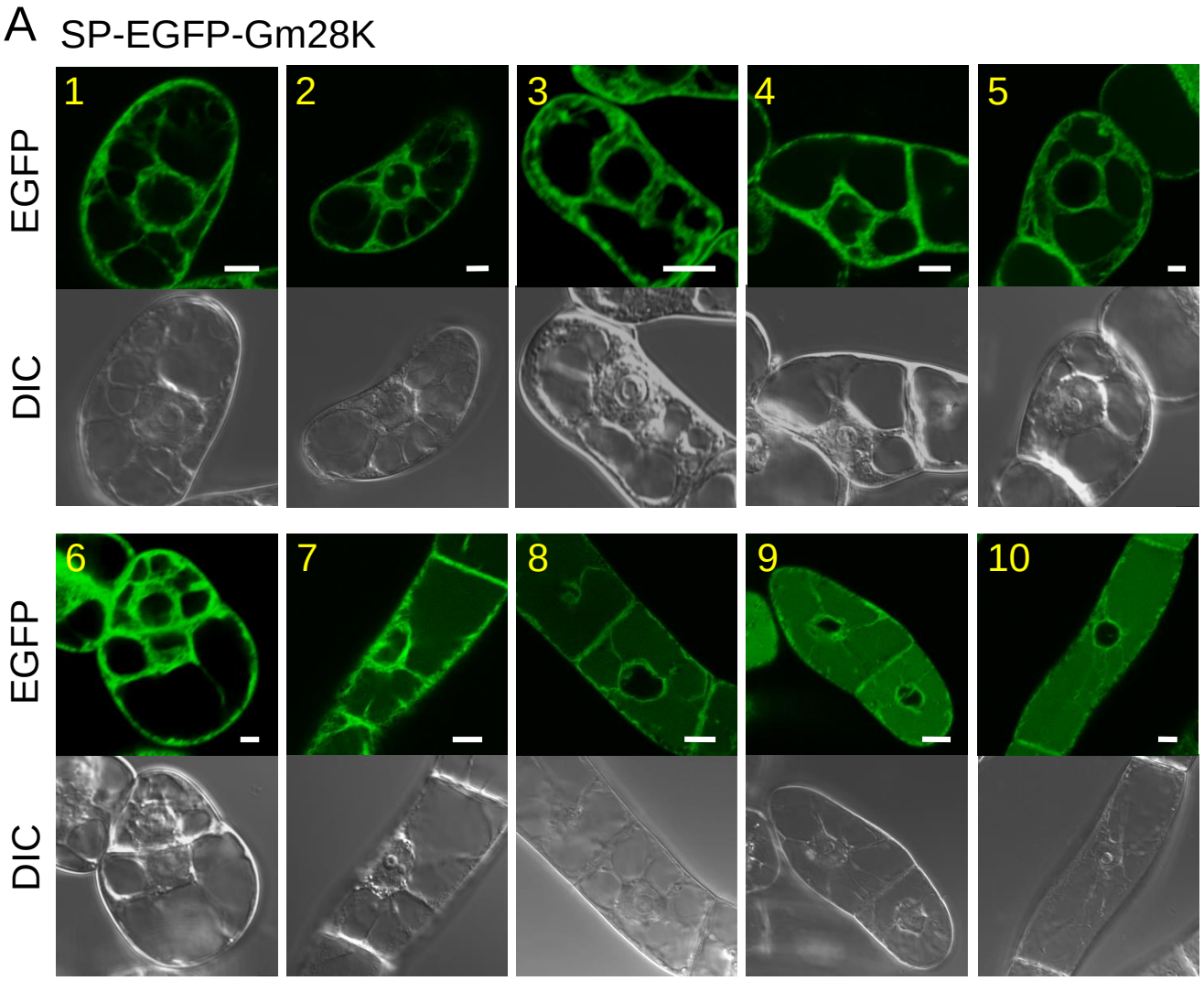




Yumioka-Figure 5



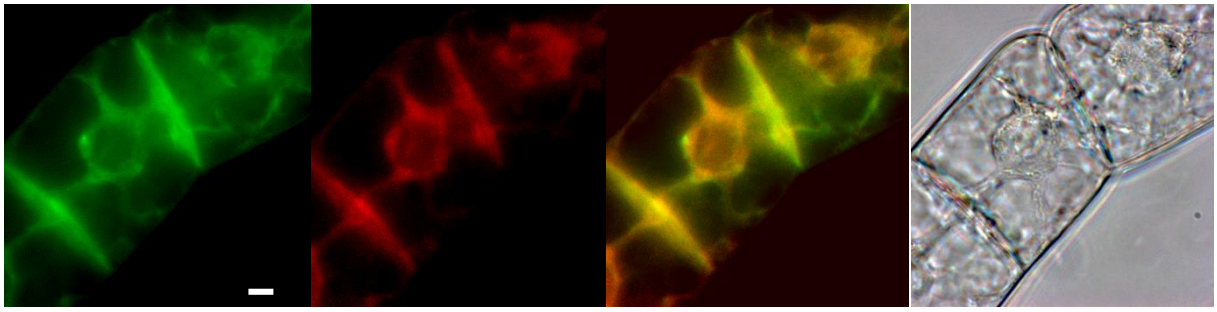
Yumioka-Figure 6



Yumioka-Figure 7

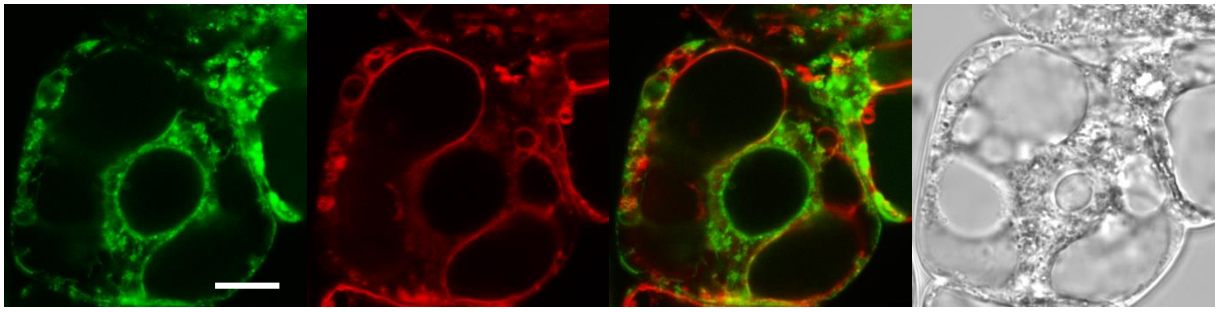
SP-EGFP-Gm23K

A



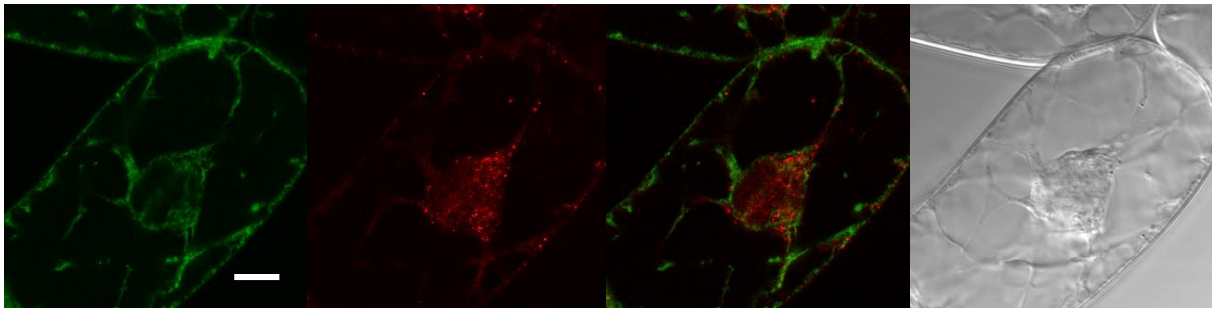
EGFP ER tracker Merge DIC

B



EGFP FM4-64 Merge DIC

C



EGFP BODIPY Merge DIC

Yumioka-Figure 8